

# Right-Fit Evidence Unit



## TEACHING AT THE RIGHT LEVEL (PEC)

COMPANY PARTNER/J-PAL

EVALUATION REPORT

Prepared by Innovation for Poverty Action's Right-Fit Evidence Unit

## Context

---

The Transforming Education in Cocoa Communities (TRECC) initiative aims to improve the living conditions of children and youth in Côte d'Ivoire by promoting quality education in cocoa-growing communities. Via its Grant Matching Mechanism round 2 (GMM2), 13 pilots-to-scale projects are being co-funded with 12 cocoa companies and implemented by 14 implementing organizations in the sectors of Early Childhood Development, Primary Education and Vocational Training.

The role of Innovations for Poverty Action (IPA) is to provide technical support to the companies and implementing agencies to design and implement sound monitoring systems to closely monitor and learn from these pilots. IPA also conducts its own independent and complementary data collection to evaluate the pilot projects. For each pilot, IPA analyzes these two sources of information—the administrative data collected by the implementing organizations through their own M&E system and IPA's independently collected data—to assess the pilot on a pre-agreed rubric. The final scale-up report will therefore be based on the evaluation rubric that was agreed upon by all partners.

This report contains IPA's analysis and recommendations on the potential scale-up of the COMPANY PARTNER pilot. The scale-up under discussion here is focused on other relevant cocoa-growing communities, as is GMM2's contractual emphasis. However, TRECC may consider whether certain pilots are feasible for future scale-up beyond such communities, for example to the regional or national level.

The report is divided into the following five sections. Each section is broken down into 2-3 components, for a total of eleven criteria:

### **1. Relevance**

- 1.1. The program is targeting important needs in the community
- 1.2. The program is aligned with donors' priorities

### **2. Results (outputs and immediate outcomes)**

- 2.1. The program delivers outputs at high quality
- 2.2. The program achieves direct outcomes
- 2.3. Beneficiary feedback about the program is positive

### **3. Costs & Operations management**

- 3.1. Costs are well-managed
- 3.2. Project management is successful

### **4. Capacity to learn, improve and innovate**

- 4.1. Project collects credible monitoring data
- 4.2. Monitoring is used to learn and improve

### **5. Sustainability**

- 5.1. Provides sustained benefit to the community
- 5.2. There are prospects for scale-up beyond GMM2

Each of these eleven components (e.g., “5.1. Provides sustained benefit to the community”) is evaluated on indicator-level criteria and assigned a color score.



Green indicates that the pilot is, in IPA’s view, consistent with the criteria required for potential scale-up



Orange indicates that IPA’s data shows the pilot is partially consistent with the criteria for scale-up and that eligibility for scale-up should be conditional on corrective measures to be taken in that area



Red indicates that the pilot does not appear to be consistent with the requirements for scale-up.

Pilots with green assessments on all 11 criteria receive an unconditional recommendation for eligibility for a scale-up proposal; pilots with only green and orange criteria (no red), and among these a majority of green criteria, receive a conditional recommendation for scale-up (i.e. conditional on the various corrective measures being proposed in the orange criteria). Pilots with any red criteria are not recommended for scale-up.

## List of Acronyms

---

<b>FGD:</b>	Focus Group Discussion
<b>GMM2:</b>	Grant Matching Mechanism round 2
<b>IPA:</b>	Innovations for Poverty Action
<b>MENET- FP</b>	Ministère de l'Éducation Nationale, de l'Enseignement Technique et de la Formation Professionnelle
<b>M&amp;E:</b>	Monitoring and Evaluation
<b>PEC</b>	Programme d'Enseignement Ciblé
<b>TaRL</b>	Teaching at the Right Level
<b>TRECC</b>	TRECC Transforming Education in Cocoa Communities
<b>IEPP</b>	Inspection de l'Enseignement Préscolaire et Primaire

# Table of Contents

---

## Contents

Project summary .....	8
Overall Timeline.....	9
General assessment and recommendation .....	10
Snapshot of specific assessment against each pre-defined evaluation criteria: .....	1
1. Relevance.....	2
1.1 The program is targeting important needs in the community .....	2
Data and methodology .....	2
Indicator 1.1.1. J-PAL's needs assessment report and IPA's baseline assessment show evidence of a need being addressed by the pilot.....	2
A majority of children do not master basic reading and math skills.....	2
Indicator 1.1.2. Beneficiaries' description of their needs links to the outcomes delivered by the pilot.....	4
Indicator 1.2.1. The pilot, as implemented, remains aligned with COMPANY PARTNER's objectives.....	5
Indicator 1.2.2. The pilot, as implemented, remains aligned with Jacobs Foundation's objectives .....	5
2. Results (outputs and direct outcomes).....	6
2.1 Delivers outputs at high quality .....	6
▪ Indicator 2.1.2. Participation rate.....	7
2.2 Achieves direct outcomes.....	8
Data and methodology .....	8
▪ Indicator 2.2.1. Changes in teachers' knowledge, behavior and practices .....	8
▪ Indicator 2.2.2 Teachers show effective use of the PEC materials during classroom observations .....	12
▪ Indicator 2.2.3 Teachers reported having a coaching session with schools' principals or inspectors.....	17
2.3. Beneficiary feedback about the program is positive.....	18
▪ Indicator 2.3.1.a Beneficiaries provide positive feedback on the delivery of outputs.....	18
▪ Indicator 2.3.1.b Teachers provided positive feedback on the improvement on their students' learning in the classroom as result of PEC methods.....	22
▪ Indicator 2.3.1.c Beneficiaries describe positive experiences with the program .....	22
3. Costs and operations management.....	24

• Indicator 3.1.1 Did the project make efficient utilization of resources?	<b>Erreur ! Signet non défini.</b>
• Indicator 3.1.2 Do the partners have a vision for cost-efficiency at scale?	<b>Erreur ! Signet non défini.</b>
3.2 Project management is successful .....	24
• Indicator 3.2.1. Intervention is implemented according to roadmap, and any changes are fully justified.....	<b>Erreur ! Signet non défini.</b>
• Indicator 3.2.2. Cooperation Partners' involvement .....	<b>Erreur ! Signet non défini.</b>
4. Capacity to learn, improve and innovate.....	24
4.1 Project collects credible monitoring data.....	25
• Indicator 4.1.1 Routine monitoring data are collected and shared on time with stakeholders	25
• Indicator 4.1.2 IPA's spot-check visits confirm the quality and accuracy of the data shared by the partner .....	26
4.2 Monitoring is used to learn and improve .....	27
• Indicator 4.2.1 Monitoring data is actionable and aligned with program management.	27
• Indicator 4.2.2 Program improvement in response to monitoring.....	27
5. Sustainability.....	28
5.1 Provides sustained benefit to the community.....	28
• Indicator 5.1.1 Indications that the animators and mentors are likely to continue with the practices or program activities .....	28
5.2 There are prospects for scale-up beyond GMM2.....	29
• Indicator 5.2.1. Potential for further scale-up, by the government of Côte d' Ivoire, the Cooperation Partner or other development actors.....	29
• Indicator 5.2.1 Enabling and financing environment .....	29
• Indicator 5.2.2 Organizational capacity to implement at scale.....	30
Brookings High-level Feedback on GMM2 TaRL Africa Project Work Plan 2017-2021).....	31
6. Annexes.....	34
Annex 1: Independent Data collection Methodology .....	34
Annex 2: ASER test and Implementation.....	37
Annex 3: Protocol to administrate ASER literacy test.....	39
Annex 4: Protocol to administrate ASER numeracy test.....	40
Annex 5: Teachers logs or registers template .....	41
Annex 6: Classroom observations components .....	42
Annex 7: Some results from IPA Independent Data Collection fieldwork.....	42
Annex 8: TRECC Pilot Criteria Matrix.....	45

## Figures

---

Figure 1: Project map	9
Figure 2: Literacy baseline assessment by IPA enumerators	14
Figure 3: Numeracy baseline assessment by IPA enumerators	14
Figure 4: Teachers' feedback on PEC as a solution to improve students' learning level	15
Figure 5: Share of reasons of mismatch among the seven mismatch schools.	22
Figure 6: Number of PEC classrooms observed vs Number of students	23
Figure 7: Inter rater agreement matrix literacy test	26
Figure 8: Inter rater agreement matrix numeracy test	27
Figure 9: Reported teachers' feedback regarding PEC approach assets	30
Figure 10: Reported teachers' feedback regarding PEC approach weaknesses	31
Figure 11: Percentage of teachers who report getting all necessary PEC materials	31
Figure 12: List of materials which were usually missing during the PEC implementation	32
Figure 13: Feedback of the beneficiaries on the improvement of their students learning.	33
Figure 14: Credible data criteria	41
Figure 15: Literacy baseline assessment by IPA's enumerators and teachers	53
Figure 16: Numeracy baseline assessment by IPA's enumerators and teachers	53
Figure 17: Literacy endline assessment by IPA's enumerators and teachers	54
Figure 18: Numeracy endline assessment by IPA's enumerators and teachers	54

## Tables

---

Table 1: Key output #1	18
Table 2 : Key output #2	19
Table 3 : Key output #3	19
Table 4: Sample covered by round of data collection	20
Table 5: Number of schools where teachers' logs were available per waves	20
Table 6: Trend of teachers' scores across classroom observations waves.	25
Table 7: Teachers scores across classroom observation waves	25
Table 8: One-sample-proportion-test on ASER literacy levels – Baseline results	27
Table 9: One-sample-proportion-test on ASER numeracy levels – Baseline results	28
Table 10: One-sample-proportion-test on ASER literacy levels – Endline results	28
Table 11: One-sample-proportion-test on ASER numeracy levels – Endline results	29
Table 13: Numbers of days of assistance provide by mentors during the last three months	30
Table 14: Teachers confidence in implementing the PEC methodology	30
Table 15: Percentage of teachers who agree to recommend PEC in CI	33
Table 16: Sample covered by round of data collection	48
Table 17: Literacy assessment by IPA enumerators – Baseline vs Endline	53
Table 18: Number recognition assessment by IPA enumerators – Baseline vs Endline	53
Table 19: Basic operations assessment by IPA enumerators – Baseline vs Endline	54
Table 20: Class observation 1 results	55
Table 21: Class observation 2 results	56

## Project summary

---

Through the Transforming Education in Cocoa Communities (TRECC) initiative, the Ministry of Education (MEN) with the support of the Abdul Latif Jameel Poverty Action Lab (J-PAL) piloted the Teaching at the Right Level (TaRL) approach—translated in French as the Programme d’Enseignement Ciblé (PEC)—in 25 formal schools in the region of Gbadijiri, Côte d’Ivoire.

The approach was implemented in Grades 3, 4 and 5—corresponding to the CE1, CE2 and CM1 in the Ivorian system—for an hour and a half per day. The goal of the program was to provide students who are lagging behind their peers with key skills in French and mathematics. The pilot targeted 25 schools, 75 teachers (three teachers per school, one per grade) and 2,566 students.

PEC aims to improve basic reading and mathematics skills. Assessment is at the core of the approach. Students are assessed at the outset and grouped according to their level in reading or mathematics. In their new groups, children are taught via engaging, fun and creative activities that are adapted to their learning level. Students are regularly assessed on their performance and can progress between groups throughout the year.

For this pilot, teachers assessed students using a literacy and numeracy assessment tool called the Annual Status of Education Report (ASER) test, which measures grade-level reading skills and tests numeracy skills that are taught in Grade 1 and 2 in formal schools.<sup>1</sup> After students were evaluated, Grade 3, 4 and 5 (CE1, CE2, CM1) teachers used the results to group students by skill level. Teachers then incorporated in their daily schedules an hour and a half of PEC activities with students grouped by skill level. PEC activities consist of 45 minutes of reading and 45 minutes of mathematics class, which all of the targeted grades conduct at the same time.

In their initial proposal, J-PAL listed five program components that implementers would need to incorporate during PEC’s implementation to be effective in Cote d’Ivoire:

1. **Assessment of the children’s level**, using simple 10-15 minutes tests administered orally, to constitute level groups according to the results of each child.
2. **Adapted didactic tools**: an adaptation of learning materials and teaching guides to the contextual realities of program implementation.
3. **Training for teachers and pedagogical supervisors** (Grades 3, 4 and 5 teachers, and all schools’ directors), so that they take ownership of the program’s principles and activities. The training includes a theoretical phase and a practical phase.
4. **Teaching tailored to the child**, with interactive, progressive and engaging activities.
5. **Regular pedagogical supervision and tutoring** to give friendly, practical advice to teachers so that they can overcome any difficulties they may have. This also allows supervisors to ensure that PEC requirements are met.

The PEC program in Côte d’Ivoire also benefited MENET-FP staff, who were trained in the approach during an observation trip to India.

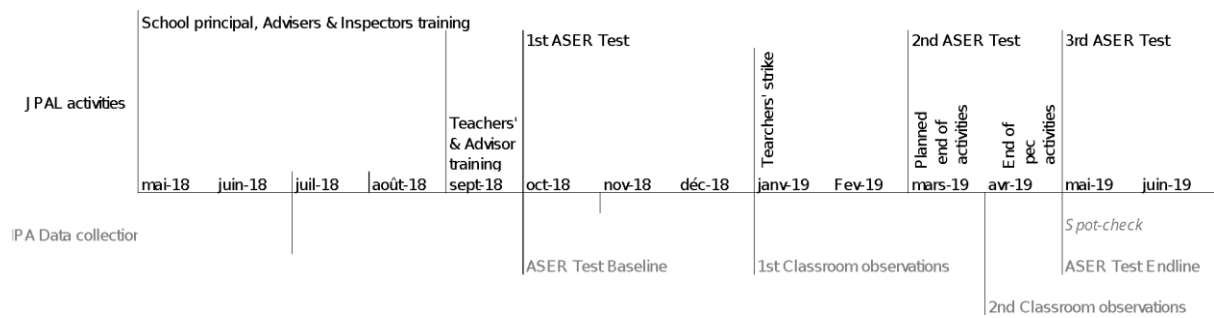
---

<sup>1</sup> ASER is used in many countries and accepted by researchers as a valid instrument for measuring students’ skills. See Annex 2 for the ASER tool used for this report.



## Overall Timeline

Figure 1: Project timeline



The pilot implementation lasted eight months, from October 2018 to May 2019, instead of six months as initially planned, because of school disruptions mainly due to teachers' strikes. Before May 18, preparatory activities were starting with the study trip in India, the training of the core team of the MENET-FP and the adaptation of the PEC material to the Ivorian context.

Figure 1: Project map



# Executive Summary

---

## General assessment and recommendation

IPA's Right-Fit Evidence unit conducted an independent process evaluation of the COMPANY PARTNER and J-PAL PEC program, which has teachers group school children for learning activities according to their level for an hour and a half per day. This pilot, implemented in 25 schools, involved training teachers on PEC principles and implementation. The pilot also trained a network of support for teachers adopting PEC, from MENET-FP pedagogical advisors to their own School directors, to provide further mentoring.

Teachers' opinions of the intervention and the training programs were collected. After teachers assessed students' skills and created groups, their accuracy was backchecked by IPA. Finally, teachers' views about the effectiveness of the six-month pilot were collected.

Overall, the PEC model as piloted by J-PAL, COMPANY PARTNER, and the MENET-FP has earned a full recommendation for scale-up, with green checkmarks on all measures. IPA's independent evaluation found strong engagement from all pilot stakeholders. Teachers were especially invested in PEC's success; stakeholders in mentorship roles were more uneven, although strong overall. This widespread buy-in bodes well for scale-up.

**1. Relevance:** IPA's assessments find that students in cocoa communities have generally low skill levels. PEC targets students with low skill levels to bring them up to grade level with group activities and instruction that meets them at their level. PEC is thus relevant for addressing a real community need. Further, donors – the Jacobs Foundation and COMPANY PARTNER– have strategic goals that PEC implementation is a close fit with. The program is relevant to the needs of beneficiaries and funders.

**2. Results: Outputs and direct outcomes:** The first output was simply to spread PEC competencies, and stakeholders generally found training on PEC useful and adequate, with some requests for more training and support. IPA backchecks - on the accuracy of the groups that teachers created - indicate that teachers correctly assessed students' skills and created relevant groups for PEC activities. During classroom observations, 100% of teachers randomly sampled were conducting PEC activities, but this may be inflated: in order to ensure observers had smooth access to schools, school inspectors were given two-week windows for observers' visits, and it is possible the inspectors passed their information on to directors and teachers. Measured several ways, teachers and mentors believed in PEC's usefulness for improving students' learning. Overall, outputs were reliably delivered and there are indications that direct outcomes proceeded smoothly, in part because stakeholders were convinced of the value of PEC.

## Recommendations for scale-up:

- Mitigation strategy for teacher absenteeism: Teachers absenteeism might be a challenge at scale. To mitigate it, IPA recommends developing a mitigation strategy: implement PEC in separate groups even if a teacher is absent, monitoring absenteeism levels, or building in interventions to reduce absenteeism during PEC hour.
- Delivering PEC materials: Learning materials contribute a lot to the effectiveness of teachers. PEC materials were not always delivered to teachers on time, an issue teacher often raised during interviews and surveys. The delivery logistics should be resolved for scale-up.

**3. Costs and Operation management:** J-PAL managed finances efficiently, with final expenses rarely deviating from initial predictions. Communication among organizations working on the pilot was close and productive. However, initial estimates suggest that scale-up may be costly compared to PEC interventions in other contexts, which are often more densely populated.

**Recommendation for scale-up:**

- Identify potential economies of scale: IPA recommends reviewing program's main components and all expenses to find opportunities for economies of scale, particularly regarding mentoring activities.
- MEN's involvement: Since the goal is for the government to implement the program at scale, IPA recommends the MENET-FP to be even more involved in taking initiative and making decisions. In particular, during the Real Time Scaling lab, the MENET-FP should take an active role to decide what is possible or not at scale.

**4. Capacity to learn, improve and innovate:** The pilot collected monitoring data on several important implementation steps. First, the attendance data on teachers' training; second, children's assessments; and third, mentors' classroom observation. On the whole, the monitoring data was credible and unbiased. Implementers had regular meetings to discuss the results and address any issues.

**Recommendation for scale-up:**

- Mentors' collection of children's assessment data: During IPA's class observations, some teachers were not able to provide the assessment logs. Those teachers reported that advisors were using the logs for monitoring purposes. At scale, IPA recommends implementing a monitoring process that ensures advisors can collect logs without depriving teachers for PEC sessions. Texting logs to advisors may be a solution.

**5. Sustainability:** Teachers' general enthusiasm for and competence in PEC after training indicates that once started, implementation can be sustained. The MEN's involvement at this early stage offers opportunities for government scale-up.

**Recommendations for scale-up:**

- Mentors' motivation and per diems: Early caution and transparency around compensation and per diems makes it easier to motivate mentors sustainably. For J-PAL, one major challenge will be to integrate an effective and sustainable monitoring system leveraging existing government practices and without excessive external incentives.
- Refresher trainings: IPA recommends having additional trainings for teachers and mentors to refresh teachers and to onboard replacements. If this proves to be too resource-intensive, text message reminders or online forums for teachers are possible alternatives.

Snapshot of specific assessment against each pre-defined evaluation criteria:


Evaluation Criteria	Assesment	Recommendations
<b>1.Relevance</b>		
1.1. The program is targeting important needs in the community	✓	
1.2. The program is aligned with donors' priorities	✓	
<b>2.Results: outputs and direct outcomes</b>		
2.1. The program delivers outputs at high quality	✓	
2.2. The program achieves direct outcomes	✓	Mitigation strategy for teacher absenteeism
2.3. Beneficiary feedback about the program is positive	✓	Make sure PEC material is always available for teachers
<b>3.Costs and operations management</b>		
3.1. Costs are well-managed	✓	Identify potential economies of scale
3.2. Project management is successful	✓	MEN's involvement
<b>4.Capacity to learn, improve and innovate</b>		
4.1. Project collects credible monitoring data	✓	Mentors' collection of assessment data
4.2. Monitoring is used to learn and improve	✓	
<b>5.Sustainability</b>		
5.1. Provides sustained benefit to the community	✓	Mentors' motivation and per diems and refresher trainings for teachers
5.2. There are prospects for scale-up beyond GMM2	✓	

# 1. Relevance

**Relevance** is divided into two components, both of which the pilot is assessed on:

1. The program is targeting important needs in the community
2. The program is aligned with donor's priorities

IPA finds that the COMPANY PARTNER pilot appears to be consistent with the requirements for scale-up in both components of Relevance. The pilot's score for each component of Relevance is determined by the pilot's performance against pre-agreed indicator-level criteria. Note also that some general indicators (such as "evidence of a need being addressed") are decomposed into indicators specific to this pilot (such as "a majority of children do not master basic reading and math skills"). The pilot is evaluated on the specific indicators, which determine its score for the general indicator.

 **1.1 The program is targeting important needs in the community**  
This project targets important and specific needs in cocoa communities, where children have low learning levels in literacy and numeracy. PEC is well-suited to addressing low learning levels. On a handful of skills, students showed a wide range of skill levels, which is also a situation in which PEC is likely to be effective addressing students' needs.

## Data and methodology

The data analyzed in this section comes from students' learning level using the ASER literacy and numeracy test administrated by IPA during the baseline data collection. The baseline survey was implemented in all 25 pilot schools. Data on reading and mathematics skills were collected using a literacy and numeracy assessment tool called the Annual Status of Education Report (ASER). Details on IPA surveys methodology, ASER score levels, and interpretation are available in Annex 1 and Annex 2.

### Indicator 1.1.1. J-PAL's needs assessment report and IPA's baseline assessment show evidence of a need being addressed by the pilot

PEC is a remedial approach well-suited to primary schools where Grades 3, 4 and 5 students lack foundational reading and mathematics skills or where students have a wide range of literacy and numeracy skill levels<sup>2</sup>. This section presents students' learning levels in targeted schools and confirms that this PEC minimum requirement held true before the program was implemented.

### A majority of children do not master basic reading and math skills

IPA baseline data shows that the vast majority of students are experiencing delays in both reading and mathematics skills.

The literacy learning assessments tested reading ability in French and placed learners into one of five levels: in descending order, Story, Paragraph, Word, Letter, and Beginner.<sup>3</sup> Only 28 percent of students in Grades 3 to 5 in our sample can read a word, while they all should be at this level at the end of Grade 2<sup>4</sup>. Overall, as presented in figure 1, students' literacy level is low.

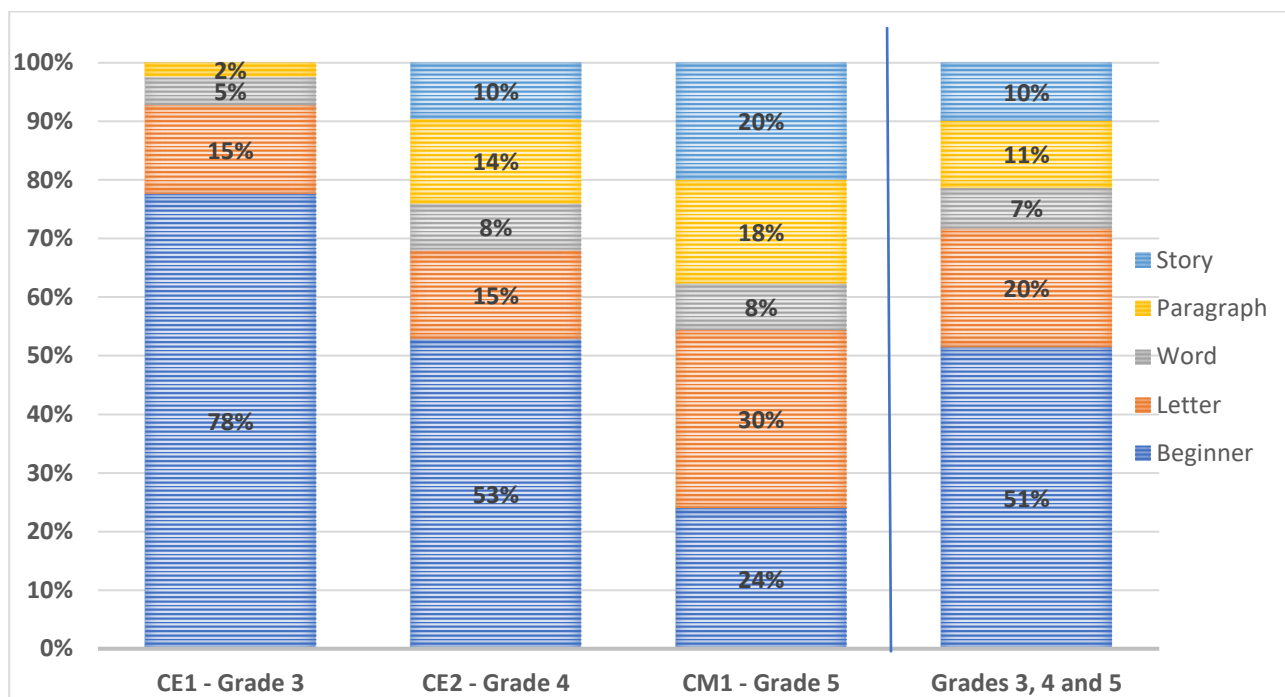
---

<sup>2</sup> More information about the context in which PEC approach applied are available at "Teaching at the Right Level" website <https://www.teachingattherightlevel.org/home/is-tarl-right-for-my-context/>

<sup>3</sup> For more detail, see Annex 2

<sup>4</sup> As part of the IMPAQ project, an adaptation workshop was conducted in April 2016 at MENET in Abidjan to ensure that the reading part of the ASER test was culturally appropriate and consistent with Ivory Coast's learning standards for each grade level in primary school ([https://www.impaqint.com/sites/default/files/files/AVSI%20FFE\\_Baseline%20FINAL\\_091916.pdf](https://www.impaqint.com/sites/default/files/files/AVSI%20FFE_Baseline%20FINAL_091916.pdf)).

Figure 1: Literacy baseline assessment by IPA enumerators



The numeracy assessments are split into two parts: number recognition and mathematical operations.

For number recognition, a learner can be placed into one of four levels, depending on the largest type of number they can recognize. Results of the numeracy test are slightly better than the literacy test, which is a standard result in the region<sup>5</sup>. As described in Figure 2, only 4 percent of students are in the Beginner level, which means all students can identify at least one-digit numbers, by the end of Grade 4. 84 percent of Grade 5 students can read 3-digit numbers, the highest level of numeracy ASER tests for. However, by Grade 3 (CE1), all students should be able to read 3-digit numbers. In Figure 2, a slight pattern is visible: many more students can read 1-digit or 3-digit numbers than 2-digit numbers. It is likely just noise (even if the national average is a perfectly even distribution, one would still expect different distributions in smaller samples), but it is possible it indicates that once students can read 2-digit numbers, it is usually easy for them to make the jump to 3-digit numbers. However, it is concerning that only 40% of Grade 3 (CE1) students can read 2-digit numbers. Students are supposed to be able read 2-digit numbers by the start of Grade 3.6

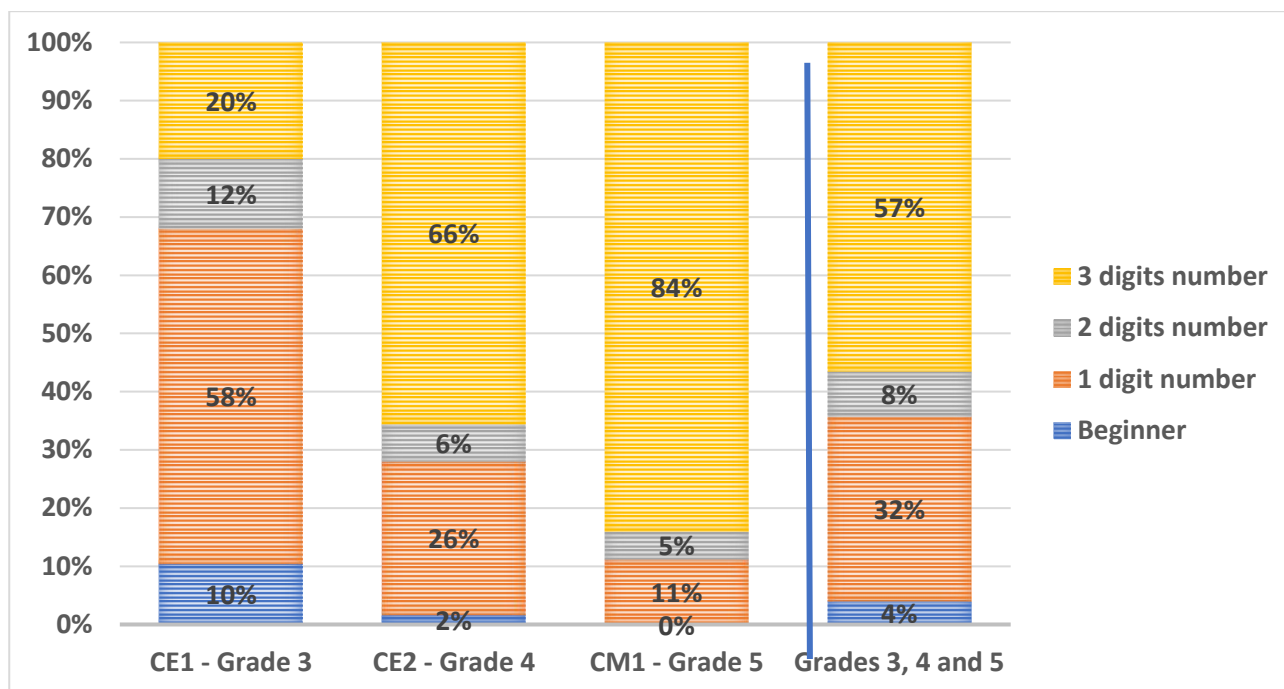
Regarding basic operations, students are more comfortable with addition and subtraction than with multiplication or division. 41 percent of students can do addition and 38 percent can do subtraction, compared to 15 percent for multiplication and 3 percent for division (see full results in Annex 7). All students in Grade 3 and above are supposed to be able to do addition and subtraction with numbers smaller than 50.<sup>7</sup>

<sup>5</sup> PASEC 2014 results "<https://www.pasec.confemen.org/wp-content/uploads/2016/12/PASEC2014-C%C3%B4te-d'Ivoire.pdf>".

<sup>6</sup> <https://www.pasec.confemen.org/wp-content/uploads/2016/12/PASEC2014-C%C3%B4te-d'Ivoire.pdf>, pg 29

<sup>7</sup> Ibid

Figure 2: Numeracy baseline assessment by IPA enumerators



Grades are likely to be homogeneous in literacy and heterogeneous in mathematics

The above results show that although students' capacity in literacy and numeracy is below grade level, there is overall improvement with each grade, suggesting that some learning is taking place – just not enough. In such a case, interventions to improve pedagogy may be valuable. If so, PEC addresses a real community need. Students' literacy levels are generally below expected grade competencies. Baseline results show that 93 percent, 68 percent and 54 percent of students in Grades 3, 4 and 5 respectively cannot read a word. The majority of students in Grades 3 and 4 are in the Beginner group, meaning they cannot reliably read a letter. Students' low literacy level shows that there is a room for PEC implementation in the targeted schools, as PEC is a tool for addressing learning lags. It provides more support and opportunities to master basic skills, as well as simple recognition that they are struggling and could benefit from more time to master the basic skills.

In numeracy, and sometimes in literacy, IPA's ASER assessment found that students have a wide range of skill levels. In most grade and subject combinations (e.g., Grade 4 literacy or Grade 3 numeracy), there are more than 10 percent of students at each skill level. The heterogeneity demonstrated in these results confirms the relevance of grouping children by level and not by grade, which is what PEC achieves.

✓ **Indicator 1.1.2. Beneficiaries' description of their needs links to the outcomes delivered by the pilot**

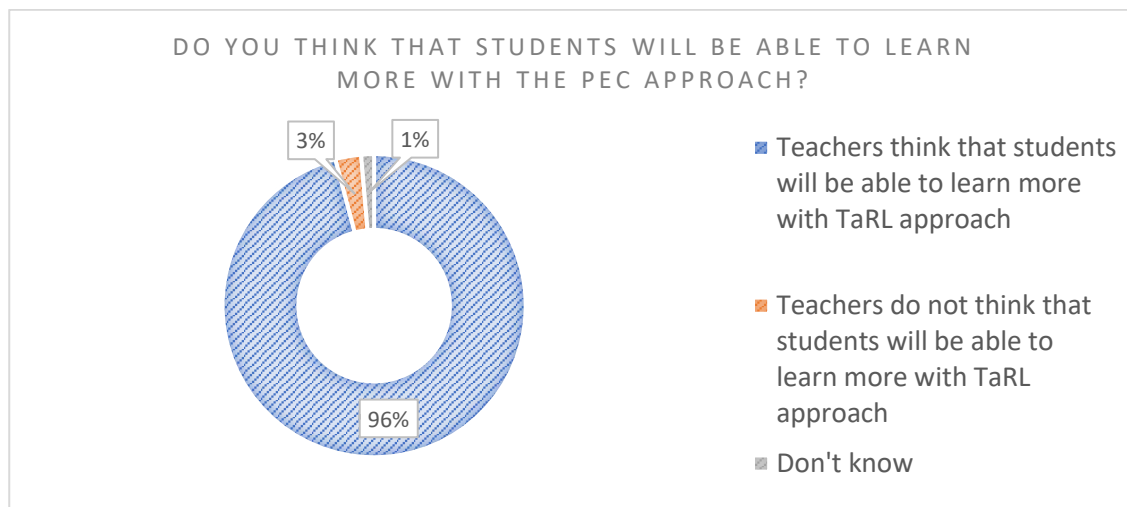
RESEN and PASEC 2014 indicated a need for pedagogy targeting students' actual learning levels. To confirm this, teachers' feedback on the relevance of PEC in their context was collected after they were trained in PEC but before they implemented it.

After teachers attended PEC training, IPA gathered all teachers' ideas on PEC as an efficient remedial education approach in primary schools before they started the pilot implementation. When considering PEC as a solution to fill students' gaps in learning, almost all teachers (93 percent) reported that students will be able to learn more with the PEC approach. There is some risk in weighting these results heavily. Teachers were only given three options (students will learn more, students will not learn more, don't



know) and had not yet implemented PEC. Teachers' overall positive response to the question after training is not firm proof that they value PEC. However, it would indicate that PEC is a poor fit for local needs if many teachers rejected PEC after training but before implementation. There is no evidence of such a rejection in this data.

Figure 3: Teachers' feedback on PEC as a solution to improve students' learning level



## **1.2 The program is aligned with donors' priorities**

The approach proposed is aligned with donors' strategic ambition and the pilot, as implemented, has strengthened their willingness to continue supporting the project.



### **Indicator 1.2.1. The pilot, as implemented, remains aligned with COMPANY PARTNER's objectives**

As reported during the Key Informant Interview, one of the key concerns in COMPANY PARTNER's choice about which project to support was the sustainability of the approach. Indeed, for several decades, COMPANY PARTNER has been committed to a sustainable development policy.

According to COMPANY PARTNER, as implemented, the project exceeds its expectations and remains aligned with its initial objectives. Indeed, the pilot increased the quality of education as well as the satisfaction of children towards school. Therefore, COMPANY PARTNER's representative confirmed that COMPANY PARTNER would stay involved in the next scale-up phase of the project. They will also continue to involve field staff, but in a lighter touch given the new outreach of the program.



### **Indicator 1.2.2. The pilot, as implemented, remains aligned with Jacobs Foundation's objectives**

The PEC pilot as implemented remains aligned with Jacobs Foundation's objectives and priorities, which are to improve the quality of education in Cote d'Ivoire. First, the pilot succeeded in showing the relevance of the TaRL approach in Cote d'Ivoire, addressing an urgent need in cocoa growing communities where children have low levels of basic literacy and numeracy skills. The Jacobs Foundation team believes that this evidence-based approach is promising in this context. Second, the project has potential for a large scale-up and fully fits into the pilot-to-scale approach. The engagement of the Ivorian Ministry of Education (Ministère de l'Education Nationale, de l'Enseignement Technique et de la Formation Professionnelle) in the pilot and in the current discussions is encouraging for the next phases of the project. The technical support of Pratham, which has deep experience scaling education programs, is also a true asset.



## 2. Results (outputs and direct outcomes)

**Results** are divided into three components to capture each type of result TRECC is looking for in a pilot. The pilot is assessed on each of these components:

1. Delivers output at high quality
2. Achieves direct outcomes
3. Beneficiaries' feedback about the program is positive

The pilot's implementation appears to be consistent with the requirements for scale-up in each component of Results. The pilot's score for each component of Results is determined by the pilot's performance against pre-agreed indicator-level criteria. Each component, including the criteria that determined the pilot's score, is discussed below.

Note also that general indicators allowing comparisons between pilots (such as "Key outputs from the proposal logframe were achieved") are decomposed into indicators specific to this pilot, such as "Schools received PEC material". For example, one such specific indicator is teachers' ability to successfully administer the ASER and group students according to skill levels. This is discussed with respect to numeracy and literacy skills at baseline and numeracy and literacy skills at endline. The pilot's success on the specific indicators determine its score for the general indicators.



### **2.1 Delivers outputs at high quality**

All outputs were achieved and high quality. PEC materials were received in all schools, albeit delayed by a few days. Data show that both mentor and teacher trainings had high participation rates.

#### **Indicator 2.1.1. Key outputs from the proposal logframe were achieved**

To measure output achievements, IPA largely collected its own data, which serves as the basis for most of this section.

##### *Schools received PEC material*

Each school should have received three complete sets of materials for mathematics and reading (for the three group levels).

Baseline data, collected on October 21, 2018, shows that almost no teacher had yet access to the necessary materials. The 25 schools received the material from October 24 to 26, two weeks after teachers have administered the first ASER test in the pilot classes. Each director dated and signed a "material sign-in sheet" when he received the PEC material in his school. J-PAL uploaded the sheet on a google drive shared with IPA.

Table 1: Key output #1

Indicator	Target	Achieved	Data source
<b>Schools that received the material</b>	25	25	J-PAL administrative data

✓ **Indicator 2.1.2. Participation rate**

✓ *Attendance rate at mentor training*

Mentoring is critical for the success of PEC's approach. Teachers' mentors are school principals, pedagogical advisors<sup>8</sup>, and primary school inspectors<sup>9</sup>. School principals provide close supervision to teachers who apply the PEC approach in their school. Pedagogical advisors and school inspectors provide monthly mentoring sessions.

Pratham, with the support of J-PAL led the mentor training – where school principals, pedagogical advisors and school inspectors are trained - on May 7 to 17, 2018. During the 11 days, participants learned about the theory of TaRL approach, and practiced the activities. Administrative data collected by J-PAL shows that 25 school principals (over 25) participated in the 11 days and one missed a day of training. The two pedagogical advisors invited attended the full training – but one was terminated during the course of the pilot and his successor was not trained - and the inspector missed a day. We consider mentors to be trained if they participated to all sessions of the mentor training. Mentors were also invited to participate *in the teachers' trainings*.

Table 2 : Key output #2

Indicator	Target	Achieved	Data source
<b>Principals trained</b>	25	24	J-PAL administrative data
<b>CP trained</b>	2	2	J-PAL administrative data
<b>IEPP trained</b>	1	1	J-PAL administrative data

✓ *Attendance rate at teacher training*

74 teachers were trained during 5 days from September 10 to 14, 2018. The training included theoretical and practical sessions. Training was facilitated by six Master trainers from the DPFC and MEN/PEC National Coordinator M. Vincent Dally, with support from Pratham's senior management.

Attendance data shared by J-PAL show that 66 teachers participated in the full training and four missed one day. At baseline, the 75 teachers reported to IPA that they attended all sessions of the training. The difference could be explained by a few (8) teachers missed one day to one hand, and one teacher did not attend all the session to other hand.

Table 3: Key output #3

Indicator	Target	Achieved	Data source
<b>Teachers trained</b>	75	66	J-PAL administrative data and IPA baseline data

<sup>8</sup> « Conseiller pédagogique » or CPP in Cote d'Ivoire

<sup>9</sup> « Inspecteur de l'enseignement préscolaire et primaire » or IEPP in Cote d'Ivoire



## 2.2 Achieves direct outcomes

The majority of the program's direct outcomes were reached. Teachers were able to group their students by learning level. However, there is a concern with their students' assessment. Most teachers were able to implement PEC pedagogy in the classes. However, we observed that in 20% of schools we visited, there was at least one teacher absent during the PEC class, which resulted in schools merging PEC groups – this would have reduced the effectiveness of the PEC approach somewhat. In addition, there is room for improvement with respect to coaching teachers on PEC.

### Data and methodology

Table 1 summarizes the different targets and sample covered during all rounds of data collection. Students' learning level assessment using ASER test were performed during the baseline and the endline data collection along with teachers' interviews. Details about the IPA independent data collection methodology are available in Annex 1.

Table 1: Sample covered by round of data collection

Round of data collection	Targets	Survey planned	Survey covered	Rate
<b>Baseline</b>	Students	375	375	100%
	Teachers	75	73	97%
<b>Classroom observations #1</b>	PEC classes	25	25	100%
<b>Classroom observations #2</b>	PEC classes	25	19	76%
<b>Endline</b>	Students	375	375	100%
	Teachers	75	73	97%

### Indicator 2.2.1. Changes in teachers' knowledge, behavior and practices

#### Teachers completed ASER test with children and can produce the results in the ASER logs

J-PAL provided teachers with a register (ASER logs)<sup>10</sup> where in which they were expected to record students' ASER results. The IPA team were able to access those files during their round of data collection.

During the different rounds of data collection, IPA enumerators always had access to the logs that reported individual ASER test results. Also, J-PAL shared with IPA on Google Drive students' ASER test results at schools' level.

Table 2: Number of schools where teachers' logs were available per waves where it was checked

	Baseline	Classroom observation 1	Classroom observation 2
<b># Schools (out of 25) where teachers logs were available and checked by IPA</b>	25	25	25

IPA encountered difficulty accessing teachers' logs in a few schools because the logs had been lent to advisors for monitoring purposes (data entry). IPA observed that some teachers were using a handwritten

<sup>10</sup> The teachers' logs' template is available at Annex 5.

list of students' groups when they had no access to the logs. For smooth implementation, teachers must have access to those logs before or during each PEC classes.

✓ *Teachers are performing PEC activities during classroom observations*

The goal of the classroom observation was to collect data to inform the quality of the program implementation. IPA did two waves<sup>11</sup> of unannounced classroom observations. For each wave of classroom observations, at least one teacher per school was observed during a PEC session. In total, IPA observers performed 25 classroom observations during the first wave and 19 classroom observations<sup>12</sup> during the second wave (Annex 1).

Note that IPA observers' goal to perform unannounced observations was not perfectly achieved during the different waves of classroom observations. In fact, there are reasons to suspect that schools' directors were informed about IPA observers' visit. For instance, it is normal for teachers and directors to refuse any interviews without authorization from their inspector. To facilitate IPA observers' integration in schools, it was agreed that IPA should inform inspectors about any surveys at least two weeks before fieldwork. IPA provided *inspectors* with survey periods one to two weeks without sharing the exact dates enumerators would visit each school. But, given directors' close relationships with inspectors, it is quite likely that the directors, and then teachers, were informed by the inspectors about the likely period during which these visits would happen, and then make sure to implement PEC during those periods only. In retrospect, it may have been possible to achieve smooth and genuinely unexpected observations by obtaining pre-emptive blanket permission from all stakeholders, including inspectors, for IPA's observations. This would also make it simple for observers to visit a school repeatedly, at different times, to see how a teacher handles each of their different classes. IPA recommends these tactics for any further roll-out of PEC.

Furthermore, besides observing teachers in the classroom, observers were assigned to collect some administrative data during their visit. As such, they often arrived at schools an hour before the start of PEC activities. The observer's early arrival could have given teachers time to adapt their daily schedule to implement PEC. However, this sort of spur-of-the-moment compliance would have required coordination with other teachers, given that PEC activities require multiple teachers. In addition, this would be challenging because it takes time for the teacher to rearrange the classroom furniture for the PEC class.

Results of the classroom observations indicate that in all cases when the IPA enumerator went to visit a classroom and was able<sup>13</sup> to observe a PEC session, a PEC session was effectively happening – Except on case during the second wave of classroom observations where implementing PEC activities with *students*

---

<sup>11</sup> Note that the delay between IPA's classroom observations and the teacher-led last ASER test were respectively three months and one week for the first and second classroom observation survey wave. The delay for implementing the first classroom observation was due to teachers' strikes, which ended with the year-end holiday.

For the first classroom observation, teacher testing ended the 15th of October 2018 and IPA performed classroom observation from January 13<sup>th</sup> to 19<sup>th</sup>, 2019. The second classroom observation took place from April 8 to 11, 2019 while teachers' ASER test was implemented the week of the 25 of March.

<sup>12</sup> Note that the rate during the second wave of observation was explained by the absence in schools of three of expected random selected teachers to be observed and the remain teachers were performing non PEC activities.

<sup>13</sup> During the first wave of classroom observations, all the 25 targeted teachers were performing PEC activities. While during the second wave, 19 out the 25 targeted teachers were performing PEC activities – Among the six remain, three were absent during the IPA enumerators' visit and three were performing non-PEC activities with their students.

*sit in rows*. The main parameters observers relied on to decide whether a PEC class was effectively implemented are:

- classroom arrangement – benches are pushed to the sides and floor is used or activities are performing outside classroom
- PEC materials are used
- PEC curriculum is being taught by teachers.

However, given the strong implementation fidelity observed on the rest of the program, we believe that teachers were implementing PEC consistently, not only when IPA made our visits. For example, we observed a high use of correct PEC materials; correct grouping of students; organization of the classroom; and positive feedback from mentors and teachers. All of this together makes us more confident that the teachers were indeed carrying out PEC hour consistently. Even if they were partially responding to advance notice, they were able to effectively implement the approach, which requires practice.

Table 6: Number of observed classrooms where PEC activities were implemented

# Observed classrooms where ..	Classroom observations 1 (out of 25)	Classroom observations 2 (out of 19)
<b>Tables are pushed to the sides and the floor is used</b>	22	16
<b>Students sit in rows</b>	-	1
<b>Activities happen outside class \ outdoors</b>	3	2

#### ✓ *Students are in the right group for their level*

To assess whether students were assigned to the group adapted to their level during PEC sessions, IPA took a two-step approach.

The first step was to check if the number of group levels created (or observed) in the school was consistent with the performance of students on the ASER test led by teachers. This tests whether the teachers were *grouping students correctly*. Indeed, students learning levels recorded in teachers' logs after the ASER test were used to organize students into several groups. So, this first step's goal will be to verify if the observed number of groups matched the expected number of group levels based on teachers logs.

The second part of the evaluation consisted of randomly selecting five students in a given PEC class, noting what level group they were allocated to in the classroom at that time, and cross-checking their observed level with the level they were supposed to be in according to the teachers' ASER logs (or teachers' registers).

Both of these facets of right-group-level-assignment is discussed in turn below.

#### ✓ *Were teachers creating the correct number of groups?*

Prior to classroom observations, data on the PEC class schedules were collected in all schools to allow IPA enumerators to perform unannounced observation of teachers' PEC sessions. In each school, besides the classroom they were assigned to observe, IPA enumerators were asked to identify the number of effective PEC sessions (or PEC group levels) implemented in each classroom selected for random observation. Then, enumerators would determine the expected number of PEC classes for that classroom based on the teachers' ASER test result logs.

Analysis of data from the second round of classroom observations shows that in 15 out of 19 schools (79 percent), the number of observed PEC classes matched the number of expected PEC classes. IPA's

enumerators investigated the reasons of the discrepancy inside the four remain schools by asking the teacher. The main reported reason (see below figure) was that *"the schools temporarily merged groups because a teacher was absent"* the day enumerators visited the schools.

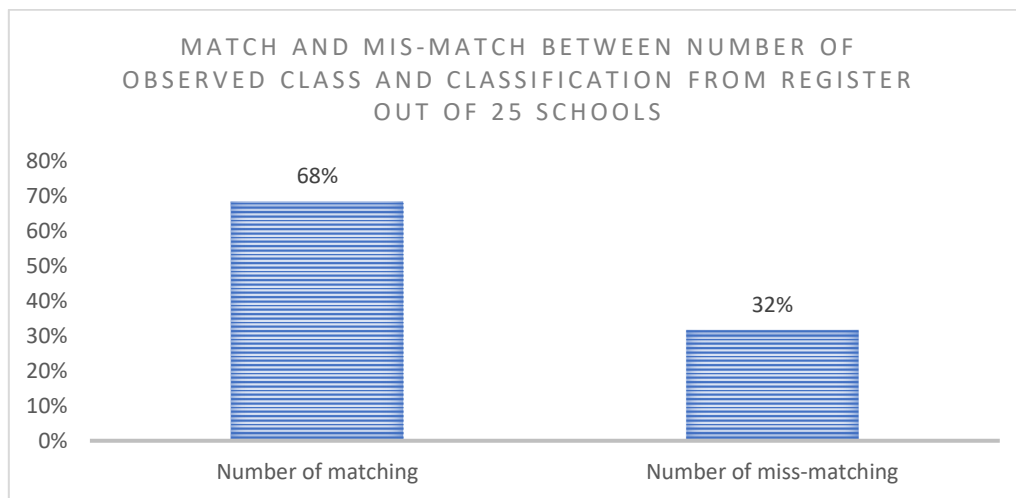
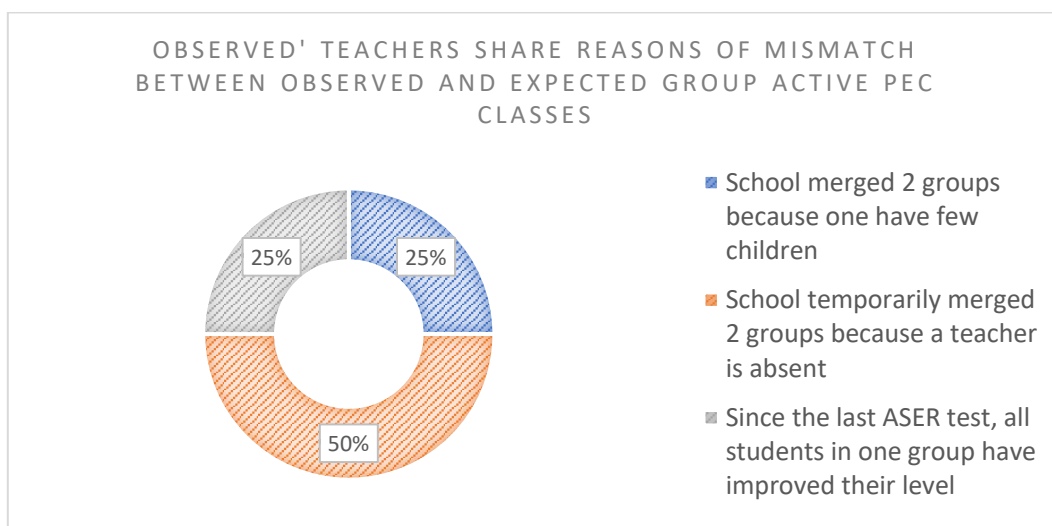


Figure 4: Share of reasons of mismatch among the seven mismatch schools.



Teachers' absence in schools caused *schools to temporarily merge group levels*, which explains some discrepancy observed in the students' assignment to the correct PEC group (see next section). Other reasons for mismatch between the number of observed PEC classes and the number of expected PEC classes were: *"School merged groups levels because one group have few children"* and *"Since the last ASER test, all students in one group have improved their level"*.

Besides the three schools, IPA observers did not performed observations given that selected teachers were absent, another take away from this result is that in 2 out 19 schools (11 percent) there was at least one teacher responsible for the PEC class who were absent. That gives an idea of teachers' attendance rate during PEC hours – at least 20 percent of teachers (5 out of 25 selected teachers) was absent in schools during PEC hours.

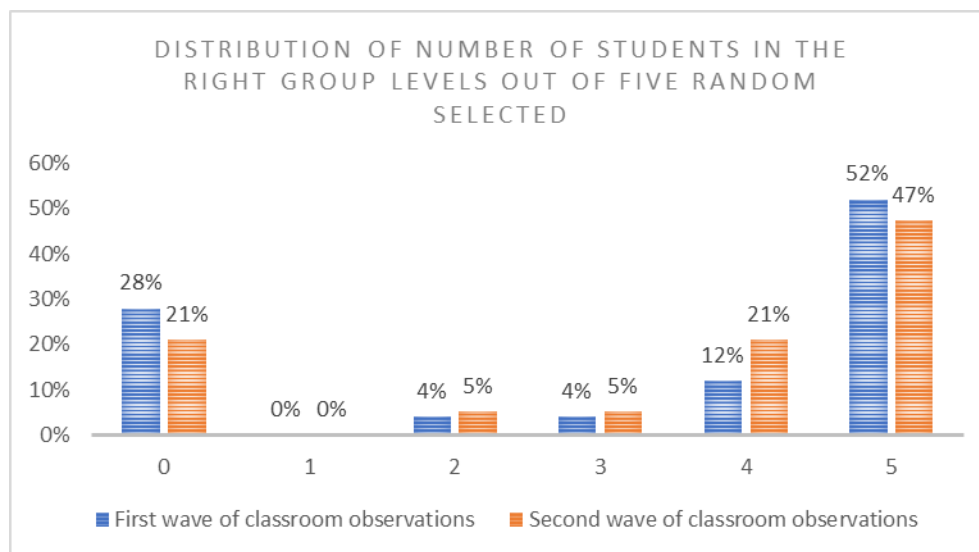
**Recommendation for scale-up:** At scale, IPA recommend adopting a mitigation strategy for teachers' absenteeism: for instance, implement PEC in separate groups even if a teacher is absent, monitoring absenteeism levels, or building in interventions to reduce absenteeism during PEC hour.

### *Are students actually in the group level they were supposed to be in?*

Results from the first classroom observation revealed that out of the 125 students randomly selected, 65 percent of students are in their reported level in teacher logs corresponded to their observed level. In 50 percent of the schools (13 schools) all randomly selected students were in the correct group, whereas in seven schools no student was in the correct group. In those seven schools, either the student logs or the grouping of students on the day were likely incorrect.

Results from the second classroom observation show that 66 out of the 95 students (69 percent) randomly selected perfectly matched – very similar to the first classroom observation.

*Figure 5: Number of PEC classrooms observed vs Number of students in their reported group levels among five random selected*



Just as teacher absences caused PEC sessions to be missed, teacher absences caused students to be assigned to the wrong group. For instance, when different group levels were merged (regardless of the reason why, but most often because the teacher was absent) the new teacher, often without access to the group list or ASER results, had to guess about the group levels the majority of students should be in.

### ✓ **Indicator 2.2.2 Teachers show effective use of the PEC materials during classroom observations**

To measure the effective use of PEC materials by teachers, J-PAL supported Ministry of Education staff on the pilot in developing a classroom observation tool that teachers' mentors should use to perform checking of teachers during their school visit (See Annex 6). The goal was to compute teachers' score between 0-10 on a composite index that captures three dimensions:

1. Students' grouping by level and progress
2. Use of materials and activities
3. Students involvement and teachers' position

IPA classroom observations' tool contains the same information as the mentors' classroom observations' tool but provide a more detailed account of activities in the classroom. Data used in this analysis were collected by IPA observers.

Results show that on average, teachers scored over 8/10 during the two waves of classroom observations, with scores similar in the two waves as displayed in Table 3. The minimum score was the same in both waves of observation, but the maximum score increased to the highest available in the rubric, a 10.

*Table 3: Trend of teachers' scores across classroom observations waves.*

	First classroom observations	Second classroom observations
Means	7.96	8.74
Standard deviation	0.93	1.15
Min	6.00	6.00
Max	9.00	10.00

Table 4 shows short analysis of teachers scores components – Note that teachers scores is a composite index filled by three components listed above. Results show that there is little change between the first and second observation the components. The only exception is that teachers improved at “*students grouping and progress component*” – the scores for this component move from 1.48 to 2.42 (out of 3) across waves, but this is largely driven by the fact that one of the three points on this score is for moving children up to a new level, which was not possible in the first observation.

Table 4: Teachers scores across classroom observation waves

	First classroom observations	Second classroom observations	Diff.	P-Value	Sig.
<b>Students grouping and progress (score out of 3)</b>	1.48	2.42	-0.94	0.0000	***
	(0.59)	(0.51)	(0.72)		
<b>Use of materials and activities (score out of 4)</b>	4.00	3.89	0.11	0.3306	
	(0.00)	(0.46)	(0.30)		
<b>Students participation &amp; Teachers position (score out of 4)</b>	2.48	2.42	0.06	0.7824	
	(0.59)	(0.77)	(0.66)		
<b>Teachers scores (score out of 10)</b>	7.96	8.74	-0.78	0.0216	**
	(0.94)	(1.15)	(1.09)		

#### ✓ Teachers administer the ASER test to students accurately

To check whether teachers accurately delivered the ASER test to students, we compared the evaluation conducted by teachers to the one performed during the independent evaluation by IPA’s enumerators. Because it is possible that teachers administered the ASER test accurately at baseline but not endline, or vice versa, the comparison is conducted for both<sup>14</sup>. There is a section for the Baseline comparison which discusses Literacy and Numeracy results, then a section for the endline comparison that does the same.

In general, we observe that teachers were able to conduct the ASER test fairly accurately, with substantial agreement between IPA’s results and the teachers’ results. Generally speaking, teachers were more likely to rate the students’ ability higher than IPA, particularly at endline.

<sup>14</sup> For IPA’s baseline and endline, the delays between teachers and IPA’ enumerators evaluations were not more than one week – which did not allow an increase of student knowledge between the two evaluations.



Note that teachers conducted their evaluation on all students in the classroom, while IPA's assessment focused on a sample of students. Thus, the analysis here consists of one-sample-test-of-proportions<sup>15</sup> to test whether the proportions of each learning level obtained after IPA assessors led their assessment is equal to the ones found with teachers' data on all students. The general rule is that we will consider that teachers accurately assess students on a given subject –literacy or numeracy – if at least half of the student levels<sup>16</sup> show statistical equality of proportions between the IPA and the teacher tests. Lastly, as an additional check on the quality of assessment at baseline, inter-rater agreement analysis is performed on IPA's sample to look at whether IPA assessors and teachers find the same level for a given student. Cohen's Kappa<sup>17</sup> test results – suitable for this kind of analysis - will be discussed.

Note that a discrepancy between IPA and the teachers' ratings could reflect an error by either IPA or the teachers. However, we consider that the IPA ratings were reliable, thanks to IPA's training and quality assurance measures.

Other than assessor error, discrepancies could be due to two other reasons. First, discrepancies could be related to the relationship between teachers and students. If students feel more stressed during the teacher's assessment than the enumerators assessment, they could have lower performance. Secondly, students could have made some progress during the two weeks between the teachers' assessment and our visit.

#### *Baseline - Teacher assessment vs independent assessment*

The first phase of teacher-led ASER tests ended on the 15th of October 2018 and the IPA independent evaluation team assessed students seven days after, from October 22nd to 26th, 2018.

#### **Literacy**

Results of the literacy baseline assessments shows that teachers generally assess children accurately (see table 5 below). This is key finding as the intervention relies on students being put into homogenous groups based on the results of these assessments and being taught at those levels.

Although IPA's assessors find comparable proportions of students in the words, paragraph and story, we observe that IPA assessors are more likely to put weak students at *beginner* level while teachers rate them more at *letters* level.

---

<sup>15</sup> The One-Sample Proportion Test is used to assess whether a proportion of a given indicators is equal to a known constant. The P-value computes along with the test can be interpreted as the probability that the difference is due to the chance. So higher it is, we cannot reject the assumption that there is equality between proportion and the constant at the given threshold. Thus, having a significant result here mean that we can reject the equality assumption. For statistical inference we will use a significance threshold of 5% ( $p$  value of 0.05).

<sup>16</sup> Literacy ASER test has five levels which are Beginner, Letters, Words and Paragraph and Story – we will look for at least three levels to be consistent. While Numeracy ASER test contains four levels: Beginner, 1-digit level, 2-digits level and 3-digits level – we will look for at least two levels to be consistent.

<sup>17</sup> The coefficient kappa of 0 occurs when observed agreement can be exactly accounted for by chance and the coefficient kappa of 1 occurs when there is complete agreement between raters. Kappa can yield a negative value when there is less observed agreement than is expected by chance.

Table 5: One-sample-proportion-test on ASER literacy levels – Baseline results

	Teachers results	IPA results	P-value	Sig.
Beginner	35%	51%	0.0005	***
Letters	42%	20%	0.0000	***
Words	9%	7%	0.4653	
Paragraph	9%	11%	0.3737	
Story	6%	10%	0.0721	*

We also performed an inter-rater agreement test with this data. For the literacy test, 58 percent of students were classified in the exact same group. As kids might perform differently during the different tests, one should expect some differences. Looking at the remaining 42 percent that were incorrectly classified, theoretically we consider that a child that was classified as able to read a letter when he can read a word is less problematic than classifying a child as being able to read a story when he can only read a letter.

For these reasons we performed an inter-rater agreement test, called the Cohen's Kappa<sup>18</sup> test. This test weighs the different scenarios, assigning a linear weight to close misses (e.g. a rank of 2 vs. a rank of 3) than to misses that are further apart (e.g. a rank of 1 vs. a rank of 5) based on the assumption that a difference of adjacent ranks is less critical than a difference that is farther apart.

Figure 6 presents this test and gives an overview of the overall performance. Green boxes indicate full agreement, meaning that the same child was allocated to the same group by teachers and IPA enumerators. A partial agreement in orange, refers to the situation where we observe a slight difference between IPA enumerators and teachers. This is particularly the case for the Letter and Word sections for teachers. In general, we see that the great majority of ratings were in agreement.

Figure 6: Inter rater agreement matrix literacy test

		IPA ASER					
		Beginner	Letter	Word	Paragraph	Story	
Teacher ASER	Beginner	106	9	1	2	0	<div>Agreement</div> <div>Partial Agreement</div> <div>Non Agreement</div>
	Letter	81	52	11	4	0	
	Word	5	11	8	15	4	
	Paragraph	1	4	6	15	10	
	Story	0	0	0	7	23	

The result of the kappa weighted test gives us a final value of 87 percent of agreement while the expected agreement is 65%, reflecting "almost perfect" agreement according to the standard in the literature.<sup>19</sup>

## Numeracy

Results of the numeracy baseline assessments shows that teachers rated hardly students than IPA assessors. Teachers found more *Beginner* and less students with *3-digits level* (see table 6 below). This is

<sup>18</sup> The coefficient kappa of 0 occurs when observed agreement can be exactly accounted for by chance and the coefficient kappa of 1 occurs when there is complete agreement between raters. Kappa can yield a negative value when there is less observed agreement than is expected by chance.

<sup>19</sup> Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". *Biometrics*. **33** (1): 159–174.

crucial finding as the intervention relies on students being put into homogenous groups based on the results of these.


Table 6: One-sample-proportion-test on ASER numeracy levels – Baseline results

	Teachers results	IPA results	P-value	Sig.
Beginner	12%	4%	0.0113	**
1-digit	32%	32%	0.9755	
2-digits	28%	8%	0.0000	***
3-digits	28%	57%	0.0000	***

Figure 7 presents results from the two sources of data, IPA and the teachers. We observe that teachers tend to be more likely to classify children as beginners compared to IPA's enumerators. The result of the kappa weighted test gives us a final value of 86 percent of agreement while the expected agreement is 63%, indicating "substantial" agreement.

Figure 7: Inter rater agreement matrix numeracy test

		IPA ASER			
		Beginner	1 digit	2 digits	3 digits
Teacher ASER	Beginner	14	13	1	6
	1 digit	1	91	7	20
	2 digits	0	13	20	61
	3 digits	0	2	1	125



Agreement

Partial Agreement

Non Agreement

#### Endline – Teacher assessment vs independent assessment

The last phase of teachers testing ended the 24 of May 2019, while IPA's independent evaluation team collected data from May 13<sup>th</sup> to 17<sup>th</sup>, 2019.

#### Literacy

Results of the literacy endline assessments shows that teachers were somewhat more likely to assess children as having a higher level of literacy than IPA's enumerators (see Table 7 below). In several cases the difference between teachers' and IPA's assessments are statistically significant. It is possible that this reflects teachers facing slightly different incentives than they did at baseline.

Table 7: One-sample-proportion-test on ASER literacy levels – Endline results

	Teachers results	IPA results	P-value	Sig.
Beginner	2%	27%	0.0000	***
Letters	7%	30%	0.0000	***
Words	26%	11%	0.0004	***
Paragraph	27%	14%	0.0038	***
Story	37%	18%	0.0001	***

## Numeracy

Results of the numeracy endline assessments shows that teachers were somewhat more likely to assess children as having a higher level of literacy than IPA's enumerators (see Table 8 below). It appears that the reliability of teachers' assessments may have declined since the baseline.

In summary, the main discrepancy is among students who are less than *3-digits level*, with teachers more likely to put children in the 2-digit level, while IPA's results put more children at the 1-digit level. Teachers would likely say that students simply gained skills in between when they tested them and when IPA tested them, but there is no way to confirm that. XX

Table 8: One-sample-proportion-test on ASER numeracy levels – Endline results

	Teachers results	IPA results	P-value	Sig.
Beginner	0%	2%	0.0000	***
1-digit	3%	11%	0.0000	***
2-digits	19%	6%	0.0013	***
3-digits	78%	80%	0.6156	

### ✓ Indicator 2.2.3 Teachers reported having a coaching session with schools' principals or inspectors

As part of the monitoring of PEC implementation, school directors and pedagogical advisors were designated as mentors, and as such were expected to provide first assistance to teachers.

During the endline, we collected data to check if mentors were performing their coaching work. When asked if they received assistance on PEC during the last three months, all interviewed teachers said that they received assistance from their school director, 99 percent of them said that they got assistance from their pedagogical advisors, and 14 percent said they got assistance from their inspectors (see Table 9).

Table 9: Assistance teachers receive from mentors during three the last months

(N=73)	Frequency	Proportion
Teachers who had not received assistance	0	0%
Teachers who benefited assistance from schools' directors	73	100%
Teachers who benefited assistance from pedagogical advisors	72	99%
Teachers who benefited assistance from inspectors	10	14%

There is wide variation in the number of days on which a teacher received a visit from a mentor. For instance, while some teachers reported that they only ever received assistance from their school director on two days, other teachers said that they received assistance on 50 different days during the last three months. Overall, schools' directors performed more support visits than pedagogical advisors or inspectors. – which is a predictable result given that teachers share the same workplace with their director. However, pedagogical advisors did more than they should which appreciate. Note that pedagogical advisors were expected to perform one unannounced visit per school per month and they provided in average 9 days of assistances during the last three months.

Table 10: Numbers of days of assistance provide by mentors during the last three months

(N=73)	Mean	Standard deviation	Min	Max
Number of assistances provide by schools' directors	18	13	2	50
Number of assistances provide by pedagogical advisors	9	9	1	36
Number of assistances provide by inspectors	1	2	0	6



### **2.3. Beneficiary feedback about the program is positive**

All outputs were achieved at the expected quality. Data show that the trainings had a very high participation rate and PEC materials were received in all schools, albeit delayed by a few days.

#### **Indicator 2.3.1.a Beneficiaries provide positive feedback on the delivery of outputs**

Data show in this session come from different interviewed IPA enumerators administrated to teachers during the baseline and the endline surveys. But some data were also picked from spot-check IPA team performed on the field. 73 teachers were interviewed during the baseline and the endline data collection among 75 expected.

Teachers were interviewed about their level of confidence and on their ability to successfully implement the PEC's methodology in their classes. We found that teachers were increasingly confident in their ability to deliver this new methodology and on students' ability to progress.

#### **Teachers provide positive feedback on the PEC materials/activities**

At the beginning of the pilot, almost all teachers (97 percent) reported that they were confident in their ability to implement the PEC methodology. Results also show that at least 96 percent of teachers think that students will be able to learn more using the PEC approach.

This percentage did not change over time. (See Table 11 below)

Table 11: Teachers confidence in implementing the PEC methodology

(N=73)	Baseline	Endline
% of teachers who are confident about their ability to implement the PEC methodology	97%	-
% teachers who think that students will be able to learn more with PEC approach	96%	97%

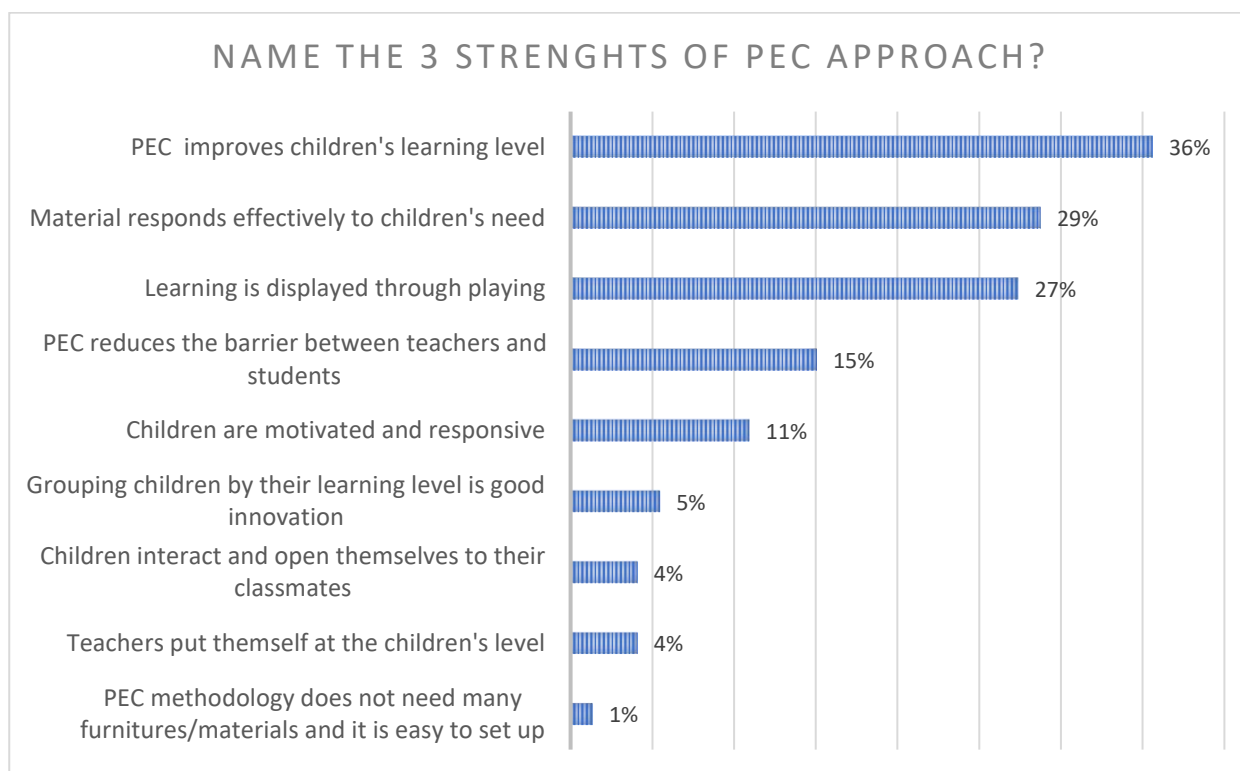
#### **What are the strengths of PEC approach?**

At endline, IPA agents asked teachers a number of open-ended questions to capture what they perceived as strengths in the PEC approach. The questions were open-ended and IPA later coded teachers' answers into themes. When asked to list the three biggest assets they found in PEC approach, teachers' most common answers thematically were "PEC improve children learning level (36 percent of respondents), "Material responds effectively to children's need" (29 percent of respondents) and, "Learning is displayed through playing" (27 percent of respondents). (See Figure 8 for more information).

Qualitative endline survey corroborates the results of the quantitative survey. Among the strengths of the PEC pedagogy's qualities facilitators cited the following:

- PEC helps children improve in school. *"PEC really helps to improve the level of children and I want to put it in all schools".*
- The use of games as a pedagogical tool. *"It's a good thing because children learn by playing*

Figure 8: Reported teachers' feedback regarding PEC approach assets



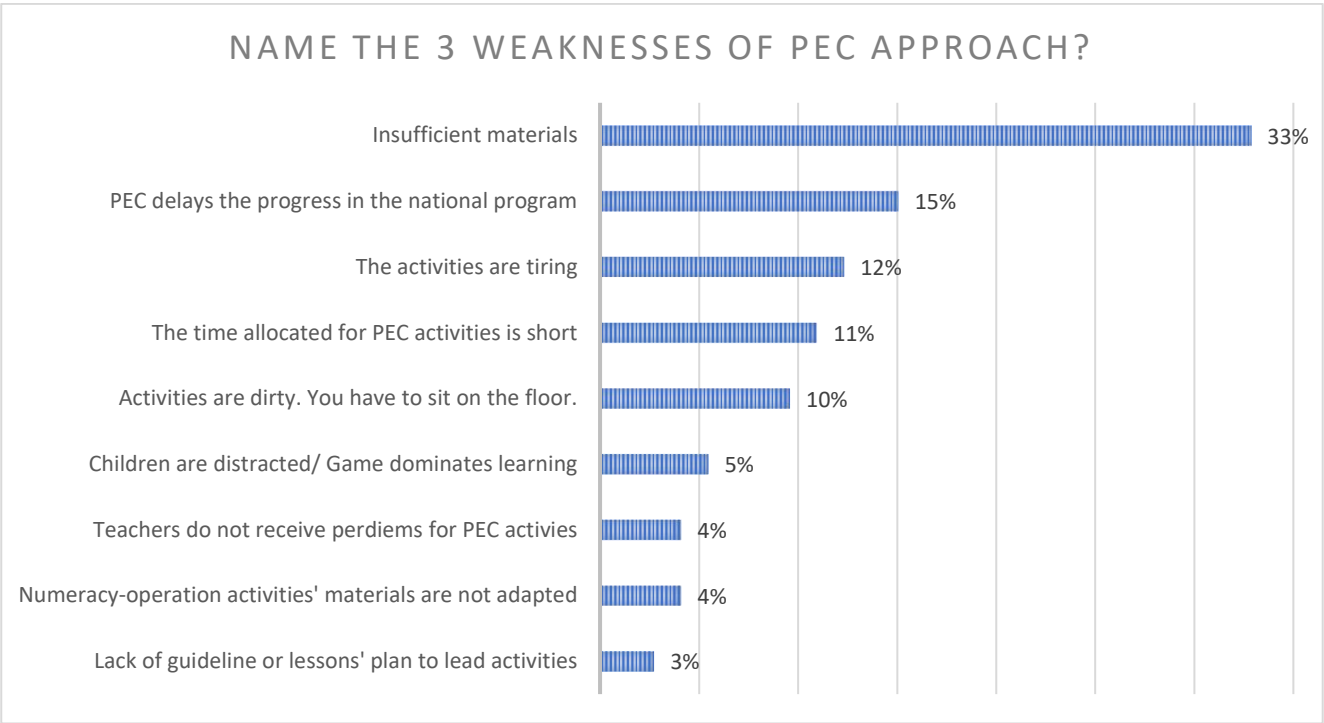
#### *What did teachers report as the PEC approach's weaknesses?*

Despite their favorable opinion about the program, another open-ended question was asked to teachers to mention three main weakness of PEC approach, if any. The main answers from teachers – again, received in a semi-structured interview and later coded thematically; the most common codes are reported here - regarding this question were *"Insufficient materials"* (33 percent of respondents), *"PEC implementation delays the progress in the national curriculum"* (15 percent of respondents), *"PEC activities are tiring"* (12 percent of respondents). These are interesting weakness that should be considered for the scale-up given that it point out some expectations issues at teachers side – like some expected perdiems to implement PEC (see figure9)

Because the national curriculum is challenging for teachers to complete in a year even when they are merely lecturing students without checking for comprehension, it is understandable that they find PEC, which requires making sure students understand lessons, to be an obstacle to completing the syllabus: comprehension takes more time than rote listening. By the same token, because teachers need more time for each lesson when implementing PEC, it is understandable that they feel there is not enough time allocated to PEC. These responses indicate that teachers are doing PEC well: if they sped through it, it would lose its benefits. However, they also highlight that it is only possible to enact meaningful pedagogical change if the MENET-FP is on board. If teachers and schools face consequences for slowing their lessons to ensure more students understand, it will be very difficult to achieve wide and consistent adoption of PEC methods. MENET-FP can facilitate scale-up by creating more flexible national syllabuses

and rewarding improvements in students’ grade level while de-emphasizing teachers’ completion of annual curriculums.

Figure 9: Reported teachers’ feedback regarding PEC approach weaknesses

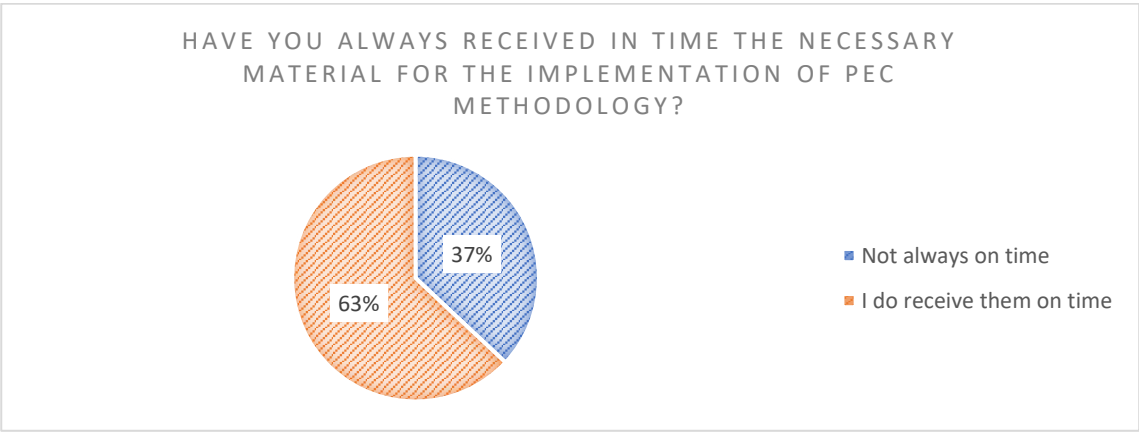


*Were PEC materials available throughout the pilot implementation?*

During the baseline survey, almost all teachers reported have not yet received PEC materials. This was not disturbing since at this time the program implementation was not started and implementers started sending materials in pilot’s schools.

Even if along the school year there was substantial improvement with PEC materials delivery to teachers, we notice that necessary materials have not reached teachers always on time. – 63 percent of teachers said that they have not received always on time necessary materials for PEC implementation during the school year.

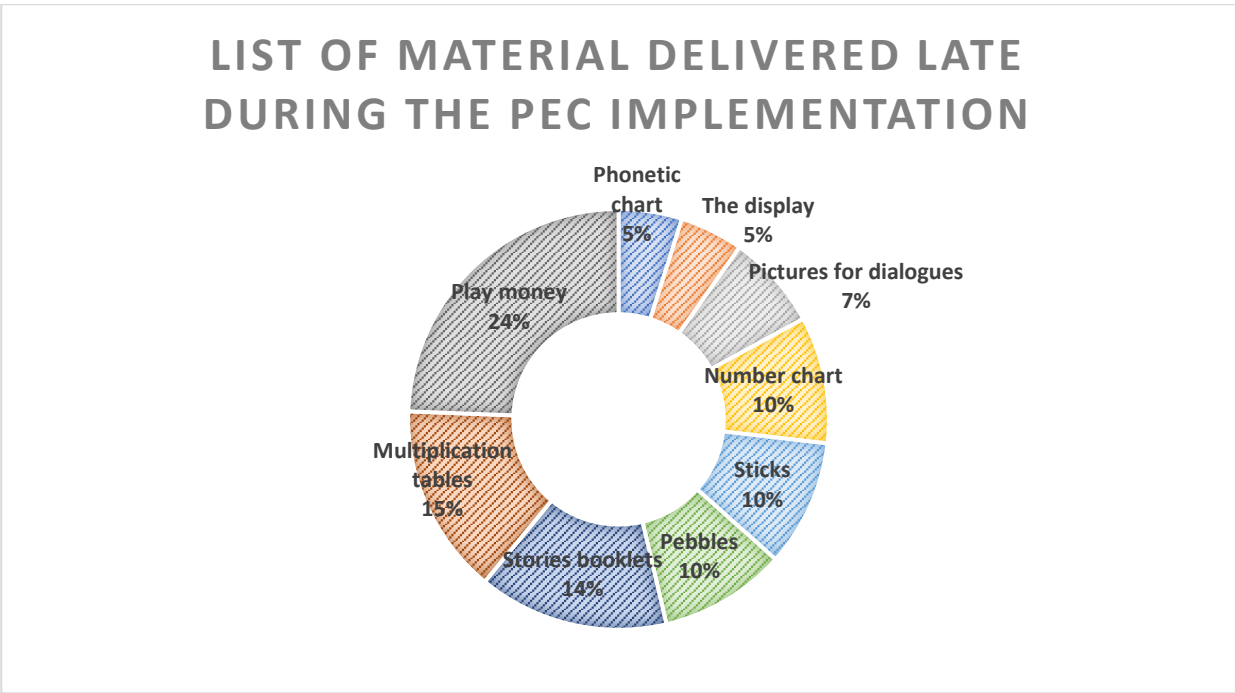
Figure 10: Percentage of teachers who report getting all necessary PEC materials during the implementation



According to the teachers, the lack of material had an impact on their performance. 89 percent of teachers, who complain about non availability of PEC materials on time, think that they could have performed better if all the materials had been available. The teachers listed eight materials which were usually missing- three of which cover 80 percent of teachers’ complaints. Among these main missing materials, we found *Play money* (37 percent of complaints), *Multiplication tables* (22 percent of complaints) and *Stories booklets* (22 percent of complaints).

See Figure 11 below for an exhaustive list of all materials which were usually missing during the pilot implementation.

Figure 11: List of materials delivered late during the PEC implementation



**Recommendation for scale-up:** Learning materials contribute a lot to the effectiveness of teachers. As above mentioned, PEC materials were not available on time, and teachers complained in many ways about this lack. Implementer should think about how to avoid PEC materials missing on teachers’ desk given that can imply the decrease of their enthusiasm since they cannot work effectively. Teachers agree to recommend PEC approach for all primary schools in Côte d’Ivoire (CI)

The percentage of teachers who are ready to recommend PEC for a full roll-out in Ivory Coast remained constant from baseline to endline – More than 95 percent of interviewed teachers are still available to recommend PEC approach to be taught in all primary schools.

During spot check visits, teachers were asked what they think would be necessary for PEC to be implemented successfully across the country. They suggested specific outfits to facilitate their work during PEC session as well as additional remuneration. They also suggested to resize the group levels, to build classroom adapted for PEC session and also ensuring materials availability. It is possible that these ideas were only came up because IPA solicited suggestions; it is not entirely clear that they reflect real bottlenecks to implementation.



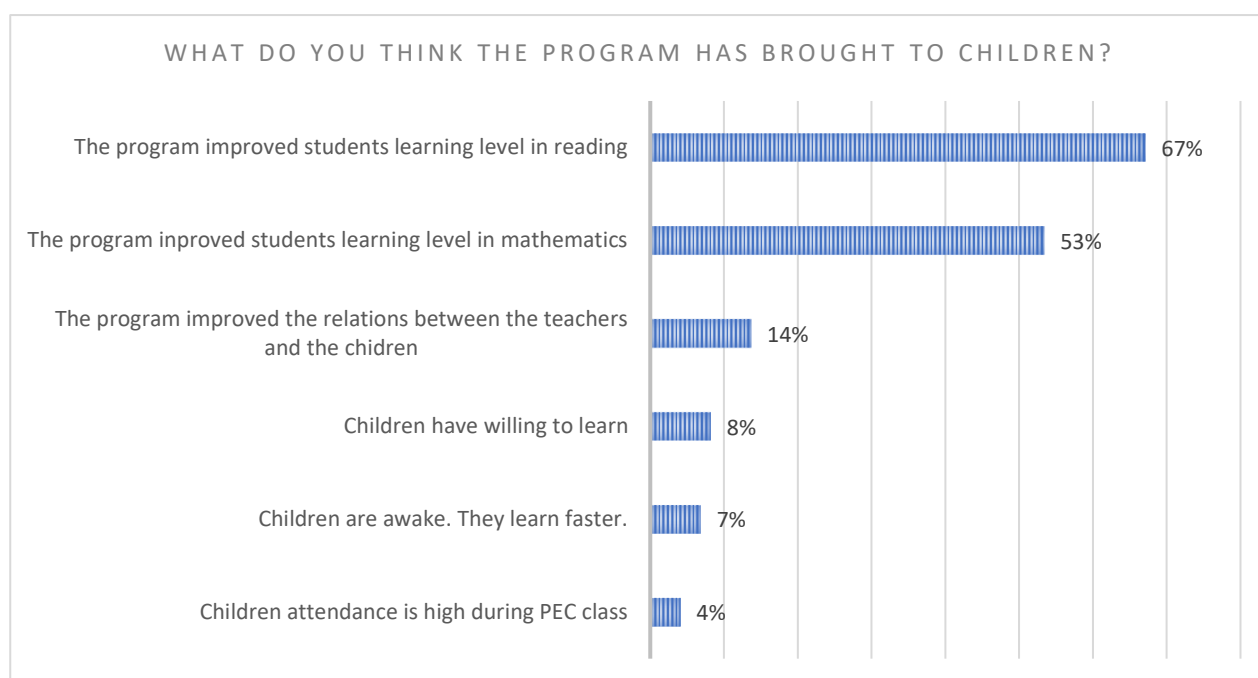
Table 12: Percentage of teachers who agree to recommend PEC in CI across waves of data collection

	Baseline	Endline
% of teachers who agree to recommend PEC to be taught in all primary schools in CI	95%	96%

✓ **Indicator 2.3.1.b Teachers provided positive feedback on the improvement on their students' learning in the classroom as result of PEC methods**

Almost all interviewed teachers gave positive feedback on the improvement of the students. For instance, 67 percent of them said the program improved students learning level in reading and 53 percent of them said the program improved students learning level in mathematics. In addition, 14 percent of them indicated that they thought that the PEC improved student-teacher relationships. More than 7 percent of the teachers reported that children are awake, willing to learn and learn faster with PEC. This is not proof that PEC had any of these effects, but it does give us insight into teachers' thoughts about the value of PEC.

Figure 12: Feedback of the beneficiaries on the improvement of their students learning.



✓ **Indicator 2.3.1.c Beneficiaries describe positive experiences with the program**

Interviews with the teachers, inspectors, pedagogical advisors and school principals show that they are satisfied with the project, including the training and coaching.

IPA conducted a spot check visit in the communities of intervention from May 6<sup>th</sup> to 11<sup>th</sup> 2019. This mission's main objective was to understand the mentoring system and the mentors' activities as well as to gain a better idea of the difficulties met by the facilitators in the implementation of the pedagogy. IPA

also used this opportunity to gather the facilitators' opinion about their training and the implementation of the PEC pedagogy. For this purpose, we conducted qualitative semi-structured interviews with:

1. 11 Facilitators to check their knowledge on the PEC, understand the difficulties they encounter in pilot implementation and understand the pedagogical follow-up they receive.
2. 7 pilot schools directors visited to verify their knowledge of CEP, evaluate the pedagogical assistance they provide to rural facilitators and understand the difficulties related to their activity.
3. 4 pedagogical advisers to verify knowledge of the CEP, evaluate the pedagogical assistance it provides to rural facilitators and Understand the difficulties related to their activity.
4. 2 primary education inspectors to understand their role in the program in addition to their knowledge of CEP

#### *Facilitators' Feedback about their training*

Facilitators had mixed opinions regarding their training, citing negatives and positives. 10 out of the 11 facilitators interviewed reported that they were satisfied with the quality of their training. *"... we had good trainers, we understood the courses"*.

5 facilitators reported that the training was too short. As a result, they described their difficulties in applying TARL in practice. As all distributed material was not presented during training a few facilitators asked more information on how to use wooden pavers."

Despite comments about the training's duration, most facilitators reported that they were comfortable with PEC activities and tools. *"I am comfortable with PEC activities, the handling of materials. There are activities that help to awaken children."*

#### *Feedback about coaching:*

The directors seemed to be very involved in their role as mentors given the number of class visits, they did. 7 out of 8 directors interviewed reported that they conducted a visit at least once a week in all 3 classes under their supervision. According to the directors, class visits help facilitators to improve and solve problems, as illustrated by the following quote: *"[...] for example, the activity on informal dialogue was complicated but today it is ok (pictures, group work), I saw the problem during an observation. For example, it was difficult for facilitators to put the division in place (classes). So, I helped them to master this activity. "*

*Facilitators also have a good perception of the coaching. They considered their supervisors to be sufficiently skilled to fulfill their mission and that supervision visits helped them improve their work. "They are more trained than us, so they give us tips that help us improve. Example: They taught us to divide high levels to be more efficient". Also, facilitators reported that some of the issues they met in the implementation of the PEC were solved by their supervisors during class observations. "The director helped me get the kids working as a group, using the sticks. He noticed it (problem) during an observation »*

### 3. Costs and operations management

---

**Costs and operations management** is divided into two components, both of which the pilot is assessed on:

1. Costs are well managed
2. Project management is successful

The pilot appears to be consistent with the requirements for scale-up in both components of costs and operations management. The pilot's score for each component is determined by the pilot's performance against pre-agreed indicator-level criteria. Each component, including the criteria that determined the pilot's score, is discussed below.

#### ✓ **3.1 Costs are well-managed**

The project management team made efficient use of the resources and spent less than budgeted to implement all the activities of the pilot. However, when we estimate the cost per child of implementing such a program in Cote d'Ivoire (using the pilot cost structure but excluding all development costs and costs related to J-PAL activities), we found that it would cost twice more per child than another TaRL program in Ghana.

No budget was attached to the scale-up proposal. Therefore, we are not able to analyze the partners' vision of cost efficiency at scale.

[Details removed]

#### ✓ **3.2 Project management is successful**

Except for the strike that impacted most of the schools in December 2018 and January 2019, the project was implemented according to plan and no significant deviation was observed from the proposal.

The cooperation between the different partners was strong.

[Details removed]

## 4. Capacity to learn, improve and innovate

**Capacity to learn, improve and innovate** is divided into two components, both of which the pilot is assessed on:

1. Project collects credible monitoring data
2. Monitoring is used to learn and improve

The COMPANY PARTNER's pilot appears to be consistent with the requirements for scale-up with respect to capacity for learning. The pilot's score for each component is determined by the pilot's performance against pre-agreed indicator-level criteria. Each component, including the criteria that determined the pilot's score, is discussed below.

### ✓ 4.1 Project collects credible monitoring data

During the pilot, J-PAL regularly collected high quality data on teachers' and mentors' attendance at trainings, assessment data of students' skill levels, and classroom observations data on teacher performance during PEC classes. The only key threat to data reliability was that teachers likely face incentives to over-mark their students at endline in order to show inspectors they have done a good job during the pilot.

#### ✓ Indicator 4.1.1 Routine monitoring data are collected and shared on time with stakeholders

During the pilot, J-PAL collected 3 main types of information regularly:

1. **Training attendance data.** Every day of the training, all training participants (mentors or teachers) were asked to sign an attendance list containing all relevant information (date, training's objective, time and place of the training etc.).
2. **Assessment data.** Teachers conducted ASER tests for all their students three times over the pilot period. Teachers documented the results in a log provided by J-PAL.
3. **Classroom observation data:** Mentors collected information regarding the facilitator's practice in the classroom during PEC activities every month.

Most PEC programs collect data on students' skill levels and the frequency of teachers' implementation of PEC lessons. It is to the pilot's credit that training attendance data was also collected.

As described in the output section, IPA encountered difficulty accessing teachers logs in a few schools because the logs had been lent to advisors for data entry for monitoring purposes. For smooth implementation, teachers must have access to those logs before or during each PEC classes

**Recommendation for scale-up:** At scale, we recommend giving clear rules for pedagogical advisors to collect the assessment data: they should give back the logs to the teachers before the beginning of the PEC classes.

J-PAL saved attendance and assessment data on a Google Drive shared with IPA for review, but we did not receive the classroom observation data. According to J-PAL, classroom observation data were not digitized. Moreover, J-PAL and MENET-FP communicated to all partners key results and updates on the pilot's progress, success and challenges, during operational and advisory meetings. Therefore, IPA has been able to draw conclusions even without seeing the classroom observation data firsthand.

✓ Indicator 4.1.2 IPA's spot-check visits confirm the quality and accuracy of the data shared by the partner

IPA conducted spot checks from May 6<sup>th</sup> to 11<sup>th</sup> 2019 in four communities. The objectives were to understand the structure of the mentoring and assess the functioning of monitoring activities through interviews with mentors and teachers.

The spot check found that teachers and mentors interviewed have a good knowledge of the PEC approach. Mentors showed a more detailed knowledge about the program than teachers, which is to be expected as mentors receive almost twice as many days of training as teachers do (11 and 6 days, respectively). Teachers' knowledge seemed adequate for implementing PEC.

Through interviews with pedagogical advisors from the MEN, we found that all have a clear vision of their role on the pilot: *"I observe the teachers during a session. Then, we exchange about the session observed. Teachers are self-critical. The principal can also provide feedback. At the end, I give some pedagogical recommendations. During each visit, I produce a bulletin in which I put all the observations for the facilitator to refer to"*. Interviews with teachers revealed that regular visits from mentors (meaning directors, pedagogical advisors, and inspectors) helped them to better conduct activities. For instance, most teachers had difficulties with the PEC approach on divisions. Among them, most declared they have understood via mentoring sessions.

IPA also interviewed the inspector about his role. But the inspector did not know his role well as he had just joined and did not participate in the mentor training. The one who was involved in the project has been assigned in other locality's inspection. So for COMPANY PARTNER's pilot, it's only the principal pedagogical advisor who is the main contact for the inspection.

Finally, we asked all parties about the difficulties they faced. Teachers and mentors reported facing difficulties with some activities. For instance, some teachers had difficulties with the division and reorganizing the classroom every day and the pupil's money handling. Also, their reported had difficulties with some PEC words which was difficult to understand for the pupils. We shared their difficulties with J-PAL during the spot-check feedback meeting. J-PAL was aware of all difficulties through their most recent monthly monitoring meeting on the field with mentors. When possible, remediation actions were taken.

The credibility of data is assessed on three axes:

- **Validity**: valid data accurately captures the core concept one is seeking to measure
- **Reliability**: implies that the same data collection procedure, implemented repeatedly to measure the same reality, will produce consistently similar data.
- **Unbiased**: data does not have systematic errors

Figure 14 shows our overall conclusions, and each data source is discussed in detail below.

Figure 14: Credible data criteria

Credible data criteria	Valid	Reliable	Unbiased
Training attendance data	✓	✓	✓
Assessment data	✓	✓	✓
Classroom observation data	✓	✓	✓

#### Training attendance data

We are confident in the quality of the attendance data for different training sessions organized by the project management team. The data were collected daily during each training session. Baseline data and discussions confirmed the participation of the different actors. Participants signed the attendance sheet, reducing the risk of errors and cheating.

#### Assessment data.

We are confident that ASER, the testing tool used by the teachers to assess the level of the students, reported meaningful results. ASER is a common test for checking students' progress through frequent, formative assessments and was carefully adapted to the context and age of the pilot's students. Moreover, all teachers were carefully trained to conduct this test. However, the endline results may be biased. Teachers could see their students' assessment as a proxy for the quality of their teaching, creating an incentive to overestimate students' progress. In fact, J-PAL has observed this before in a PEC program in India.

#### Classroom observation data

We are confident in the tool used for the classroom observation data. It is an adaptation of a widely used tool. Mentors were carefully trained for 11 days and then had a refresher training focus on their role as a mentor. Therefore, we believe the way they implemented the test is reliable. There is no reason to think that mentors would introduce systematic bias into the data.

## ✓ 4.2 Monitoring is used to learn and improve

Collecting monitoring information during PEC program is crucial for success. During the pilot, regular review meetings and feedback loops ensured that information collected was acted upon.

### ✓ Indicator 4.2.1 Monitoring data is actionable and aligned with program management

During the pilot, J-PAL collected data on training attendance, teacher ASER assessments, and classroom observation. As previously noted, this information goes beyond ordinary PEC program monitoring by including attendance data. In our judgement, these indicators capture the key program management data required for corrective action to ensure the project is correctly implemented. Examples of how this was done are in the following section.

### ✓ Indicator 4.2.2 Program improvement in response to monitoring

PEC teaching requires frequent use of data by instructors and program staff, and the pilot showed implementers making appropriate use of the data they collected. First, teachers used assessment data to

group children into learning level groups. Information collected through oral one-on-one assessment also connects the instructor to the problems each child is facing (because they can see when and how children make mistakes during the test). This enables instructors to target the areas where the child needs support and motivates the instructor when the child improves.

The next instance of data regularly informing PEC programming is through mentors providing immediate individual feedback to teachers. Separately, the pilot implementation team invited all mentors to monthly meetings to celebrate what is going well in their schools and share difficulties they have seen teachers face during their observations. Then, potential solutions were discussed and, when possible, decided, with the support of the pilot coordination team. If necessary, the problem was also discussed in operational meetings. During the pilot, this process helped improve class management, implementation of a division calculation activity and other specific activities that were not well understood by the teachers. For instance, many teachers did not know how to use the “student money” material proposed as one of the activities of the approach.

## 5. Sustainability

**Sustainability** is divided into two components, both of which the pilot is assessed on:

1. Provides sustained benefit to the community
2. There are prospects for scale-up beyond GMM2

The COMPANY PARTNER pilot appears to be consistent with the requirements for scale-up in both components of Sustainability. The pilot’s score for each component is determined by the pilot’s performance against pre-agreed indicator-level criteria. Each component, including the criteria that determined the pilot’s score, is discussed below.

### ✓ **5.1 Provides sustained benefit to the community**

Overall the enthusiasm of teachers and mentors is encouraging for the sustainability of the approach. However, it is important to ensure that teachers overcome the difficulties they face to implement PEC in their classes and to evaluate the feasibility of such extensive monitoring at scale.

#### ✓ **Indicator 5.1.1 Indications that the animators and mentors are likely to continue with the practices or program activities**

Overall, most teachers reported having a positive opinion about the PEC approach because they think it helps the children. Most of them would recommend the approach to be implemented in all schools in the country.

However, some difficulties faced by teachers during the pilot, if not solved at scale, might cause them to abandon PEC over time. In May 2019, some teachers interviewed by IPA mentioned that their training was too short, impacting their ability to put the methodology into practice. Indeed, they still had difficulties applying the PEC approach in their classes because they found it tiring, especially in classes that are overcrowded. *“The numbers in our classes are plethoric and this makes it difficult for the teachers to apply the PEC”,* one teacher reported. In theory, PEC should be useful when teachers are burdened by large classes, as it encourages students to work more independently. However, this does not seem to be teachers’ view. Since applying a new approach is often more tiring, this feedback is not entirely surprising.

**Recommendation for scale-up:** We recommend having additional trainings to accompany teachers and help them better understand the PEC pedagogy and the manner to perform some activities. The training should emphasize the advantages of the PEC approach with a large number of children and how to better use classroom space.

Overall, teachers and mentors were satisfied with mentoring activities. Mentoring activities are a key ingredient for the success of a PEC program. For instance, some teachers reported they did not understand how to group children by level during the teachers training, but mentors accompanied them in class and taught them how to do it.

However, the pilot identified two threats to the implementation of mentoring activities. First, school directors lack time. In addition to their duties as school directors most directors are in charge of a class that they teach. *"I don't have time because I'm holding a CM2 class so between my program and mentoring, I'm running out of time", said one.* Secondly, pedagogic advisors and inspectors struggle to conduct class observations because of difficulties commuting between classes placed under their supervision. It seems that this is due to the combination of the fact that the communities are sometimes hard to access and because they lack adequate means of transportation. *"It is difficult to move because my bike is old, I pay the fuel [sic; the program design called for advisors to be paid for transport costs, but complaints were still received], but my bike is old it breaks down a lot". "Areas are sometimes difficult to access".*

Moreover, it is not clear that intense monitoring is scalable in terms of budget and engagement of the mentors over time.

**Recommendation for scale-up:** One challenge of the scale up will be to maintain the effectiveness of the program and ensure its sustainability. It is important that the model is fully adaptable and does not generate exorbitant costs that will be difficult to bear if implemented on a national scale.

## ✓ **5.2 There are prospects for scale-up beyond GMM2**

Building on the strong governmental buy-in of the projects offers prospects for scale-up beyond the company funding exists.

### ✓ **Indicator 5.2.1. Potential for further scale-up, by the government of Côte d'Ivoire, the Cooperation Partner or other development actors**

The government of Côte d'Ivoire recognizes the challenges remaining on the way to universal, quality education. The MENETFP has established a framework for addressing these challenges in the Education Sector Plan 2016-2025. By 2025, the MENETFP aims at increasing the quality of primary education. PEC objectives are directly aligned with the government's intention to develop a program for "Reading, Writing and Arithmetic for All". In this context, the good performance of this pilot constitutes a future opportunity in line with national strategic objectives.

Moreover, the MENETFP has shown interest in the PEC approach. A working group including various high-level actors within the MENETFP was created to reflect how to scale up programs like the PEC. Some representatives had participated in a conference about PEC approach in the September 2018 in Johannesburg. The MENETFP was an engaged partner during the pilot. Finally, the PEC approach was also tested in another TRECC pilot, with the support of the government, in bridging schools.

UNICEF has also expressed strong interest in scaling PEC.

### ✓ **Indicator 5.2.1 Enabling and financing environment**

As described in the first section, the PEC, as implemented, is aligned with the priority of the different donors including JF and COMPANY PARTNER.

The PEC program in Côte d'Ivoire was implemented through the existing MENET-FP system and human resources. The largest in-kind contribution of the government has been the personnel. Moving forward, the scope of this is expected to be expanded to continue to leverage government resources. In this sense



the current project is fully aligned the MENETFP ambition to increase the quality of education in a cost-efficient way as it relies and mobilized already existing staff.

✓ **Indicator 5.2.2 Organizational capacity to implement at scale**

The conclusion of the PEC pilot marks the transition of responsibilities from J-PAL Europe to TaRL Africa. TARL Africa is essentially a joint venture between Pratham and JPAL Africa will establish a local office in Abidjan. IPA acknowledges that a local office with Francophone Management will be a clear asset for the future of the project, and likely necessary for scale-up. Overtime, a local office could diminish the costs involved in travel during scale-up. Still, it is a challenging process and TARL should weigh the administrative costs carefully.

Formalizing the leverage and support from government systems will be critical at scale. It is one of the key activities proposed by TaRL for the first year of the scale-up. We believe the critical point will be on supporting the monitoring and intense mentoring system.

# Brookings High-level Feedback on GMM2 TaRL Africa Project Work Plan 2017-2021

---

August 2, 2019

---

Overall, CUE felt this draft scaling proposal was very strong and aligned closely with our understanding of plans for the TRECC GMM2 extension phase and the primary education Real-time Scaling Lab. In particular, CUE noted the following strengths that should help facilitate scaling and sustainability:

- The first activity is focused on designing and formalizing this plan for the extension phase together with the Ministry of National Education (MENET), TRECC, and other relevant ministries, to ensure alignment and formalize the leverage and support of the government system. CUE fully agrees that this is a critical step in the coming months and is glad to see it formally included as the first project activity.
- Government ownership of the process is included as a key output of the extension phase, and there is a strong focus on building government capacity to deliver the program, collect data, and conduct monitoring and evaluation activities.
- There is a strong focus on adaptation and learning from experiences in the pilot phase, as well as openness to experimentation and learning while doing. CUE feels this approach is very closely aligned with the Real-time Scaling Lab approach and aims and believes the scaling lab could help further support this process of testing and adaptation.
- The plan includes an openness to incorporating learning from other approaches tried and tested in the context where appropriate. CUE knows the government is particularly interested in pursuing an integrated model, based on PEC but also drawing from other approaches, so it is excellent to see openness to this moving forward. At the same time, we envision both accepting adaptations while maintaining the core components of PEC that lead to impact may prove a delicate balancing act, and believe the lab might be able to help support this process.
- It was excellent to see the inclusion of the development of a cost model that ensures the program will be mainstreamed into the government system in a cost-effective manner (and great to learn more about this activity).
- Additionally, CUE found the reflections from Company 2 on the process to date to be incredibly useful and feels they could be instructive for TRECC's work moving forward more broadly.
- In Annex 5, which reports on the site visit, it was excellent to see included as one of the key recommendations "the establishment of a multidisciplinary working group within the Ministry to develop an integrated program to improve learning outcomes. The organization of working sessions bringing together all the structures involved in these projects in order to feed the reflection on the scaling up of these two pilots with the support of Brookings Institution, which could constitute an additional asset." We would be keen to engage in further discussions with the TaRL Africa team to ensure an integrated and supportive process moving forward.

In the following section, CUE has included several remaining questions and suggestions of areas for further consideration for the proposed scaling plan:

- CUE recognizes that thus far, the development of the plans for the PEC extension phase and the plans for the Real-time Scaling Lab and the "Reading, Writing, and Arithmetic for All" initiative have been happening as parallel processes, rather than as a single effort. As such, currently this draft scaling plan details the extension phase of the PEC pilot in Côte d'Ivoire, expanding a "pure PEC" model to additional schools and communities. At the same time, conversations with the MENET team have focused more on using the TRECC extension phase to pilot and adapt a

government-developed model, the “Reading, Writing, and Arithmetic for All” initiative, based strongly on PEC but also drawing learning from other pilots. CUE knows that the meetings planned between the TaRL team and MENET for August 2019 are intended to bring these efforts together and to gain clarity and consensus on plans moving forward and how these plans will be communicated externally. CUE feels a critical part of the discussions should include determining which of these two options is the correct framing and branding approach for this next phase of work and ensuring consensus and common messaging between all of the stakeholders involved.

- CUE sees many clear links and areas of alignment between this scaling plan and the work of the Real-time Scaling Lab and looks forward to working more closely with the TaRL Africa team and collaborating further. At the same time, CUE feels it would be beneficial to have additional discussion around planned activities between the different players, to ensure that everyone is coordinated and aligning efforts, feeding into each other’s work rather than being duplicative. With the current draft plan, CUE sees the potential for duplication of efforts in certain areas, and therefore it would be useful to discuss these areas further and determine how we can work together more closely to synergize activities and support each other’s work.
- It is excellent that the first project activity is focused on further developing and formalizing plans for the extension phase together with the MENET and TRECC. As part of this process, it would be good to incorporate the launch of the Primary Education Real-time Scaling Lab, planned for the end of September, into the activities as well, given the shared objective to introduce plans for the extension phase with key stakeholders and solicit feedback and buy-in into the process. As discussed with MENET and TRECC, the current intention is to share the draft scaling plan of the government’s “Reading, Writing and Arithmetic for All” initiative with the lab participants (a diverse group of key government, non-governmental, and private sector stakeholders) in advance of the September event, and then use the lab launch to formally present the draft plan to these key stakeholders, solicit their questions and feedback, and discuss potential adaptations and refinements of the plan. Given that the intention for this event is not to have the consultation be an empty exercise but to really take into account the participants’ input and use it to improve the plan, it is important to build in time and capacity for both the event itself and follow up adaptation into the project activities.
- Given that different versions of PEC were piloted, will one or both approaches be included in the extension phase? Is this something that has already been decided or is it a topic of conversation for the planning sessions with the MENET? It was CUE’s understanding that the government is potentially interested in moving forward with both models, employing the different approaches in different contexts, as well as exploring the possibility of starting PEC activities in grade one, but it would be useful to have additional clarity on this.
- The objective of Activity 7 is described as “documenting and enabling learning and innovation through the extension phase. Led by TaRL Africa, the learning and innovation plan will be built into overall implementation, and progress and findings will be shared across stakeholders to inform the scale up process in country and to provide insights for other country contexts.” There seem to be strong synergies and potential areas of overlap between this intended activity and the proposed activities of the Real-time Scaling Lab. A core aspect of the Real-time Scaling Lab is to provide a structured opportunity to bring together this group of key stakeholders over the next two years to reflect on the process of piloting the government’s “Reading, Writing and Arithmetic for All” initiative through TRECC’s extension phase, then based on data and learning from the extension phase, identify “change ideas” or tweaks to the current model to test, and document the overall process so that ongoing insights and key learnings are shared within Côte d’Ivoire and beyond. As such, it would be helpful to learn more details about plans for TaRL Africa’s work documenting the process and sharing learning to inform scaling, so that we can

ensure the lab is aligned and builds on their work and is not duplicative. It would be excellent to explore ways to work more closely on these proposed activities where it makes sense.

- Additionally, with Activity 8 Planning for Scale, it is fantastic to see planning for scale beyond the extension phase as a discrete activity. This also aligns with a core objective of the Real-time Scaling Lab. It would be great to explore how these efforts can be integrated or feed into each other in the most efficient and beneficial way possible for the government.
- Similarly, the plan mentions an operational supervision team made up of the government project leadership, TaRL Africa country manager and coordinator, and representatives from the chocolate companies and IPA that convenes monthly to keep everyone updated on activities and to make high level decisions on next steps. Here too, there seems to be close alignment and potential overlap with proposed plans for the Real-time Scaling Lab, including routine convenings of the diverse stakeholders involved to reflect on progress made, discuss adaptations and course corrections based on real-time learning, and agree on next steps collectively. It would be helpful to have more information and detail about this proposed team, to better understand either how its' work and data might feed into the work of the Real-time Scaling Lab or whether there may be ways to combine its actions with some of the lab activities to avoid replication.
- Under the sustainability section, the plan to build leaders of practice is mentioned, which seems very important and could be integral to the ongoing expansion and sustainability of the effort. However, CUE noted this was not included earlier in the proposal, as part of an activity or target.

## 6. Annexes

---

### **Annex 1: Independent Data collection Methodology**

#### Introduction

IPA led the independent Data Collection (IDC) in order to inform the Evaluation Matrix in the context of the TRECC-GMM2 project. This is complementary to the administrative data collected by the M&E teams of each partners in the pilot. The objective of Independent Data Collection is to better understand the context in which the pilots are being implemented, as well as to learn and to provide feedback and recommendations to TRECC, implementers and companies.

This document presents the protocol (or the Survey's Methodology) guiding IPA's Independent Data Collection in the framework of PEC implementation in Côte d'Ivoire. The Survey Methodology covers four main sections: (i) the presentation of different waves of data collection; (ii) the sampling and the sample calculation; and (iii) the data collection tools and response rate.

#### 1.1. Waves of data collection

Three to four data collection rounds/waves were planned. Ideally, IPA planned to run one before the beginning of the intervention (baseline), one or two during the implementation (follow-up) and one at the end of the pilot (end-line). Depending of the purpose of each round of data collection, IPA aimed to administer student a test, to interview teachers, and do classroom observation.

In the framework of PEC implementation, teachers should assess the students three times during the life of the pilot. It was initially targeted that the first assessment of students by teachers would happen before the implementation of PEC pedagogy in classrooms. It was also anticipated that the second teacher's assessment would be held in midway and, the last assessment will be done at the end of the pilot's implementation.

IPA's IDC was scheduled for two weeks after teachers had led their students' evaluation, where circumstances allowed. This choice was made to mitigate the increase of students' knowledge between teacher's evaluation and IPA evaluation given that the two assessment use the same test – which is ASER<sup>20</sup> test.

- Baseline data collection

The purpose of the baseline is to provide detailed information on students' learning levels and characteristics at the beginning of the program. For this reason, IPA waited for teachers to finish implementing their own ASER tests and create the different groups. This phase of testing ended the 15 of October 2018 and IPA started collecting data 7 days after from October 22<sup>nd</sup> to 26<sup>th</sup>, 2018.

Teachers were interviewed during the baseline to gather their feedback and their confidence in the PEC approach after having attended PEC's training.

Schools' level questionnaire was also implemented to collect information around PEC implementation in schools such names of teachers responsible of PEC, PEC hours in school and teachers logs (register after having led-ASER test) – these data was I useful, for instance, when planning unannounced schools visit.

---

<sup>20</sup> More information on this test are available here ""

- Classroom observations

Two waves of classroom observations visits were organized during the pilot's implementation. The first was organized from January 13<sup>th</sup> to 19<sup>th</sup>, 2019 and the second took place from April 8<sup>th</sup> to 11<sup>th</sup>, 2019.

Classroom observations consisted in unannounced visit of schools during the reported PEC hours to observe the teacher's implementation of PEC with their students. The goal was to gain information about the quality of the program implementation.

- Endline data collection

The endline data was collected from May 13<sup>th</sup> to 17<sup>th</sup>, 2019.

The targeted respondents for this last survey were students and teachers. Students level was be assessed through ASER test and teachers were asked about the feedback concerning the implementation of the PEC

## 1.2. Sampling and sample calculation

The pilot targeted 25 public primary schools selected in the region of Gabiadji. It targeted grades 3, 4 and 5 (corresponding to CE1, CE2 and CM1 in the Ivorian education system). One grade 3 class, one grade 4 class, and one grade 5 class from each school was included in the pilot. As described in the proposal, the schools were chosen according to four criteria – low level of literacy and numeracy, high average rate of students and teacher attendance, willingness of schools' staff to implement the new PEC methodology and inclusion of schools in private cocoa company region of activity.

- Students' sample

The sample computation for students follows two lines of reasoning. Firstly, we want to interview enough students to have a reliable test evaluating whether teachers are assessing the students properly (we calculate it with a power calculation for Cohen's Kappa and found 103 students to be surveyed). Secondly, we need a large enough sample of students to have an accurate estimate of baseline characteristics, including students' baseline test scores simply using the population ( $n=2,625$ ) at a 95% confidence level with a margin of error of  $\pm 5\%$  gives a sample size of 336 at baseline. This is larger than the 103 calculated above. So, we used 336 as our intended sample size. As a result, in each class, 15 students were randomly selected by IPA's agents to be surveyed.

The same assumption will be maintained for the endline data collection. So, the endline survey will target also 15 students by grade – which gives 375 students to be assessed.

- Teachers' sample

Given the small number of teachers who participate in the pilot, we planned to survey all of them. In total, 75 teachers were interviewed during the baseline and the endline data collection

- Classroom observations sample

Classroom observations were made without intervention in the class and enumerators visit in schools were unannounced. In practice, one teacher was randomly selected per school. Which gave 25 classrooms to be observed during each wave of classroom observations.

### 1.3. Data collection

Data collection tools were designed by IPA and conducted with students and teachers.

- Data collected and tools

#### *Students' interview*

Students were assessed to gather their learning level in literacy and mathematics. Data on students' learning was collected using a literacy and numeracy assessment tool called the Annual Status of Education Report (ASER). Each test took about 10-15 minutes for both sections and was administered to students one-on-one by a trained field enumerator.

During the baseline, a short questionnaire was delivered to each student at the end of ASER test. It lasted around 10 minutes and it includes questions on the way students managed their time outside of school. The aim of this short questionnaire was to assess the learning environment of children.

#### *Teachers' interview*

The goal of teachers' interviews changes a bit from baseline to endline.

During the baseline, the teachers' questionnaire covered attendance to teachers' training, satisfaction toward the PEC training and teachers' self confidence in implementing PEC with their students.

During the endline, teachers were asked for self-assessment of their implementation of PEC, their feedback on PEC and assistance they received from mentors.

Teachers interviews usually lasted 30-45 minutes.

#### *Classroom observations*

Observations of PEC's classroom usually lasted 20 minutes. The classroom observation was made without intervention or disruption of teacher class. During this observation, IPA agents collected data on whether students attend school, how classrooms are organized, and how teachers implemented PEC methodology.

IPA classroom observations tools include the teachers-mentors observations tool developed with J-PAL's support.

- Sample covered and response rate

The table below presents the summary of sample covered during all rounds of data collection.

*Table 13: Sample covered by round of data collection*

Round of data collection	Targets	Survey planned	Survey covered	Rate
<b>Baseline</b>	Students	375	375	100%
	Teachers	75	73	97%
<b>Classroom observations #1</b>	PEC classes	25	25	100%
<b>Classroom observations #2</b>	PEC classes	25	19	76%
<b>Endline</b>	Students	375	375	100%
	Teachers	75	73	97%

## **Annex 2: ASER test and Implementation**

- Presentation of the test

The ASER tool tests basic reading and numeracy skills that are taught in grade 1 and 2. The choice was made to use the ASER to assess the students given its advantages compared to other tests. The ASER is reliable to measure grade level reading skills, more appropriate for French speaking country (like Côte d'Ivoire), and less complex to develop and administrate.

### *Literacy test*

The literacy learning assessments tested the ability to read in French and placed learners into one of five levels:

1. Beginner. At this level, a learner cannot identify sounds or letters correctly.
2. Letter level. A learner can identify sounds or letters but cannot read words.
3. Word level. A learner can read words but cannot read sentences coherently.
4. Simple paragraph level. A learner can read sentences that use simple language but cannot read a short story.
5. Story level. A learner can read a short story containing simple language.

### *Numeracy/Mathematics test*

The numeracy assessments are split into two parts. The first part tests number recognition and the second part tests math's operations.

For number recognition, a learner can be placed into one of four levels:

1. Beginner. At this level, a learner is unable to name even one-digit numbers.
2. One-digit level. A learner can name one-digit numbers but not two-digit numbers.
3. 2-digits level. A learner can name two-digit numbers but not three-digit numbers.
4. 3-digits level. A learner can name three-digit numbers. Learners should be at this level by the end of grade 2.

For math's operations, learners are tested on two sets of operations. First, they are given some two-digit by two-digit addition and subtraction sums. Second, learners are tested on two-digit by one-digit multiplication and division sums. Depending on their results they can be placed into one of five levels:

1. Beginner. These learners are unable to recognize a one-digit number (So, they cannot do either addition, subtraction, multiplication or division given that they have not prerequisite for that) or they have none either addition, subtraction, multiplication or division levels (see definition below).
2. Addition. These learners can do addition. Learners are expected to reach this level by the end of grade 2.
3. Subtraction. These learners can do subtraction. Learners are expected to reach this level by the end of grade 2.
4. Multiplication. These learners can do multiplication.
5. Division. These learners can do multiplication and division.



- IPA/IDC ASER tests sample

IPA 's ASER tests was developed with the support of J-PAL and MENET-FP experts—who received the expertise to developed ASER test which were used by teachers from Pratham team. The fact to have IDC tests were validated by MEN's experts ensures to IPA to have reliable tests in the sense that students should perform in the same way regardless of the assessor (IPA enumerators or Teachers).

- IPA implementation of ASER test

IPA implementation of ASER test follows the same protocol as describe in the Literacy and Numeracy PEC Manuals provided to teachers by COMPANY PARTNER—See Annex 3 and Annex 4 for details for the ways to administrate ASER test.

The test was programmed in the electronic tablets in such way assessors were only able to gather learners' answers at each step of the test step but cannot provide the final score or level. The learners score was automatically calculated by the device and was unknown to the assessors. This choice was made to minimize assessors bias in learners final score.

The test was delivered during one-on-one friendship's discussions with students in respect of "*IPA Child Safeguarding Policy*". Besides of schools' staff consent to implement the survey, students' assent was required to administrate the test. In addition, enumerators were trained to reassure children and help them to perform at their best during the test.

## Annex 3: Protocol to administrate ASER literacy test

### Comment tester la lecture :

#### Commencer par le niveau « Paragraphe »

Demander à l'enfant de lire un des deux paragraphes.

Laisser l'enfant choisir le paragraphe lui-même / elle-même. Si l'enfant n'arrive pas à choisir, lui désigner un des deux paragraphes.

Lui demander de le lire. Ecouter attentivement comment il / elle lit.

S'il fait des erreurs, lui laisser une seconde chance.

#### L'enfant n'est pas au niveau « Paragraphe » s'il / elle :

- Lit le texte comme un enchaînement de mots, plutôt que comme une phrase.
- Lit le texte de façon saccadée, en s'arrêtant souvent.
- Lit le texte de façon fluide mais en faisant plus de 3 erreurs.

#### L'enfant est au niveau « Paragraphe », s'il / elle :

- Lit le texte comme une phrase, plutôt que comme un enchaînement mots.
- Lit le texte de façon fluide et aisée, même s'il / elle lit lentement.
- Lit le texte avec 3 erreurs ou moins

Si l'enfant n'est pas au niveau « Paragraphe », lui demander de lire des mots.

Si l'enfant peut lire un paragraphe, lui demander de lire l'histoire.

#### Tester le niveau « Mot »

Demander à l'enfant de lire 10 mots parmi la liste de mots.

Laisser l'enfant choisir les mots. S'il / elle ne choisit pas, lui montrer des mots à lire.

L'enfant est au niveau « Mot » s'il / elle :

- Lit au moins 6 mots sur 10 avec aisance.

#### L'enfant est au niveau « Histoire » s'il / elle :

- Lit le texte comme une phrase, et non comme un enchaînement de mots.
- Lit le texte de façon fluide et aisée. L'enfant peut lire lentement.
- Lit le texte avec 3 ou moins de 3 erreurs.

## **Annex 4: Protocol to administrate ASER numeracy test**

Comment tester les mathématiques :

**Instructions pour la partie 1 : « Reconnaissance des nombres »**

Commencez avec les nombres à 3 chiffres. Si l'enfant est capable de reconnaître 4 nombres à 3 chiffres, il est au niveau 3, et vous pouvez passer à la partie 2 du test.

Si l'enfant n'est pas capable de reconnaître les nombres à 3 chiffres, demandez-lui de reconnaître les nombres à 2 chiffres. Si l'enfant est capable de reconnaître 4 nombres à 2 chiffres, il est au niveau 2, et vous pouvez passer à la partie 2 du test.

Si l'enfant n'est pas capable de reconnaître les nombres à 2 chiffres, demandez-lui de reconnaître les chiffres. Si l'enfant est capable de reconnaître 4 chiffres, il est au niveau 1, et vous pouvez passer à la partie 2 du test.

S'il n'est pas capable de reconnaître les chiffres, il sera au niveau débutant.

**Instructions pour la partie 2 : « Opérations »**

*Il faut demander à l'enfant d'essayer toutes les opérations en commençant par l'addition, puis la soustraction, puis la multiplication et enfin la division. Dans chaque section, le facilitateur notera si l'enfant « peut faire » ou « ne peut pas faire » ce type d'opérations.*

Commencez par le niveau « addition ». Si l'enfant est capable de résoudre au moins deux opérations, inscrivez « peut faire addition ». S'il résout moins de deux opérations, inscrivez « ne peut pas faire addition ».

Une fois les exercices d'addition terminés, demandez-lui de faire les soustractions. Si l'enfant est capable de résoudre au moins deux opérations, inscrivez « peut faire soustraction ». S'il résout moins de deux opérations, inscrivez « ne peut pas faire soustraction ».

Une fois les exercices de soustraction terminés, demandez-lui de faire les multiplications. Si l'enfant est capable de résoudre au moins deux opérations, inscrivez « peut faire multiplication ». S'il résout moins de deux opérations, inscrivez « ne peut pas faire multiplication ».

Enfin, demandez à l'enfant de faire les divisions. Si l'enfant est capable de résoudre au moins deux opérations, inscrivez « peut faire division ». S'il résout moins de deux opérations, inscrivez « ne peut pas faire division ».

Si l'enfant ne peut pas faire ni addition, ni soustraction, ni multiplication ni division, il est débutant en opérations.

# Annex 5: Teachers logs or registers template

Ce formulaire doit être complété par l'enseignant

Nom de l'école:	
Code de l'école:	
Nom de l'enseignant:	
Classe (CE1, CE2, CM1):	
Période du test	du _____ au _____

Nom de l'élève	Sexe		Français					Mathématiques Reconnaissance des nombres				Mathématiques Opérations simples							
			Cochez le niveau le plus élevé atteint par l'élève (1 seul niveau)					Cochez le niveau le plus élevé atteint par l'élève (1 seul niveau)				Pour chaque opération cochez "peut faire" ou "ne peut pas faire"							
	H	F	Débutant	Lettres	Mots	Paragraphe	Histoire	Débutant	Niveau-1 (1 à 9)	Niveau-2 (10 à 99)	Niveau-3 (100 à 999)	Addition		Soustraction		Multiplication		Division	
												Débutant	Peut faire	Ne peut pas faire	Peut faire	Ne peut pas faire	Peut faire	Ne peut pas faire	Peut faire
1																			
2																			
3																			
4																			
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
no. TOTAL d'élèves																			

Notez les noms des élèves absents	
-----------------------------------	--

## Annex 6: Classroom observations components

<p>Signature du mentor (conseiller pédagogique / directeur) _____</p> <p>Signature de l'enseignant _____</p> <p>Date _____</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="2">Total de points</th> </tr> <tr> <td>Regroupement et progrès</td> <td></td> </tr> <tr> <td>Matériel et activités</td> <td></td> </tr> <tr> <td>Implication de l'élève</td> <td></td> </tr> </table>	Total de points		Regroupement et progrès		Matériel et activités		Implication de l'élève	
Total de points									
Regroupement et progrès									
Matériel et activités									
Implication de l'élève									

	Des améliorations sont nécessaires	Bien, mais des améliorations sont possibles	Très bonne mise en place de PEC
Regroupement et progrès des élèves	Regroupez les élèves en fonction de leur niveau. (0-1 point)	Assurez-vous d'envoyer les élèves dans un groupe plus avancé à mesure qu'ils progressent. (2 points)	Félicitations ! Très bonne mise en œuvre du PEC (3 points)
Matériel et activités	Développez et utilisez les outils PEC Assurez-vous que les activités se déroulent d'abord avec le groupe entier, puis en petits groupes et ensuite au niveau individuel Assurez-vous que dans chaque activité, les enfants écoutent, parlent, lisent, écrivent et manipulent (0-1 point)	Assurez-vous d'utiliser les outils PEC. Fixez des objectifs réalisables pour aider les enfants à progresser au niveau suivant Assurez-vous que la disposition de la salle de classe facilite le déroulement des activités (2-3 points)	Félicitations ! Très bonne mise en œuvre du PEC (4 points)
Implication de l'élève et position du facilitateur	Encouragez les élèves à participer Utilisez une gamme d'activités différentes. Donnez des consignes claires pour chaque activité Asseyez-vous près des élèves. (0-1 point)	Assurez-vous que les activités encouragent les élèves à participer. Donnez des consignes claires pour chaque activité Asseyez-vous près des élèves. (2 points)	Félicitations ! Très bonne mise en œuvre du PEC (3 points)

Total de points de la dernière leçon observée	/ 10
Total de points de la leçon d'aujourd'hui	/10

Conseils pour l'enseignant

## Annex 7: Some results from IPA Independent Data Collection fieldwork

Table 14: Literacy assessment by IPA enumerators – Baseline vs Endline

	CE1 - Grade 3		CE2 - Grade 4		CM1 - Grade 5		Grades 3, 4 and 5	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Beginner	78%	50%	53%	21%	24%	11%	55%	41%
Letters	15%	24%	15%	38%	30%	27%	24%	25%
Words	5%	10%	8%	10%	8%	12%	7%	13%
Paragraph	2%	9%	14%	15%	18%	18%	9%	12%
Story	0%	8%	10%	16%	20%	31%	5%	9%

Table 15: Number recognition assessment by IPA enumerators – Baseline vs Endline

	CE1 - Grade 3		CE2 - Grade 4		CM1 - Grade 5		Grades 3, 4 and 5	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Beginner	10%	6%	2%	1%	0%	0%	3%	2%
1-digit	58%	22%	26%	6%	11%	5%	34%	19%
2-digits	12%	9%	6%	8%	5%	2%	11%	11%
3-digits	20%	62%	66%	85%	84%	94%	52%	68%

Table 16: Basic operations assessment by IPA enumerators – Baseline vs Endline

	CE1 - Grade 3		CE2 - Grade 4		CM1 - Grade 5		Grades 3, 4 and 5	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Beginner in operation	88%	46%	38%	18%	22%	12%	51%	34%
Can do addition	6%	50%	53%	75%	73%	85%	41%	59%
Can do subtraction	9%	43%	54%	74%	67%	80%	38%	55%
Can do multiplication	1%	24%	18%	44%	38%	56%	15%	25%
Can do division	0%	6%	4%	19%	10%	35%	3%	9%

Figure 13: Literacy baseline assessment by IPA's enumerators and teachers

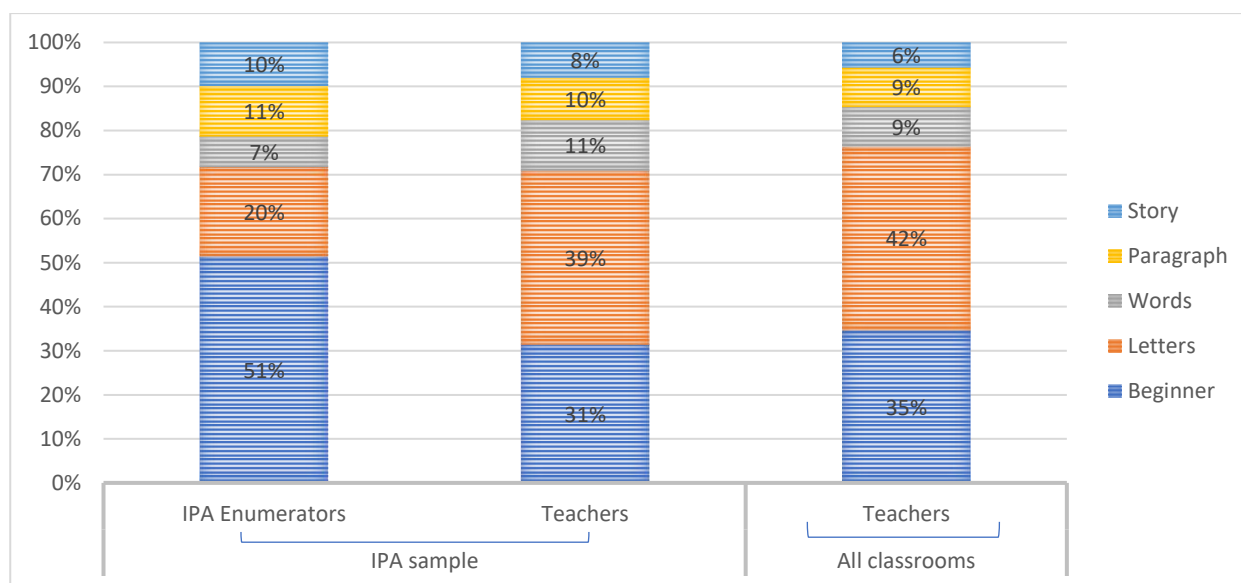


Figure 14: Numeracy baseline assessment by IPA's enumerators and teachers

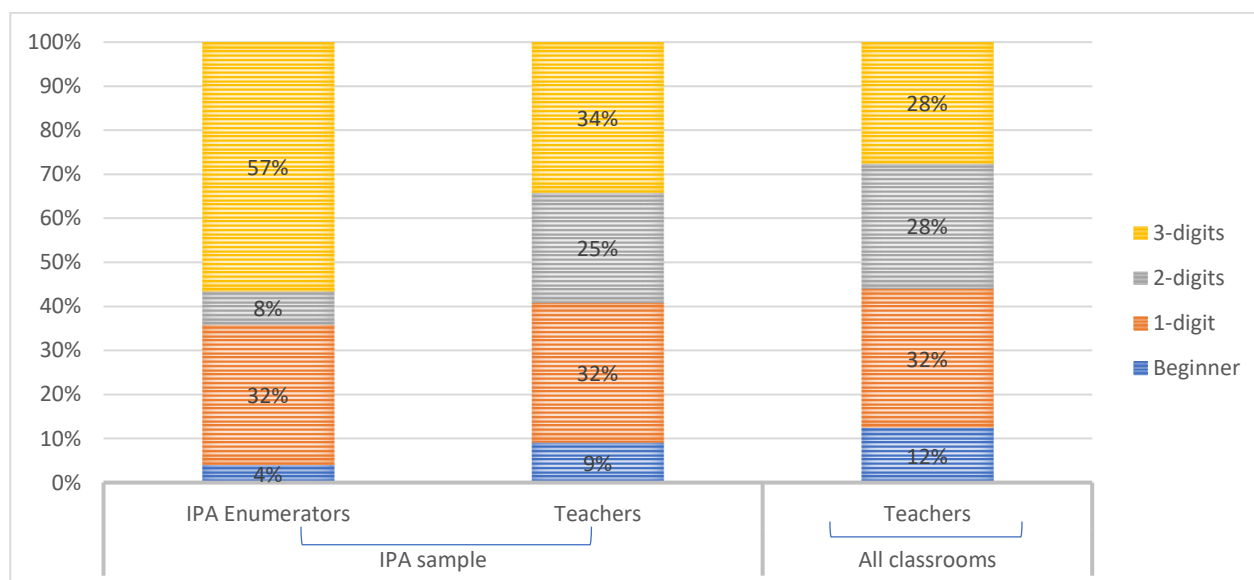


Figure 15: Literacy endline assessment by IPA's enumerators and teachers

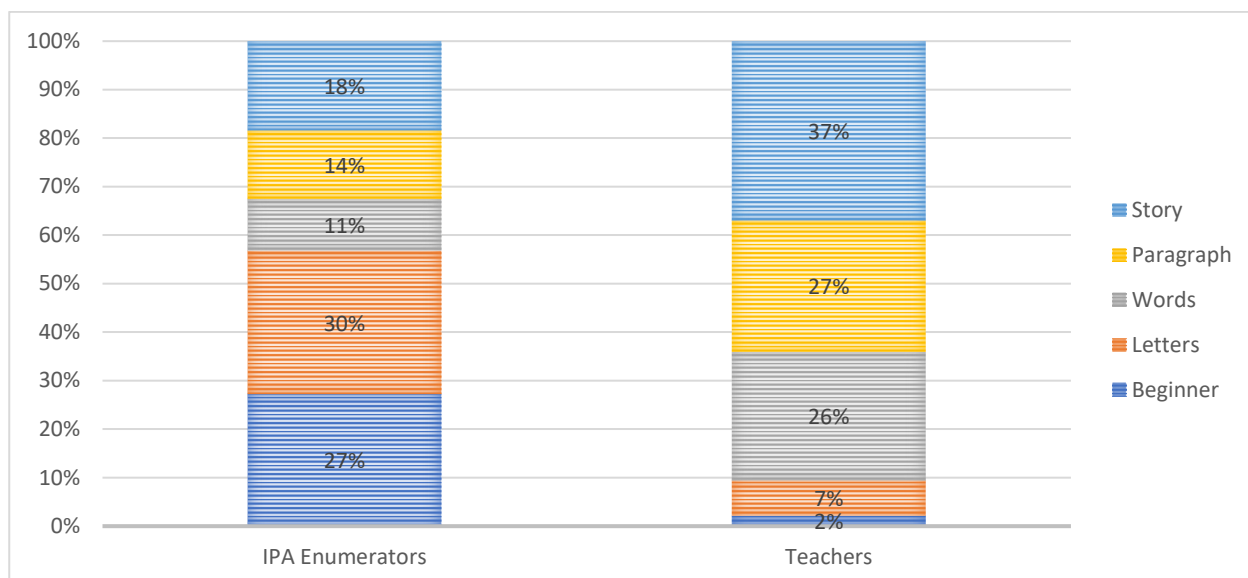


Figure 16: Numeracy endline assessment by IPA's enumerators and teachers

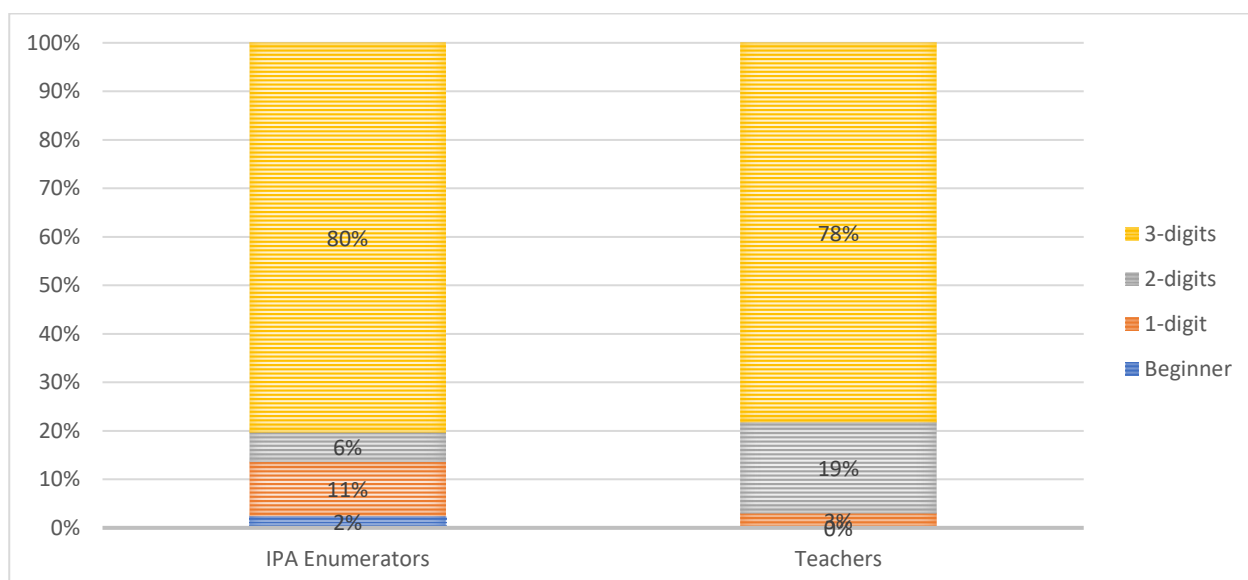


Table 17: Class observations 1 vs Class observations 2 -Results

	Results from previous classes' observation		Results from current classes' observation	
	N	Prop.	N	Prop.
<u>Numbers of TaRL classes animated during the class observations</u>				
% Schools where 1 TaRL level was implemented	25	12%	19	21%
% Schools where 2 TaRL level was implemented	25	24%	19	16%

% Schools where 3 TaRL level was implemented	25	64%	19	63%
% of schools where numbers of observed TaRL's group of levels match with register classification	-	-	19	79%

Table 18: Class observation 1 results

Table 10: Classes observations results		
	Frequency	Proportion
% of TaRL classes where students were grouped by level and not by grade	24	96%
% of TaRL classes where the grouping of student was well done	13	52%
% of students actively involved in the activities?		
0%-49%	0	0%
50%-69%	9	36%
70%-100%	16	64%
Teachers scores		
N	25	
Mean	7.12	
Sd	0.88	
% Schools where 1 PEC level was implemented	3	12%
% Schools where 2 PEC levels were implemented	6	24%
% Schools where 3 PEC levels were implemented	16	64%

## Annex 8: TRECC Pilot Criteria Matrix



This is the source of the report's indicators.

The Transforming Education in Cocoa Communities (TRECC) initiative aims at improving the living conditions of children and youth in Côte d'Ivoire by promoting quality education in cocoa-growing communities. Via its Grant Matching Mechanism round 2 (GMM2), 9 pilots-to-scale projects are being co-funded with 10 cocoa companies and implemented by 14 implementing organizations in the sectors of Early Childhood Development, Primary Education and Vocational Training.

The role of Innovations for Poverty Action (IPA) is to provide technical support to the companies and implementing agencies to design and implement sound monitoring systems to closely monitor and learn from these pilots. In parallel, IPA conducts its own independent and complementary data collection. IPA will use these two sources of information – the administrative data collected by the implementing organizations through their own M&E system and the independent data collection – to feed into an evaluation matrix to assess each pilot.

Each evaluation matrix describes how IPA will use the data to make recommendations on the potential scale-up of the pilots to other relevant cocoa-growing communities. In addition, TRECC may consider whether certain pilots are feasible for future scale-up beyond such communities, for example to the regional or national level, though this is not a central focus of this evaluation matrix given the existing contractual arrangements on GMM2.

The evaluation matrix comprises five sections which will be the basis on which to recommend eligibility of a pilot team to submit a proposal for potential scale-up. The five criteria are Relevance; Results (outputs and immediate outcomes); Costs & Operations management; Capacity to learn, improve and innovate; and Sustainability. For each criterion, we describe the key evaluation questions, which will be common across all pilots.

This document presents the evaluation matrix tailored to the J-PAL and COMPANY PARTNER pilot, with the five criteria and the qualitative and quantitative indicators to that will be used to assess each criteria. IPA will work with COMPANY PARTNER and J-PAL to finalize the indicators, targets and data sources for the pilot, drawing on their logical framework.

We will use a "traffic light" color system to provide an overall assessment against each of the five criteria: green will mean that the pilot is compliant with the criteria requirement for potential scale-up, red will mean that it is not, and orange will mean that it does partially comply and that eligibility for scale-up should be conditional on corrective measures to be taken. At the final evaluation stage:

- pilots with green assessments on all 11 criteria will receive an *unconditional* recommendation for eligibility for a scale-up proposal;
- pilots who have only green and orange criteria (no red), and among these a majority of green criteria, will receive a *conditional* recommendation for scale-up – i.e. conditional on corrective measures in response to findings from monitoring and evaluation.
- pilots who have only green and orange criteria (no red), and among these a majority of orange criteria, will not be recommended for scale-up.
- pilots with any red criteria will not be recommended for scale-up

During implementation, IPA will provide work-in-progress traffic-light assessments to the pilots as part of our quarterly updates (ideally during advisory calls), for criteria for which there is enough data already available for an interim assessment. This will allow opportunities for course correction and improvement. Before sharing our

work-in-progress assessments, if any are orange or red light assessments, we will discuss them with the implementing partners first.

1. Relevance					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
<b>1.1.</b> Targets an important need in the community	<b>1.1.1.</b> Is there any evidence of this being an important need in the community? <b>1.1.2.</b> Is the intervention model appropriate to respond to the identified need? <b>1.1.3.</b> If relevant, do the project's targeting criteria allow to reach the group that needs the intervention the most in the community?	<b>1.1.1 J-PAL's needs assessment report and IPA baseline shows evidence of the need being addressed</b> <ul style="list-style-type: none"> <li>A majority of children do not master the basic reading and math skills (ASER test) which are a prerequisite to successfully follow CE and CM curricula.</li> <li>The above indicator (basic skills) is sufficient to justify the need for TaRL. Classroom heterogeneity confirms the relevance of grouping children by level and not by grade –yet TaRL can also be implemented in homogeneous classrooms where children are failing to master basic skills.</li> </ul>	<b>1.1.1 Key government stakeholders' description of community shows evidence of the need being addressed</b> Interviews with government relevant representatives confirm shortcomings in French and mathematics achievement  <b>1.1.2. Beneficiaries' description of their needs links to the outcomes delivered by the pilot</b> <ul style="list-style-type: none"> <li>Teachers or principals believe that variation in student levels within the classroom is a challenge to student learning (at least when asked directly about this)</li> </ul>	<b>Administrative data</b> <ul style="list-style-type: none"> <li>Project targeting criteria</li> <li>Needs assessment report</li> </ul> <b>Independent evaluation</b> <ul style="list-style-type: none"> <li>Quantitative and qualitative interviews with students and teachers</li> <li>Qualitative interview with government relevant stakeholders</li> <li>ASER tests at baseline</li> </ul>	<b>Green:</b> <ul style="list-style-type: none"> <li>Pilot addresses specific important needs that were among targeted population.</li> <li>The intervention as implemented focuses on the objectives initially agreed, or agreed changes.</li> </ul> <b>Orange:</b> <ul style="list-style-type: none"> <li>Pilot did not fully address important identified needs and/or partially reached the targeted population</li> <li>The implementation strayed somewhat from the initial agreed objectives</li> <li>Feasible remedial measures are identified to address these issues</li> </ul> <b>Red:</b> <ul style="list-style-type: none"> <li>Pilot does not address important needs in the community and/or does not serve the targeted population</li> <li>The implementation shows that program objectives are no longer relevant to those originally agreed.</li> </ul>

1. Relevance					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
					<ul style="list-style-type: none"> <li>No feasible remediation identified for these issues</li> </ul>
1.2. Aligns with the priorities of the donors	1.2.1. Does the pilot still align with the objectives of TRECC and COMPANY PARTNER?		<p><b>1.2.1. The pilot, as implemented, remains aligned to the objectives originally agreed and any agreed changes have been documented</b></p> <ul style="list-style-type: none"> <li>Interview with program management team confirms that the program goals remain aligned with the priorities originally described in the proposal, which is to improve achievement on the basics in French and mathematics for students from CE1 to CM1 by adapting the teaching to the actual level of each child.</li> </ul>	<p><b>Independent evaluation</b></p> <ul style="list-style-type: none"> <li>Interviews with program management including TRECC and COMPANY PARTNER</li> <li>Desk Research</li> <li>Review of initial proposal</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>Pilot addresses specific important needs that were among targeted population.</li> <li>The intervention as implemented focuses on the objectives initially agreed, or agreed changes.</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>Pilot did not fully address important identified needs and/or partially reached the targeted population</li> <li>The implementation strayed somewhat from the initial agreed objectives</li> <li>Feasible remedial measures are identified to address these issues</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>Pilot does not address important needs in the community and/or does not serve the targeted population</li> <li>The implementation shows that program objectives are</li> </ul>

1. Relevance					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
					no longer relevant to those originally agreed. No feasible remediation identified for these issues

2. Results (outputs and direct outcomes)					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
<b>2.1.</b> Delivers outputs at high quality	<b>2.1.1.</b> Has the pilot produced measurable outputs with the required quality? <b>2.1.2.</b> Did the expected number of beneficiaries engage in sustained way?	<b>2.1.1. Key outputs from the proposal logframe were achieved</b> <ul style="list-style-type: none"> <li>– # of the teachers completed the training</li> <li>– # of school mentors trained</li> <li>– # of schools that received TARL material (a school is considered to have received TARL material if it received 3 complete sets of materials for calculation and reading (for 3 level groups))</li> </ul> <b>2.1.2. Participation rate</b> <ul style="list-style-type: none"> <li>– Teacher attendance rate for training courses (total number of teachers who attended all training sessions)</li> <li>– Attendance rate during mentor training (total number of mentors who attend all trainings)</li> </ul>	<b>2.1.1. Positive feedback about the quality of key outputs</b> <ul style="list-style-type: none"> <li>– Feedback from teachers, principals and mentors about training received</li> </ul>	<b>Administrative data</b> <ul style="list-style-type: none"> <li>o Review of the quarterly &amp; annual reports and routine monitoring reports</li> <li>o Analysis of project monitoring data</li> </ul> <b>Independent Evaluation</b> <ul style="list-style-type: none"> <li>o Results from spot-checks</li> <li>o Qualitative interviews with teachers, principals and mentors</li> </ul>	<b>Green:</b> <ul style="list-style-type: none"> <li>o Most of the outputs and key immediate outcomes were achieved, and with the expected quality</li> <li>o A high and sustained participation rate was recorded among beneficiaries, who provided positive feedback</li> </ul> <b>Orange:</b> <ul style="list-style-type: none"> <li>o Some outputs and/or key immediate outcomes were not achieved, or not with the expected quality</li> <li>o Below expectation participation rate was recorded among beneficiaries who provided</li> </ul>

					<p>mixed feedback</p> <ul style="list-style-type: none"> <li>o Feasible remedial measures are identified to address these issues</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>o Outputs and key immediate outcomes were only partially achieved and/or low quality</li> <li>o Low participation rate was recorded among beneficiaries who provided negative feedback</li> <li>o No feasible remediation identified for these issues</li> </ul>
<p><b>2.2.</b> Achieves direct outcomes</p>	<p><b>2.2.1.</b> Have changes been observed/self-reported in the knowledge, attitudes, behavior or practices directly targeted by the interventions?</p>	<p><b>2.2.1.a Changes in teachers' knowledge, behavior and practices have been observed over time<sup>21</sup></b></p> <ul style="list-style-type: none"> <li>– % of teachers who have completed the ASER test with children and can produce the results in the register</li> <li>– % of students in the right group for their level (based on a random sample) during classroom observation% of observed teachers present and conducting PEC activities during TARL hour during classroom observation</li> <li>– % of observed teachers who show</li> </ul>		<p><b>Administrative data</b> Pre and post-test evaluation of teachers' learning conducted by J-PAL on every day of the training</p> <p><b>Independent evaluation</b></p> <ul style="list-style-type: none"> <li>o Quantitative interviews with teachers</li> <li>o ASER tests with sample of students at baseline and midline</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>o Most of the outputs and key immediate outcomes were achieved, and with the expected quality</li> <li>o A high and sustained participation rate was recorded among beneficiaries, who provided positive feedback</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>o Some outputs and/or key immediate outcomes</li> </ul>

<sup>21</sup> Any changes observed over time cannot be attributed to the program without a counterfactual; but the absence of such changes might indicate that the program is not operating as planned.

		<p>effective use of the TARL materials during classroom observations (scoring at least 8/10 on the classroom observation tool</p> <ul style="list-style-type: none"> <li>– % of teachers who reported having had a coaching session with schools’ principals or inspectors in the last four weeks</li> <li>– Teachers administer the ASER test to students accurately<sup>22</sup></li> </ul>			<p>were not achieved, or not with the expected quality</p> <ul style="list-style-type: none"> <li>○ Below expectation participation rate was recorded among beneficiaries who provided mixed feedback</li> <li>○ Feasible remedial measures are identified to address these issues</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>○ Outputs and key immediate outcomes were only partially achieved and/or low quality</li> <li>○ Low participation rate was recorded among beneficiaries who provided negative feedback</li> <li>○ No feasible remediation identified for these issues</li> </ul>
<b>2.3.</b> Beneficiary feedback about the program is positive	<b>2.3.1.</b> How satisfied are the beneficiaries with the intervention? <b>2.3.2.</b> How meaningful is the intervention to	<p><b>2.3.1.a Beneficiaries provide positive feedback on the delivery of outputs</b></p> <ul style="list-style-type: none"> <li>– % of interviewed teachers who provided a positive feedback on the TARL materials/activities</li> </ul>	<b>2.3.1.c Beneficiaries describe positive experiences with the program</b> Interview with the teachers, inspectors and	<b>Independent evaluation</b> <ul style="list-style-type: none"> <li>○ Focus group discussions with beneficiaries</li> <li>○ Quantitative Interviews with teachers,</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>○ Most of the outputs and key immediate outcomes were achieved, and with the</li> </ul>

<sup>22</sup> To be measured as the average inter-rater reliability calculated as Cohen’s kappa, between IPA’s assessed ratings for each child and the teachers’ ratings for each child, where a kappa score of 0.41-0.6 indicates “moderate” agreement. A score below 0.41 would indicate only “fair” agreement, meaning that the teachers’ test scores did not agree with IPA’s scores as closely as should be expected (previous tests have found reliability measures of at least 0.6 on the Cohen’s Kappa rating).

	beneficiaries' lives?	<ul style="list-style-type: none"> <li>– % of teachers who would be likely to recommend TARL for all classes</li> <li>– % of inspectors and school principals who provided positive feedback on the TARL methodology</li> <li>– % of teachers who can name a useful piece of advice they received from the coaching by mentors and school principals</li> </ul> <p><b>2.3.1.b Beneficiaries provide positive feedback on the main immediate outcomes</b></p> <ul style="list-style-type: none"> <li>- % of teachers who provided positive feedback on the improvement on their students' learning in the classroom as result of TARL methods</li> </ul>	<p>school principals shows that they are satisfied with the project, including the training and coaching</p> <p><b>2.3.2. Beneficiaries report that the pilot was meaningful for them</b></p> <p>During the interviews with the teachers, school directors or principals , they can describe how the program had a positive effect on student learning</p>	inspectors and school principals	<p>expected quality</p> <ul style="list-style-type: none"> <li>o A high and sustained participation rate was recorded among beneficiaries, who provided positive feedback</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>o Some outputs and/or key immediate outcomes were not achieved, or not with the expected quality</li> <li>o Below expectation participation rate was recorded among beneficiaries who provided mixed feedback</li> <li>o Feasible remedial measures are identified to address these issues</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>o Outputs and key immediate outcomes were only partially achieved and/or low quality</li> <li>o Low participation rate was recorded among beneficiaries who provided negative feedback</li> <li>o No feasible remediation</li> </ul>
--	-----------------------	---	--	----------------------------------	--



					identified for these issues
--	--	--	--	--	-----------------------------

3. Costs and operations management					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions

<p><b>3.1. Costs are well managed</b></p>	<p><b>3.1.1.</b> Did the project make efficient utilization of resources? <b>3.1.2.</b> Does the implementing partner have a vision for cost-efficiency at scale within GMM2?</p>		<p><b>3.1.1. Budget expenditures show good value for money</b> The expenditure indicates cost-efficient use of money There are indications that cost-efficiency was continuously sought during the pilot – avoidance of unnecessary high costs, efficiencies found, etc. <b>3.1.2. Partner’s vision for scaling</b> COMPANY PARTNER and J-PAL team can explain how they will achieve a more cost-efficient model at greater GMM2 scale.</p>	<p><b>Administrative data</b></p> <ul style="list-style-type: none"> <li>○ Pilot Financial proposal</li> <li>○ Pilot Quarterly and annual financial reports</li> <li>○ GMM2 Scale-up Proposal</li> </ul> <p><b>Independent Evaluation</b></p> <ul style="list-style-type: none"> <li>○ Interview with J-PAL</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>○ Good cost and project management, in line with the proposal</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>○ Some signs that costs could have been better managed resourced better allocated; some delays were observed; costs significantly exceeded initial budget; some activities initially planned were not implemented because of lack of time or resources</li> <li>○ Most issues were justified and feasible remedial measures are identified to address these issues if pilot was scaled-up</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>○ Signs that costs were not well managed and some resources were wasted; costs significantly exceeded budget without justification; important delays occurred without justification;</li> </ul>
---	---	--	---	--	--

					<p>key activities were not implemented by lack of resources or time.</p> <ul style="list-style-type: none"> <li>o No feasible remediation identified for these issues</li> </ul>
<p><b>3.2. Project management is successful</b></p>	<p><b>3.2.1.</b> Has the project been implemented as planned, with any changes being justified or outside the partners' control?</p> <p><b>3.2.2.</b> Have the cooperation partners been involved in the pilot management according to agreed aspects?</p>		<p><b>3.2.1. Intervention is implemented according to roadmap, and any changes are fully justified</b></p> <p>The pilot was implemented according to the roadmap, and significant deviations were fully justified and approved.</p> <p><b>3.2.2. Cooperation Partners' involvement</b></p> <p>COMPANY PARTNER participated as agreed in the management and implementation of the pilot, or changes in this role have been documented.</p>	<p><b>Administrative data</b></p> <ul style="list-style-type: none"> <li>o Review of the proposal including planned end dates of keys activities and planned budget</li> <li>o Review of the quarterly &amp; annual reports including realized end dates of key activities and realized spending</li> <li>o Review of the pilot relevant meetings reports (operational, advisory, steering and review meetings) reports</li> </ul> <p><b>Independent evaluation</b></p> <ul style="list-style-type: none"> <li>o Interviews with J-PAL</li> <li>o Interviews with COMPANY PARTNER</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>o Good cost and project management, in line with the proposal</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>o Some signs that costs could have been better managed resourced better allocated; some delays were observed; costs significantly exceeded initial budget; some activities initially planned were not implemented because of lack of time or resources</li> <li>o Most issues were justified and feasible remedial measures are identified to address these issues if pilot was scaled-up</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>o Signs that costs were not well managed and some resources were wasted; costs</li> </ul>

					<p>significantly exceeded budget without justification; important delays occurred without justification; key activities were not implemented by lack of resources or time.</p> <p>No feasible remediation identified for these issues</p>
--	--	--	--	--	---

4. Capacity to learn, improve and innovate					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
<b>4.1.</b> Project collects credible monitoring data	<b>4.1.1.</b> Is actionable monitoring data collected? <b>4.1.2.</b> Are the routine monitoring data credible and reliable?	<b>4.1.1. Routine monitoring data are collected and shared on time with the stakeholders</b> Monitoring data are collected and analyzed by program management and shared with IPA monthly as defined in the monitoring plan. <b>4.1.2. IPA spot check visits confirm the quality and accuracy of the data shared by partner</b> IPA spot check confirms the data quality; percentage discrepancies found in the data shared by COMPANY PARTNER/J-PAL compared with IPA independent data is not more than 10% for variables with fixed, stable values (such as the number of training sessions held) and are not statistically significant for variables that may vary in surveying (such as estimates of the numbers of attendees).	<b>4.1.1.b. Monitoring data is actionable and aligned with program management</b> - Routine monitoring data that was collected that provides actionable insight on program management. - J-PAL identified and documented project risks IPA observed in spot checks and acted on them in a timely manner.	<b>Administrative data</b> <ul style="list-style-type: none"> <li>Review of the monitoring plan and routine quarterly reports</li> <li>Analysis of the project routine monitoring data</li> <li>RAID Template</li> </ul> <b>Independent evaluation</b> <ul style="list-style-type: none"> <li>Results of spot-check visits</li> <li>Review of the pilot relevant meetings reports (operational, advisory, steering and review meetings) reports</li> </ul>	<b>Green:</b> <ul style="list-style-type: none"> <li>Credible and reliable data was regularly collected by the partner and shared according to the agreed calendar</li> <li>Data were analyzed and used to learn and take corrective measures to improve the implementation</li> </ul> <b>Orange:</b> <ul style="list-style-type: none"> <li>Signs that data was not always credible or reliable and/or was not collected and shared according to agreed calendar</li> <li>Data was collected but was only sometimes used for learning and improvement or not documented</li> <li>Most issues were justified and feasible remedial measures are identified to address these issues</li> </ul> <b>Red:</b> <ul style="list-style-type: none"> <li>There was little data collected and shared and/or data was not reliable nor credible</li> <li>If data was collected, it was rarely used to take corrective measures or documented</li> <li>No feasible remediation</li> </ul>

					identified for these issues
<b>4.2.</b> Monitoring is used to learn and improve	<b>4.2.1.</b> Is the J-PAL demonstrating willingness to learn, innovate and incorporate monitoring feedback? <b>4.2.2.</b> Did the J-PAL test some key assumptions in their theory of change?		<b>4.2.1. Program improvement in response to monitoring</b> Significance of appropriate changes in program management or delivery that can be linked to monitoring findings <b>4.2.2. Use of data to refine the ToC</b> J-PAL shows that they have refined their understanding of the Theory of Change through use of their own data, and documented these refinements.	<b>Administrative data</b> <ul style="list-style-type: none"> <li>Review of the monitoring plan and routine reports</li> <li>Analyses of the project routine monitoring data and reports</li> <li>GMM2 scale-up proposal</li> </ul> <b>Independent evaluation</b> <ul style="list-style-type: none"> <li>Interviews with J-PAL</li> </ul>	<b>Green:</b> <ul style="list-style-type: none"> <li>Credible and reliable data was regularly collected by the partner and shared according to the agreed calendar</li> <li>Data were analyzed and used to learn and take corrective measures to improve the implementation</li> </ul> <b>Orange:</b> <ul style="list-style-type: none"> <li>Signs that data was not always credible or reliable and/or was not collected and shared according to agreed calendar</li> <li>Data was collected but was only sometimes used for learning and improvement or not documented</li> <li>Most issues were justified and feasible remedial measures are identified to address these issues</li> </ul> <b>Red:</b> <ul style="list-style-type: none"> <li>There was little data collected and shared and/or data was not reliable nor credible</li> <li>If data was collected, it was rarely used to take corrective measures or documented</li> </ul>

5. Sustainability					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
5.1. Provides sustained benefit to community	5.1.1. Are there any signs that the intervention from the pilot will continue to benefit the beneficiaries/community members over time?		<p><b>5.1.1. Indications that the community members are likely to continue with the practices or program activities</b></p> <p>Interview with teachers, school principals and inspectors shows that they are equipped to overcome challenges to their continued use of the practices adopted during the program</p> <p>Evidence of actions taken by J-PAL to sustain the pilot in the communities after the end of the pilot</p> <p>Evidence of any provisions made by schools or IEP to continue the program activities (financial sustainability, self-funding, or in-kind support).</p>	<p>Independent evaluation</p> <ul style="list-style-type: none"> <li>Field observations</li> <li>Qualitative interviews with the beneficiaries, community members and key informants</li> <li>Interview with the J-PAL</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>Strong signs of sustained benefits in the community</li> <li>Promising prospects for scale-up beyond the company funding</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>Some signs of sustained benefits in the community and ways of improvement are identified</li> <li>Few potential prospects can be explored for scale-up beyond the company funding</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>No sign of benefits or only short-term benefits</li> <li>No prospect for scale-up beyond the company funding was identified</li> </ul>

5. Sustainability					
Evaluation Criteria	Evaluation Questions	Quantitative indicators	Qualitative indicators	Data source and collection method	Assessment definitions
5.2. There are prospects of scale-up beyond GMM2	5.2.1. Are there indications that there is potential for further scale-up of the pilot approach in some way, by the government of Côte d'Ivoire, the Cooperation Partner or other development actors?		<p><b>5.2.1.1. Evidence of government/partners buy-in</b> Level of participation of government representatives in implementation or monitoring Government staff confirm the alignment of the pilot with government priorities</p> <p><b>5.2.1.2. Enabling environment</b> Interviews with relevant stakeholders show that the national and or local environment is suitable for pilot scale-up</p> <p><b>5.2.1.3. Financing</b> Interviews with TRECC, the Cooperation Partner and/or other donors indicate the potential for government, other donors or the Cooperation Partner to fund the scale-up</p> <p><b>5.2.1.4. Organizational capacity to implement at scale</b> J-PAL can demonstrate capacity to operate at larger scale, or plans to handover the program to a partner with this capacity</p>	<p>Independent evaluation</p> <ul style="list-style-type: none"> <li>Keys informant interviews with implementing Côte-D'Ivoire government/Cocoa company relevant representatives</li> <li>Interview with TRECC team</li> <li>Interview with Brookings Institution team</li> </ul>	<p><b>Green:</b></p> <ul style="list-style-type: none"> <li>Strong signs of sustained benefits in the community</li> <li>Promising prospects for scale-up beyond the company funding</li> </ul> <p><b>Orange:</b></p> <ul style="list-style-type: none"> <li>Some signs of sustained benefits in the community and ways of improvement are identified</li> <li>Few potential prospects can be explored for scale-up beyond the company funding</li> </ul> <p><b>Red:</b></p> <ul style="list-style-type: none"> <li>No sign of benefits or only short-term benefits</li> <li>No prospect for scale-up beyond the company funding was identified</li> </ul>



---