



Impacts of Targeted Instruction with Mobile-Based Support on Teachers and Students in Rural Côte d'Ivoire

Sharon Wolf* 

University of Pennsylvania

Amy Ogan*

Carnegie Mellon University

Carnegie Mellon University-Africa

Kaja K. Jasińska* 

University of Toronto

Haskins Laboratories

Samuel Kembou* 

Jacobs Foundation

University of Lausanne

Vikram Kamath Cannanure

Carnegie Mellon University

Hannah Whitehead

Julianna Corbett 

University of Toronto

Fabrice Tanoh

Université Peleforo Gon Coulibaly

Brooke Wortsman

Joshua Lim

Hui Zhao

University of Toronto

Anahita Kumar

University of Pennsylvania

We report results from a school-randomized trial examining a targeted instruction program (“Teaching at the Right Level”) in Côte d’Ivoire implemented by teachers ($N = 167$ schools, 303 teachers, and 3,808 students). The program modestly improved literacy and numeracy learning levels ($d_w = .13$, $q = .013$ and $.27$, $q = .001$) but had no statistically significant impacts on functional literacy and numeracy ($d_w = .08$, $q < .096$ for both). After 2 years, 63 and 21% of children scored at the two lowest levels of literacy and numeracy, respectively. Learning gains were concentrated among children at the bottom of the

*These authors contributed equally.

SHARON WOLF is an associate professor in the Human Development and Quantitative Methods Division at the University of Pennsylvania Graduate School of Education. Her research examines the role of developmentally informed interventions in the home and at school on children's learning and development.

AMY OGAN is an associate professor in the Human-Computer Interaction Institute at Carnegie Mellon University, with an appointment at Carnegie Mellon University Africa. She is an educational technologist focusing on ways to make learning experiences more engaging, effective, and culturally relevant in emerging economies. Her research encompasses areas such as context and infrastructure in education, engendering rapport and relationships with social technologies to aid learning, and supporting teacher professional development through novel technologies.

KAJA K. JASIŃSKA is an associate professor in the Applied Psychology and Human Development Department at the University of Toronto. Her research examines how childhood experiences shape neurocognitive development and learning, focusing on understudied environments characterized by high poverty-related risks.

SAMUEL KEMBOU is the global lead for learning and evidence at the Jacobs Foundation. His work focuses on how education systems generate and use evidence to improve child learning and development, particularly in low- and middle-income countries. Trained as an economist and former lecturer at Lausanne University, his research and practice sit at the intersection of impact evaluation, education policy, and system-level reform in Africa.

VIKRAM KAMATH CANNANURE is a postdoctoral researcher at Saarland University in Saarbrücken, Germany. His research sits at the intersection of human-computer interaction, artificial intelligence, and education. He advances a Thoughtful AI approach that designs artificial intelligence systems to support reflection, learning, and inclusion rather than automation alone. His work focuses on conversational and AI-supported systems for education, health, and civic awareness in under-resourced contexts.

HANNAH WHITEHEAD is a PhD candidate in the Developmental Psychology and Education Program at the Ontario Institute for Studies in Education. Her research examines how experience shapes the neurodevelopment of complex skills such as reading and mathematics.

JULIANA CORBETT was a master of education student in developmental psychology and education at the University of Toronto. Her research examines how caregiving and educational environments shape children's socioemotional and academic development, particularly in contexts of adversity.

FABRICE TANOI is an assistant professor of economic sociology in the Faculty of Social Sciences, Department of Sociology, at Péléfero Gon Coulibaly University in Korhogo, Côte d'Ivoire. His research focuses on issues related to children's education and women's socioeconomic empowerment in relation to household livelihoods in cocoa-producing communities. He is a scientific advisor at the Laboratory for Education (Ed-Lab) of the Ministry of National Education of Côte d'Ivoire and works on a number of specific issues such as school dropout and teacher motivation and retention.

BROOKE WORSTMAN is a PhD student in the School and Clinical Child Psychology Program at the University of Toronto. Her research explores the mechanisms that promote educational attainment in Côte d'Ivoire using ecological systems and risk and resilience frameworks.

JOSHUA LIM is a PhD student in the Curriculum and Pedagogy Program with a specialization in comparative, international, and development education at the University of Toronto. His research interests center on the effects of poverty and child labor on child development.

HUI ZHAO was a master of education student in developmental psychology and education at the University of Toronto.

ANAHITA KUMAR is a research specialist at the Consortium for Policy Research in Education at the University of Pennsylvania School of Education. She holds a PhD in human development from the University of Pennsylvania. Her work applies impact evaluation and quantitative methods to large-scale survey and experimental data to study how poverty and structural disadvantage shape children's learning, health, and socioemotional development across global contexts.

distribution. Quantitative and qualitative data highlighted the importance of distinguishing between fluency and decoding or computational skills, teacher program knowledge, and barriers to and enablers of quality implementation.

KEYWORDS: randomized, controlled trial, sub-Saharan Africa, targeted instruction, teacher professional development

Introduction

Despite large increases in school enrollment around the world over the past 30 years, functional literacy and numeracy skills remain out of reach for many children in low- and middle-income countries. This is particularly true in Sub-Saharan Africa, where nine in 10 children are not at the level of reading a simple sentence by the end of primary school (UNICEF, 2021). One systemic challenge among many that schools face is the wide range of ability levels within a single classroom. With high rates of both underage and overage enrollment and grade repetition (Jasińska & Guei, 2022) and large differences in children's informal home learning environments, students have a wide range of skills within a single classroom (Wagner, 2017). As a result, teachers tend to teach to the top or average skill level in the class, leaving many students disengaged from the learning process (Wagner, 2017).

In such situations, targeting instruction to a child's learning level—rather than to their age or grade—has achieved acclaim as one of the most cost-effective approaches to improving learning outcomes (Angrist et al., 2023; Banerjee et al., 2023). This approach groups children across age and grade levels based on their knowledge levels for part of the school day, separately for literacy and numeracy. In small groups, children are taught at their level of knowledge to advance their learning. “Teaching at the Right Level” (TaRL), one such targeted instruction approach, is one of the most studied educational programs from a causal evidence framework.

Targeted instruction programs build on remedial education programs, which are designed to improve the performance of low-achieving students. Remedial education programs have many forms, including summer school, grade retention and intensive reading support, and individual tutoring (Eide & Showalter, 2001). Recent research on the state of Michigan's remedial education program—which included both intensive reading support and grade retention—showed that the program raised students' reading scores in the next school year by 0.045 standard deviations (Berne et al., 2025). The study found that the reading-support component drove the improvements rather than grade retention. Similarly, Jacob and Lefgren (2009) found that in Chicago public schools, summer school increased academic achievement in reading and math by up to 2 years, but grade retention had mixed effects based on children's age. Both studies pointed to the utility of intensive, student-centered support for low-achieving students.

A recent review of 14 impact evaluations of targeted instruction programs in seven geographic areas concluded that “[t]argeted instruction delivers average learning gains of 0.42 SD when taken up” (Angrist & Meager, 2023, p. 1). Importantly, this review showed substantial heterogeneity in effect sizes on learning outcomes across the studies, ranging from 0.08 to 0.70 standard deviations (Angrist & Meager, 2023), suggesting that there is still much to be learned about how to incorporate targeted instruction in classrooms successfully and with high fidelity as well as how to maximize effectiveness. The authors identified two key factors that explained the variation—dosage (measured by child attendance) and government teachers versus volunteer implementers, with much larger effects for treatment on the treated analyses and when programs were implemented by volunteers compared with teachers. This volunteer effect pertains to the effectiveness of educational programs when they are implemented at scale by local and national governments. Many previous evaluations of targeted instruction programs have been implemented by nongovernmental organizations, which can provide more oversight and support throughout a program to both staff and participants (e.g., Banerjee & Chavan, 2016). In Kenya, the same program had different impacts on teachers and students when it was implemented by a nongovernmental organization versus the government (Bold et al., 2018). Thus, although targeted instruction programs are considered a “great buy” (Banerjee et al., 2023, p. 20), the impacts of such programs when implemented by governments at a larger scale are limited; two studies suggested that government-implemented programs can have small impacts on children's learning outcomes (Banerjee et al., 2016; Beg et al., 2021). More evidence on optimizing government-led programs is critical to scale national efforts in sustainable ways.

The literature on teacher professional development (PD) points to some areas that can be incorporated. First, the duration of the training beyond 1 academic year appears to be important. In a study of elementary school teachers

involved in a language arts PD program, Cohen et al. (2016) found that teachers who participated in 2 years of PD compared with 1 year used the targeted practices more frequently and with higher levels of sophistication. Second, ongoing support and feedback are critical to effectively implement programs learned in a one-off training. Coaching has been widely demonstrated as one way to increase teacher take-up of PD programs and improve instructional quality and student learning (Kraft et al., 2018). In low-resource rural contexts, where travel to individual schools can be challenging and very time-consuming, virtual communication could enable low-cost PD at scale. An evidence review from developing countries concluded that simple but effective strategies to enhance teachers' practices and motivation have been demonstrated using phone-based supports (McAleavy et al., 2018).

An additional and critical gap in the research on targeted instruction programs is the need for valid external assessments. Traditionally, programs such as TaRL have been assessed using an internal tool called the Annual Status of Education Report (ASER) test, which is also used for grouping students according to ability. Aside from the non-peer-reviewed study by Banu Vagh (2012), there has been scarce research examining the validity of ASER classifications in measuring children's literacy development, and what has been done has been conducted in Indian contexts. Studies using data from Pratham's "Read India Program" suggested strong associations between ASER classifications and results from other standardized literacy assessment tools (e.g., Early Grade Reading Assessment [EGRA] and Dynamic Indicators of Basic Early Literacy Skills [DIBELS]) (Banu Vagh, 2012; UNESCO Institute for Statistics (UIS), 2016). However, the few other studies that exist comparing ASER levels with fluency batteries adapted from standardized assessment tools reveal inconsistencies (USAID, 2018). Likewise, results from an intervention evaluation assessment yielded similar discrepancies when comparing ASER levels with oral reading fluency scores (de Oca et al., 2022). The concise nature of the ASER 5-item assessment may contribute to these inconsistencies as well as variations in demographics, socioeconomic status, language of implementation, and different implementation progress (USAID, 2018). Although the ASER assessment has merit for efficient testing, research should explore program efficacy with established assessment tools.

In this study we used a school-randomized trial to assess the impacts of an adapted targeted instruction program in Côte d'Ivoire implemented by the Ministry of Education and Literacy in partnership with TaRL Africa. The Programme d'Enseignement Ciblé (PEC) was introduced in Côte d'Ivoire as a small pilot project in 2018 to address the learning crisis in the country and over the next few years was scaled across regions. Based on qualitative formative research, for this study, we supplemented the PEC program with an online electronic support system—DIA—that provided a platform for teachers to ask questions and receive answers about the program over 2 academic years. We contribute to the previous impact-evaluation literature on targeted

instruction programs by assessing a government-implemented program in a country vastly under-represented in educational research in low- and middle-income countries. Further, we expand the outcomes examined to include comprehensive measures of literacy and numeracy, advancing most previous studies that almost exclusively focused on learning-level groupings rather than more extensive measures of functional literacy and numeracy skills. We also examine potential mechanisms through teacher surveys and interviews to understand teachers' program knowledge and perspectives on implementing the program.

Rural Ivorian Educational Context

Côte d'Ivoire is a West African country with a population of 25.7 million people with a life expectancy of 57 years (World Bank, 2021). The country currently ranks 166 of 192 countries on the Human Development Index (a composite index of life expectancy, education, and per capita income) and is the largest producer and exporter of cocoa in the world (United Nations Development Programme [UNDP], 2024). In rural cocoa-producing communities, poverty is prevalent (World Bank, 2019), with many households living on \$1–2 a day (Côte d'Ivoire Institut National de la Statistique, 2015).

Although the Ivorian government has committed to expanding educational access through universal basic education, teaching quality and learning outcomes are very low, particularly in rural cocoa-growing regions (see, e.g., Jasińska & Guei, 2022; Jasińska et al., 2023). Côte d'Ivoire ranks among the bottom 30 countries globally (Angrist et al., 2021), with large inequalities between urban and rural regions (Programme on the Analysis of Education Systems [PASEC], 2020). In rural areas, this is partly due to very large class sizes and little ongoing PD and training for teachers (World Bank, 2018), leading to poor teacher motivation and performance, as well as other issues such as learning in a language not spoken in the home (Ball et al., 2022, 2024) and exposure to child labor (Kembou et al., 2023).

Côte d'Ivoire is highly linguistically diverse, with >70 Ivorian languages spoken as well as several Burkinabe and Malian languages due to the large migrant population (Jasińska et al., 2023). The official language is French, which is also the language of instruction in schools. Côte d'Ivoire has had several bilingual education initiatives, although bilingual education is not widespread throughout the country (for a detailed commentary, please see Ball et al., 2022, 2024; Jasińska et al., 2023). Most Ivorians learn French as their second language in school and speak one or more Ivorian languages. In 2015, the Ivorian state enacted a compulsory schooling policy, mandating that all children between the ages 6 and 16 years enroll in school. Enrollment in primary school is high (net enrollment rate = 91.1%), although few children transition to secondary school (net enrollment rate = 43.4%; UNESCO UIS, 2018). Only 33.1% of second grade students achieve minimum proficiency in reading

according to 2019 Programme on the Analysis of Education Systems of Francophone West African states (PASEC, 2020), with Côte d'Ivoire ranking 12th out of 14 countries surveyed. Even among children attending primary school, the average fifth grader is only able to read a few words (Sobers et al., 2023).

Teachers in Côte d'Ivoire face challenging working conditions, including staff shortages, overcrowded classrooms, and under-resourced schools (Education International, 2010). Close to half of teachers (42%) in the country are on fixed-term or temporary contracts, which pay less and are tenuous (Evans et al., 2022). Teacher burnout is a critical but understudied aspect of the Ivorian teaching context. Teacher absenteeism is high and estimated to be responsible for the loss of ~25% of teaching time (Albán Conto, 2021). The role of teacher burnout is central to improving teaching practice and student outcomes (e.g., Pianta et al., 2005) and to how teachers take up new programs (see, e.g., Brown et al., 2010; Jennings et al., 2013), particularly in under-resourced contexts (Wolf et al., 2019).

Côte d'Ivoire's PEC + DIA Program

The PEC program was codeveloped by TaRL Africa and the Ivorian Ministry of National Education and Literacy with the goal of improving teaching and students' foundational math and French skills for third, fourth, and fifth grade students. The program was piloted initially by the government with light-touch support from TaRL Africa in 50 schools and showed improvements in students' skills over the school year (Curtiss Wyss & Perlman Robinson, 2021). Teachers first performed a baseline test to group students by learning level based on their proficiency. Then the expected practice was that each day teachers split the children by learner groups across grade levels and conducted small-group activities in dedicated 45-minute slots of French and mathematics skills, where children were taught "at their level." These activities were intended to be child centered and playful. Teachers tested the students again during the middle and at the end of the year to evaluate their progress and regrouped them if needed. Teachers also received mentoring from trained government workers.

The program was embedded in the Ivorian education system and used the stakeholders in the ministry to implement and monitor the intervention, including in-classroom visits by district-level pedagogic advisors and school directors, who were senior schoolteachers appointed by the inspector to manage the school. In the 2021–22 school year, the program was scaled to many more schools throughout the country. The training for teachers was intended to be conducted for 1 week at the start of each school year (but did not actually occur until mid-November), with a week of training in the first year followed by a refresher training the following school year for the sample included in this study.

To supplement the in-person program, our team developed a chatbot to deliver expert knowledge to teachers throughout the year as needed (see Cannanure [2023] for design and implementation details). The chatbot was pilot tested and designed to be usable in low-resource contexts (see Cannanure et al., 2020) and was named *DIA*. DIA was implemented through Facebook Messenger (the preferred platform for teachers at the time) over the course of the study. It is a human–chatbot (humbot) hybrid system that organically learns topic-specific knowledge and local language from user interactions. The DIA platform was a central repository for documents such as the training manuals and assessments, which often were lost on paper. The platform used a human–artificial intelligence (AI) hybrid chatbot to provide answers to the questions teachers had about the PEC goals and techniques, and it served community functions such as allowing teachers to share and receive classrooms stories about the PEC implementation and to connect with other teachers implementing the program. Pedagogic advisors had access to DIA so that they could monitor the questions teachers asked and the discussions teachers posted to the platform.

This Study

This study focused on a research–practice–policy partnership between TaRL Africa, the Ivorian Ministry of Education, and academic researchers. In addition to examining impacts on children’s literacy and numeracy learning levels, we conducted in-depth assessments of children’s functional literacy and numeracy skills employing a widely used global assessment tool. Additionally, an unanswered question in the literature on targeted instruction programs is how child-, classroom-, teacher-, and school-level factors may or may not moderate program impacts. For example, targeted instruction programs sort children by ability groups based on skill levels, addressing learner heterogeneity by regrouping students into smaller, more manageable groups. Such programs could be more impactful in classrooms with high levels of heterogeneity in skill levels, but how classroom baseline-level heterogeneity may impact programs is unknown. Additionally, programs that address heterogeneity in skill levels should reduce classroom heterogeneity over time because the lowest-performing students receive targeted instruction that allows them to catch up to their more advanced peers. However, to the best of our knowledge, impacts of targeted instruction programs on learner variability have not been examined previously. Moreover, teacher burnout and school resources may moderate treatment impacts. Over 2 years of program implementation, we asked:

1. What are the impacts of PEC + DIA on children’s literacy and numeracy outcomes in CE1, CE2, and CM1 (equivalent to third, fourth, and fifth grades), and how do they depend on the measure of literacy and numeracy used?

2. Do impacts on children's literacy and numeracy outcomes vary based on child, classroom, teacher, and school characteristics?

Further, we considered potential mechanisms of program effects by examining knowledge of the targeted instruction program measured multiple times over the course of the first and second years of implementation. We posed two additional exploratory questions using non-experimental methods:

1. For treatment teachers, how does program knowledge change over the course of implementation? And does program knowledge relate to student outcomes?
2. What are teachers' implementation experiences, and how do they describe barriers to and enablers of implementing the program well?

Methods

Participants and Procedures

Our sample included students, teachers, and school directors sampled from participating schools in the fall of 2021 (beginning of the school year) and the spring of 2023 (end of the following school year). Data were collected in French and local language by trained enumerators in schools through direct assessments with children and direct surveys administered to teachers and school directors. Enumerators completed a 5-day training workshop and were examined on their skills prior to deployment in the field. Enumerators spent about 15 minutes with each child at baseline and 30 minutes with each child at endline to conduct more comprehensive assessments. Teacher surveys were ~40 minutes and school director surveys were ~20 minutes and were conducted at the school or in the nearby town. Research approval was received from the Ministry of Education, and ethics approval was obtained from the University of Toronto, University of Pennsylvania, University of Lausanne, and Carnegie Mellon University.

School Sample

One hundred and forty communities in three cocoa-growing regions of Côte d'Ivoire—Meagui, Daloa, and Bouafle—were part of a larger randomized, controlled trial testing the dual impacts of quality education (PEC program) and poverty reduction (cash transfers) on children and their families. Villages from the regions were included in the study if they had a school and were affiliated with a partnering cocoa co-op. We did not report on any data within Meagui ($n = 34$) because schools in that region had been saturated with the PEC intervention before baseline and thus were removed from the education experimental arm. An additional 61 villages (in the other two regions) were randomly selected from the full list of schools and added to

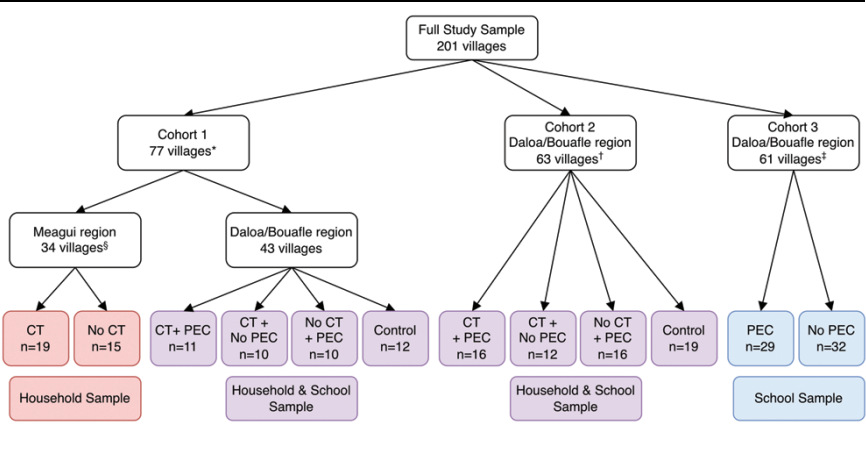


Figure 1. Study sample and randomization.

Note. Randomization of villages occurred at the cohort level.

*Eighty villages were originally randomized to all four conditions. After randomization, three villages opted out of the study, resulting in a cohort 1 sample size of 77.

†After randomization, three villages opted out of the cash-transfer study. Their PEC/no PEC treatment assignment was retained.

‡After the Meagui region was saturated with the PEC intervention, cohort 3 was added to increase study power. Cohort 3 was only randomized to PEC or no PEC (control) and did not participate in the cash-transfer portion of the study. No household sample was collected in cohort 3.

§After randomization, the Ivorian Ministry of Education decided to implement the PEC program in the Meagui region. Meagui was saturated with PEC, and PEC/no PEC treatment assignment was abandoned; only the cash-transfer (CT)/no-cash-transfer (no CT) randomization was retained (e.g., CT + PEC and CT + No PEC were reassigned to CT, and No CT + PEC and No CT + No PEC were reassigned to No CT). No school sample was collected in Meagui.

increase statistical power (see Figure 1). The original 140 communities were then randomly assigned to one of four conditions in the larger randomized, controlled trial before baseline data collection (i.e., unconditional cash transfer and PEC + DIA and unconditional cash transfer, PEC + DIA, and control), and the 61 added villages were randomly assigned to one of two conditions (PEC + DIA or control). All schools in a village were assigned to the same condition, and data were collected from one randomly selected school in the village if the village had multiple schools. Due to three villages opting out of the cash-transfer study (remaining only in the education arm) and the exclusion of villages in the Meagui region, the final number of villages/schools reported in this paper is 167 (106 of the original 140 communities and 61 added communities; $n_{\text{treatment}} = 82$, $n_{\text{control}} = 85$).

Student Sample

Within each school, seven to eight children were randomly sampled from each class roster in grades CE1, CE2, or CM1, equivalent to grades 3, 4, and 5 in North America, respectively, for data collection (although all children in PEC classrooms received the intervention). Cross-sectional data were collected from CE1 and CE2 students at baseline ($N = 2,481$) and CE1, CE2, and CM1 students at endline ($N = 3,808$) across the 167 schools.

At baseline, in the fall of year 1 (2021), we sampled 1,230 children from CE1 (590 girls and 640 boys) and 1,251 children from CE2 (609 girls and 642 boys). At endline, in the spring of year 2 (2023), we drew from the current class rosters to randomly sample children from CE1 ($n = 1,246$; 613 girls and 633 boys), CE2 ($n = 1,245$; 552 girls and 693 boy), and CM1 ($n = 1,317$; 658 girls and 659 boys).

With 3,800 children nested in 167 schools and assumptions of 80% power at the 5% significance level, an average intraclass correlation of child learning outcomes across schools of $\rho = 0.20$, and an R^2 value of 0.2, given baseline assessments, we were powered to detect a minimum detectable effect size of 0.175.

Teacher Sample

Two to three teachers ($N = 303$) who were responsible for implementing the PEC program were randomly sampled from each respective school for data collection at baseline and endline and followed over the 2 years of the project. We also conducted an embedded qualitative study with in-depth interviews with 12 teachers (two of whom had additional responsibilities as directors of their school) equally distributed across the two regions during the middle and at the end of the first year. An analysis of the frequency of chat-bot usage was conducted, and interview participants were randomly selected across active, moderate, and inactive users. Demographics of this sample were broadly representative of teachers in the region and are provided in the Supplementary Materials in the online version of the journal.

Measures

ASER Survey

The ASER is an assessment tool that measures the status of children's learning outcomes and enrollment (Pratham, 2020) as well as foundational literacy and numeracy skills (Banu Vagh, 2012). It was developed to be used as a community-based assessment and has been used internationally in India, Pakistan, and Ghana (Adil et al., 2022; Banerjee et al., 2016; Duflo et al., 2020; Jamil & Saeed, 2018). The ASER was adapted for French from Hindi and English versions, and enumerators administered the tests with each child in French.

Literacy Groups. Based on children's reading ability, they were placed in one of five categories: beginner, letter, word, paragraph, and story. Children were first asked to read one of two paragraphs (comprising four short sentences; e.g., "*Moussa voit un chat. Le chat est perdu. Moussa le caresse. Il lui donne du lait*" [Moussa sees a cat. The cat is lost. Moussa caresses him. He gives him milk]). If the child was able to read the paragraph, they were asked to read a short story (eight sentences). For children who were not able to read the paragraph, they were asked to read five words from a list (e.g., *banane, soleil, farine, and roi*). Children were able to select the words they wished to read. If the child was able to read the words without error, they were given a second opportunity to read a paragraph. For children who were not able to read at least four of five selected words, they were asked to read five letters. Children were able to select the letters they wished to read. If children were able to read all five letters, they were given a second opportunity to read five words. This was not timed; students could take as long as they needed.

If a child was not able to read four of five letters, they were placed in the "Level 1: Beginner" category. Children who were able to read four or more letters but not four words were placed in the "Level 2: Letter" category. Children who were able to read four or more words but not a paragraph were placed in the "Level 3: Word" category. Children who could read a paragraph but not a story were placed in the "Level 4: Paragraph" category. Finally, children who were able to read the story were placed in the "Level 5: Story" category.

Numeracy Groups. The mathematics portion of the ASER includes two parts: (a) number recognition and (b) operations. Based on children's ability to recognize numbers and perform mathematics operations, they were placed in one of five categories: "Level 1: Beginner," "Level 2: Recognition of single-digit numbers," "Level 3: Recognition of two-digit numbers," "Level 4: Subtraction," and "Level 5: Division." Children were first asked to solve subtraction problems. If the child was able to solve at least two subtraction problems, they were asked to solve two division problems. In both subtraction and division tasks, if the child was unable to solve the first problem, they moved on to the second problem. If successful on the second problem, they were able to reattempt the first problem. If the child was unable to solve two subtraction problems, they were asked to read eight two-digit numbers. If the child was unable to read six of the eight numbers presented, they were asked to read eight one-digit numbers. This was not timed.

Children who were unable to read at least six one-digit numbers were placed in the "Level 1: Beginner" category. Children who were able to read six one-digit numbers but unable to read at least six two-digit numbers were placed in the "Level 2: One-digit" category. Children who were able to read at least six two-digit numbers but unable to complete the subtraction operations were placed in the "Level 3: Two-digit" category. Children who

were able to complete subtraction operations but unable to complete division operations were placed in the “Level 4: Subtraction” category. Finally, children who were able to complete division operations were placed in the “Level 5: Division” category.

Functional Literacy Skills

The EGRA is a widely used global assessment measuring early literacy skills (RTI International, 2015). We used the French version of the EGRA, which has been used in Côte d'Ivoire and other Francophone countries in Sub-Saharan Africa (Ball et al., 2022; Gove & Wetterberg, 2011; Jasińska et al., 2022a, 2022b; Sobers et al., 2023; Sprenger-Charolles, 2008; Whitehead et al., 2024) and included the following subtasks: letter reading, word reading, and pseudoword reading. These tasks were previously reported to be reliable (letter-reading Cronbach's $\alpha = 0.88$; word-reading Cronbach's $\alpha = 0.86$; pseudoword-reading Cronbach's $\alpha = 0.863$) and valid in an independent sample of Ivorian primary school students (Sobers et al., 2023). The EGRA was not designed specifically for Ivorian French, but a group of native Ivorian linguists and education researchers confirmed that the Senegalese French EGRA (Gove & Wetterberg, 2011) was appropriate to use in this study. Children were given 60 seconds to complete each task. All children completed the same tasks regardless of age or grade. Total scores were calculated by creating a composite score of subtasks by averaging subtask means.

Letter Reading. Children were asked to identify 100 letters or combinations of letters (e.g., “on”) as quickly and accurately as possible (Gove & Wetterberg, 2011; RTI International, 2015). During a small practice portion of the assessment, the experimenter provided scaffolding and feedback. Items were selected based on frequent use in Ivorian French. Each letter or combination of letters was used once. Items were marked as correct if the child was able to correctly name the letter or identify the letter sound. Incorrect items were marked as such, and experimenters concluded the task if the child failed to correctly read the first 10 items or after 60 seconds (Gove & Wetterberg, 2011; RTI International, 2015).

Word Reading. Children were asked to read 50 familiar French words (e.g., *carte*, *papa*) as quickly and accurately as possible (Gove & Wetterberg, 2011; RTI International, 2015). During the practice portion, the experimenter provided scaffolding and feedback. Incorrect items were marked as such, and experimenters concluded the task if the child failed to correctly read the first 10 words or after 60 seconds (Gove & Wetterberg, 2011; RTI International, 2015).

Pseudoword Reading. Children were asked to read 50 mono- or bisyllabic words that do not exist in French but conform to French syllabic and

phonotactic structure (e.g., *donné, toche*) as quickly and accurately as possible (Gove & Wetterberg, 2011; RTI International, 2015). During the practice portion, the experimenter provided scaffolding and feedback. Incorrect items were marked as such, and experimenters concluded the task if the child failed to correctly read the first 10 words or after 60 seconds (Gove & Wetterberg, 2011; RTI International, 2015). Pseudoword reading assesses the ability to decode novel words. This task accounts for orthographic difficulty; pseudowords were as similar as possible to the difficulty of real French words from the word-reading task.

Functional Numeracy Skills

The Early Grade Math Assessment (EGMA) is a widely used global assessment measuring early numeracy skills (RTI International, 2014). We used the French version of the EGMA, which has been used previously in Côte d'Ivoire (RTI International, 2014; Whitehead et al., 2024) and includes the following subtasks: missing value, addition, and subtraction. For the addition and subtraction tasks, children were given 60 seconds to complete as many items as possible. Children were told that they could skip over items to which they did not know the answers. Total scores were calculated by creating a composite of subtasks by averaging subtask means.

Missing Value. Children were presented with four sets of numbers with one value missing and asked to identify the missing number in the set. Children were given a score out of four based on how many correct responses they provided.

Addition. Children were given 15 single- and double-digit addition problems and asked to complete as many as possible within the given time frame. Children were scored based on the number of correctly completed items.

Subtraction. Children were given 15 single- and double-digit subtraction problems and asked to complete as many as possible within the given time frame. Children were scored based on the number of correctly completed items.

Teacher Burnout at Baseline

Teachers reported on their levels of burnout at baseline using the Maslach Burnout Inventory–Educators Survey (Maslach et al., 1993). The full scale includes 22 questions measuring three domains of burnout (i.e., emotional exhaustion, personal accomplishment, and depersonalization). We found, consistent with previous studies in West Africa (Lee & Wolf, 2019; Oh & Wolf, 2023), that the depersonalization items had very low variation in our sample. Thus, we focused on the other two domains of burnout: emotional exhaustion (nine items, $\alpha = 0.61$) and lack of personal accomplishment (six items, $\alpha = 0.71$).

Teacher Program Knowledge

TaRL Africa created a 20-question multiple-choice questionnaire to evaluate teacher knowledge of the PEC methods. The knowledge questions comprised four sections of five questions each for teaching French (e.g., “What activity should be conducted only by students in the ‘word and paragraph’ level group?” Correct answer: “Fix the error”), teaching math (e.g., “During the activity ‘The circle of numbers,’ the pebbles to take into account after the launch are?” Correct answer: “The stones inside the circles”), ASER administration (e.g., “The ASER test must be performed?” Correct answer: “Individually”), and mentoring (e.g., “What is the role of the sector’s mentor?” Correct answer: “The mentor supports the facilitators in the implementation of PEC activities and guarantees the proper functioning of the program”). The questions consisted of multiple-choice and true or false questions, and the test was scored out of 20 points (mean = 8.4, SD = 2.3, range = 2–14 points at baseline).

Teacher Interviews

Interviews lasted ~45 minutes and were conducted in the schools. The interview protocol explored (a) conversational agent perceptions and barriers to usage, (b) the influence of PEC or technology on teachers’ aspirations, (c) PEC support from colleagues and trainers, and (d) PEC knowledge access and barriers. The U.S.-based research team designed the initial protocol, and the nongovernmental organization and the Ivorian research team contextualized and revised the questions. One member of the Ivorian team conducted the interviews.

School Characteristics

School directors answered a set of questions about the school’s history, management practices, community participation, infrastructure, and human resource and administrative processes. Descriptive results are summarized in the Supplementary Materials in the online version of the journal (see Supplementary Table S1). Factor analyses revealed three underlying factors: school infrastructure, teaching quality, and administration quality (see Supplementary Materials in the online version of the journal for additional details.)

Analytic Approach

Baseline Equivalence

We first conducted a baseline equivalency analysis to test that the randomization yielded treatment and control groups that were statistically equivalent and tested whether the mean values for a set of teacher and student

Table 1
Baseline Descriptive Statistics and Equivalence Between
Treatment and Control Groups

Statistic	Control %/mean	<i>b</i>	SE	<i>p</i> Value
Student % (<i>n</i> = 2,451)				
Child female	50.8	−0.20	0.08	.012*
Child grade (CE1)	50.3	−0.06	0.08	.450
ASER literacy group				
Beginner	45.3	−0.37	0.18	.043*
Letter	44.6	0.15	0.13	.270
Word	5.5	0.44	0.23	.052**
Paragraph	3.5	0.19	0.28	.497
Story	1.2	0.09	0.71	.904
ASER numeracy group				
One-digit	22.9	−0.21	0.14	.138
Two-digit	48.8	−0.07	0.09	.433
Three-digit	25.4	0.26	0.13	.054**
Subtraction	2.6	0.47	0.31	.121
Division	0.3	−1.36	1.12	.225
Teacher (<i>n</i> = 297)				
Female (%)	17.2	0.04	0.05	.421
CE2 (%)	53.1	−0.04	0.06	.517
Self-efficacy	1.98	0.02	0.12	.876
Burnout, emotional exhaustion	2.44	0.05	0.10	.634
Burnout, personal accomplishment	2.03	−0.13	0.11	.257
Perceived social support	4.06	0.09	0.13	.492

Note. ASER = Annual Status of Education Report. Students and teachers were nested in 167 schools. *b* coefficient represents the treatment–control group difference.

p* < .05; *p* < .1.

characteristics differed by treatment group (see Table 1). We found no differences in teacher characteristics at baseline but did find some differences in child characteristics. Specifically, children in the treatment group were less likely to be female, slightly less likely to be in the beginner ASER literacy group, and marginally statistically more likely to be in the word ASER literacy group. Finally, students in the treatment group were marginally statistically more likely to be in the three-digit ASER numeracy group. We controlled for baseline child-level characteristics in our models.

Impact Analysis

Our impact analysis is regression based, with different model specifications and covariates based on the student or teacher level and outcome. For ASER outcomes, given the ordinal nature of the group status, we used

multilevel cumulative link models, which account for ordinal outcome variables while also accounting for nesting. Multilevel proportional odds ordinal regression models predicting children's literacy and numeracy grouping level at endline with random intercepts for classroom were estimated. Thus

$$\text{logit}[P(Y_{ij} \leq k)] = \theta_k - (\beta_0 + \beta_1 X_{ij} + \dots + \beta_p X_{ij} + u_j) \quad (1)$$

where Y_{ij} is the ordinal outcome (e.g., literacy or numeracy level) for student i in classroom j ; θ_k are the thresholds for ordinal categories; β_0 is the fixed intercept; $\beta_1, \dots, \beta_p$ are the coefficients for student- or teacher-level covariates X_{ij} ; and u_j is the random intercept for classroom j .

For EGRA/EGMA outcomes, we used multilevel regression models with an ordinary least squares estimator given the continuous operationalization of scores. Students were nested within schools with classroom fixed effects for both analyses; control variables included child age, sex, and baseline classroom ASER group mean and standard deviation for each respective outcome. Where age data were missing ($n = 524$), we imputed mean age for grade. Thus

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \dots + \beta_p X_{ij} + u_j + \varepsilon_{ij} \quad (2)$$

where Y_{ij} is the continuous EGRA/EGMA score for student i in school j ; β_0 is the fixed intercept; $\beta_1, \dots, \beta_p$ are the coefficients for covariates X_{ij} (student- and classroom-level covariates); u is the random intercept for school j ; and ε_{ij} is the residual error term. To adjust for multiple comparisons, we used sharpened false discovery rate q values. The false discovery rate controls for the proportion of rejections that are "false discoveries" (type I errors) and allows for a comparison of the original p values with sharpened q values (Anderson, 2008).

To test for impact variation by child and teacher characteristics, we added a cross-level interaction term between treatment status and the moderator; for school characteristics, we included an interaction term between treatment status and the school characteristic.

Potential Explanatory Mechanisms

To examine potential mechanisms of program effects using nonexperimental methods, we examined teacher program knowledge for treatment teachers at four time points: before the training as a pretest (fall 2021), immediately after the training (fall 2021), at the end of the first school year (spring 2022), and at the end of the second school year (spring 2023) as immediate and delayed post-tests using multivariate t tests for mean comparisons.

Further, we conducted an embedded qualitative study with a subset of teachers. We used qualitative analysis to analyze (a) the interviews and (b) the chatbot log data; qualitative analysis for the interviews aimed at explaining the quantitative data in a mixed-methods approach (Bowen et al., 2017). Our interview protocol had questions to discover (a) PEC support from colleagues

and trainers, (b) PEC knowledge access and barriers, (c) the influence of PEC or technology on teachers' aspirations, and (d) chatbot perceptions and barriers to usage. We combined the interview data from all stages into a single dataset. We transcribed and translated the interview data from French to English. We then proceeded using an iterative inductive (Clarke & Braun, 2017), consensus-based (Hammer & Berland, 2014) thematic analysis approach to data analysis (Corbin & Strauss, 1990, Corbin & Strauss, 2008). The sixth author reviewed the transcripts and formed initial codes. The second and sixth authors together reviewed the codes and considered the relations between them, categorizing them into themes using the lens of the main research questions. The initial themes were then discussed together with the quantitative results alongside the Ivorian research team and were iteratively modified, with additional codes added as needed based on insights across these discussions.

Results

Descriptive Statistics on Child Learning Outcomes

Literacy ASER scores showed that at baseline the vast majority of children (85%) were in the bottom two learning levels. Specifically, 40% of students scored in the beginner group for literacy, 45% in the letter group (they could identify letters but not read a word), 5% in the word group (they could read a word but not beyond that), 3% in the paragraph group, and only 1% could read a story. For numeracy ASER scores, at baseline nearly three quarters of children were in the bottom two learning levels. Specifically, 23% score in the beginner group, 49% in the level 1 group, 25% in the level 2 group, 3% in the addition/subtraction group, and 0% in the division/multiplication group.

At endline, the proportion of children in higher learning levels increased in both treatment and control groups. There was a high percentage of low scores (i.e., 51.9% of the sample scored <10% correct) on the EGRA at endline, indicating overall low learning levels in our sample. In contrast, the percentage of low scores on the EGMA at endline was low (7.6%). At endline, children from higher ASER baseline groups had higher endline literacy and numeracy EGRA and EGMA scores, respectively (literacy: $r = .27$, $p < .001$; numeracy: $r = .22$, $p < .001$). Boys had higher endline numeracy scores ($M_{\text{boys}} = 42.9$ vs $M_{\text{girls}} = 38.0$; $p < .001$), but no gender differences were observed for endline literacy scores ($M_{\text{boys}} = 16.6$ vs $M_{\text{girls}} = 16.4$; $p = .637$).

Program Impacts

Research Question 1: Impacts on Student Learning

Impacts on ASER Learning Levels. We first assessed impacts on literacy and numeracy outcomes by examining the distribution of children across

Table 2
Impacts on Students and Teachers

Outcome	<i>b</i>	SE	<i>p</i> -value	<i>q</i> -value	<i>d_w</i>
ASER literacy group	0.38	0.14	.008	.013	.133
ASER numeracy group	0.67	0.13	.000	.001	.267
EGRA score	1.53	1.12	.175	.096	.082
EGMA score	1.82	1.21	.135	.096	.081

Note. ASER = Annual Status of Education Report; EGRA = Early Grade Reading Assessment; EGMA = Early Grade Math Assessment. Students and teachers are nested in 167 schools across three regions. EGRA is a composite of subtask scores on letter reading, word reading, and pseudoword reading. EGMA is a composite of subtask scores on missing values, addition, and subtraction. Student models for ASER outcomes were computed using multilevel cumulative link models, controlling for age, grade, gender, and baseline classroom level. Student models for EGRA/EGMA outcomes were computed using multilevel ordinary least squares regression models, controlling for age, grade, gender, and baseline classroom level. *q* Values represent sharpened false discovery rate *q* values (Anderson, 2008). *d_w* is the computed standardized effect size accounting for the clustering of children in schools, computed using the conditional unstandardized regression coefficient (*b*) and the residual variance estimates from the unconditional model at the school and child levels (Hedges, 2009).

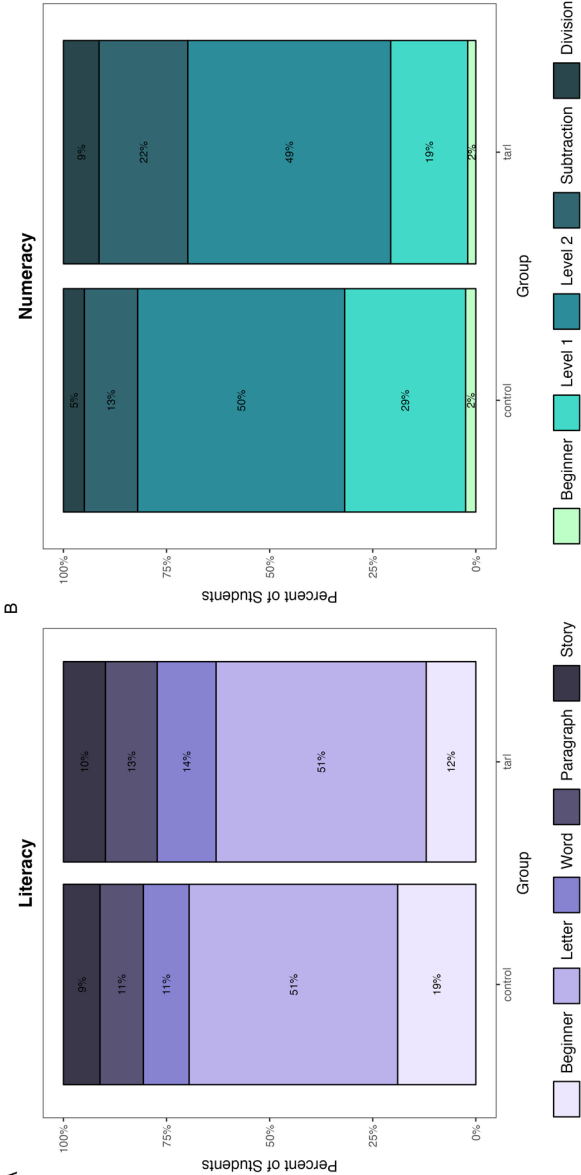
q* < .05; *q* < .001; ****q* < .10.

ASER levels across the three grade levels (see Table 2). We found that PEC + DIA had positive impacts on both ASER literacy (*d_w* = .133, *q* = .013) and numeracy groups (*d_w* = .267, *q* = .001). For literacy, children were 6 percentage points less likely to be in the beginner group for literacy and 3 percentage points more likely to be in the word group, 1 percentage point more likely to be in the paragraph group, and 1 percentage point more likely to be in the story group (the top three groups, respectively). For numeracy, children were 10 percentage points less likely to be in the beginner group (i.e., cannot identify two-digit numbers) and 8 and 2 percentage points more likely to be in the subtraction and division groups, respectively (the top two levels). The distributions are shown in Figure 2. Additional information on how treatment status impacted the odds of being assigned to each learning level is presented in the Supplementary Materials in the online version of the journal.

Impacts on functional literacy and numeracy skills. The impacts on learning levels indicate the levels into which students were categorized as a result of the program. We conducted more extensive assessments of functional literacy skills using EGRA and numeracy skills using EGMA. We found only marginally statistically detectable impacts on functional literacy (*d_w* = .082, *q* = .096) or on numeracy skills (*d_w* = .081, *q* = .096), although coefficients are positive (see Table 2). Importantly, we were underpowered to detect the size of the effects found on EGRA/EGMA. Associations between

A

B



C

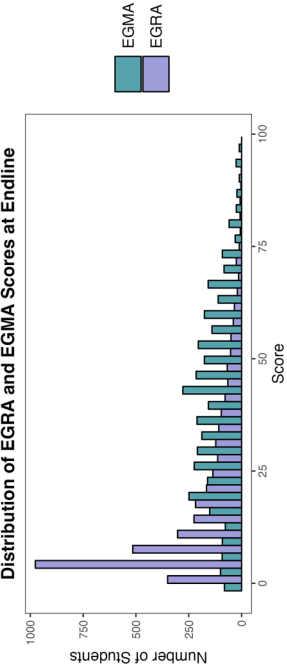


Figure 2. Distribution of classroom-level ASER ability groupings for literacy and numeracy at endline.

learning outcomes and teacher and school factors are reported in Supplementary Materials in the online version of the journal.

EGRA/EGMA and ASER levels were highly correlated (EGRA: $r = .86$, $p < .001$; EGMA: $r = .66$, $p < .001$). However, only impacts on ASER were statistically detectable. Both EGRA and EGMA scores were generally lower than ASER levels. For example, for numeracy skills, children who were categorized at the subtraction level on the ASER scored, on average, only 55% on the EGMA (i.e., sequencing, addition, and subtraction composite) and only scored 35% of items correct on the subtraction subtest of the EGMA. Similarly, for reading skills, children who were categorized as story-level readers on the ASER scored, on average, only 50% on the EGRA (i.e., letter, word, and nonword reading composite; see Figure 3).

Further analyses examining the impacts of treatment on sequencing, addition, and subtraction separately revealed a significant treatment effect on sequencing ($b = 0.039$, $SE = 0.017$, $p = .020$) but not on addition ($b = 0.010$, $SE = 0.015$, $p = .484$) or subtraction ($b = 0.006$, $SE = 0.012$, $p = .612$) EGMA subtests. Similarly, examining treatment impacts on letter, word, and nonword reading separately revealed a marginally significant impact on letter reading only ($b = 0.020$, $SE = 0.011$, $p = .064$) but not on word ($b = 0.013$, $SE = 0.013$, $p = .340$) or nonword ($b = 0.008$, $SE = 0.011$, $p = .468$) EGRA subtests. We conducted additional analyses to examine whether decoding accuracy (number of correct items out of number of attempted items) and reading fluency (number of attempted items in 60 seconds) may inform the discrepancy between EGRA/EGMA and ASER scores. There was a significant effect of treatment on EGRA accuracy ($b = 0.038$, $SE = 0.018$, $p = .0392$) but not fluency ($b = 0.695$, $SE = 0.661$, $p = .295$). Parallel analyses for EGMA (number of correct answers out of attempted problems and number of attempted problems in 60 seconds) did not reveal any significant treatment effects.

Research Question 2: Impact Variation by Student, Teacher, Classroom, and School Characteristics

In exploratory analyses, we examined impact variation by child (sex), teacher (baseline burnout), classroom (average baseline classroom skill level and baseline classroom skill heterogeneity), and school quality (infrastructure, teaching, and administration). In separate models, we added interaction terms between our specified moderators and treatment status to the impact models on ASER levels. We found no evidence that the program's impacts differed based on the child and classroom factors examined (see Supplementary Table S3 in the online version of the journal). Additionally, we did not find that PEC + DIA reduced (nor increased) learner heterogeneity for literacy outcomes, although PEC + DIA did increase learner heterogeneity for numeracy outcomes (see Supplementary Materials in the online version of the journal).

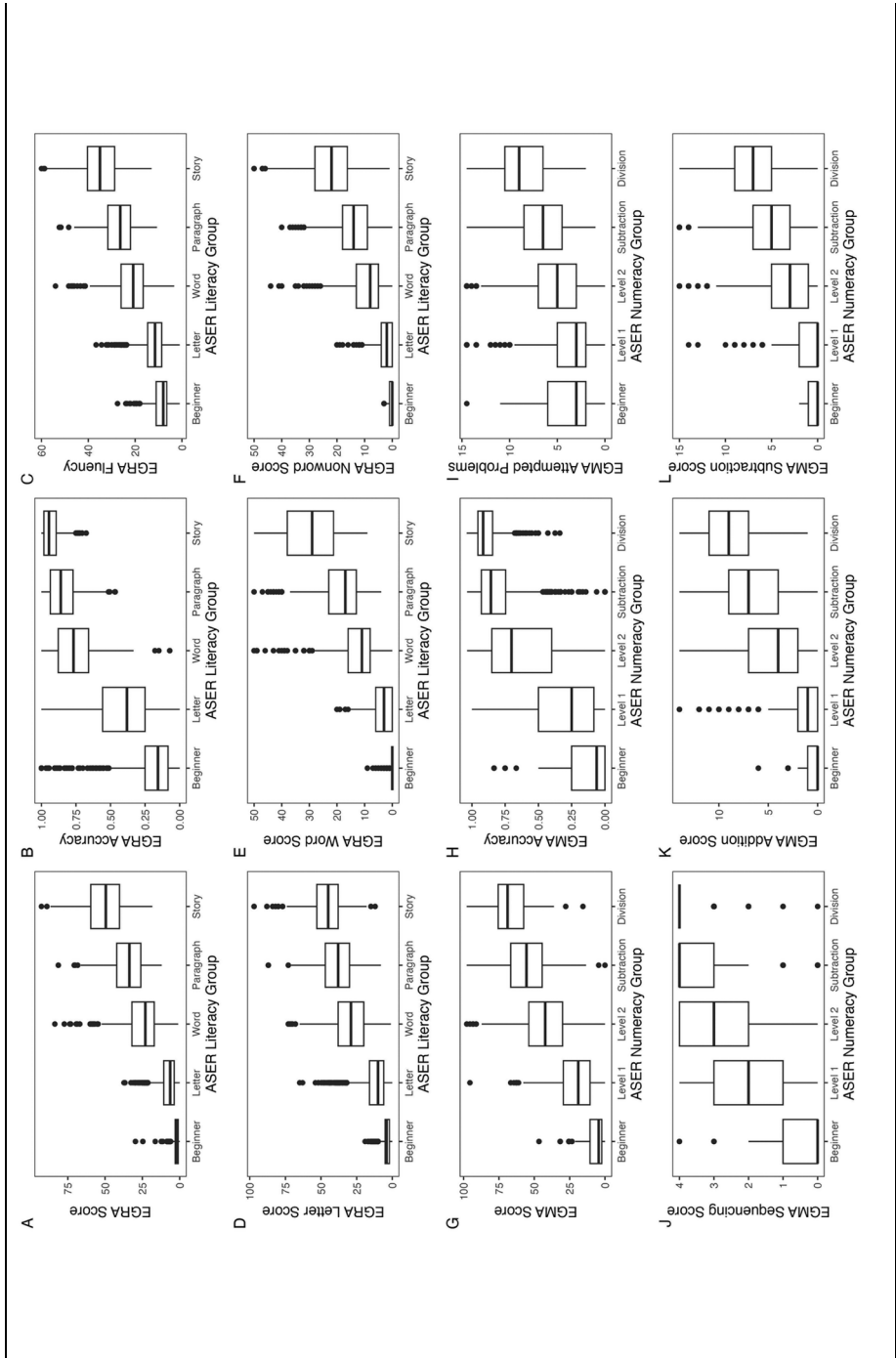


Figure 3. EGRA and EGMA reading scores by ASER ability groupings.

Note. ASER = Annual Status of Education Report; EGRA = Early Grade Reading Assessment; EGMA = Early Grade Math Assessment.

We did find significant interaction effects by teacher burnout. The impact of PEC + DIA on numeracy skills was significantly greater for teachers who reported a lower level of burnout, specifically lower levels of lack of personal accomplishment (Treatment status $\times$ burnout – lack of personal accomplishment: $b = -0.259$, $SE = 0.112$, $p = .021$). Treatment impacts of PEC + DIA on student outcomes (ASER level) were nearly half in classrooms where teachers reported high burnout at baseline (+ 1 SD above mean: $b = 0.525$, $SE = 0.185$, $p = .004$) compared with classrooms where teachers reported low burnout (–1 SD below mean: $b = 1.027$, $SE = 0.182$, $p < .001$).

To examine whether school quality factors moderated treatment impacts, we first conducted a factor analysis of our school quality variables, including variables related to school infrastructure (e.g., building quality, electricity, latrines, water source, canteen, and school supplies), teaching force (e.g., teacher experience, qualifications, training, compensation, enthusiasm, and absences), educational resources and materials, class size, and school administration (e.g., school has a written code of conduct, staff performance evaluation, PD activities, and school inspector visits). Factor analyses and results are reported in the Supplementary Materials in the online version of the journal. We used the resulting three factors (i.e., infrastructure, teaching, and administration) in a moderation analysis.

School infrastructure moderated treatment impacts (Treatment status $\times$ infrastructure – ASER literacy group: $b = 0.395$, $SE = 0.216$, $p = .068$; EGRA: $b = 3.869$, $SE = 1.522$, $p = .012$; ASER numeracy group: $b = 0.550$, $SE = 0.187$, $p = .003$; EGMA: $b = 3.310$, $SE = 1.757$, $p = .062$). Treatment impacts were largest in schools with better infrastructure. Teaching quality also moderated PEC + DIA treatment impacts on literacy scores (Treatment status $\times$ teaching quality – ASER literacy group: $b = -0.540$, $SE = 0.207$, $p = .009$). Treatment impacts were largest in schools with lower-quality teaching. Administration quality was not found to be related to literacy or numeracy scores nor moderated treatment impacts.

Potential Mechanisms

Research Question 3: Teacher Program Knowledge Over Time

Program knowledge was assessed for treatment-group teachers. Knowledge of PEC + DIA increased significantly between the baseline and first assessment (immediately after the teacher training; $t(119) = 25.15$, $p < .001$). Program knowledge declined at the next assessment at the end of that academic year (midline vs baseline after assessment; $t(124) = -19.01$, $p < .001$) and remained at the same level 1 year later at endline (vs midline: $t(142) = -1.37$, $p = .174$). Although knowledge declined from baseline (post-training assessment) to endline, teacher program knowledge was overall higher at endline than it was prior to PEC training ($t(116) = 7.55$, $p < .001$), indicating that teachers retained some knowledge.

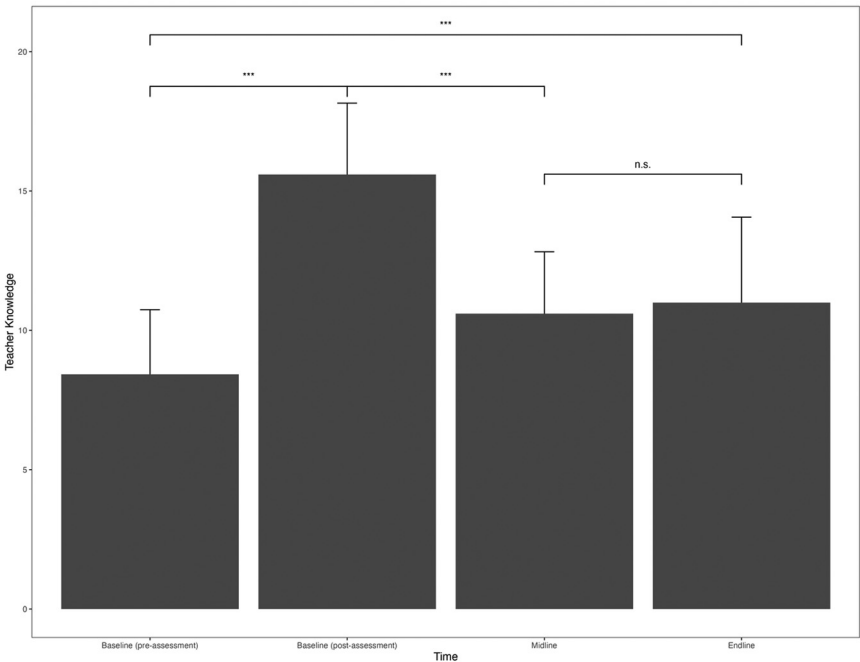


Figure 4. Treatment teachers' knowledge levels over time.

Note. Multivariate *t*-tests showed statistically significant differences between baseline and post-training ($p < .001$) and post-training and midline ($p < .001$) and no statistical differences between midline and endline ($p = .174$).

Teacher program knowledge immediately after completing the PEC training (baseline post-training assessment) was significantly associated with student outcomes at endline (ASER literacy level: $b = 0.05$, $SE = 0.02$), $p = .014$; ASER numeracy level: $b = 0.041$, $SE = 0.016$, $p = .014$; EGRA: $b = 0.85$, $SE = 0.31$, $p = .007$; EGMA: $b = 0.84$, $SE = 0.37$, $p = .024$). Program knowledge at midline and endline was not associated with student outcomes (see Figure 4).

Additional insights into teacher program knowledge may be gleaned from the logs of the DIA chatbot, where 146 unique questions were recorded over the duration of the project. The most common question asked was about how to implement the PEC French activities (11.6%), how to implement the ASER assessment (11.6%), and how to prepare session plans (9.6%), all critical elements of program implementation that are central to the success of the program. Regarding the ASER assessment, the crux of the program used to group children by learning levels, teachers asked questions about grouping students after the ASER test. Teachers also asked questions when they found students

who did not fit the traditional TaRL model (e.g., “Where to classify the students who do not recognize the numbers but manage to divide?”). Teachers also asked questions about specifics of ASER implementation. For example, teachers wanted to know if they should use the same instrument for baseline and midline for a particular student (e.g., “Should we use the same number of the ASER tool for the student in this midterm [assessment?]”). Lastly, teachers asked questions about the different documents to implement and record the ASER test (e.g., “I want to know how to complete the ASER end-test document?”). These all point to knowledge gaps that teachers had in implementing the program with high fidelity that might need to be further supported with training or coaching.

Research Question 4: Qualitative Results—Perceptions and Barriers of the Program

In semistructured interviews, teachers reported on a wide range of experiences in their training and implementation of the PEC + DIA program, including positive experiences and challenges integrating the new practices with traditional teaching methods and workloads. Our analysis uncovered four main themes that shed light on the quantitative results: teachers’ positive perceptions of PEC for students’ learning, the tension between PEC and other required teaching methods, lack of training support and resources, and the critical value of support from their colleagues. All themes were supported by data from the majority of teachers.

Positive Perceptions of the Program on Student Growth. Teachers thought the program activities helped students be more actively involved in their learning and reduced barriers between teachers and students. One teacher (J3) stated:

[PEC] is face to face, that is to say that the teacher is no longer the master, he is no longer on his platform, there is no office and you are together [with children], you play, there is no longer any difference because the teacher puts himself on the same level as the children and they have a certain confidence in you. . . . they forget your rank and then they break free by talking.

In addition, teachers thought that the program allowed children to play and learn more freely than the traditional curriculum, such as in this statement (J7):

It’s learning through play because I myself like to play with my students, and so this means that the children have no barriers with us; we almost sit on the floor . . . and also it is a way of animating [them].

Teachers also described seeing visible improvements in children’s learning. One teacher (J9) stated, “Today thanks to PEC, the children who could not

read before, manage to read syllables, letters, words and many arrive to make sentences.” A second teacher (J12) described how PEC was more adaptable to children’s individual levels:

Before, we had a fixed program . . . without thinking about the level of the children. With the PEC, it allows us to go back to take a little of everything that the children missed at the base. . . . When you see children with PEC today, they manage to assimilate certain things. . . . There have been progressions: at the beginning, there were small documents that they could not read, . . . but thanks to the PEC, they can read certain documents; there are words they understand and manage to read.

Overall, teachers were in agreement that PEC provided positive experiences in the classroom and led to improved learner outcomes. This finding is particularly important given the simultaneous feelings of tension and work overload described below.

Tensions Between PEC and Traditional Teaching Methods. Despite perceiving significant improvements in outcomes, a primary challenge for teachers was a feeling of disconnect between the PEC methods and their classical training. Teachers viewed the new program as a disruption to the normal curriculum. One of the reasons was that the activities within the PEC program involved teachers doing playful, student-centric activities with students instead of the more typical teacher-led instruction from the front of the classroom. Teachers thought it was challenging to adapt PEC activities to their teaching style and context. As teacher J2 noted:

What was difficult was above all the environment needed to carry out the activities. Because the framework here does not conform to the practice of PEC activities since it is an activity that is done outside and we . . . have no place set up for this, no place to make the children sit . . . with the rain, it bothers us in the practice of the activities.

Besides the challenging environment for this new teaching approach, teachers stated that the PEC program was too time-consuming when implemented beside traditional teaching, as illustrated by teacher J9, who had expected that government-led changes to their teaching practice should make the teachers’ jobs easier: “The PEC has not lightened our workload. . . . Classical education has its methodology, and the PEC has its methods, but the two compete for [teacher] well-being and improving children’s performance.”

This highlights the fact that teachers viewed the daily PEC practice as an additional activity they had to implement rather than as a way to support student learning for the existing curriculum. As teacher J7 said:

What was difficult for me was always preparing lesson plans. . . . we have 8 to 10 sheets per day, and we still have to make a weekly lesson

plan for PEC. Then we come to class for the classic classes, and at 10:15 a.m. we leave for PEC classes.

In another example, teacher J11 reported, “We already have our files for the classic lessons and then we were asked to make this [additional] kind of lesson plan.” This feeling of PEC being extra work made teachers feel that the program was imposed on them.

These perceptions of additional workload outside of teaching time may have contributed to teacher burnout and were mitigated only by their appreciation of the playful, barrier-breaking nature of the PEC activities that improved their enjoyment of being in the classroom.

Lack of Support for Training and Implementation. In addition to the tension between the two pedagogic programs, teachers reported feeling overloaded with learning new PEC practices, resource allocation, and incorporating PEC into their classrooms. Teachers said that the training they received for PEC felt discontinuous and did not prepare them enough for implementing the program well over the year in comparison with their traditional teaching. Teachers were taught the entire PEC curriculum in a week-long training, which involved many activities in both French and math, both of which were very different from traditional classroom teaching. The teachers expressed a desire for additional training to help them understand the program more deeply. As one teacher (J12) stated:

We received training that lasted two weeks, there we cannot retain everything at the same time and therefore additional training and practice are needed. . . . it helps you to work more and to better understand what you did not understand during the training.

This concern about insufficient training was widely shared by teachers in our sample and in all informal conversations we held with teachers throughout the study.

Teachers also reported that after the training, implementation challenges remained. For instance, teachers described not having all the required material or financial resources to carry out the activities in their session plan despite the intent of the program to ensure that all schools had these materials. While government schools generally have upward of 60 students per class, some teachers mentioned that the implementation of PEC further increased the overload in classrooms. Specifically in rural schools, students were initially at very low literacy levels, as assessed by the ASER test. Thus, level 1 literacy classes needed to accommodate a very large number of students. Teacher J3 illustrated this effect in her class: “Well, here the difficulty is in terms of numbers because we have too many students per level. . . . at the beginning at level 1 there were more than 80 students.” Directors sometimes addressed this issue by scheduling multiple lower-level classes and combining the upper levels,

which typically were quite small, into one room with a single teacher. This workaround alleviated some aspects of the overload but often introduced other issues such as lack of training on how to support groups at two different PEC levels at the same time. Teachers mentioned that these challenges led to a lack of motivation to implement the training program.

Support from Mentors and Colleagues. Finally, access to mentors for guidance and for answering questions related to PEC was an important component of overcoming the above-mentioned challenges. Some found help from mentors to be highly relevant and timely. For example, teacher J4 noted, "It's the counselor who spends a lot. He is our physical and direct point of contact. When we have difficulties, we turn to him." Teachers also frequently sought help from their colleagues, particularly when mentors were not available. Teacher J8 said, "Between us colleagues from the same school, we help each other with the difficulties we encounter."

At the same time, many teachers expressed frustration that mentors did not always respond to them and that mentorship was inconsistent. One of the difficulties was cancellations or interruptions of the weekly meetings where teachers and the director were supposed to gather to discuss planning the PEC activities for the week. The PEC administrators recommended that teachers convene every Friday and discuss the challenges of the current week while planning their PEC activities for the upcoming week. However, teachers often would have administrative meetings during these days, interrupting the PEC workflow. As one teacher (J11) said, "[We meet] at the end of each week, but it can vary since we do it with the Directors and they are constantly called in for inspection." The rural location of the schools meant traveling up to an hour into town for administrative meetings, which would disrupt teaching activities and caused challenges. Teacher J7 said, "Very often we may not finish the implementation in the week and then there are meetings in the weeks that have nothing to do with the lessons. . . . they do not take these meetings into account."

Teachers also complained of a lack of responsiveness from mentors. They acknowledged that this may have been because of the scarcity of mentors available. The structure of mentorship in the program was hierarchical. Although there were formally designated pedagogic advisors, as one teacher (J12) stated, "During the various meetings, the director gives the floor to everyone and everyone tells what he encounters as a problem. . . . we present the difficulties and the solutions at the same time. . . . we make a small report of all that we have as a problem." Advisors also were not always able to visit the schools due to infrastructure challenges, leading to infrequent and inconsistent mentorship. One teacher (J2) said, "The adviser makes the effort to come at least once a month. . . . The inspector in particular has not yet arrived at our school."

Finally, teachers had positive views of the questions and stories shared through the DIA platform, which was intended to act as an additional support

when mentors were not available. Teachers highlighted how questions answered through the social media chatbot were a supportive resource. One teacher (J12) stated, “When I have difficulties, I send them to the DIA platform and in return DIA gives me answers to be able to achieve my objectives.” Teachers also found value in learning from each other through stories shared on the chatbot. Teachers were able to make connections with other community members through the stories; they also mentioned that the stories inspired them to implement PEC. Teacher J9 summarized the benefits of the technology:

The anecdotes that some colleagues share [on DIA], which are often funny. In PEC, there is also informal dialogue; therefore, usually, with the anecdotes of certain colleagues [from DIA], you can inspire yourself to tell your story to the children, too. The little stories and anecdotes are important because you can draw inspiration from them to move forward.

Discussion

We presented findings from an impact evaluation of a government-led targeted instruction program in Côte d'Ivoire implemented over 2 years with a supplemental electronic support program. Our results show that PEC may have the potential to improve children's foundational literacy and numeracy skills, but the size of the impact was small and concentrated among the lowest learners. While these effect sizes are in line with those of other targeted instruction programs designed to be implemented at scale by government personnel (e.g., Banerjee et al., 2016), given very low learning outcomes at baseline in our sample and in Cote d'Ivoire in general, whether these impacts are practically meaningful is unclear. We further documented several potential barriers teachers faced to implement the program and suggested that future iterations of the program could be larger if these barriers are addressed. Despite targeted instruction programs being categorized as a “great buy” for improving children's learning outcomes in low- and middle-income countries (Banerjee et al., 2023, p. 20), our findings suggest that there is much more to be learned on how to optimize the impacts of such programs through better support of teachers.

Moving beyond the question of whether the program improved children's academic learning levels—as is commonly assessed in targeted instruction program evaluations—we also considered impacts on more extensive measures of functional literacy and numeracy skills and teacher outcomes. At endline, most of the children (>60%) were still categorized as not being able to read at the word level after 2 years of the program. More important, when examining more extensive, in-depth measures of functional literacy and numeracy skills, we found weaker evidence of effects and only about

half the size in magnitude and only marginally statistically significant. The discrepancy in impacts on literacy and numeracy levels (ASER) and more extensive assessments of functional literacy and numeracy skills (EGRA/EGMA) is worth considering further. We note that functional literacy and numeracy scores were generally lower than ASER levels (e.g., children who were categorized as story-level readers on the ASER were only able to read 50% of letters, words, and nonwords on the EGRA). This suggests that ASER levels may not adequately represent learners' skills and may in fact overestimate learners' true skills.

A key difference between the ASER and EGRA is ASER's lack of a measure of reading fluency, an important component of skilled reading. Detailed analysis of impacts on the EGRA suggest that the PEC + DIA program may have improved decoding skills only but not reading fluency. Well-established models of reading (i.e., a simple view of reading) posit that skilled reading comprehension is a product of both decoding ability and oral language skills (Gough & Tunmer, 1986). Reading fluency mediates the link between decoding and reading comprehension (e.g., Cadime et al., 2017; Pikulski & Chard, 2005; Silverman et al., 2013) and thus constitutes an important path to fluent and proficient reading.

Results also differed between the ASER mathematics assessment and the EGMA measure of numeracy. In fact, this discrepancy presents a complex challenge of interpretation because the ASER assessment contains relatively more difficult mathematics problems (e.g., double-digit subtraction with borrowing and division with remainders) than the EGMA (whose most advanced questions are single-digit addition and subtraction problems). To understand this finding, the research team reviewed observations given by the enumerators about the assessments. Observations indicated that in the ASER assessment, students used pencil and paper to perform written calculations and review their responses. In the EGMA assessment, given the known time constraint and the simpler problems, many students appeared to view the test as a challenge of their mental math abilities and attempted the problems quickly and without writing. This led to frequent errors on routine calculations that may (similarly to our EGRA findings) be due to a lack of mathematical fluency rather than a lack of basic computational skills or facts (Russell, 2000). Validation studies examining the ASER assessments have been relatively limited in the existing literature; future studies could consider a Rasch modeling approach as a way to combine both assessments. See the Supplementary Materials in the online version of the journal for a review of use of ASER and EGRA and a discussion of the validity of the ASER.

Overall, this suggests that future evaluations of targeted instruction programs also should include more extensive measures of functional skills in addition to learning levels to understand more fully the practical impacts (or lack thereof, potentially) of such programs. It also has implications for the types of activities included in such programs, suggesting a need for

more emphasis on fluency once accuracy has been achieved. This finding is a very important addition to the literature, particularly given that nine in 10 children in Sub-Saharan Africa complete primary school without being able to read a simple sentence (World Bank, 2022). The findings elevate implementation challenges that should be addressed to help such programs reach their potential.

Interestingly, we did not find any differences in treatment effects based on child gender, grade level (primary 3 vs primary 4), and baseline classroom level or heterogeneity. This is encouraging from an equity perspective—that is, inequalities did not increase among the targeted, relatively disadvantaged population. Although the program was designed for heterogeneous classrooms, we did not find greater impacts for classrooms with a wider range of learner skills at baseline. However, the program increased classroom heterogeneity (for numeracy specifically), suggesting that further longitudinal research should carefully examine whether impacts are equitable over time. Attention to who benefits from educational programs, as well as how to ensure that the most marginalized benefit, are critical areas for research to continue to consider.

Importantly, we found that teacher and school factors moderated treatment impacts. These exploratory analyses showed that, first, program impacts were significantly smaller in classrooms where a teacher reported high level of burnout at baseline, suggesting that teachers with lower burnout may have had more psychological resources to be able to implement the program effectively. Our findings are in line with previous research in the region showing negative effects of teacher burnout on children's learning (e.g., Oh & Wolf, 2023) and extend that research to highlight the potential role burnout can have in hampering efforts to improve learning outcomes. Second, impacts were larger in schools with better infrastructure, which may be a possible mechanism for increased attendance and thus increased exposure to the program. Third, impacts were larger in schools with lower teaching quality at baseline (a composite of teacher experience, qualifications, training, compensation, enthusiasm, and absences). This is counterintuitive but perhaps points to less qualified and experienced teachers being more willing to embrace new approaches and thus implement the program with higher fidelity. Alternatively, this may suggest that talented teachers are already implementing higher-quality teaching practices and that their students benefit less from new approaches. Additional research is needed to replicate and understand these findings further.

Finally, the communities in which this study took place were multilingual. It is probable that children's early language experiences (i.e., whether they are monolingual, bilingual, or multilingual or whether they speak French at home) contribute to their academic outcomes. An examination of home language environments was outside the scope of this study, but previous research on a TaRL intervention in Morocco did suggest that learning in a first or second

language matters. Binaoui et al. (2023) found that children had greater improvements in Arabic reading, their first language, than in French reading, their second language. Significant improvements in French were limited to children who already had basic literacy in French, suggesting that the language(s) children speak and the language(s) in which TaRL is implemented are important considerations if the goal is to improve reading outcomes.

Improving Targeted Instruction Programs Through Better Support of Teachers

An important finding from Angrist and Meager's (2023) meta-analysis was that targeted instruction programs implemented by teachers had much smaller effect sizes than those implemented by volunteers. It is possible that volunteers are more motivated to implement the programs with high fidelity and see this as their sole task as volunteer educators. For programs to be sustainable and scalable, it will be critical to train teachers to deliver programs effectively, integrate them into existing practice, and continuously inform the implementation with evidence.

The qualitative findings shed light on some of the potential challenges that teachers faced. For example, teachers viewed the PEC program as extra work as opposed to transforming the way they teach and part of what they do each day, which also may shed light on findings from Angrist and Meager's (2023) paper. This suggests that more effort to support teachers in using this program as part of the school day and in aligning its implementation with the curricular material teachers need to get through may help ensure greater fidelity of implementation for future programs. Further, despite having positive perceptions of the program and seeing it work to improve learning for some students, teachers overwhelmingly reported that they saw the PEC program as an addition to their existing workload (rather than a substitution to other approaches) and reported tension between and some resistance to the program methods with traditional teaching methods. They reported not having enough training or materials to implement the activities and were resistant to the idea of sitting on the floor with children during the PEC activities, an approach very different from traditional teacher-led upfront instruction. Indeed, earlier implementations of the TaRL program in India between 2008 and 2010 in Bihar and Uttara-Khand states reported disappointing results when the program was implemented by government teachers during the school year due to low compliance and teachers' perceptions that the program competed with other priorities (Beery, 2017).

Another contribution of this study was to assess teacher knowledge of the PEC program implemented over the course of 2 years after the training. Indeed, pedagogic content knowledge is a core part of teacher education and performance (see, e.g., Grossman, 2009; Grossman & Thompson, 2008) yet is rarely assessed in impact evaluations. Our results showed that the initial

training did improve teachers' knowledge of the program activities and its purpose but that program knowledge declined over time. Teachers did not have additional follow-up training during the first year of the program but did receive a refresher training at the start of year 2. Importantly, these declines in knowledge were in the context of teachers often receiving visits from mentors and supervisors and having access to the DIA chatbot to ask questions. Moreover, teacher program knowledge immediately after the initial training was significantly related to students' learning outcomes. This points to a critical area of need for ongoing PD that is effective in future iterations of this program, including future research on targeted instruction programs that examine the validity of such teacher program knowledge measures.

The findings, along with those from another study with the teachers in our sample (Whitehead et al., 2025), point to several policy implications for targeted interventions in low- and middle-income countries. These include addressing declines in teacher knowledge over time with refresher trainings and ongoing coaching targeted toward solidifying knowledge and practice; aligning the program and training with the existing curriculum so that teachers do not view it as an additional workload; and teaching components of the science of learning to teachers to ensure that aspects of skilled reading, such as decoding, fluency, and comprehension, are familiar and understood deeply in terms of how certain targeted instruction activities support specific skill building. Teachers then would be able to appropriately select PEC activities based on students' skill gaps. This can foster teacher autonomy, a key area of concern teachers raised (Whitehead et al., 2025), better positioning them to implement the program as they see best fit for each student.

Limitations and Conclusions

Our study has several limitations that should be considered when interpreting the results. First, we were unable to conduct classroom observations or collect any systematic data on the program implementation quality for the full sample of teachers, limiting our ability to understand whether and how the PEC program changed the quality of classroom instruction, a key mediating factor. Second, teacher burnout at baseline had acceptable but low levels of internal consistency, suggesting that a bigger focus on measuring teacher burnout in this context is warranted and important given the structural challenges teachers face. Third, we did not have student attendance data and were unable to conduct treatment on the treated analyses to understand how program exposure may have moderated child-level impacts. This is a big limitation given how important this factor is in explaining impacts of targeted instruction programs in other studies (Angrist & Meager, 2023). Fourth, we collected student endline data nearly 2 years after the initial training was conducted with teachers and were unable to collect data on students in the

first year of the program. Thus, we have only a limited picture of impacts in the year of the training.

Fifth, we were unable to track individual children over time and instead tracked *classrooms*. Thus, we did not have information on impacts on students in the first versus second year of the program's implementation, even if they were exposed for more than 1 year. Further, we were unable to examine whether there was differential attrition between treatment and control groups. It is possible that the treatment could have had an impact on attrition, for example, by increasing the likelihood that weaker students stayed in school given the additional support they received. Finally, there was an imbalance at baseline on some child characteristics across the two treatment arms. We addressed this by controlling for these baseline variables in our impact analysis, but this still was a notable limitation.

Nonetheless, our findings confirmed that there is promise in targeted instruction programs to improve learning outcomes in primary school but also suggested that further implementation research is needed to understand how to optimize these programs, particularly considering very low learning levels in many of the countries in which they are implemented (Angrist & Pischke, 2021). For example, are the impacts observed in studies to date truly large enough to help students obtain functional literacy and numeracy skills by the end of primary school? Are programs implemented by volunteers sustainable and scalable to transform education systems at national levels, and if not, what more can be done to ensure that teachers are properly supported to implement targeted instruction programs in more effective ways? These questions are critical considering that teacher-led targeted instruction programs in formal school settings—the model of program delivery in Côte d'Ivoire—have much smaller impacts (Angrist & Meager, 2023).

Targeted instruction programs have been cited as a cost-effective approach to addressing global learning, but data on program costs are sparse. Estimates of the cost per child in Côte d'Ivoire based on the pilot phase of the PEC program were found to be twice as much as a similar program in Ghana, suggesting that scale-up of the PEC program in Côte d'Ivoire may be costly compared with similar programs in other, more densely populated contexts (Innovation for Poverty Action Right-Fit Evidence Unit, 2020). The Brookings Institute's Real-Time Scaling Lab has conducted a cost analysis of program scale-up, but cost per child estimates have not been reported externally (Curtiss Wyss & Perlman Robinson, 2021). The cost study of the PEC program identified training as the most expensive component, with training of teachers and mentors constituting 80% of total PEC implementation expenses in 2020.

Our results showed that teachers' knowledge of the PEC program predicted students' learning outcomes, but teachers' knowledge diminished over the duration of the implementation, and teachers frequently cited the need for longer and more in-depth training; balancing the need for more training with the cost of the program will be a challenge. In Côte d'Ivoire, the

Ivorian Ministry of National Education and Literacy has determined that implementing the PEC program at scale is only affordable with hybrid training models that integrate digital materials (Oulai, 2021, as cited in Curtiss Wyss & Perlman Robinson, 2021), of which the DIA platform can play an integral component. Indeed, the Real-Time Scaling Lab's report identified technology-enabled hybrid training as a requirement for scaling PEC nationally. The government has now trained 7,000 teachers on this approach and is scaling up the delivery of PEC, albeit without the DIA platform.

While promising, many questions remain about targeted instruction programs if they are to reach the transformational potential for education that many hope. In future research, we recommend delving into and distinguishing the specific components of targeted instruction programs that contribute to learning outcomes. It also will be important to examine how children's home environments (including food insecurity, quality of caregiving, language background, and resources for learning) influence learning and school attendance (and hence their exposure to school-based interventions) to understand the extent to which home factors may moderate the success of targeted instruction programs.

Acknowledgments

This article reflects contributions from many organizations and individuals. We thank the dedicated staff at Innovations from Poverty Action Cote d'Ivoire, including Romuald Anago, Romaric Ekpinda, Armand Delga, Bienvenu Gbehe, and Arthur Sagot-Duvaurox and the data-collection supervisors and enumerators. We also thank the Ivorian Ministry of National Education and Literacy for their support of this project and TaRL Africa and the Jacobs Foundation as thought partners throughout the study. We also thank the dedicated teachers, children, and families who participated in this study.

ORCID iDs

Sharon Wolf  <https://orcid.org/0000-0002-8076-8399>

Kaja K. Jasińska  <https://orcid.org/0000-0002-8851-1627>

Samuel Kembou  <https://orcid.org/0000-0001-6197-8625>

Julianna Corbett  <https://orcid.org/0009-0002-4603-2625>

Funding

The Child Learning and Education Facility, SUCDEN Foundation, and the Chocolonely Foundation provided direct support for this project.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Adil, F., Nazir, R., & Akhtar, M. (2022, September). Investigating the impact on learning outcomes through the use of EdTech during COVID-19: Evidence from an RCT in the Punjab province of Pakistan. *Frontiers in Education*, 7, 993265. <https://doi.org/10.53832/edtechhub.0067>
- Albán Conto, C. (2021). *Time to teach: Teacher attendance and time on task in primary schools*. UNICEF.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495. <https://doi.org/10.1198/016214508000000841>
- Angrist, N., & Meager, R. (2023). *Implementation matters: Generalizing treatment effects in education* (EdWorkingPaper No. 23-802). Annenberg Institute at Brown University.
- Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592(7854), 403–408. <https://doi.org/10.1038/s41586-021-03323-7>
- Angrist, N., Ainomugisha, M., Bathena, S. P., Bergman, P., Crossley, C., Cullen, C., Letsomo, T., Matsheng, M., Panti, R. M., Sabarwal, S., & Sullivan, T. (2023). *Building resilient education systems: Evidence from large-scale randomized trials in five countries* (NBER Working Paper No. w31208). National Bureau of Economic Research.
- Ball, M. C., Bhattacharya, J., Zhao, H., Akpé, H., Brogno, S., & Jasińska, K. K. (2024). Effective bilingual education in Francophone West Africa: Constraints and possibilities. *International Journal of Bilingual Education and Bilingualism*, 27(6), 821–835. <https://doi.org/10.1080/13670050.2023.2290482>
- Ball, M. C., Curran, E., Tanoh, F., Akpé, H., Nematova, S., & Jasińska, K. K. (2022). Learning to read in environments with high risk of illiteracy: The role of bilingualism and bilingual education in supporting reading. *Journal of Educational Psychology*, 114(5), 1156. <https://doi.org/10.1037/edu0000723>
- Banerjee, A., Andrab, T., Banerji, R., Dynarski, S., Glennerster, R., Grantham-Mcgregor, S., Muralidharan, K., Piper, B., Saavedra Chanduvi, J., Yoshikawa, H., Ruto, S., & Schmelkes, S. (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?* World Bank Group. Retrieved February 28, 2026. <https://policycommons.net/artifacts/4310772/2023-cost-effective-approaches-to-improve-global-learning/5121024/>
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of “Teaching at the Right Level” in India* (NBER Working Paper No. 22746). National Bureau of Economic Research.
- Banerjee, R., & Chavan, M. (2016). Improving literacy and math instruction at scale in India’s primary schools: The case of Pratham’s Read India program. *Journal of Educational Change*, 17(4), 453–475. <https://doi.org/10.1007/s10833-016-9285-5>
- Banu Vagh, S. (2012, October). Validating the ASER testing tools: Comparisons with reading fluency measures and the read India measures. *Annual Status of Education Report*. https://img.asercentre.org/docs/Aser%20survey/Tools%20validating_the_aser_testing_tools__oct_2012__2.pdf
- Beery, S. (2017). *Teaching at the right level. Evidence in Practice*. Pratham Education Foundation.

- Beg, S., Fitzpatrick, A., & Lucas, A. M. (2021). *Improving public sector service delivery: The importance of management*. IZA Institute of Labor Economics. https://conference.iza.org/conference_files/edu_2021/lucas_a26750.pdf
- Berne, J. S., Jacob, B. A., Weiland, C., & Strunk, K. O. (2025). *The impacts of grade retention policy with minimal retention* (NBER Working Paper No. w33764). National Bureau of Economic Research.
- Binaoui, A., Moubtassime, M., & Belfakir, L. (2023). The effectiveness of the TaRL approach on Moroccan pupils' mathematics, Arabic, and French reading competencies. *International Journal of Education and Management Engineering*, 13(3), 1–10. <https://doi.org/10.5815/ijeme.2023.03.01>
- Bold, T., Kimenyi, M., Mwabu, G., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, 168, 1–20. <https://doi.org/10.1016/j.jpubeco.2018.08.007>
- Bowen, P. W., Rose, R., & Pilkington, A. (2017). Mixed methods—Theory and practice. Sequential, explanatory approach. *International Journal of Quantitative and Qualitative Research Methods*, 5(2), 10–27.
- Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology*, 102(1), 153–167. <https://doi.org/10.1037/a0018160>
- Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., do Céu Cosme, M., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing*, 30, 591–611. <https://doi.org/10.1007/s11145-016-9691-3>
- Cannanure, V. K. (2023). *DIA: Supporting teacher professional din Low infrastructure settings* [Doctoral dissertation, Carnegie Mellon University].
- Cannanure, V. K., Brown, T. X., & Ogan, A. (2020, June). *DIA: A human AI hybrid conversational assistant for developing contexts* [Conference session]. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, pp. 1–5.
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Cohen, J., Schuldt, L. C., Brown, L., & Grossman, P. (2016). Leveraging observation tools for instructional improvement: Exploring variability in uptake of ambitious instructional practices. *Teachers College Record*, 118(11), 1–36. <https://doi.org/10.1177/016146811611801105>
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Sage.
- Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3–21. <https://doi.org/10.1007/bf00988593>
- Côte d'Ivoire Institut National de la Statistique. (2015). *Enquete sur le niveau de vie des menages en Côte d'Ivoire (ENV 2015)*. Cote d'Ivoire Institut National de la Statistique.
- Curtiss Wyss, M., & Perlman Robinson, J. (2021). *Improving children's reading and math at large scale in Côte d'Ivoire: The story of scaling PEC*. Brookings Institute. https://www.brookings.edu/wp-content/uploads/2021/10/CDI-Final-Report_ENG_Web.1.pdf
- de Oca, M. M., Hill, J., Aber, L., Dolan, C. T., & Gjicali, K. (2022). The impact of attending a remedial support program on Syrian children's reading skills: Using BART

- for causal inference (arXiv:2208.13906). *arXiv*. <https://doi.org/10.48550/arXiv.2208.13906>
- Duflo, A., Kiessel, J., & Lucas, A. (2020). *Experimental evidence on alternative policies to increase learning at scale* (NBER Working Paper No. No. w27298). National Bureau of Economic Research.
- Education International. (2010). Côte d'Ivoire: Teachers face long hours, large classes. *Worlds of Education*, 33, 1–2. <https://www.ei-ie.org/en/item/20724:cote-divoire-teachers-face-long-hours-large-classes>
- Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20, 563–576. [https://doi.org/10.1016/s0272-7757\(00\)00041-8](https://doi.org/10.1016/s0272-7757(00)00041-8)
- Evans, D. K., Yuan, F., & Filmer, D. (2022). Teacher pay in Africa: Evidence from 15 countries. *World Development*, 155, 105893. <https://doi.org/10.1016/j.worlddev.2022.105893>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. <https://doi.org/10.1177/074193258600700104>
- Gove, A., & Wetterberg, A. (2011). *The early grade reading assessment: Applications and interventions to improve basic literacy* (pp. 1–37). RTI Press.
- Grossman, P. (2009). Research on pedagogical approaches in teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education* (pp. 437–488). Routledge.
- Grossman, P., & Thompson, C. (2008). Learning from curriculum materials: Scaffolds for new teachers? *Teaching and Teacher Education*, 24(8), 2014–2026. <https://doi.org/10.1016/j.tate.2008.05.002>
- Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, 23(1), 37–46. <https://doi.org/10.1080/10508406.2013.802652>
- Hedges, L. V. (2009). Effect sizes in nested designs. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 337–356). Sage.
- Innovation for Poverty Action Right-Fit Evidence Unit. (2020). *Teaching at the Right Level (PEC) evaluation report*. https://trecc.wpenginepowered.com/wp-content/uploads/2020/07/PEC2_Formal-schools_Endline-report_IPA.pdf
- Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3), 33–58. <https://doi.org/10.1257/app.1.3.33>
- Jamil, B. R., & Saeed, S. (2018). *Ins and outs of rolling out Teaching at the Right Level (TaRL) in Pakistan*. UNESCO.
- Jasińska, K. K., & Guei, S. (2022). 13. Ivory Coast: Promoting learning outcomes at the bottom of the pyramid. In D. A. Wagner, N. M. Castillo, & S. Grant Lewis (Eds.), *Learning, marginalization, and improving the quality of education in low-income countries*. Open Book Publishers. <https://doi.org/10.11647/OBP.0256>
- Jasińska, K. K., Ball, M.-C., & Guei, S. (2023). Literacy in Côte d'Ivoire. In M. Joshi (Ed.), *Handbook of literacy in Africa* (pp. 235–254). Springer International. https://doi.org/10.1007/978-3-031-26250-0_12
- Jasińska, K. K., Zinszer, B., Xu, Z., Hannon, J., Seri, A. B., Tanoh, F., & Akpé, H. (2022b). Home learning environment and physical development impact children's executive function development and literacy in rural Côte d'Ivoire. *Cognitive Development*, 64, 101265. <https://doi.org/10.1016/j.cogdev.2022.101265>
- Jasińska, K. K., Akpe, Y. H., Seri, B. A. D., Zinszer, B., Agui-Kouadio, R. Y., Mulford, K., Curran, E., Ball, M.-C., & Tanoh, F. (2022a). Evaluating bilingual children's native language abilities in Côte d'Ivoire: Introducing the Ivorian Children's Language

- Assessment Toolkit for Attié, Abidji, and Baoulé. *Applied Linguistics*, 43(6), 1116–1142. <https://doi.org/10.1093/applin/amac025>
- Jennings, P. A., Frank, J. L., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2013). Improving classroom learning environments by Cultivating Awareness and Resilience in Education (CARE): Results of a randomized controlled trial. *School Psychology Quarterly*, 28(4), 374–390. <https://doi.org/10.1037/spq0000035>
- Kembou, S., Wolf, S., Jasinska, K., & Ogan, A. (2023). *Child labor activities and schooling decisions in rural Côte d'Ivoire*. Penn Graduate School of Education: Consortium for Policy Research in Education (CPRE). CPRE Working Paper. <https://repository.upenn.edu/server/api/core/bitstreams/ba313348-4427-476c-9c1f-267e5c5af197/content>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Lee, S. S., & Wolf, S. (2019). Measuring and predicting burnout among early childhood educators in Ghana. *Teaching and Teacher Education*, 78, 49–61. <https://doi.org/10.1016/j.tate.2018.10.021>
- Maslach, C., Jackson, S., Stefanile, C., & Sirigatti, S. (1993). *MBI: Maslach burnout inventory: Manuale*. Organizzazioni Speciali.
- McAleavy, T., Hall-Chen, A., Horrocks, S., & Riggall, A. (2018). *Technology-supported professional development for teachers: Lessons from developing countries*. Education Development Trust.
- Oh, J., & Wolf, S. (2023). Negative effects of teacher burnout on student executive function and social-emotional outcomes. *Educational Psychology*, 43(4), 1–22. <https://doi.org/10.1080/01443410.2023.2205067>
- Oulai, D. (2021) *Les simulations des coûts de la mise à l'échelle au niveau national du PEC: Projection des effectifs scolarisés, des besoins de formation des enseignants et estimation des coûts de l'extension du PEC au niveau national: Rapport Final*. https://www.brookings.edu/wp-content/uploads/2021/10/BRO_CDI-Report-lay-out_FRE_WEB.1.pdf
- Programme on the Analysis of Education Systems (PASEC). (2020). *Synthèse PASEC2019. Qualité des systèmes éducatifs en Afrique subsaharienne francophone: Performances et environnement de l'enseignement-apprentissage au primaire*. www.pasec.confemen.org
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers. *Applied Developmental Science*, 9(3), 144–159. https://doi.org/10.1207/s1532480xads0903_2
- Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510–519. <https://doi.org/10.1598/RT.58.6.2>
- Pratham. (2020). *Annual status of education report*. ASER. <https://asercentre.org/>
- RTI International. (2015). *Manuel pour l'évaluation des compétences fondamentales en lecture*. <https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGRA-French-Final.pdf>
- RTI International. (2014). *Early Grade Mathematics Assessment (EGMA) toolkit*. <https://shared.rti.org/content/early-grade-mathematics-assessment-egma-toolkit#>
- Russell, S. J. (2000). Developing computational fluency with whole numbers. *Teaching Children Mathematics*, 7(3), 154–154. <https://doi.org/10.5951/tcm.7.3.0154>
- Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading*, 17(2), 108–133. <https://doi.org/10.1080/10888438.2011.618153>

- Sobers, S. M., Whitehead, H. L., N'Goh, K. N. A., Ball, M. C., Tanoh, F., Akpé, H., & Jasińska, K. K. (2023). Is a phone-based language and literacy assessment a reliable and valid measure of children's reading skills in low-resource settings? *Reading Research Quarterly*, 58(4), 733–754. <https://doi.org/10.1002/rq.511>
- Sprenger-Charolles, L. (2008). *Early Grade Reading Assessment (EGRA): Results from Senegalese primary school students learning to read in French and in Wolof—Report for the World Bank*. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/894321468203669271>
- United Nations Development Programme (UNDP). (2024). *Human development report 2023–2024*. UN.
- UNESCO Institute for Statistics (UIS) (2016). *Understanding what works in oral reading assessments: Recommendations from donors, implementers and practitioners*. UNESCO Institute for Statistics.
- UNESCO UIS. Côte d'Ivoire (2018).
- UNICEF. (2021). *The state of the global education crisis: A path to recovery—A joint UNESCO, UNICEF and WORLD BANK report*. UNESCO.
- USAID. (2018). Analysis of early grade reading assessment in India (No. REQ-ASIA-16-00017). https://pdf.usaid.gov/pdf_docs/PA00TQ7X.pdf
- Wagner, D. A. (2017). *Learning as development: Rethinking international education in a changing world*. Routledge.
- Whitehead, H. L., Ball, M. C., Brice, H., Wolf, S., Kembou, S., Ogan, A., & Jasińska, K. K. (2024). Variability in the age of schooling contributes to the link between literacy and numeracy in Côte d'Ivoire. *Child Development*, 95(2):e93–e109. <https://doi.org/10.1111/cdev.14018>
- Whitehead, H. L., Kouassi, H., Yeboue, A., Tanoh, F., Hajjaj, B., Kembou, S., Wolf, S., Jasińska, K. K., & Ogan, A. (2025). “And then there are the realities of the field”: Teacher insights on the implementation of a targeted instruction program in rural Côte d'Ivoire. Teaching and Teacher Education.
- Wolf, S., Aber, J. L., Behrman, J. R., & Tsinigo, E. (2019). Experimental impacts of the Quality Preschool for Ghana interventions on teacher professional well-being, classroom quality, and children's school readiness. *Journal of Research on Educational Effectiveness*, 12(1), 10–37. <https://doi.org/10.1080/19345747.2018.1517199>
- World Bank. (2019). *Que la route soit bonne: Améliorer la mobilité urbaine à Abidjan*. World Bank. <https://documents1.worldbank.org/curated/en/624341549322162402/pdf/133722-FRENCH-WP-P168565-PUBLIC-4-2-2019-15-48-37-CtedIvoireeightheco nomicupdatereport.pdf>
- World Bank. (2018). There is no learning without prepared, motivated learners. In *World development report 2018: Learning to realize education's promise* (pp. 107–130). World Bank. <https://openknowledge.worldbank.org/server/api/core/bitstreams/2e0789b8-6328-5590-a1c2-300eb15032e9/content>
- World Bank. (2021). Data Cote d'Ivoire. <https://data.worldbank.org/country/cote-divoire>
- World Bank. (2022). The state of global learning poverty: 2022 update. <https://thedocs.worldbank.org/en/doc/e52f55322528903b27f1b7e61238e416-0200022022/original/Learning-poverty-report-2022-06-21-final-V7-0-conferenceEdition.pdf>

Manuscript received July 30, 2024

Final revision received January 22, 2026

Accepted February 6, 2026