

The **LEGO** Foundation

LEGO Foundation Play Accelerator

Partners in Play Ghana Endline Report

September 24, 2024

Prepared for
The LEGO Foundation Play Accelerator Research Initiative

Prepared by
RTI International¹

Jonathan Stern, Rachel Jordan, Chris Cummiskey, Robert Momanyi, Tabitha Nduku, Emily Staskowicz, Jennifer Collett, Sheetal Digari and Alyse Spiehler
3040 East Cornwallis Road
Research Triangle Park, NC 27709-2194

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. Clients rely on us to answer questions that demand an objective and multidisciplinary approach—one that integrates expertise across the social and laboratory sciences, engineering, and international development. We believe in the promise of science, and we are inspired every day to deliver on that promise for the good of people, communities, and businesses around the world. For more information, visit www.rti.org.

With the support of the LEGO Foundation.

¹ RTI International is a registered trademark and a trade name of Research Triangle Institute.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	6
Overview	6
Methodology	6
Implications and Recommendations	8
LTP in the Classroom	8
LTP Enabling Environment	9
Perceptions of LTP	10
Student Holistic Outcomes	10
Conclusions	11
1. INTRODUCTION	12
1.1 Overview of Activity	12
1.2 Purpose and Objectives of Endline	12
1.3 Project Background and Context	13
Approach to Teacher Training and Support	14
Implementation Models	14
Advocacy and Institutionalization of LTP	15
2. EVALUATION DESIGN, INSTRUMENTS, AND SAMPLE	15
2.1 Evaluation Design	15
2.2 Instruments	16
2.3 Samples	18
Quantitative Sample	18
Qualitative Sample	20
2.4 Training and Data Collection	20
Quantitative Training and Data Collection	20
Qualitative Training and Data Collection	22
3. FINDINGS	22
3.1 Highlights	22
3.2 Teachers' Practice, Attitudes, and Support	24
Timed Lesson Observation	24
Post-Lesson Observation Indicators	29
Learning Environment (Student Survey)	32
Stakeholder Perceptions of LTP	34
Teacher Self-Efficacy and Motivation	36
3.3 Students' Literacy, Numeracy, and SEL Outcomes	38
Assessment of SEL	38
Student Literacy and Numeracy Outcomes	42
3.4 LTP-Enabling Environment	45
Program Strengths and Challenges	45
Perceptions of LTP	46
Support to Teachers	48
School Leadership	50
Home Environments	52
Government Support of LTP	56
4. CONCLUSION	60
4.1 Measurement Considerations	60
4.2 Implications and Recommendations	60
LTP in the Classroom	60
LTP Enabling Environment	61
Perceptions of LTP	62
Student Holistic Outcomes	62

4.3 Conclusions	62
ANNEX A. ADDITIONAL TABLES	64
ANNEX B. ATTRITION ANALYSIS	68
ANNEX C. INSTRUMENTS	71

LIST OF FIGURES

Figure 1. Evaluation Design	16
Figure 2. Generalized Theory of Change for LtP Interventions	16
Figure 3. Student Activity during Lessons (Percentage of Observed Class Time), Baseline vs. Endline	25
Figure 4. Teacher Actions during Lessons (Percentage of Observed Class Time), Midline vs. Endline	26
Figure 5. Use of Materials during Lessons (Percentage of Observed Class Time), Midline vs. Endline	27
Figure 6. Student Grouping during Lessons (Percentage of Observed Class Time), Baseline vs. Endline.....	28
Figure 7. Mean Scores of Teachers on the “Attitudes toward LtP” Scale, Baseline vs. Endline	36
Figure 8. Mean Scores of School-Level Stakeholders on the “Attitudes toward LtP” Scale: Baseline v. Endline	37
Figure 9. Teachers’ Responses to “I Feel Valued by School Management,” Baseline vs. Endline	38
Figure 10. Teachers’ Responses to “I Feel Valued by School Management,” Baseline vs. Endline	39
Figure 11. Percentage of Caregivers Who Discussed with the Teacher How Their Child Was Doing: Baseline vs. Endline.....	56
Figure 12. Percentage of Caregivers Who Reported Helping Their Child with Homework in the 3 Days Prior to the Study.....	57
Figure 13. Percentage of Caregivers Who Reported Having Craft Supplies at Home	57
Figure A-1. Teachers’ Responses to “I Feel Valued by Parents,” Baseline vs. Endline.....	67
Figure A-2. Teachers’ Responses to “I Find My Job Rewarding,” Baseline vs. Endline	68
Figure A-3. Teachers’ Responses to “I Have Support I Need from the School Administration to Effectively Implement LtP,” Baseline vs. Endline	69
Figure A-4. Teachers’ Responses to “I Regularly Use Play-Based Teaching Methods in the Classroom,” Baseline vs. Endline	70
Figure A-5. Teachers’ Responses to “LtP Methods are Too Difficult to Implement in the Classroom,” Baseline vs. Endline	70

LIST OF TABLES

Table ES-1. School Sample	7
Table ES-2. Respondent Counts by Treatment Group at Endline—P3.....	8
Table ES-3. Final Count of Interview Respondents	8
Table 1. Target Respondents at Endline.....	17
Table 2. School Sample	18
Table 3. Respondent Counts by Treatment Group at Endline—P3.....	18
Table 4. Count and Percentage of Baseline Students Assessed at Midline and Endline	19
Table 5. Count of Sampled Students by Grade; Baseline, Midline, and Endline.....	19
Table 6. Count and Percentage of Baseline Students Assessed at Midline and Endline	19
Table 7. Final Count of Interview Respondents	20
Table 8. Student Agency (Percentage of Observed Class Time When Students Have Some Choice in Learning Activity or How the Approach Assigned Activities), Midline vs. Endline	29

Table 9. Observed Grade 1 Classroom Climate (Percentage of Observed Class Time), Baseline vs. Endline.....	29
Table 10. Percentage of Grade 1 Lessons in Which Target Instructional Practices Were Observed (with Difference from Baseline), Endline Only	30
Table 11. Percent of Students Reporting Each Learning Environment Experience, Grade 1 Baseline vs. Grade 4 Endline	33
Table 12. Overall Learning Environment Score, Baseline vs. Endline	35
Table 13. Percent of Grade 3 Teachers Who Agreed with the More “Playful” Statement at Endline [Change from Grade 1 Baseline]	35
Table 14. Percent of Caregivers Who Agreed with More “Playful” Statement at Endline [Change from Baseline]	37
Table 15. Observed Student Behaviors, Baseline vs. Endline	40
Table 16. Student Responses to Relationship Questions, Baseline vs. Endline	41
Table 17. Student Relationship Score.....	41
Table 18. Percentage of Students Choosing Each Response to Conflict Resolution, Baseline vs. Endline.....	42
Table 19. Mean Negative Attribution Scores, Midline vs. Endline	43
Table 20. Student Confidence Scores, Baseline vs. Endline.....	43
Table 21. Percentage of Students Identifying with Confident Person, Baseline vs. Endline	43
Table 22. Student Creativity Scores, Baselines vs. Endline	44
Table 23. EGRA Results by Subtask and Treatment Group, Endline	44
Table 24. EGMA Results by Subtask, Endline.....	46
Table 25. Percent of Teachers Who Reported Receiving Support.....	50
Table 26. Percent of Teachers Who Said They Received Play-Based Learning Materials	51
Table 27. Who Provides Teachers with Coaching and Mentoring	51
Table 28. Percent of Teachers Who said Coaching Improved Their Skills as a Teacher	51
Table 29. What about Coaching Made You Change Your Teaching? (Endline Only).....	52
Table 30. Percent of Teachers Who Said They Discuss Play-Based Strategies during Teacher Meetings Sometimes, Most of the Time, or Always	52
Table 31. Percent of HTs Who Felt P3 was Successful in Improving Teacher Practice at Their School.....	53
Table 32. In What Ways Has the LtP Program Been Successful? (Endline)	53
Table 33. What Role Have HTs Played in Helping P3 Improve Learning (Endline)	54
Table 34. How Do HTs Know If a Teacher in Their School Is Not Performing as Expected	54
Table 35. What Action Do School Directors Take if a Teacher is Not Performing as Expected	54
Table 36. Activities that Caregivers Participated in within the Month Prior to the Study	55
Table C-1. Attitudes Scale (Teachers).....	76
Table C-2. Attitudes Scale (Caregivers).....	78
Table C-3. Timed Classroom Observation Sheet (Minute 2–26)	79
Table C-4. Post-observation Checklist.....	81

ABBREVIATIONS

AAM	assessor accuracy measure
COVID-19	coronavirus disease 2019
CPD	continuous professional development
DEO	District Education Officer
DHT	Deputy Head Teacher
DTO	District Training Officer
DTST	District Teacher Support Team
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
GALOP	Ghana Accountability for Learning Outcomes Project
HT	Head Teacher
LOS	lesson observation sheet
LtP	learning through play
MOE	Ministry of Education
NTC	National Teaching Council
P3	Partners in Play Program
PTA	parent-teacher association
RtP	Right to Play
SBM	school-based mentor
SBMT	sector-based mentor trainer
SEL	social and emotional learning
SISO	School Improvement Support Officer
SMS	short message service
UNICEF	United Nations Children's Fund

EXECUTIVE SUMMARY

Overview

The Play Accelerator grant presents a unique opportunity to examine five large-scale educational interventions focused on improving learning through play (LtP) in Bangladesh, Ghana, Kenya, Rwanda, and Vietnam, respectively. One aspect of this activity is determining if and how these five programs change the practice and attitudes of key education stakeholder and whether or not this leads to improved holistic learning outcomes through increasing playful pedagogies. This in turn will generate much-needed rigorous evidence on implementing playful pedagogies successfully through government teacher professional development systems at scale.

This report presents detailed endline findings from the Partners in Play program (P3) implemented by Right to Play (RtP) in Ghana, based on data collected in February and May 2024.

Methodology

The goal of this report is to provide comprehensive endline estimates for P3 in Ghana. These results will serve two purposes: (1) to provide an estimate of program impact since baseline and (2) to provide the LEGO Foundation with data-driven recommendations on program design and implementation to inform future activities.

The endline study comprises two datasets: longitudinal, quantitative data collected at three timepoints (2021, 2022, and 2024); and qualitative data collected at endline only.

The quantitative analysis presented in this report focuses on three main comparisons, addressing the following research questions:

1. Change in attitudes and practice among teachers from baseline to endline, across treatment and comparison groups.

For this analysis, we interviewed and observed teachers assigned to the cohort of students being studied, in Grade 1 at baseline, Grade 2 at midline, and Grade 3 at endline² to address two research questions:

- What is the state of teachers' pedagogical practices in the classroom (including LtP) in treatment and comparison schools at endline? How does this compare to teachers' practices at baseline?
- What are the perceptions of LtP among teachers? How do these compare to their perceptions at baseline?

2. Change in reading, math, and social and emotional learning (SEL) outcomes of students in Grade 4 at endline, compared to Grade 1 (baseline), across treatment and comparison groups.

For this analysis, we assessed the same students at baseline, midline, and endline to address one research question:

² Although students were in Grade 4 at endline, we chose to observe Grade 3 teachers, as this was the latest grade students completed in full.

- What are the endline levels of student literacy, numeracy, and SEL among students at the beginning of Grade 4? Does the progression of students' literacy, numeracy, and SEL scores from baseline (end of Grade 1) differ by treatment group?

3. Changes in the teaching and learning environment from baseline to endline, across treatment and comparison groups.

For this analysis, we interviewed head teachers (HTs), School Improvement Support Officers (SISOs), Curriculum Leads (Teacher Champions), caregivers, District Training Officers (DTOs), and District Education Officers (DEOs) at baseline, midline, and endline to address two research questions:

- What are the enabling factors associated with implementing LtP pedagogical methods in classrooms? What are the barriers associated with implementing LtP pedagogical methods in classrooms?
- What are the perceptions of LtP among school-, district-, and community-level stakeholders at endline? How do these compare to stakeholders' perceptions at baseline?

We conducted qualitative interviews with education officials working at the national and district levels, and a small sample of HTs, teachers, and caregivers. The qualitative analysis presented here aims to answer four sub-questions related to the enabling environment for LtP:

- What elements of the program were most effective in supporting teachers' conceptions of LtP? How did these elements inform their practice in the classroom?
- What, if any, communications, expectations, resources, and support for LtP are delivered by the government?
- What are the design and impact of stakeholder engagement strategies to encourage the integration of LtP pedagogies into classrooms and homes?
- What methods have been used to secure buy-in and commitments to further scale LtP into teacher professional development?

Findings from the qualitative analyses are presented alongside quantitative findings on the enabling environment for LtP.

RTI worked with The Open University's Centre for the Study of Global Development—which led the P3 internal evaluation baseline—to select the sample for this evaluation. Due to the inclusion of 10,000 lower-performing schools in the GALOP program prior to this study, the population of potential control schools was very small, and risked being inherently different (and higher performing) than the schools receiving different arms of implementation. As such, it was determined that identified schools would serve as a non-intervention comparison group but did not constitute a proper control group. Despite this distinction, potential advantages to this “comparison” group should be considered when interpreting results. Additionally, and for this reason, the analysis of student performance on reading and math assessments does not include the comparison group—instead student reading and math scores are compared across intervention groups (with the direct implementation group compared to less intensive interventions (i.e., indirect and GALOP)). The sample sizes and treatment groups are displayed in Table **ES-1**. Additional details on these groups are provided in the report.

Table ES-1. School Sample					
Region	Direct	Indirect	GALOP	Comparison	Total
Greater Accra	10	10	10	10	40
Northern Region	36	29	27	25	117
Volta	11	9	9	11	40
Total	57	48	46	46	197

A final count of respondents per instrument and treatment group in Ghana for the quantitative data collection is presented in **Table ES-2**.

Table ES-2. Respondent Counts by Treatment Group at Endline—P3				
Instrument	Comparison	Indirect	GALOP	Direct
Grade 1 Teacher Interviews	8	6	10	13
Grade 3 Teacher Interviews	45	45	46	55
Caregiver Interviews	92	93	91	111
HT Interviews	45	48	46	57
Coach/SISO Interviews	18	6	15	15
Trainer/DTO Interviews			11	
Teacher Meeting Facilitator/Curriculum Lead Interviews	24	44	45	57
DED Interviews			11	
Classroom Observations—Grade 1 Literacy	8	6	10	14
Classroom Observations—Grade 1 Numeracy	8	6	10	14
Classroom Observations—Grade 3 Literacy	25	23	27	38
Classroom Observations—Grade 3 Numeracy	21	25	19	17
Coaching Observation	5	6	10	11
Teacher Meeting Observation	1	0	7	9
SEL Assessments	451	518	485	605
EGRA and EGMA	451	518	485	605

A final count of respondents for the qualitative data collection is presented in **Table ES-3**.

Table ES-3. Final Count of Interview Respondents		
Geographic Areas	Category of Interviewee	Count of Interviews
National	MOE Staff –Schools & Instruction; Early Childhood Education; Partnerships & Affiliations; Training; Inspectorate; Assessment; NTC	7
Greater Accra South, Weija Gbawe Districts	District staff: DED, DTO	4
	SISO	2
	HTs	6
	Teachers	6
	Parents and caregivers	6

Implications and Recommendations

LTP in the Classroom

Teachers continued to incorporate many basic best teaching practices in their lessons (e.g., student monitoring and support, material preparation, good instructional pace, making connections and sharing ideas, inclusionary practices). However, there was little change in the use of playful practices in the classroom from baseline to endline, across all four groups. There were, however, some notable shifts, particularly in direct implementation schools.

Teaching Approach. In direct implementation schools, there was a significant increase in the proportion of time that students spent reading during observed lessons. Teachers spent the largest proportion of time explaining concepts, modeling activities, and monitoring students' understanding, while students spent the largest proportion of time listening and writing. ***We recommend that teachers build on the additional reading time, in order to continue to engage students with print-based materials. Additionally, the MOE could learn from the direct implementation approach to ensure that students have more opportunities to read in class.***

Material Availability. In both direct and indirect treatment schools, the percentage of teachers who said they received play-based materials increased significantly. The availability of play materials, however, was far more prevalent in direct schools (73%) than in indirect schools (28%). The percentage of observed lesson time during which books were used also increased significantly in direct implementation schools. Overall, despite a focus on materials under P3, there was considerable variation in their availability and use across classrooms. Without access to materials, teachers' implementation of programs will continue to be limited. ***We recommend careful examination of the conditions driving greater availability of materials in direct schools compared to other implementation groups, to help inform planning that ensures classrooms are provided with both traditional materials (e.g., teachers' guides) and play-based materials.***

Student Engagement. There was some evidence of increased student engagement under P3, with teachers noting that active engagement and a child-centered approach was at the heart of the program. ***We recommend building on the program's successes and continuing to increase children's opportunities for meaningful engagement through problem solving, reflection, and critical-thinking activities.***

Group Work. Although the majority of teachers (between 62 and 75%) noted the importance of group work for student learning, neither the use of nor attitudes towards small and large group work changed significantly from baseline to endline, across classes. This may be in part attributed to differences between instruction in grades 1 and 3. ***We recommend providing teachers with explicit guidance, informed by consultative discussions with teachers, on how to increase the use of group work and/or learning corners to improve social interaction and iteration among students. This is particularly valuable in overcrowded classrooms, in which students are at risk of being left behind by whole class instructional approaches.***

Student Agency. Although student agency is a major component of the discussion around LTP, nearly all class time was dedicated to activities initiated and directed by teachers. Opportunities for students to select from choices, choose roles, set goals, make rules, or choose when to move around were virtually non-existent. However, almost one-half of lessons included a chance for students to develop and share ideas with classmates of their teacher. Whether student-initiated or teacher-initiated and student-directed, ***we recommend building on teachers' existing practice of asking students to share ideas, to***

encourage students to share ideas about the direction of their learning. This, in turn, provides students with increased levels of autonomy over how they engage in an activity or, at a minimum, with improved opportunities for asking questions, expressing opinions, and practicing skills.

LTP Enabling Environment

Learning Environment. While students increasingly reported incidences of physical and verbal aggression by their teachers at endline in Indirect, GALOP and comparison schools, this figure did not increase in direct implementation schools – showing a significant impact on the use of yelling and physical punishment. Although these are systemic issues, ***we recommend further investigating how direct implementation schools were able to abate this growing issue and provide a more supportive environment for children.***

Teacher Support Staff. In direct implementation and GALOP schools, significantly more teachers said coaching helped them improve – and most often cited coaches’ friendliness, confidence-building, and use of play as influential factors on their teaching. However, while a variety of actors are engaged to support teachers in P3, but they do not always appear to work together. For example, head teachers claimed that they give additional support to struggling teachers themselves, and teachers who received coaching most often said it came from the head teacher - despite the availability of coaches/mentors as well as teacher meetings. ***We recommend ensuring that support staff understand one another’s roles and that they work together to offer complementary support to teachers.***

Ministry Stakeholders. P3 has made strides in sensitizing ministry officials on the benefits of incorporating LtP in classrooms. However, it was also reported in qualitative interviews that MOE staff at the national level didn’t always understand the nuances of the P3 programming itself. ***We recommend continuing to leverage relationships with officials to clearly communicate expectations around learning through play, including continued sensitization of playful and interactive classrooms for teachers. In particular, the high level of engagement with, and involvement of, district level staff should be continued, promoted and leverages as the level of buy-in and understanding fostered at this level of the system is a unique strength of P3.***

Caregivers. At endline, P3 showed a significant increase in parental engagement with teachers and classrooms. ***We recommend continuing to engage caregivers in school-based activities (through PTAs/SMCs) with an explicit focus on children’s learning. Additionally, further guidance could be provided to households on how to develop low-cost or homemade toys, crafts, and play material in order to engage in LtP practices at home.***

Perceptions of LTP

Attitudes Toward Play. Caregivers, teachers and education stakeholders generally supported LtP in theory but were less supportive of a student-centered, “free” classroom environment. This finding may have stemmed in part from a misconception that student agency and more student-centered approaches will lead to a less disciplined and more chaotic classroom. ***We recommend continuing the work of orienting key stakeholders on the benefits of learning through play in the classroom and understanding their concerns and objections in order to suggest solutions for improving attitudes toward specific playful practices.***

Student Holistic Outcomes

SEL. There was minimal impact on SEL scores for P3 students at endline, with observed confidence actually going down for students in all groups. This means that despite improvements in some teacher practices, there were not significant gains in SEL, aside from relationships. ***We recommend introducing more agency and group work for children in classrooms, which would be expected to increase their confidence, collaboration, conflict resolution, and relationships.***

Literacy and Numeracy. The EGRA and EGMA results represent endline estimates for students in Grade 4. As expected, large gains were found from baseline to midline across all treatment groups. Additionally, the overall impacts for direct implementation schools were typically positive relative to indirect and GALOP schools. Interestingly, the most consistent impacts were found for students in direct schools relative to those in indirect schools. These findings are very promising given the short timeline of the program. ***We recommend continuing to focus on supports for teachers and caregivers, along with materials, in order to promote the best opportunity for continued improvement in literacy and numeracy outcomes.***

Conclusions

Overall, direct implementation schools in P3 led to some meaningful shifts at the classroom, school and community level. This is particularly impressive given the relatively short timeframe of the program, combined with the complexities of teacher behavior change. For example, while there remains a continued overreliance on the use of oral materials in classrooms, there was a considerable shift toward a greater use of books in direct implementation schools—which provided students with more time for engaging with text.

Additionally, teachers in direct implementation schools reported higher levels of feeling valued, than their comparison school counterparts. The program also produced improvements in parental engagement with teachers and classrooms—and led to high levels of ‘pro-play’ attitudes for school-level stakeholders in direct schools. These may not have been the large-scale ‘playful pedagogy’ shifts that the program was hoping for, but they are essential first steps toward greater active engagement for students, creating a foundation for additional changes and improvements in the future.

Some of these changes were able to be achieved due to Right to Play’s adaptive approach to program implementation. During the program’s life cycle, a number of concerns and challenges were identified (e.g., teacher commitment to implementing new practices in the classroom; teacher ability to create learning materials; lack of structured lesson planning). As a response, RtP developed targeted approaches to address these issues. However, there are additional challenges that students, teachers and parents face at a system/structural level (e.g., teacher transfer, lack of infrastructure, high student-teacher ratios, etc.). These challenges cannot be overcome by programs alone but also require the support of government staff, along with funding and policy changes.

The Ministry of Education in Ghana has made it clear that they are interested in moving toward greater use of child-centered approaches, like those touted in Right to Play’s P3 activity. This shift can be seen by their interest in expanding the reach of LtP approaches from several hundred direct and indirect implementation schools to nearly 10,000 schools under the GALOP program. In addition, there appears to be strong support for LtP by MOE staff at the national and district levels.

However, there remains a mixed understanding among central ministry staff regarding what ‘learning through play’ actually means, as well as the specifics of how RTP’s intervention is intended to work. Therefore, while there is broad ‘buy-in’ for the P3 activity and LtP more broadly, it will be important to continue to orient and sensitize national, district, community, and school level stakeholders to the intricacies of LtP programming in schools. This can also help to garner additional support and funding, which will be required to ensure that effective practices can be integrated and sustained into a system beyond the life cycle of programs such as P3.

1. INTRODUCTION

1.1 Overview of Activity

The Play Accelerator grant presents a unique opportunity to examine five large-scale educational interventions focused on improving learning through play (LtP). One aspect of this activity is determining if and how these five programs (implemented separately in Bangladesh, Ghana, Kenya, Rwanda, and Vietnam) change the practice and attitudes of key education stakeholders and whether these changes lead to improved holistic learning outcomes through increasing the use or incorporation of playful pedagogies. This, in turn, will generate much-needed rigorous evidence on implementing playful pedagogies successfully through government teacher professional development systems at scale. A key component of this research is a longitudinal study aimed at measuring the learning experiences and holistic outcomes of children attending Grades 1–4 in schools participating in one of the five LEGO Play Accelerator Partner Programs, as compared with selected control schools, where rollout was delayed. These student-level findings are complemented with classroom observations and education stakeholder and caregiver interviews.

This report presents detailed endline findings from the Partners in Play Program (P3), implemented by Right to Play (RtP) Ghana. These findings are based on data collected in February and May 2024. The design of this evaluation and methodology used is detailed in **Section 2**. Findings from quantitative and qualitative data collected at baseline and endline are presented together in **Section 3**, followed by conclusions and recommendations based on the findings in **Section 4**. Additional tables of data collected, but not included in the primary analyses, are presented in **Annex A**. Because of the longitudinal nature of the study, an attrition analysis is presented in **Annex B**. An overview and examples of the data collection tools can be found in **Annex C**.

1.2 Purpose and Objectives of Endline

The goal of this report is to provide comprehensive endline estimates for P3 in Ghana. These results will serve two purposes: (1) provide an estimate of the program's impact since baseline and (2) provide the LEGO Foundation with data-driven recommendations on program design and implementation to inform future activities.

The endline study comprises two datasets: (1) longitudinal, quantitative data collected at three timepoints (2021, 2022, and 2024) and (2) qualitative data collected at endline only.

The quantitative analysis presented in this report focuses on three main comparisons, addressing the following research questions:

1. Change in attitudes and practice among teachers from baseline to endline, across treatment and comparison groups.

For this analysis, we interviewed and observed teachers assigned to the cohort of students being studied, in Grade 1 at baseline, Grade 2 at midline, and Grade 3 at endline³ to address two research questions:

- What is the state of teachers' pedagogical practices in the classroom (including LtP) in treatment and comparison schools at endline? How does this compare to teachers' practices at baseline?

³ Although students were in Grade 4 at endline, we chose to observe Grade 3 teachers, as this was the latest grade students completed in full.

- What are the perceptions of LtP among teachers? How do these compare to their perceptions at baseline?
- 2. Change in reading, math, and social and emotional learning (SEL) outcomes of students in Grade 4 at endline, compared to Grade 1 (baseline), across treatment and comparison groups.**

For this analysis, we assessed the same students at baseline, midline, and endline to address one research question:

- What are the endline levels of student literacy, numeracy, and SEL among students at the beginning of Grade 4? Does the progression of students' literacy, numeracy, and SEL scores from baseline (end of Grade 1) differ by treatment group?
- 3. Changes in the teaching and learning environment from baseline to endline, across treatment and comparison groups.**

For this analysis, we interviewed head teachers (HTs), School Improvement Support Officers (SISOs), Curriculum Leads (Teacher Champions), caregivers, District Training Officers (DTOs), and District Education Officers (DEOs) at baseline, midline, and endline to address two research questions:

- What are the enabling factors associated with implementing LtP pedagogical methods in classrooms? What are the barriers associated with implementing LtP pedagogical methods in classrooms?
- What are the perceptions of LtP among school-, district-, and community-level stakeholders at endline? How do these compare to stakeholders' perceptions at baseline?

We conducted qualitative interviews with education officials working at the national and district levels, and a small sample of HTs, teachers, and caregivers. The qualitative analysis presented here aims to answer four sub-questions related to the enabling environment for LtP:

- What elements of the program were most effective in supporting teachers' conceptions of LtP? How did these elements inform their practice in the classroom?
- What, if any, communications, expectations, resources, and support for LtP are delivered by the government?
- What are the design and impact of stakeholder engagement strategies to encourage the integration of LtP pedagogies into classrooms and homes?
- What methods have been used to secure buy-in and commitments to further scale LtP into teacher professional development?

Findings from the qualitative analyses are presented alongside quantitative findings on the enabling environment for LtP.

1.3 Project Background and Context

RtP's philosophy encompasses the following key components:

- A learning environment must be friendly and safe for both teachers and students. To this end, building positive rapport with students that addresses students' social and emotional needs is critical.

- A classroom should be a positive learning environment with positive interactions between teachers and students. Teachers should do little talking while children are provided resources and materials to explore. This exploration leads to learning. Materials used to facilitate this learning should be low- or no-cost to support sustainability.
- Involving parents and caregivers is critical to sustain the RtP methodology. RtP works closely with parent-teacher associations (PTAs) to support program implementation.
- The Annual Status Educational Restructure supports P3's monitoring and evaluation. Through this evaluation, additional monies were given to train district mentors to support SISOs in monitoring and mentoring teachers in the field.

Approach to Teacher Training and Support

To support the delivery of RtP approach, P3 uses a cascade model to facilitate continuous professional development (CPD) to teachers. First, the program recruited staff in the three regions of operation (Greater Accra, Volta, and the Northern Region) which formed the program team. This team co-created an LtP training manual and other tools, which were used to train and support teachers to better engage students, build rapport with students, and help students feel comfortable in the classroom. Trainings were conducted with District Teacher Support Teams (DTSTs), which comprised the SISO, the Early Childhood Coordinator, Head of Supervision, Training Officer, Subject Specialists, and Girl Child Officers.

The DTST were, in turn, supported to train teachers. This was done at three levels. The first level of training is the foundational level, which includes life skills, SEL, and ways to incorporate the physical environment into learning. The second level of training, the learning outcomes level, encompasses areas of literacy, numeracy, content area subjects, and how to apply play-based methods of learning across these content areas. The third level of training, the positive learning environment level, covers how to create safe spaces for children to learn and how to manage behavior in productive, proactive, and positive ways. All three training levels occurred once a year.

Implementation Models

P3 is implemented through three models: direct implementation, indirect implementation, and in Ghana Accountability for Learning Outcomes Project (GALOP) schools.

Direct

In direct implementation schools, P3 was involved in supervising, training, and monitoring, while also supporting the DTST to implement the RtP methodology. This support came from two key roles employed by P3: (1) the project coordinators coordinated activities between the RtP field coordinators and schools, while (2) the field coordinators had direct interaction with the schools and provided direct support to the DTST to implement training at the school. RtP provided PTA leaders in direct schools with information about P3, which they were to relay to caregivers. Direct schools also received financial support for mentors, materials distribution, and post-training follow-up support. Although the SISO's goal is to visit one school each day, this is not feasible due to limited resources. As such, RtP mentors working for the district were also facilitated by P3 to provide support. During observations and mentor visits, any gaps identified were addressed during Professional Learning Community meetings, which were held in the schools on Wednesdays.

Indirect

In indirect implementation schools, the DTST were trained, and in charge of monitoring and supporting schools in implementing the RtP methodology. The goal of this variation in support was to determine the efficacy of RtP's role in P3 monitoring and development.

GALOP

GALOP targets the 10,000 lowest performing schools in Ghana, with additional support provided from the Ministry of Education (MOE), jointly funded by the World Bank, the U.K. Foreign, Commonwealth & Development Office (formerly the Department for International Development), and Global Partnership for Education. As part of their effort to improve students' performance in GALOP schools, the MOE requested training for a subset of GALOP schools. RtP employed the indirect approach (described above) to deliver these trainings.

Advocacy and Institutionalization of LtP

RtP handed over the LtP materials it developed to the districts, with the expectation that the DTST team will continue monitoring and engaging teachers on the methodology. P3 trainings are also expected to continue at the school level, through the established professional learning communities.

At the national level, the National Teaching Council (NTC) has a repository of trainings that are housed online and which contribute to teachers' CPD. RtP is endeavoring to include the LtP trainings developed under P3 in this repository.

2. EVALUATION DESIGN, INSTRUMENTS, AND SAMPLE

2.1 Evaluation Design

This evaluation uses a quasi-experimental design that is longitudinal at the student and teacher level. The study includes three data collection time points: baseline, midline, and endline. Although the initial design was intended to cover 3 years of implementation, coronavirus disease 2019 (COVID-19) pandemic delays and school closures led to a shortened time frame. The evaluation timeline now consists of a baseline conducted in 2021, a midline conducted in 2022, and an endline conducted in 2024.

Baseline data were collected in June 2021, when students were in the middle of grade 1. Due to shifts in the school calendar, midline data were then collected in November 2022, at the end of the school year. Endline data were collected at a similar point in the school year as baseline – when most sampled students were in the middle of Grade 4 (May/June 2024).

At the student level, the study collected baseline data from students at the end of Grade 1. These same students were tracked and assessed at the end of Grade 2, at midline, and again halfway through Grade 4, at endline. At the teacher level, baseline data collection occurred with Grade 1 teachers. At midline and endline, interviews and observations were conducted with baseline teachers who were still teaching Grade 1 at the school. Unfortunately, many Grade 1 teachers were reassigned by the time the endline data were collected. Interviews and observations were also conducted with Grade 2 teachers (midline) and Grade 3 teachers (endline)⁴. These teachers reflect the “typical” experience of teachers

⁴ Although our assessed students were starting Grade 4 at endline, we interviewed and observed Grade 3 teachers because they were the last to give the students a full year of instruction.

receiving the program. Due to the longitudinal nature of the design, we visited the same schools at each time point.

Any differences between implementation groups at baseline are accounted for in impact estimates by adjusting for baseline covariates in regression models.

As will be discussed in Section 2.3, schools from each of the three implementation groups, as well as designated comparison schools were selected from three regions, representing four main languages of instruction: Greater Accra (Asante-Twi and Ga); Northern Region (Dagbani), and Volta (Ewe). As **Figure 1** shows, potential impacts for teacher level effects and SEL were measured by comparing changes in mean scores from baseline to endline for each of the treatment groups, to the difference from baseline to endline for the comparison group (difference in differences). However, for student literacy and numeracy outcomes, comparison schools were dropped from the analyses (as shown in **Figure 1b** and explained in Section 2.3).

Figure 1a. Initial Evaluation Design

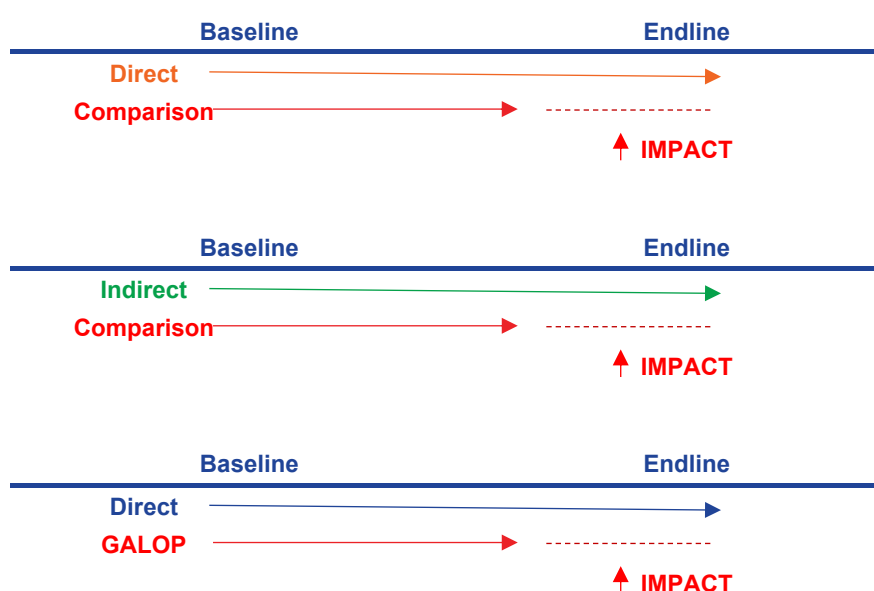
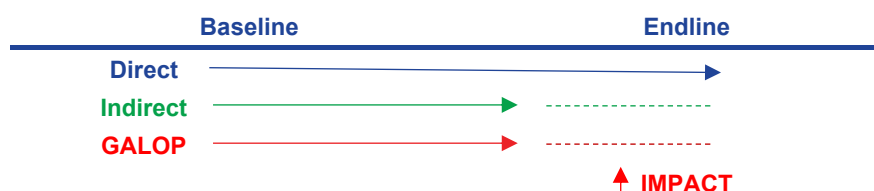


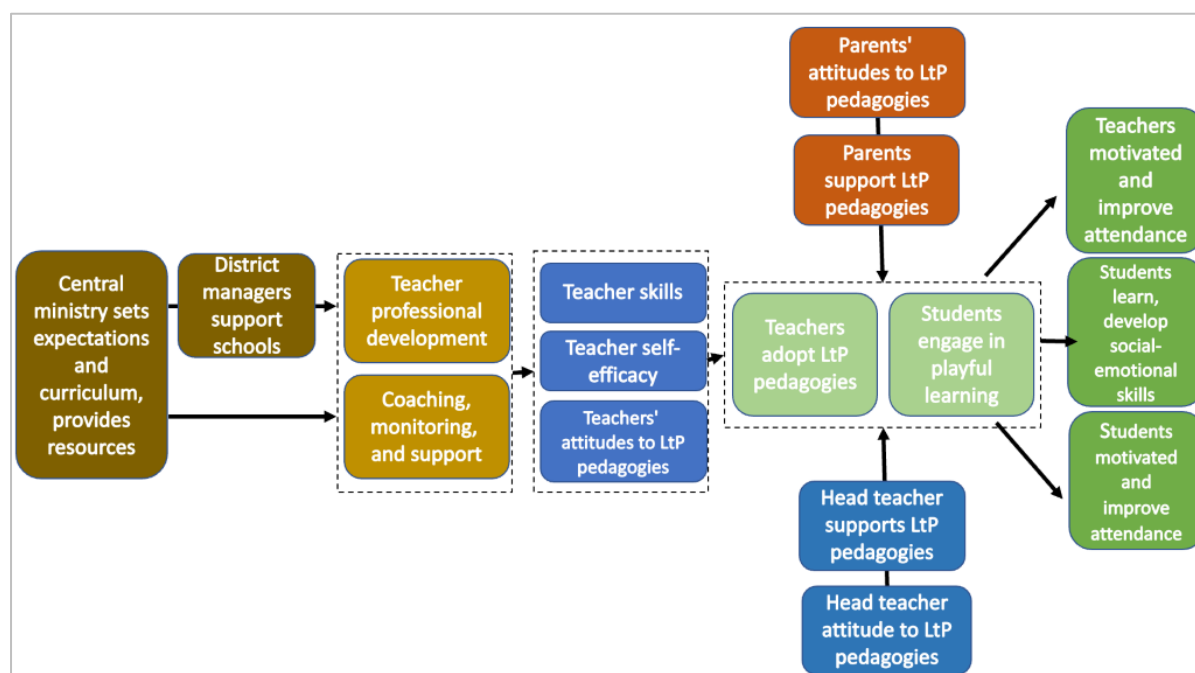
Figure 1b. Evaluation Design for EGRA and EGMA



2.2 Instruments

The instruments used in this study were developed to measure components in the LtP theory of change (**Figure 2**), as well as relationships among components.

Figure 2. Generalized Theory of Change for LtP Interventions



For this study, RTI developed and employed classroom observations, student assessments of holistic outcomes (i.e., SEL, Early Grade Reading Assessment [EGRA], Early Grade Mathematics Assessment [EGMA]), and fixed-response survey instruments for teachers, HTs, coaches, mentors, teacher meeting facilitators, trainers, and district education officials. Program staff designed these instruments to be consistent across all five Play Accelerator grant programs. Each program received draft instruments and had their feedback incorporated in the final versions (see **Annex C** for details on the instruments). **Table 1** outlines the target respondents for each instrument in Rwanda.

Table 1. Target Respondents at Endline

Instrument	Respondent
Teacher Interview and Attitudes Scale	Grade 1 teacher from baseline Grade 3 teachers (2 per school: Grades 1 and 3)
Caregiver Interview and Attitudes Scale	Grade 1 caregivers (2 per school)
HT Interview and Attitudes Scale	HTs (1 per school)
Coach Interview/Observation	SISO (1 per circuit)
Trainer Interview	DTO (1 per district)
Teacher Meeting Facilitator Interview/ Observation	Curriculum Leads/Champion Teachers (1 per school)
District Education Official Interview	District Education Director (DED; 1 per district)
Classroom Observations	Grade 1 (1 to 2 observations, if available) Grade 3 (1 observation, literacy or numeracy)
SEL Assessment	Students from Grade 1 baseline (16 per school: 8 girls, 8 boys)

2.3 Samples

Quantitative Sample

Although it was not possible to randomly assign treatment and comparison groups for the LEGO Play Accelerator programs, the approach included a random selection of treatment schools along with a random, matched sample of comparison schools.

P3's direct and indirect intervention activities were implemented in seven districts across three regions (Greater Accra, Volta, and Northern Region). Overall, the program includes three distinct approaches to implementation that are included in this evaluation: (1) Direct implementation (100 schools within 7 districts; includes intensive training for all teachers, as well as HTs and district officials), (2) Indirect implementation (256 schools within 7 districts; includes condensed training for three teachers per school), and (3) GALOP schools (4,000 schools, including those in 5 P3 districts that received master training and direct district-level support from P3; 6,000 additional GALOP schools received indirect support through the master training only). The proposed design for this evaluation included 50 schools from each of these three treatments, along with 50 comparison schools.

To streamline our work and maximize the value of the data collected, RTI worked with the United Kingdom Open University's Centre for the Study of Global Development—which led the P3 internal evaluation baseline—to select the sample for this evaluation. Open University's multistage cluster sampling approach led to a final sample of 75 schools for each of the three treatment groups, as well as 75 comparison schools (which were selected from nearby districts in each of the three regions). Our evaluation sample consisted of 50 randomly selected schools from each of these four groups (three treatment groups and one comparison). This random selection was stratified across the three regions (with 30 Northern schools, 10 Volta schools, and 10 Greater Accra schools in each group). The final sample included schools from 11 districts across the three regions.

Due to the inclusion of 10,000 lower-performing schools in the GALOP program prior to this study, the population of potential control schools was very small, and risked being inherently different (and higher performing) than the schools receiving different arms of implementation. As such, it was determined that identified schools would serve as a non-intervention comparison group but did not constitute a proper control group. Despite this distinction, potential advantages to this “comparison” group should be considered when interpreting results. Additionally, and for this reason, the analysis of student performance on reading and math assessments does not include the comparison group—instead student reading and math scores are compared across intervention groups (with the direct implementation group compared to less intensive interventions (i.e., indirect and GALOP)).

The intended sample consisted of 200 schools, evenly distributed across the four groups. However, due to logistic complications during data collection, the final sample numbers varied slightly, as shown in **Table 2**.

Table 2. School Sample					
Region	Direct	Indirect	GALOP	Comparison	Total
Greater Accra	10	10	10	10	40
Northern Region	36	29	27	25	117
Volta	11	9	9	11	40
Total	57	48	46	46	197

A final count of respondents, per instrument and treatment group is presented in **Table 3**.

Table 3. Respondent Counts by Treatment Group at Endline—P3				
Instrument	Comparison	Indirect	GALOP	Direct
Grade 1 Teacher Interviews	8	6	10	13
Grade 3 Teacher Interviews	45	45	46	55
Caregiver Interviews	92	93	91	111
HT Interviews	45	48	46	57
Coach/SISO Interviews	18	6	15	15
Trainer/DTO Interviews		11		
Teacher Meeting Facilitator/Curriculum Lead Interviews	24	44	45	57
DED Interviews		11		
Classroom Observations—Grade 1 Literacy	8	6	10	14
Classroom Observations—Grade 1 Numeracy	8	6	10	14
Classroom Observations—Grade 3 Literacy	25	23	27	38
Classroom Observations—Grade 3 Numeracy	21	25	19	17
Coaching Observation	5	6	10	11
Teacher Meeting Observation	1	0	7	9
SEL Assessments	451	518	485	605
EGRA and EGMA	451	518	485	605

Students were included in the endline analysis if they were assessed at both baseline and endline (regardless of whether they were present for midline data collection). **Table 4** shows the count and percentage of students according to when they were assessed. In total, 69.2% of the students initially assessed at baseline were also assessed at endline.

Table 4. Count and Percentage of Baseline Students Assessed at Midline and Endline				
Baseline	Midline	Endline	Number of Students	Percentage (Out of All Students Assessed at Baseline)
			441	14.8%
			476	16.0%
			142	4.8% ^a
			1,917	64.4% ^a

^a Indicates students included in the baseline to endline comparison presented in this report.

In tracking as many students as possible from baseline, the study team assessed students regardless of what grade they were in and documented their current grade during each round of data collection. As **Table 5** shows, at endline, 8.2% of students initially assessed in

Grade 1 in 2021, who should have matriculated to Grade 4 in school year (SY) 2024, remained in Grade 1 ($n = 2$); Grade 2 ($n = 17$); or Grade 3 ($n = 151$).

Table 5. Count of Sampled Students by Grade; Baseline, Midline, and Endline					
Time of Data Collection	No. Students in Grade 1	No. Students in Grade 2	No. Students in Grade 3	No. Students in Grade 4	Total
Baseline	2,976	0	0	0	2,976
Midline	68	2,300	25	0	2,393
Endline	2	17	151	1,889	2,059

Similar to the student sample, Grade 1 teachers were included in the analyses presented in this report if they were confirmed to be the same teacher, still teaching Grade 1 at endline—regardless of whether the teacher was interviewed, observed, or teaching Grade 1 at midline. **Table 6** shows the count and percentage of teachers according to when they were interviewed and observed. In total, only 18.8 % of the Grade 1 teachers initially interviewed and observed at baseline were still assigned to their same grade and school at endline. As such, the primary teacher interviews and observations presented in this report are from Grade 3 teachers at endline.

Table 6. Count and Percentage of Baseline Students Assessed at Midline and Endline				
Baseline	Midline	Endline	Number of Teachers	Percentage (Out of All Teachers Observed at Baseline)
			90	45.7%
			70	35.5%
			11	5.6% ^a
			26	13.2% ^a

^a Indicates teachers eligible for the baseline to endline comparison.

Qualitative Sample

For qualitative interviews with MOE officials, P3 staff recommended key respondents who had interacted regularly with the program. Additionally, at RTI's request, the program recommended districts and schools with past evidence of a strong fidelity in program implementation. The overall district- and school-level samples included two district offices and between two and four schools per district (six schools total). At each school, staff requested an interview with the HT, one teacher receiving the program (typically Grade 2) and one parent or caregiver of a child receiving the program. A final count of qualitative interview respondents is presented in **Table 7**.

Table 7. Final Count of Interview Respondents		
Geographic Areas	Category of Interviewee	Count of Interviews
National	MOE Staff –Schools & Instruction; Early Childhood Education; Partnerships & Affiliations; Training; Inspectorate; Assessment; NTC	7
Greater Accra South, Weija Gbawe Districts	District staff: DED, DTO	4
	SISO	2
	HTs	6
	Teachers	6

Table 7. Final Count of Interview Respondents		
Geographic Areas	Category of Interviewee	Count of Interviews
	Parents and caregivers	6

2.4 Training and Data Collection

Quantitative Training and Data Collection

With support from the data collection firm subcontractor, DevtPlan, consultants and a team from RTI facilitated a residential training for 59 assessors in Accra, Ghana, from May 6 to 10, 2024, at the University of Ghana, Legon Campus. During the training, participants were divided into three groups based on their previous experience and skillsets: (1) classroom observers, (2) EGRA and EGMA assessors, and (3) SEL interviewers. Most assessors were recruited at baseline and midline and trained in the same groups across timepoints to ensure consistency in content mastery.

Trainees participated in an assessor accuracy measure (AAM), which evaluated their ability to accurately score items. Participants watched videos and used a tablet to score as if they were conducting the assessment or observation. The scores were then compared to a gold standard scoring, which had previously been completed by a trainer scoring the same video. Problematic items identified through the AAM exercise were addressed and improved upon during the training. AAM scores and observations conducted during two practice school visits (on Days 3 and 4 of training) were used to select the top performing assessors to participate in fieldwork.

Average scores on the final AAM were as follows:

- Classroom Observation—96.0% agreement
- SEL—90.0% agreement
- EGMA subtasks (all assessors)—95.2% agreement
- EGRA subtasks (Ga language assessors)—96.0% agreement
- EGRA subtasks (Dagbani language assessors)—97.0% agreement
- EGRA subtasks (Ewe language assessors)—97.0% agreement.

Quantitative data were collected from May 13 to May 31, 2024. In advance of school visits, a team of 12 field officers were trained as “runners,” who were tasked to visit schools and update the database of students. Runners were divided into three groups, with each group assigned to one of the three study regions. Training for runners in the Greater Accra and Volta regions was held on May 3, 2024, at the DevtPlan Consult office in Accra. The Northern Region runners were trained on May 3 at AgriSearch Lead office in Tamale. The runners visited each school to

- Make introductions to the HT,
- Identify students who were sampled for the assessments during baseline,
- Document names of sampled students who left the school for particular reasons,
- Identify teachers and HTs who were interviewed during baseline, and
- Work with the HT to request that caregivers attend interviews during the planned data collection days.

Recognizing the risk to the sample posed by the absence of students selected at baseline on the day of endline data collection and other factors, assessors were trained on how and when to replace students, as given below.

- If a student had **transferred**, assessors were advised to replace them. The replacement was supposed to be done with students (1) for whom the team already had consent on file from baseline and (2) who had attended that school the whole of SY2023. If there were not enough students with consent already obtained, an EGRA or EGMA assessor randomly sampled a new student from the attendance register and obtained verbal consent over the phone.
- If a student was **absent**, teams were advised to do their best to find the student and conduct the assessments. This included waiting for the student to come to school or arranging to visit the student at home (after completing data collection at the school).
- If a student was **repeating a grade**, assessors were advised to assess them and make a note in the “Additional Comments” section of the EGRA or EGMA.

During data collection, student assessments, classroom observations, and caregiver interviews were prioritized during morning hours to adhere to schools’ timetables and caregivers’ schedules. School-level stakeholder interviews were conducted during the school day when respondents were available. DEO, trainer, and coach interviews were conducted at each respondent’s convenience at district offices or other agreed-upon meeting places.

Qualitative Training and Data Collection

Qualitative training for two qualitative researchers and two note takers was conducted from April 1–5, 2024, by a qualitative research expert consultant, with support from DevtPlan. During the training, participants were introduced to the five interview protocols to use for the qualitative study.

On Day 1 of training, participants read, discussed, and practiced conducting the interviews. Day 2 included a school visit to pilot the school-level tools and a visit the program offices. Based on program interviews conducted, all instruments were then adapted to the context and the approach being studied. Although the intent of the questions and what they aimed to measure did not change, terms were revised, and sub-questions were added to gather as much meaningful data as possible. Days 3, 4, and 5 were used as pilot days in selected schools and for tool revisions based on learnings from the piloting. Data was collected April 8–19, 2024.

3. FINDINGS

3.1 Highlights

This section describes endline findings in detail, organized by the three main study components presented in Section 1.2. Overall, the impacts on teacher and school level changes for P3 intervention schools was very promising and provides strong evidence of effective implementation. Key findings, by treatment group, are outlined below:

Direct Implementation

- The direct implementation group saw the greatest impacts on teaching practice, school environment, school-based support and caregiver involvement. This group also saw the most intensive intervention, including direct support to training, supervision and monitoring through field coordinators, direct information-sharing with PTA executives, and financial support for mentors, materials distribution, and post-training follow-up support.
- At endline, children in direct implementation schools are spending significantly more time actively reading and the use of books during instruction has increased significantly.
- While grade 4 (endline) students in all groups are spending more time listening during lessons than they did in grade 1 (baseline), this shift to passive learning has been mitigated in direct implementation classrooms, where the increase in “listening time” is significantly lower than in comparison schools. Hand in hand with this, direct implementation teachers significantly increased the amount of time they spend monitoring students while they undertake learning activities on their own or in groups.
- Overall, student reports of yelling or physical punishment by adults in their school increased in grade 4 (endline), but this negative trend was mitigated in the direct implementation group, where reporting of these behaviors was similar to when students were in grade 1 (baseline).
- Teachers in direct implementation schools report significant increases in three areas of support: receipt of play-based materials, SMS or other remote support, and the positive impact of coaching on their teaching.
- In terms of the enabling environment, head teachers’ attitudes towards learning through play improved significantly from baseline in direct implementation schools. Additionally, the percentage of caregivers who say they meet with their child’s teacher and who volunteer in class also increased significantly.
- Student performance gains in reading and numeracy were stronger in direct implementation schools than they were in indirect implementation or GALOP schools.

Indirect Implementation

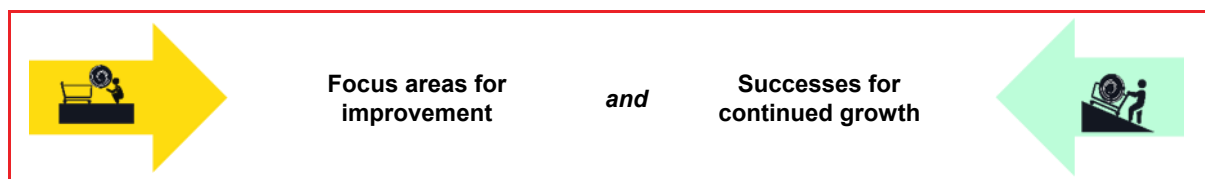
- In the indirect implementation group, many of the functions directly supported by P3 in the direct implementation group are instead carried out almost entirely by the DTST (with P3 training). This group saw fewer impacts than the direct implementation group, though areas where significant changes were observed (in training, dedicated resources, and caregiver engagement) may point to which aspects of the program are most sustainable.

- The number of lessons observed in which students both have a dedicated resource and use it to solve a problem (including more traditional materials, manipulatives, charts, reference books) increased significantly in indirect schools. While the overall prevalence of materials was much lower than in direct implementation schools, the percentage of teachers who received materials also increased significantly in indirect implementation schools.
- By endline, the percentage of indirect implementation teachers who received training increased significantly.
- The percentage of caregivers who volunteered in class increased significantly in indirect implementation schools.

GALOP Schools

- In support of GALOP, RTP provided an intervention approach similar to that of the indirect implementation group – the key difference being that this was part of a larger activity implemented by the MOE, explicitly targeting the lowest performing schools. Impacts are mixed, and it may be difficult to parse out what changes are linked to the P3 intervention and where they are the result of other programming.
- In grade 3 (endline), teachers spend more time on non-academic management tasks than their grade 1 counterparts (at baseline) in all groups. However, this increase is significantly less in GALOP schools, suggesting more time is dedicated to academics. At the same time, opportunities for students to use manipulatives, or give a physical response, reduced significantly from baseline in GALOP schools.
- At endline, significantly more parents and caregivers in GALOP schools are meeting with teachers to discuss how their child is doing. At the same time, significantly fewer teachers feel parents welcome LTP in GALOP schools.
- In terms of student outcomes, children's relationships score improved significantly in GALOP schools, specifically in the number of children who had someone they could tell secrets to.

The remainder of this section describes the baseline results in detail while highlighting key findings using icons on the margins, as follows:



The first two subsections present quantitative data only. The third subsection presents quantitative and qualitative data.

3.2 Teachers' Practice, Attitudes, and Support

At endline, teacher interviews and classroom observations were conducted with two groups of teachers: (1) Grade 1 teachers who were first interviewed and observed at baseline and (2) Grade 3 teachers who had most recently provided a full year of instruction to the study's student cohort. At endline, very few Grade 1 teachers from baseline were still teaching Grade 1 at the same school. As such, the focus of this section is the comparison between Grade 1 teachers at baseline to Grade 3 teachers at endline.

Observations at endline were conducted for the full duration of the lesson, which ranged from 16 minutes to 50 minutes with a median duration of 34 minutes for both math and reading lessons.

Lesson observations for each Grade 3 teacher were conducted in either a reading or math class. Each observation employed two tools: (1) a timed lesson observation tool and (2) a post-observation checklist. Results are averaged at the teacher level and disaggregated by treatment group throughout this section.

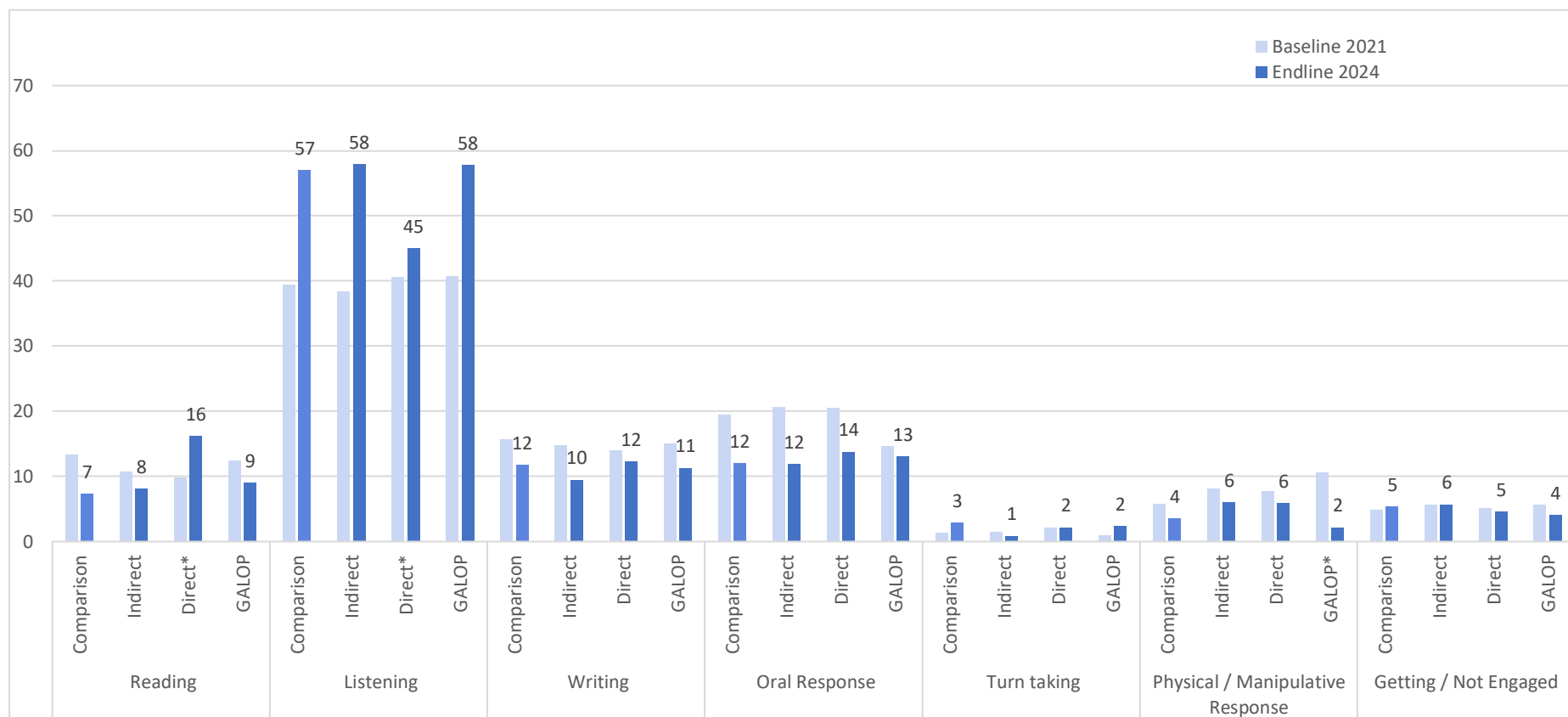
Timed Lesson Observation

Student Engagement and Activity

Across all treatment groups, students spent the largest amount of class time listening, followed by speaking (oral response), reading, and writing (**Figure 3**). In direct implementation schools **the percentage of class time that students spent actively reading increased significantly from baseline. In line with this, while the percentage of class time students spent listening in direct implementation schools did increase from Grade 1 baseline, it increased by significantly less than in comparison schools.**

No significant changes were observed in the opportunities for students to provide more play-based responses, especially through turn-taking. Further, **the percentage of class time students spent giving a physical response or working with manipulatives decreased significantly from baseline in GALOP schools.**

Figure 3. Student Activity during Lessons (Percentage of Observed Class Time), Baseline vs. Endline



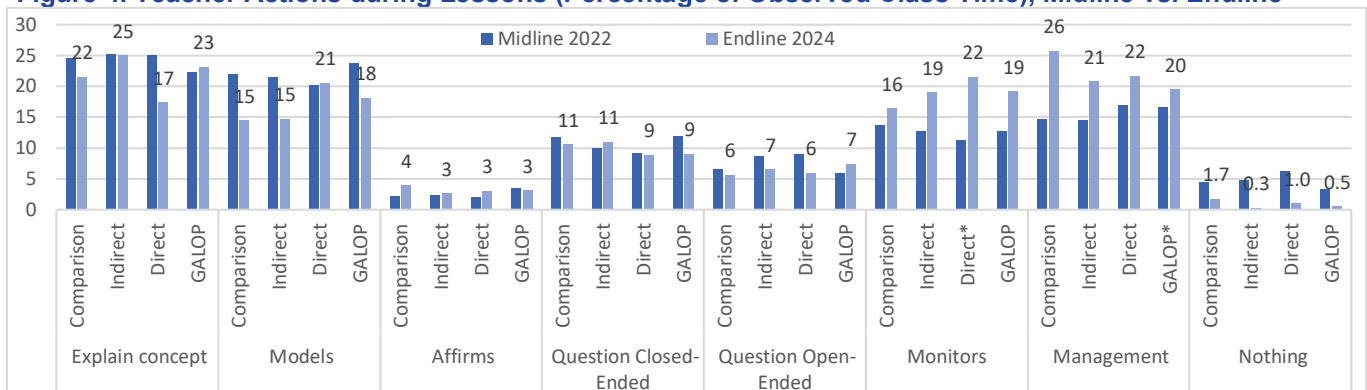
* $p > 0.05$; "Getting" refers to taking out books or other supplies.

Teacher Practice

Due to changes in the observation tool related to this construct that were made after baseline, we compared endline data to midline data for teacher practice. As **Figure 4** shows, teachers spent the most class time on management activities (i.e., taking attendance, giving general instructions, marking books); explaining concepts; and monitoring students as they practiced skills. **Although the percentage of time teachers spent on management tasks increased significantly in all groups, this increase was significantly lower in GALOP schools when compared to comparison schools** (which saw the greatest increase in non-academic management time of any group). **The percentage of time teachers in Direct implementation schools spent monitoring as students worked on their own increased significantly from baseline when compared to comparison.** Another promising finding is the relatively frequent use of close- and open-ended questioning by teachers, observed for between 15% and 18% of class time.



Figure 4. Teacher Actions during Lessons (Percentage of Observed Class Time), Midline vs. Endline

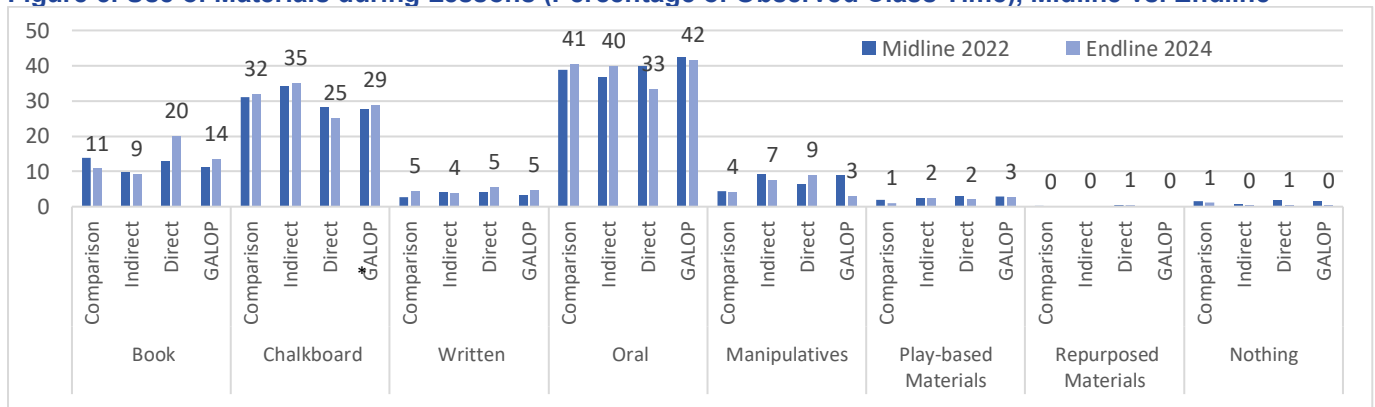


*p>0.05

Materials Use

Again, due to changes to the observation tool after baseline, findings on the use of materials (**Figure 5**) compare midline to endline. **During roughly one-third of lessons, teachers used no materials, other than speaking** (see “oral” in **Figure 5**). **Chalkboards were employed as the main material for learning between 25% and 35% of class time.** The percentage of lesson time during which books were used increased significantly in direct implementation schools, when compared to comparison. **Use of manipulatives and play-based materials in schools remained low overall at endline.**

Figure 5. Use of Materials during Lessons (Percentage of Observed Class Time), Midline vs. Endline



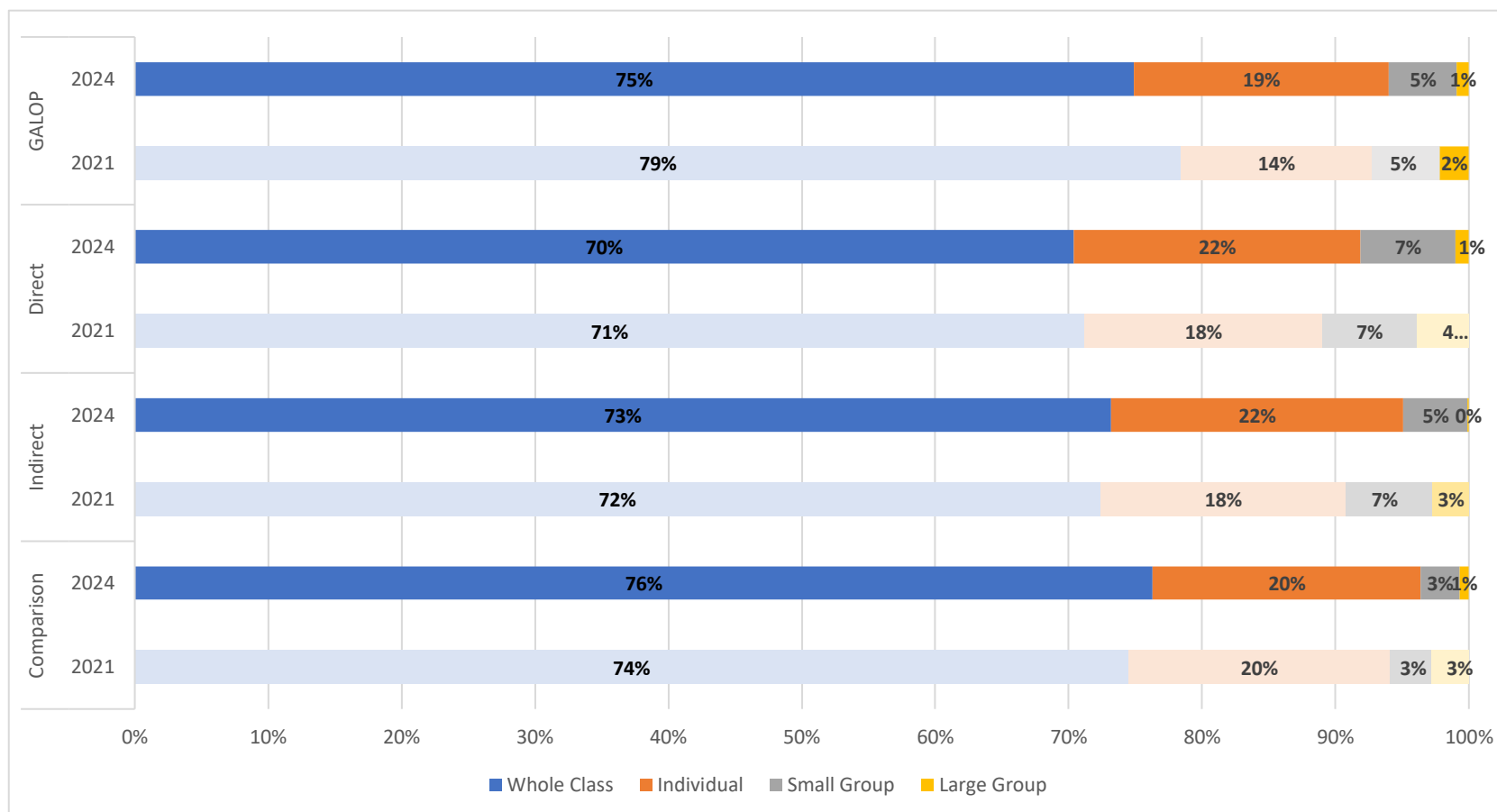
*p>0.05

Learning Environment



At endline, teachers continued to rely on whole-class teaching for about 75% of instructional time (Figure 6). Across all groups, no significant changes were noted with the use of small groups, large groups, and individual work.

Figure 6. Student Grouping during Lessons (Percentage of Observed Class Time), Baseline vs. Endline



Assessors also measured the percentage of class time that was “student-centered,” i.e., when students have some choice through planned or self-directed learning, rather than simply following directions (**Table 8**). This can include allowing students to assign roles, decide how to solve a problem, or choose a writing topic or story to read. Again, due to changes in the tool, these data were compared to midline. **At endline, no significant changes were detected and virtually no class time was allocated for activities that were either directed or initiated by students. This is consistent with the use of the whole-class approach and teachers spending much of class time on management activities.**

Table 8. Student Agency (Percentage of Observed Class Time When Students Have Some Choice in Learning Activity or How the Approach Assigned Activities), Midline vs. Endline

	Midline	Endline
	Grade 2	Grade 3
Comparison	2.9	1.2
Indirect	3.1	0.4
Direct	2.3	1.1
GALOP	4.2	0.3

Assessors also measured classroom climate during observed lessons, focusing on four main criteria: (1) perseverance (e.g., recognition of effort or encouragement of goal setting), (2) positive verbal/physical signs (e.g., use of positive language or physical evidence of joy), (3) negative verbal/physical signs (e.g., use of negative language or physical punishment), and (4) ordinary signs (e.g., tone is neither noticeably positive nor negative). **Table 9** shows the percentage of time when the classroom climate was observed to be explicitly positive or negative. **The percentage of class time in indirect schools where an explicitly positive climate was observed (through positive words or actions) decreased significantly from baseline, when compared to the comparison group. Most of these positive interactions observed at baseline in Grade 1 had been teacher-initiated.**

Table 9. Observed Grade 1 Classroom Climate (Percentage of Observed Class Time), Baseline vs. Endline

	Positive Climate		Negative Climate	
	Baseline	Endline	Baseline	Endline
	Grade 1 2021	Grade 3 2024	Grade 1 2021	Grade 3 2024
Comparison	17.1	9.7	0.8	0.1
Indirect*	21.8	7.2	1.0	0.3
Direct	19.4	11.2	0.8	0.1
GALOP	14.3	9.7	0.6	0.2

*p>0.05

Post-Lesson Observation Indicators

At the end of each timed classroom observation, assessors completed a checklist to determine whether a range of activities occurred at least once during the lesson. These indicators were divided into six construct categories:

1. Integrated pedagogies

2. Application (use of games)
3. Responsiveness (to student behaviors)
4. Resources (availability and use)
5. Efficiency (maximized instructional time)
6. Student-directed (ownership or autonomy).

Table 10 shows the prevalence of play-related pedagogical activities occurring throughout the observed lessons (with high prevalence in green, moderate prevalence in yellow, low prevalence in pink, and very low prevalence in gray).

Pedagogies that integrate some aspect of play (category 1) were observed in more than half of the lessons. The most frequently observed of these pedagogies were giving students a chance to make connections to their prior knowledge or experiences and to solve problems or think critically.

In terms of student games (category 2), the majority of lessons included some kind of energizer or song, although far fewer lessons included symbolic, pretend, or physical play.

Nearly all lessons included some evidence of teachers' responsiveness to students (category 3). Teachers in all groups also had their instruction materials ready and made sure the instruction pace was lively (category 5).

The percentage of lessons in which students had a dedicated resource (category 4) that they used to solve a problem (e.g., manipulatives, math charts, dictionary), increased for all groups and increased significantly more in indirect schools (when compared to comparison). There was also a significant increase in the percentage of indirect lessons of students using any resource that was dedicated for their use.

In comparison schools, the percentage of lessons in which students were given a choice (category 6) decreased significantly when compared to GALOP schools, which saw a very small increase. However, the percentage of GALOP lessons where students were given a choice of how long to take on an activity decreased significantly.

Table 10. Percentage of Grade 1 Lessons in Which Target Instructional Practices Were Observed (with Difference from Baseline), Endline Only

Observed (with Difference from Baseline), Ending Only					
	High prevalence (75%–100%)	Moderately prevalent (50%–74%)	Low prevalence (25%–49%)	Very low prevalence (0%–24%)	BOLD Indicates statistically significant differences
	Comparison		Indirect	Direct	GALOP
1. Integrated Pedagogies					
The teacher asked questions or raised topics that got students to respond to each other, not to them	58.7 [20.5]	63.3 [32.1]	71.4 [25.4]	63 [33.8]	
Student(s) persisted with the activity after the rest of the class had started to do something else (i.e., engage).	47.8 [28.7]	49.0 [32.9]	60.7 [39.5]	41.3 [30.1]	
Student(s) started the task again or adjusted when they thought of a better idea/approach (i.e., iterative).	37.0 [22.3]	32.7 [10.1]	51.8 [30.5]	37 [21.2]	
Student(s) had a chance to make connections between the new experiences/content to their prior knowledge or experiences (i.e., meaningful).	69.6 [26.9]	85.7 [44.9]	87.5 [31.7]	84.8 [34.2]	

Table 10. Percentage of Grade 1 Lessons in Which Target Instructional Practices Were Observed (with Difference from Baseline), Endline Only

	High prevalence (75%–100%)	Moderately prevalent (50%–74%)	Low prevalence (25%–49%)	Very low prevalence (0%–24%)	BOLD Indicates statistically significant differences
	Comparison	Indirect	Direct	GALOP	
Students problem-solved or had a chance to think critically or reflect about a topic.	84.8 [12.9]	85.7 [26.6]	83.9 [10.5]	76.1 [22.2]	
2. Application: Students have generalized games					
Songs or energizers to start or divide activities.	69.6 [-9.1]	67.3 [-9]	78.6 [-10.8]	73.9 [-7]	
Students played games with obvious rules.	19.6 [2.7]	24.5 [-6.7]	25.0 [-15.7]	39.1 [15.5]	
Students had physical play.	15.2 [-3.9]	32.7 [5.8]	33.9 [-10.3]	39.1 [6.5]	
Students used objects to represent something else.	2.2 [-6.8]	12.2 [-0.7]	14.3 [-0.8]	13 [1.8]	
Students had symbolic play.	13.0 [6.3]	12.2 [-0.7]	17.9 [1.9]	17.4 [14]	
3. Responsiveness: The teacher adapts to student behaviors					
The teacher gave students hints, suggestions, or redirection when they needed it.	91.3 [5.9]	98.0 [10.9]	100.0 [8.8]	93.5 [4.7]	
The teacher circulated, leaned in, or redirected to support students, as needed.	76.1 [9.8]	87.8 [21.1]	92.9 [20.3]	82.6 [14.1]	
The teacher asked the class or a group for a solution to a student question or problem.	41.3 [-7.0]	57.1 [7.7]	60.7 [0.5]	45.7 [8.6]	
The teacher included students who did not volunteer to answer (i.e., inclusion).	82.6 [15.2]	83.7 [8.4]	87.5 [1.7]	80.4 [7.4]	
4. Resources: Materials that are needed for implementation are used					
Students used a resource that was dedicated for their use (e.g., book, workbook, slate, tablet).	39.1 [3.2]	67.3* [30.8]	69.6 [21.9]	63 [11.4]	
Students used something to create something else (e.g., art supplies, items in nature, objects repurposed).	2.2 [-1.2]	6.1 [4.0]	7.1 [-0.8]	8.7 [4.2]	
Students used a resource to solve a problem (e.g., manipulatives, math charts, dictionary).	30.4 [2.3]	42.9* [24.6]	42.9 [11.0]	32.6 [3.4]	
Instruction used an item designed for interaction between teacher and students or between students (e.g., read aloud, pocket chart, counters, puzzle).	58.7 [12.6]	57.1 [10.9]	75 [16.6]	54.3 [6]	
Students used physical space for learning activities (e.g., counting large steps for measurement, sit in a circle for sharing).	21.7 [8.3]	30.6 [10.2]	28.6 [-5.1]	28.3 [0.2]	
Students used stations/corner work/learning centers.	8.7 [3.1]	20.4 [14.0]	23.2 [3.7]	13 [-6.1]	
5. Efficiency: Instructional time is maximized					
The teacher had materials ready at the start of an activity or the class.	87.0 [8.3]	85.7 [-4.6]	85.7 [3.4]	89.1 [0.4]	
The pace of the instruction was lively.	89.1 [1.5]	89.8 [10.2]	87.5 [-4.5]	87 [4.9]	
6. Student-centered					

Table 10. Percentage of Grade 1 Lessons in Which Target Instructional Practices Were Observed (with Difference from Baseline), Endline Only

Observed (with Difference from Baseline), Ending Only					
	High prevalence (75%–100%)	Moderately prevalent (50%–74%)	Low prevalence (25%–49%)	Very low prevalence (0%–24%)	BOLD Indicates statistically significant differences
	Comparison		Indirect	Direct	GALOP
Students developed and shared ideas with either classmates or the teacher.	34.8 [-10.2]	46.9 [1.8]	46.4 [-13.7]	41.3 [4.2]	
Students had a chance to select from choices (state of play: exploring).	0 [-19.1]	10.2 [-5.9]	7.1 [-13.2]	8.7* [0.8]	
Students chose collaborators or roles.	0 [-2.2]	8.2 [1.7]	3.6 [-12.4]	10.9 [0.8]	
Students made or changed rules for an activity.	0 [0]	2.0 [-0.1]	1.8 [-3.5]	2.2 [2.2]	
Students chose how long to do an activity (state of play: owning).	0 [0]	4.1 [-2.4]	1.8 [-6.2]	2.2* [-10.2]	
Students set goals or helped to plan how to use time.	0 [-1.1]	2.0 [2.0]	1.8 [-1.8]	0 [-1.1]	
Students chose when to move around.	2.2 [-3.4]	6.1 [-3.6]	3.6 [-6.2]	6.5 [-2.5]	

**p > 0.05*

Learning Environment (Student Survey)

Students were asked 10 questions about the learning environment (see **Table 11**) with each response allowing a score of 0,1, or 2. Although endline values likely hit a “ceiling effect,” the percentage of students in indirect implementation and GALOP schools who reported that children ask questions in class increased significantly from baseline.

Overall reports of verbal and physical punishment were high. **Interestingly, the percentage of students who reported an adult yelling at a child or using physical punishment increased in all treatment and comparison groups except in direct implementation schools, where reports of teachers yelling stayed the same, while reports of physical punishment decreased (both impacts were significant when compared to comparison).**



Table 11. Percent of Students Reporting Each Learning Environment Experience, Grade 1 Baseline vs. Grade 4 Endline

	Learning Environment Question	Treatment Group	Baseline	Endline	Difference	Impact
1	Does your teacher praise you when you give the right answer?	Comparison	96.8	99.1	2.3	—
		Indirect	96.3	99.0	2.7	0.5
		Direct	97.3	99.5	2.2	-0.1
		GALOP	98.1	99.4	1.2	-1.0
2	Can you talk to at least one teacher in the school about things that are bothering you?	Comparison	77.0	82.0	5.1	—
		Indirect	66.4	78.2	11.8	6.7
		Direct	70.9	78.3	7.5	2.4
		GALOP	69.3	81.0	11.7	6.7
3	Do children in your class ask the teacher questions in lessons?	Comparison	76.7	95.1	18.4	—
		Indirect	65.2	94.8	29.5	11.2*
		Direct	73.5	93.7	20.2	1.8
		GALOP	64.3	96.9	32.6	14.2*
4	Last week, did some children in your class ask the teachers for help with their work?	Comparison	69.8	80.9	11.2	—
		Indirect	70.1	89.0	18.9	7.7
		Direct	71.7	78.7	7.0	-4.2
		GALOP	62.2	80.4	18.2	7.0
5	Do your teachers encourage you to try out new things?	Comparison	88.0	92.7	4.6	—
		Indirect	81.6	85.7	4.2	-0.5
		Direct	88.7	94.9	6.2	1.5
		GALOP	85.5	91.8	6.3	1.6
6	Does your teacher get angry or yell when you get the answer wrong?	Comparison	60.9	69.0	8.0	—
		Indirect	51.7	64.9	13.2	5.2
		Direct	46.8	58.0	11.3	3.3
		GALOP	54.8	63.7	8.9	0.9
7	Do your teachers help those who do not understand?	Comparison	94.1	99.3	5.2	—
		Indirect	93.2	96.5	3.3	-1.9
		Direct	93.3	99.5	6.2	1.0
		GALOP	91.9	99.8	7.9	2.7
8	Did an adult in your school scream or yell at a child?	Comparison	62.1	83.4	21.3	—
		Indirect	61.6	75.5	13.9	-7.4
		Direct	63.6	63.6	0.1	-21.2*
		GALOP	60.4	75.9	15.5	-5.8
9	Is it easy for students to share their thoughts during class/in class?	Comparison	74.9	78.5	3.5	—
		Indirect	70.7	80.9	10.2	6.7
		Direct	68.9	75.4	6.5	2.9
		GALOP	67.6	74.6	7.0	3.5
10	Did an adult or older student in your school hit a child or use physical punishment?	Comparison	70.4	78.9	8.5	—
		Indirect	66.6	76.3	9.7	1.1
		Direct	69.6	63.3	-6.2	-14.8*
		GALOP	68.7	73.8	5.1	-3.4

An overall learning environment score was calculated for midline and endline, with a range from 0 (least supportive) to 10 (most supportive) (Table 12). At endline, **the percentage of children who reported that children in their class ask the teacher questions increased significantly in indirect implementation and GALOP schools.** The percentage of children who reported that an adult screamed, yelled, hit, or used physical punishment decreased significantly in direct schools when compared to comparison schools. **Overall, the learning environment score improved significantly in direct and GALOP schools.**

Table 12. Overall Learning Environment Score, Baseline vs. Endline

	Baseline	Endline	Difference	Impact
Comparison	58.8	59.7	0.9	—
Indirect	57.6	60.1	2.5	1.7
Direct	58.2	62.8	4.5	3.6*
GALOP	56.6	61.5	4.9	4.1*

*p > 0.05

Stakeholder Perceptions of LtP

The research team interviewed teachers, HTs, caregivers, SISOs, DTOs, and Curriculum Leads about their perceptions and attitudes toward LtP. To do this, the stakeholders were presented with two statements and asked which they agreed with more. The statements were similar; however, one statement was more positive toward play-based learning, or “pro-play.”

Attitudes Toward Learning through Play—Teachers

Table 13 shows the percentage of teachers who agreed with the more “pro-play” of the two statements at endline. The number in brackets shows the change in agreement levels from baseline. Green boxes indicate an increase in the percentage of teachers selecting the more pro-play statement, and pink boxes signify a decrease. **At endline, the percentage of GALOP teachers who felt parents welcome play-based approaches decreased significantly when compared to teachers in comparison schools. Across all groups, there was relatively lower agreement with explicitly student-centered statements (e.g., “It is important for children to be free to talk,” “Children learn best from their peers,” and “children can choose their own learning activities in the classroom”).**

Table 13. Percent of Grade 3 Teachers Who Agreed with the More “Playful” Statement at Endline [Change from Grade 1 Baseline]

		Increase		Decrease	
#	Items	Comp.	Indirect	Direct	GALOP
1	Parents welcome more play-based approaches to learning in the classroom. (vs. Parents prefer that I stick to traditional teaching approaches in the classroom.)	78 [22]	77 [8]	91 [16]	66* [-11]
2	Play is an important part of learning for children in primary school. (vs. Play is only important for learning for preschool-aged children.)	83 [4]	79 [-9]	88 [-1]	79 [-7]
3	It is more important that the students are engaged. (vs. It is more important that the teacher is engaged.)	85 [20]	87 [15]	84 [19]	81 [11]
4	Children learn best by exploring new things on their own. (vs. Children learn best when new things are explained by the teacher.)	74 [13]	72 [-7]	77 [5]	79 [11]

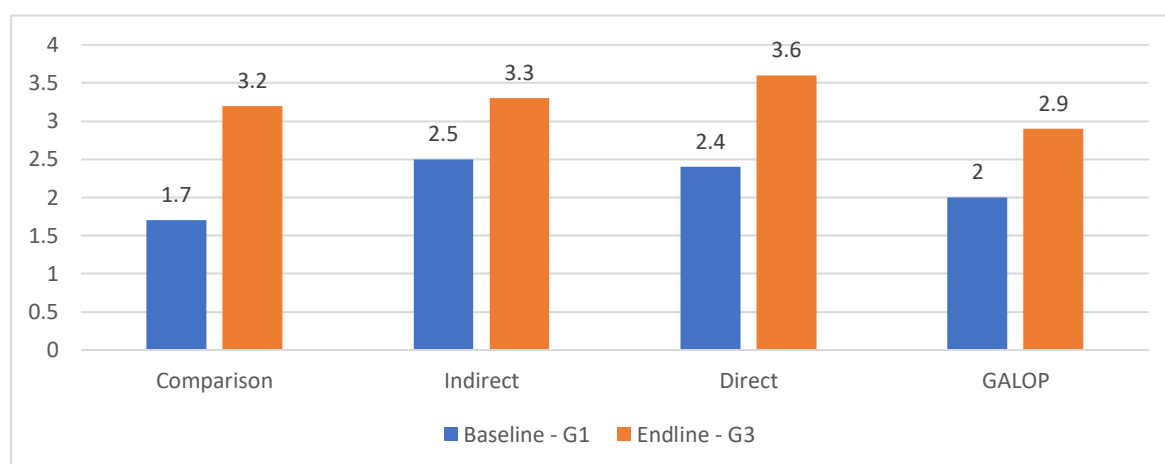
Table 13. Percent of Grade 3 Teachers Who Agreed with the More “Playful” Statement at Endline [Change from Grade 1 Baseline]

#	Items	Increase		Decrease	
		Comp.	Indirect	Direct	GALOP
5	Children learn literacy/language best in small groups. <i>(vs. Children learn literacy/language best as a whole class.)</i>	67 [17]	72 [10]	75 [-1]	72 [9]
6	If children answer their own questions, then they will be more likely to understand. <i>(vs. If the teacher answers children's questions they will be more likely to get the correct answer.)</i>	60 [10]	66 [-5]	60 [8]	57 [-7]
7	It is important for children to be free to talk. <i>(vs. It is important for the class to be orderly.)</i>	72 [26]	68 [10]	59 [4]	66 [29]
8	Children learn best from their peers. <i>(vs. Children learn best from the teacher.)</i>	67 [28]	68 [17]	51 [13]	66 [24]
9	Children can choose their own learning activities in the classroom. <i>(vs. Teachers are responsible for choosing the learning activities for his/her students)</i>	20 [1]	18 [5]	37 [19]	27 [7]
10	If children do not understand, then they should try to find the answer independently before asking the teacher. <i>(vs. It is better if children tell the teacher when they don't understand.)</i>	18 [-11]	18 [3]	19 [-7]	13 [0]
Number of Interviews Conducted		45	46	55	46

* $p > 0.05$

An overall playfulness attitude-scale score was further calculated for teachers and other school-level stakeholders. This score derives from the number of items on which the respondent selected the more “pro-play” statement, with possible scores ranging from –10 (least supportive of LtP) to 10 (most supportive of LtP) (**Figure 7**). Although no significant changes between groups was detected, the average attitude score among teachers increased slightly at endline in all groups. No scores averaged below 0.

Figure 7. Mean Scores of Teachers on the “Attitudes toward LtP” Scale, Baseline vs. Endline

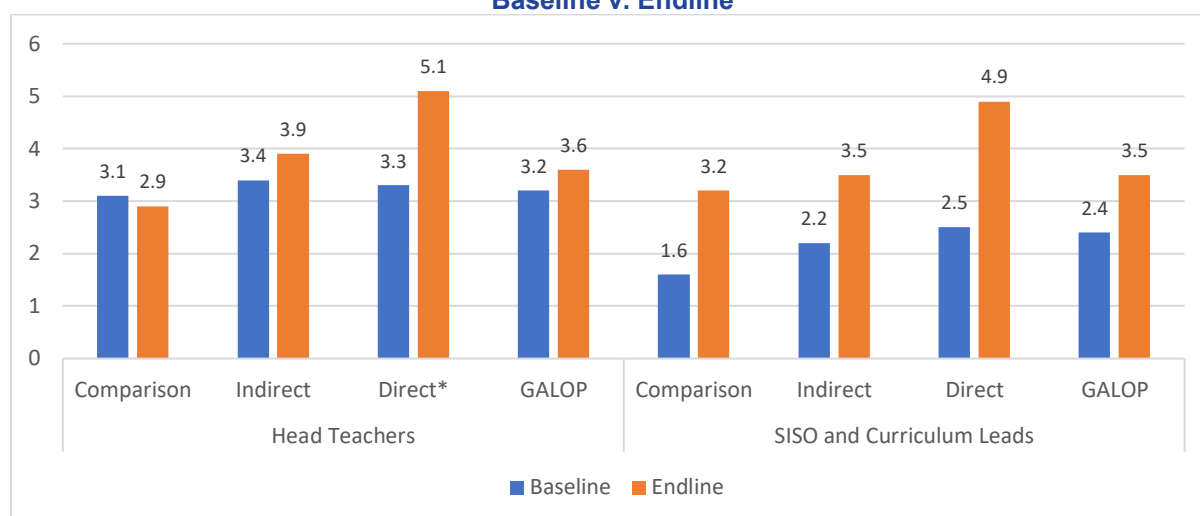


Attitudes Toward LtP: HTs, SISOs, and Curriculum Leads

The same set of items was administered with HTs, SISOs, and Curriculum Leads. **HTs' attitude scores in direct implementation schools became significantly more “pro-play” since baseline compared to comparison head teachers (Figure 8).**



Figure 8. Mean Scores of School-Level Stakeholders on the “Attitudes toward LtP” Scale: Baseline v. Endline



* $p > 0.05$

Attitudes Toward Learning through Play–Caregivers

A second, simplified survey of attitudes was given to caregivers, who were also asked to select their preference between two statements. The survey results are presented in **Table 14**. While agreement with the more “pro-play” statement increased across all items among caregivers in direct schools, no significant impacts were detected at endline.

Table 14. Percent of Caregivers Who Agreed with More “Playful” Statement at Endline [Change from Baseline]

		Increase		Decrease	
#	Items	Comp.	Indirect	Direct	GALOP
1	Teachers should use play to promote learning in the classroom.	70 [-1]	69 [-8]	82 [7]	74 [-6]
2	I would be happy if teachers use play to support learning in the classroom.	71 [-2]	70 [-2]	77 [7]	63 [-10]
3	It is more important to teach creativity than academic subjects.	68 [4]	60 [-2]	69 [7]	66 [2]
4	Children learn best when they are having fun.	49 [1]	52 [2]	64 [11]	58 [3]
5	Teachers should create a friendly classroom environment.	48 [-4]	60 [1]	59 [9]	61 [-2]
6	It is important for children to be free to talk.	42 [23]	55 [24]	53 [23]	55 [26]
Number of Interviews Conducted		92	93	111	91

Teacher Self-Efficacy and Motivation

During interviews, several statements about feelings of motivation and efficacy in implementing LtP were posed to teachers, asking if they strongly disagreed, disagreed, agreed, or strongly agreed with each statement. No significant changes were detected on these items, although the percentage of teachers in direct schools who strongly agreed that they felt valued by school management did increase by 20 percentage points from baseline (**Figure 9**). Further, as **Figure 10** shows, virtually all teachers felt they understood expectations for their students’ learning this year, and most (88%) were optimistic that their students would reach these expectations. Interestingly, while 98% of teachers agreed or strongly agreed that using LtP made their job easier, 33% also felt their workload increased

as a result of the program and 47% said that if given the chance, they would leave teaching to pursue another career.

Additional figures displaying teachers' responses to questions about their motivation and feelings of self-efficacy can be found in **Annex A**.

Figure 9. Teachers' Responses to "I Feel Valued by School Management," Baseline vs. Endline

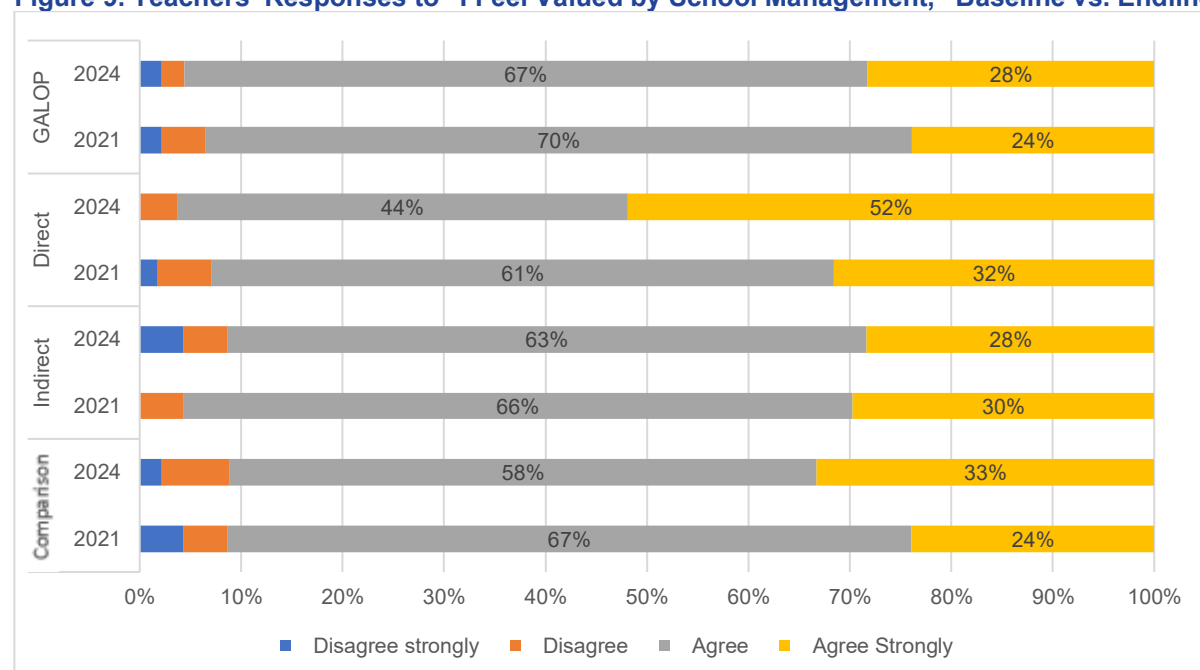
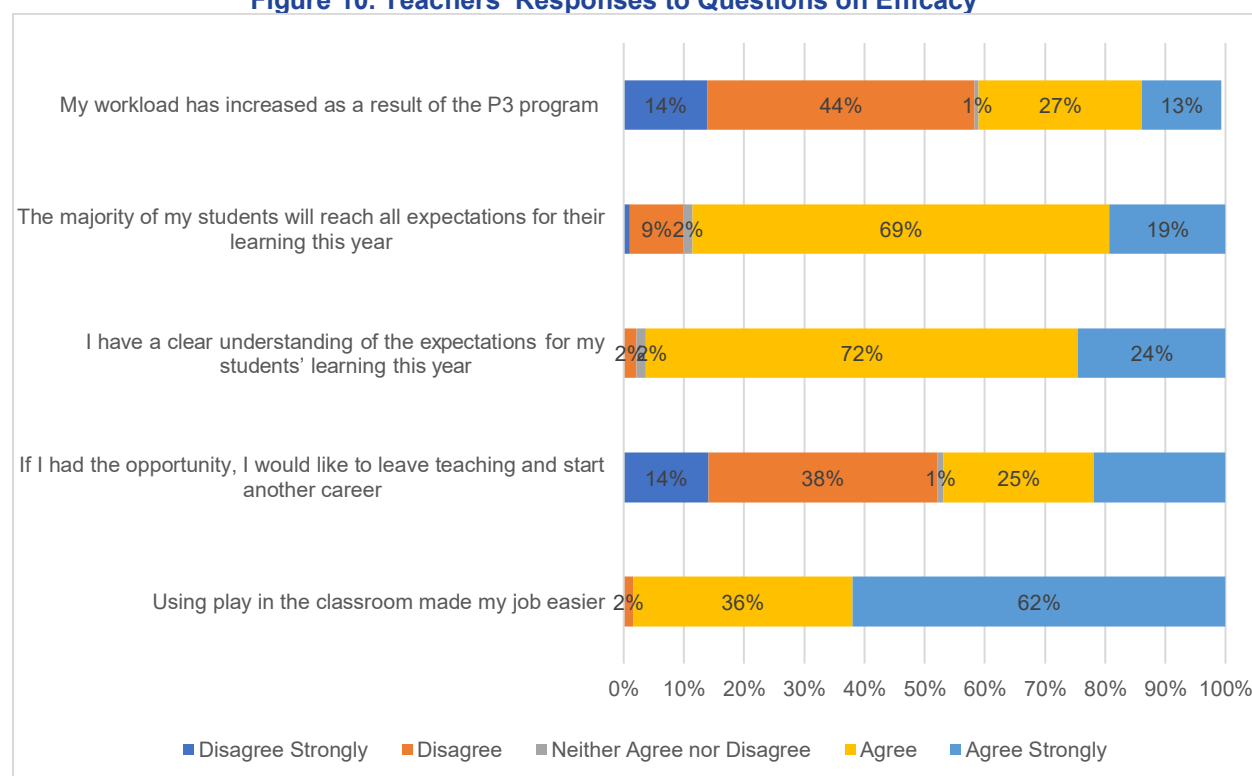


Figure 10. Teachers' Responses to Questions on Efficacy



3.3 Students' Literacy, Numeracy, and SEL Outcomes

Assessment of SEL

In the tables presented below, the difference column compares the change in mean score from baseline to endline for each group. To test significance, we compared (1) the change in mean score from baseline to endline for the direct, indirect, and GALOP treatment groups to (2) the change in mean score from baseline to endline for the comparison group (difference in differences). This calculation adjusts for differences at baseline and for typical changes that take place as students age. An asterisk indicates a significant difference in differences between the comparison group to the three treatment groups, with 95% confidence ($p > 0.05$).

Due to refinements to the SEL tool following baseline data collection, some comparisons are made between midline and endline. This applies specifically to measures of negative attribution. All other comparisons are between baseline and endline. Student assessments were conducted in Asante-Twi, Dagbani, Ewe, and Ga. Data are aggregated across all student assessment languages.

Observed Student Behavior during Assessments

During the SEL assessment, three student behaviors initially observed at baseline were again observed by the assessor at endline (see **Table 15**), providing a measure of students' displayed confidence and positive emotions. Observation scores are distributed on a scale of 0–100, with 100 representing the most confident behaviors. At endline, **the percentage of students in GALOP schools who looked assessors in the eye when listening to or responding to questions decreased significantly when compared to the comparison schools.** No other impacts were detected.

Table 15. Observed Student Behaviors, Baseline vs. Endline

Behavior (Variable)	Treatment Group	Baseline Mean	Endline Mean	Difference	Impact
Does the child greet you confidently?	Comparison	51.4	34.3	-17.1	—
	Indirect	55.2	39.7	-15.6	1.5
	Direct	52.1	37.1	-15.0	2.1
	GALOP	56.4	37.7	-18.7	-1.6
Does the child respond to questions confidently?	Comparison	63.7	61.6	-2.0	—
	Indirect	63.4	59.3	-4.1	-2.1
	Direct	64.6	60.2	-4.4	-2.4
	GALOP	65.5	58.6	-6.9	-4.9
Does the child look you in the eye when listening to or responding to questions?	Comparison	67.8	71.6	3.8	—
	Indirect	62.9	69.6	6.7	2.9
	Direct	67.1	70.1	2.9	-0.8
	GALOP	67.9	64.8	-3.1	-6.9*
Total observed behavior score	Comparison	60.9	55.8	-5.1	—
	Indirect	60.5	56.2	-4.3	0.8
	Direct	61.3	55.8	-5.5	-0.4
	GALOP	63.3	53.7	-9.6	-4.5

* $p < 0.05$

Relationships

Students were presented a series of scenarios, and asked if they have a friend that they could turn to in each one (**Table 16**). At endline, the number of students in both treatment

and comparison schools who could name one or more friends increased from baseline. **The percentage of GALOP students who could name at least one friend whom they could tell secrets to increased significantly more from baseline**, when compared to the comparison group.



Table 16. Student Responses to Relationship Questions, Baseline vs. Endline					
Question (Variable)	Treatment Group	Baseline	Endline	Difference	Impact
Do you talk to any of your friends when you feel sad? If yes, whom do you talk to?	Comparison	73.5	87.6	14.1	—
	Indirect	75.7	89.4	13.7	-0.5
	Direct	76.7	87.6	10.9	-3.2
	GALOP	69.2	87.6	18.4	4.3
When you have an argument/fight with another child, do you go to any of your friends?	Comparison	70.1	81.6	11.5	—
	Indirect	76.0	86.3	10.3	-1.2
	Direct	71.1	82.0	10.9	-0.6
	GALOP	64.3	82.9	18.6	7.0
If you are working on something difficult, do you ask any of your friends for help?	Comparison	78.5	92.0	13.5	—
	Indirect	78.8	92.3	13.4	-0.1
	Direct	78.2	89.6	11.4	-2.1
	GALOP	75.1	92.6	17.5	4.0
Do you have a friend who helps you learn? If yes, who helps you learn?	Comparison	81.3	94.2	13.0	—
	Indirect	78.3	92.3	14.0	1.1
	Direct	82.8	93.1	10.3	-2.7
	GALOP	76.6	93.6	17.1	4.1
Do you have a friend whom you tell your secrets to? If yes, who?	Comparison	67.4	79.8	12.4	—
	Indirect	70.4	80.1	9.7	-2.7
	Direct	61.4	74.9	13.4	1.0
	GALOP	54.7	79.4	24.7	12.3*

*p < 0.05

An overall student relationship score was also calculated, and **increased significantly from baseline in the GALOP group, when compared to the comparison group (Table 17)**.



Table 17. Student Relationship Score				
Treatment Group	Score based on number of friends named in different proposed situations			
	Baseline	Endline	Difference	Impact
Comparison	73.9	87.1	13.2	—
Indirect	75.6	88.1	12.5	-0.7
Direct	73.5	85.4	11.9	-1.3
GALOP	67.9	87.2	19.3	6.1*

*p < 0.05

Conflict Resolution

Children listened to five scenarios in which there was potential conflict with another student and were asked what they would do in that situation. For each scenario their response was placed into one of eight categories in four levels. At endline, the most common response

from Grade 4 students across all scenarios was to command the other child or to appeal to authorities.

Table 18 shows the average percentage of times across all five scenarios children chose each type of response to conflict at baseline and endline. Across groups, we see a consistent pattern: as students get older, they are less likely to respond to the scenarios presented with verbal or physical aggression or to disengage. Instead, they are more likely to try to reason with the other child, command the other child, or appeal to authorities.

At endline, the percentage of students in comparison schools who chose physical or verbal aggression decreased by significantly more than the indirect implementation group and the percentage of indirect implementation students who chose to do nothing or disengage decreased by significantly more than the comparison group. It is worth noting, however, that overall means at endline were largely the same between groups.

Table 18. Percentage of Students Choosing Each Response to Conflict Resolution, Baseline vs. Endline

Behavior (Variable)	Treatment Group	Baseline	Endline	Difference	Impact
Percentage of times students chose verbal or physical aggression	Comparison	24.7	14.8	-10.0	—
	Indirect	24.4	19.1	-5.3	4.7*
	Direct	23.5	16.7	-6.9	3.1
	GALOP	22.4	15.3	-7.1	2.8
Percentage of times students chose to do nothing, disengage, or wait	Comparison	28.5	28.7	0.2	—
	Indirect	31.7	27.1	-4.6	-4.8*
	Direct	29.6	27.1	-2.5	-2.7
	GALOP	31.3	28.6	-2.6	-2.8
Percentage of times students chose to “command” the other child or to appeal to authorities (tell the teacher)	Comparison	31.6	45.9	14.3	—
	Indirect	31.3	43.3	12.0	-2.2
	Direct	31.6	45.0	13.4	-0.9
	GALOP	35.1	45.4	10.3	-4.0
Percentage of times students chose to use reasoning or try to help the other child	Comparison	11.2	10.5	-0.7	—
	Indirect	10.1	10.2	0.2	0.9
	Direct	11.6	10.9	-0.6	0.1
	GALOP	9.3	10.5	1.2	1.9

*p < 0.05

For each scenario, students were asked why the other child in the scenario behaved the way they did. The explanation was coded according to whether the motive of the other student was seen as intending negative outcomes toward the student being interviewed (e.g., “she doesn’t like me”), neutral (e.g., “she was distracted”), or positive (e.g., “she was trying to help me”). The purpose of this subtask was to measure the frequency with which students’ exhibited negative attribution (perceiving intentional harm from another even when other explanations are viable). **Table 19** shows the mean percent of times students demonstrated negative attribution across the five scenarios. Although negative attribution decreased from midline to endline in all groups, no significant differences were detected.

Table 19. Mean Negative Attribution Scores, Midline vs. Endline

Treatment Group	Midline	Endline	Difference	Impact
Comparison	24.6	14.3	-10.3	—
Indirect	24.7	15.6	-9.2	1.1
Direct	23.7	14.3	-9.4	0.8
GALOP	25.6	16.0	-9.7	0.6

Confidence

Students were read descriptions of two children—a child who shows confidence and a child who does not—and asked which one they identified with. **Table 20** shows the percentage of children identifying with the confident child across the seven questions. At endline, confidence scores increased across all groups. No significant differences were detected between groups.

Table 20. Student Confidence Scores, Baseline vs. Endline

Treatment Group	Percentage of Times Children Identify with the Confident Option (Out of 7)			
	Baseline Mean	Endline Mean	Difference	Impact
Comparison	55.9	58.2	2.3	—
Indirect	54.1	59.9	5.8	3.5
Direct	54.9	57.1	2.2	-0.1
GALOP	54.4	58.1	3.7	1.5

Confidence scores were higher at endline (when children were older) for some scenarios, such as being trusted to look after younger children and leading other children to sing a song, while confidence scores decreased in other scenarios, such as starting conversations with adults and students' sharing their thoughts even if they know that they are different from others (**Table 21**).

Table 21. Percentage of Students Identifying with Confident Person, Baseline vs. Endline

Question	Average Percentage Identifying with Confident Person	
	Grade 1 Baseline	Grade 4 Endline
X often feels confident (not afraid) when going into a room full of people.	53.7%	50.7%
When children in the class have a question they get X to ask the teacher.	56.1%	66.9%
X can lead other children to sing a song.	54.2%	65.3%
X starts conversations with adults they don't know well.	41.1%	33.4%
X likes to read stories out to the whole class.	57.4%	62.4%
X is trusted to look after younger children.	72.1%	88.0%
X shares his/her thoughts even if they are very different from others'.	54.9%	42.4%

Creativity

Children were asked to imagine different uses for two everyday objects, a spoon and a pencil. To enable a more effective comparison, creativity scores were scaled to a range of 0–100, with 0 representing no evidence of creativity (zero suggestions for alternative uses for everyday objects) and 100 representing the highest level of creativity on the distribution (10 or more uses). **Table 22** shows students' average creativity scores (the number of different uses they named) by treatment group. Although creativity scores decreased from in all groups, differences were not statistically significant.

Table 22. Student Creativity Scores, Baselines vs. Endline				
Treatment Group	Different Uses Students Suggest for Two Everyday Items (Combined)			
	Baseline Mean	Endline Mean	Difference	Impact
Comparison	37.1	32.4	-4.7	0
Indirect	35.2	28.6	-6.6	-1.9
Direct	36.5	29.3	-7.2	-2.5
GALOP	38.3	31.9	-6.4	-1.7

* $p < 0.05$

Student Literacy and Numeracy Outcomes

The program administered shortened versions of the EGRA and EGMA in Ghana. The EGRA consisted of letter sound questions, invented word questions, an oral reading passage, and reading comprehension questions in the four national languages: Ewe, Ga, Dagbani, and Asante-Twi. All six subtasks on the EGMA (i.e., number discrimination, missing number, basic addition, basic subtraction, addition level 2, and subtraction level 2) were also administered in these languages.

For each subtask, two measures are displayed in the following tables: (1) either fluency or the percentage correct overall and (2) the percentage of students with zero scores. For each measure, there are four estimates: baseline mean scores (2021 baseline), endline mean scores (2023 endline), the change in scores from baseline to endline (mean gain or loss), and the impact estimate for each treatment group (impact). The impact estimate was calculated using a regression model that controls for time (baseline to endline) and the treatment group. The impact measurements for the indirect and GALOP treatment groups are each relative to the direct implementation group, which is why there are impact measures listed for each of the treatment groups, except the direct group, as the point for comparison. Due to potential structural advantages in the comparison schools (as a result of the limited selection pool), the comparison group is not included in this analysis.

The EGRA results are presented in **Table 23**. As expected, moderate to large gains were seen for all subtasks across all groups. **Mean reading scores in the direct group increased significantly more than in the indirect group on letter sound reading, invented word reading, and the oral reading passage. The percentage of students scoring zero decreased significantly more in the direct group than in both the indirect and GALOP groups on the letter sounds and oral reading subtasks.**

Table 23. EGRA Results by Subtask and Treatment Group, Endline					
Reading Subtask	Measure	Estimate	Indirect	Direct	GALOP
			$n = 515$	$n = 605$	$n = 485$
Letter Sounds	Fluency (correct)	Baseline Mean	4.7	6.2	5.6
		Endline Mean	20.6	24.3	22.1



Table 23. EGRA Results by Subtask and Treatment Group, Endline					
Reading Subtask	Measure	Estimate	Indirect	Direct	GALOP
			<i>n</i> = 515	<i>n</i> = 605	<i>n</i> = 485
	letters per minute)	Mean Gain/Loss	15.9	18.1	16.6
		Impact	-3.0*	-	-1.9
	Percentage of students with zero scores	Baseline Mean	62.7	53.2	58.6
		Endline Mean	18.3	12.7	17.9
		Mean Gain/Loss	-44.4	-40.5	-40.6
		Impact	4.2*	-	4.3*
	Invented Words Fluency (correct words per minute)	Baseline Mean	0.4	0.3	0.2
		Endline Mean	5.2	6.7	6.7
		Mean Gain/Loss	4.9	6.4	6.5
		Impact	-1.6*	-	0.1
	Percentage of students with zero scores	Baseline Mean	96.5	96.0	97.3
		Endline Mean	67.1	59.2	62.9
		Mean Gain/Loss	-29.4	-36.9	-34.4
		Impact	7.7*	-	3.2
Oral Reading Passage	Reading fluency (correct words per minute)	Baseline Mean	0.3	0.2	0.1
		Endline Mean	6.5	8.4	8.2
		Mean Gain/Loss	6.2	8.2	8.0
		Impact	-2.0*	-	-0.2
	Percentage of students with zero scores	Baseline Mean	96.5	95.9	96.1
		Endline Mean	68.5	56.0	62.3
		Mean Gain/Loss	-28.0	-39.9	-33.8
		Impact	12.3*	-	6.2*
	Percentage correct	Baseline Mean	0.2	0.0	0.1
		Endline Mean	8.9	10.9	10.5
		Mean Gain/Loss	8.7	10.9	10.4
		Impact	-2.2	-	-0.6
Reading Comprehension	Percentage of students with zero scores	Baseline Mean	99.2	100.0	99.6
		Endline Mean	83.8	80.6	81.4
		Mean Gain/Loss	-15.4	-19.4	-18.1
		Impact	3.4	-	1.2

*p < 0.05

EGMA results are presented in **Table 24**. Across all subtasks, average math scores improved from baseline to endline in all groups. Mean scores for the direct group increased by significantly more than in the indirect group on two lower order subtasks: quantity discrimination and missing number. Students' performance on the basic subtraction subtask was significantly higher in the GALOP group when compared to the direct group.

Table 24. EGMA Results by Subtask, Endline					
Math Subtask	Measure	Estimate	Indirect	Direct	GALOP
			<i>n</i> = 515	<i>n</i> = 605	<i>n</i> = 485
Quantity Discrimination	Percentage correct	Baseline Mean	37.8	41.3	39.4
		Endline Mean	67.5	73.1	73.5
		Mean Gain/Loss	29.7	31.9	34.1

Table 24. EGMA Results by Subtask, Endline

Math Subtask	Measure	Estimate	Indirect	Direct	GALOP
			<i>n</i> = 515	<i>n</i> = 605	<i>n</i> = 485
Missing Number	Percentage of students with zero scores	Impact	-4.8*	-	0.7
		Baseline Mean	17	11.7	16.3
		Endline Mean	4.1	3.5	3.1
		Mean Gain/Loss	-12.9	-8.3	-13.2
	Percentage correct	Impact	0.4	-	-0.5
		Baseline Mean	13.2	14.4	13.1
		Endline Mean	35.6	39.9	41.3
		Mean Gain/Loss	22.4	25.4	28.2
	Percentage of students with zero scores	Impact	-3.4*	-	2.3
		Baseline Mean	29.2	23.6	31.1
		Endline Mean	4.4	3.6	4.9
		Mean Gain/Loss	-24.7	-20	-26.2
Basic Addition	Fluency (correct addition problems per minute)	Impact	0.6	-	0.9
		Baseline Mean	1.7	2.3	2.4
		Endline Mean	8.7	9.1	9.1
		Mean Gain/Loss	6.9	6.9	6.8
	Percentage of students with zero scores	Impact	-0.1	-	-0.1
		Baseline Mean	64.5	56.5	56.5
		Endline Mean	8.7	9.1	7.4
		Mean Gain/Loss	-55.8	-47.4	-49.1
	Fluency (correct subtraction problems per minute)	Impact	-1.1	-	-1.7
		Baseline Mean	0.8	1.1	1.0
		Endline Mean	5.3	5.7	6.4
		Mean Gain/Loss	4.5	4.6	5.4
Basic Subtraction	Percentage of students with zero scores	Impact	-0.4	-	0.7*
		Baseline Mean	80.9	75.5	81.2
		Endline Mean	29	25.5	20.2
		Mean Gain/Loss	-51.9	-50.1	-61
	Percentage correct	Impact	3.1	-	-5.6*
		Baseline Mean	3.4	5.5	4.4
		Endline Mean	31.5	34.7	35.5
		Mean Gain/Loss	28.1	29.2	31.1
	Percentage of students with zero scores	Impact	-2.0	-	1.4
		Baseline Mean	91.1	87.3	88.5
		Endline Mean	39.8	35.0	32.8
		Mean Gain/Loss	-51.4	-52.2	-55.7
Addition Level 2	Percentage correct	Impact	3.7	-	-2.5
		Baseline Mean	0.6	1.1	1.6
		Endline Mean	16.6	18.3	18.4
		Mean Gain/Loss	16	17.2	16.9
	Percentage of students with zero scores	Impact	-1.3	-	-0.1
		Baseline Mean	98.1	97.0	95.5
		Endline Mean	16.6	18.3	18.4
		Mean Gain/Loss	16	17.2	16.9
	Percentage correct	Impact	-1.3	-	-0.1
		Baseline Mean	98.1	97.0	95.5
		Endline Mean	16.6	18.3	18.4
		Mean Gain/Loss	16	17.2	16.9
Subtraction Level 2	Percentage correct	Impact	-1.3	-	-0.1
		Baseline Mean	98.1	97.0	95.5
		Endline Mean	16.6	18.3	18.4
		Mean Gain/Loss	16	17.2	16.9

Table 24. EGMA Results by Subtask, Endline

Math Subtask	Measure	Estimate	Indirect	Direct	GALOP
			<i>n</i> = 515	<i>n</i> = 605	<i>n</i> = 485
	Percentage of students with zero scores	Endline Mean	65.8	62.8	60.2
		Mean Gain/Loss	-32.2	-34.2	-35.3
		Impact	2.6	-	-2.2

**p* < 0.05

3.4 LtP-Enabling Environment

This section presents findings from quantitative interviews conducted with teachers, HTs, SISOs, Curriculum Leads, parents, and caregivers. Tabulations at endline are compared to those at baseline, and differences over time are compared between the indirect, direct, and GALOP schools and comparison schools.

This section also includes more detailed findings from qualitative interviews conducted with program staff, central government and district-level education officials, HTs, teachers, and parents and caregivers.

Program Strengths and Challenges

P3's design is outlined in Section 1.3, above. In addition to the key components of training, mentoring, monitoring, support, and PTA engagement undertaken through and alongside the DTST (discussed in Section 1.3), a key component of the RtP methodology emphasized in interviews with program staff was the Reflect-Connect-Apply approach. Through this approach, the teacher first introduces a playful activity related to something that the student already knows (reflect). Then, the teacher helps the student make connections between the playful activity and new content being taught that day (connect). Finally, the student practices using the knowledge built during the connect phase in other learning activities (apply). For example, P3 staff explained that if a teacher is teaching the letter sound "p," they will begin with a song that uses and reiterates the /p/ sound to expose children to this sound in a playful way. Then, the teacher will tell the students explicitly they are going to learn about the sounds of the letter p. Finally, the teacher will help the student identify the /p/ sound in new words.

Staff reported a number of challenges to implementing P3, including the following:

- Teachers lack the ability to create learning materials and resources.
- Teachers had poor understanding of the instructional content (i.e., mathematical understandings, etc.). Therefore, some of the teachers struggled to understand the content behind what they teach.
- Availability and punctuality of teachers during training.
- Lack of supervision by the HTs to support teachers' development of lesson notes/plans.
- Teacher transfers across districts leading to high attrition (this was also noted in the study's longitudinal teacher sample).
- Commitment and attitudes of teachers that relied on traditional learning models.
- Large student population and lack of space in classrooms to organize activities.
- Lack of infrastructure, including limited numbers of desks for students in the classrooms.

- Lack of planning. Some teachers do not plan ahead to infuse the play-based methodology into their teaching. Teachers copy lesson notes or plans without thinking critically about how the RtP methodology can be implemented to enhance these lesson notes and plans.
- Timing. Teachers do not have enough time to use play-based methodologies to teach the content in the classroom.
- Leadership and accountability issues.

P3 staff also noted a number of strategies they have undertaken to mitigate these challenges, as follows:

- For teachers who were not planning ahead, RtP developed additional training that included developing work schemes and lesson plans.
- RtP specifically supported teachers on how to structure group work and form junior/peer leaders to lead group sessions.
- RtP encouraged “co-creation” of learning materials to support development of more materials.
- Teacher champions were introduced. They acted as ambassadors for the schools and would help to introduce playful activities.
- To support fundraising, in some schools, PTAs could generate provisional funds to buy furniture.

Perceptions of LtP

In qualitative interviews, all school-level stakeholders (HTs, teachers, and caregivers) seemed to understand LtP as a child-centered play-based pedagogy that positively impacts learning outcomes. Caregivers made sense of this approach by observing their children’s progress in retaining the lessons or by observing students completing homework. For instance, a caregiver from Greater Accra South explained,

“Learning through play helps children to easily understand what a teacher is teaching in class and learners always have [a] memory of lessons that were delivered through [a] play approach.”

Teachers and HTs understood and interpreted the LtP approach by focusing on the usage of activities, materials, and games to engage children in a joyful way that develops their academic and social faculties. There was a strong and distinct relationship between conceptions of LtP and infusing manipulatives across instruction. A teacher from Weija Gbawe explained LtP as follows:

“Learning through play, I will say, is when learners are actively involved, there is collaboration, and they get to manipulate objects. Thinking critically and coming out with ideas builds their confidence.”

Teachers also drew a connection to the learning environment, as a teacher from Greater Accra South explained,



“I used activities like [a] positive environment [or an] environment must be ‘print rich.’ That is designing classroom walls with colorful learning materials that attract learners. Environments must be safe for teaching and learning.”

Teachers said they were confident that LtP-based pedagogical approaches are an opportunity for students to be active, engage, and enjoy the learning process. A teacher from Weija Gbawe explains students’ engagement as follows:

“The learners get to relate the activity to what they’ve learned... It helps the learners to agree and disagree in groups and to get the learning outcome. It got the learners engaging with each other and getting involved. They also had fun and connected with the activities.”

Teachers expected such engagement to enhance students’ learning outcomes. A teacher from Greater Accra South shared,

“Because the activities are based on the learner’s strength and they do it themselves, they pick [up] the concept very fast.”

Conversely, prior to LTP’s introduction as a pedagogical method, most of the teachers recalled that their teaching did not engage learners in the classrooms. Before LtP, their pedagogical approach included mostly lecturing and limited or no use of activities in the classroom. In some cases, teachers also confirmed that learning outcomes from such pedagogical practices were not great. For example, a teacher from Greater Accra South shared,

“During lessons, I did all activities without involving learners. I talked about the subject then [gave] learners assignments to do. [For] example, I am teaching English with noun[s] as the topic for discussion. I enter the class and copy notes on noun[s] on the blackboard and ask learners to write the notes in their notebooks. After that, I give assignment to learners...[the] learning outcome was not good because learners did not really understand what was [being] taught.”

There was a noteworthy shift in how teachers described changes in their instruction after engaging with the LtP training. After the introduction of LtP, teachers described a change from their conservative, teacher-led pedagogical approach to a child-centered approach. A teacher from Weija Gbawe explained,

“I start the lesson with a starter and a song with the learners. The song is related to numeracy...I used them to get the learners involved. It makes them understand the lesson better. Because, when using the old method, they were not getting the lesson. But with the training I have received from RtP am able to teach the learners to get a better understanding.”

Teachers interviewed indicated how they use LtP practices daily in their classrooms. They hoped to continue this practice with the use of available resources and materials. A teacher from Greater Accra South suggested,

“I will infuse or integrate play-base[d] methods or RtP pedagogies in all my lesson [deliveries]. The concept of learning through play has come to stay.”

Another teacher from Weija Gbawe echoed similar sentiments in response to continuing LtP,

“The activities and the games...has become part of us. And these things have come to stay. I mean the LtP methodologies. We also make sure new teachers acquire the skills that we have through [professional learning communities].”

Support to Teachers

The percentage of indirect implementation teachers who reported receiving training increased significantly from baseline, when compared to comparison schools. This is likely in part due to the training and support given to the DTSTs, whose primary role was to cascade training to these schools. **Additionally, although, overall, it remained low (16%), the percentage of direct teachers who received short message service (SMS) or other remote support increased significantly from baseline, when compared to comparison schools (Table 25).**

	Comparison		Indirect		Direct		GALOP	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Training	67%	58%	57%	89%*	95%	91%	83%	80%
Coaching	39%	51%	38%	57%	51%	75%	52%	72%
SMS/Remote Support	11%	2%	4%	4%	4%	16%*	17%	9%
Teacher Meetings	44%	80%	60%	94%	70%	93%	57%	91%

* $p > 0.05$

As Table 26 shows, the proportion of teachers who said they received play-based materials also increased significantly in indirect and direct schools. Although both groups saw increases, it is worth noting that the frequency of receiving play-based learning materials is much higher in direct implementation schools (with only approximately one-quarter of teachers reporting having received play-based materials in Indirect and GALOP schools).

	Baseline	Endline
Comparison	11%	7%
Indirect	11%	28%*
Direct	46%	73%*
GALOP	33%	28%

* $p > 0.05$

The majority of teachers across treatment groups said the coaching and mentoring they receive comes from an HT (Table 27). There were no significant changes from baseline on who provides coaching and mentoring.

Table 27. Who Provides Teachers with Coaching and Mentoring								
Comparison	Indirect		Direct		GALOP			
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
HT	44%	43%	33%	44%	25%	30%	56%	32%
SISO or District Official	11%	9%	22%	8%	19%	14%	12%	11%
Fellow Teacher or Champion Teacher	7%	3%	11%	%	3%	2%	9%	3%

* $p>0.05$

When asked about the impacts of coaching and mentoring, the percentage of teachers who said that coaching improved their skills significantly increased in direct implementation and GALOP schools, when compared to the comparison group (Table 28).

Table 28. Percent of Teachers Who said Coaching Improved Their Skills as a Teacher		
	Baseline	Endline
Comparison	78%	80%
Indirect	72%	84%
Direct	59%	88.5*
GALOP	49%	81.6*

* $p>0.05$

When asked what about the coaching helped them improve, teachers most often cited that the person providing coaching or mentoring was supportive, gave them confidence, helped them increase “play” in the classroom, and observed and provided feedback (Table 29).

Table 29. What about Coaching Made You Change Your Teaching? (Endline Only)				
	Comparison	Indirect	Direct	GALOP
The coach was very supportive/friendly	71%	73%	85%	82%
The coach gave me the confidence to improve my teaching	86%	76%	83%	84%
The coach helped me increase play in the classroom	66%	76%	83%	71%
Coach observes my teaching and provides useful feedback	83%	70%	73%	66%
Coach and I discuss progress from last observation	51%	51%	54%	53%
Coach models instruction	40%	27%	42%	37%

When asked about their participation in professional learning communities (PLCs), the percentage of teachers who reported discussing play-based strategies during these meetings sometimes, most of the time, or always, increased across groups, with the highest increase in the proportion of teachers happening in the direct implementation group (although no significant differences were detected; Table 30).



Table 30. Percent of Teachers Who Said They Discuss Play-Based Strategies during Teacher Meetings Sometimes, Most of the Time, or Always

	Baseline	Endline
Comparison	41%	71%
Indirect	55%	83%
Direct	77%	93%
GALOP	61%	87%

During qualitative interviews, teachers spoke positively about RtP and district staff, and saw them as support to understand and infuse the LtP methodology. Importantly, teachers said they were continuing to use the same activities that they learned during the training sessions. A teacher from Weija Gbawe explained the components of training as follows:

“We’ve had the foundation training, lesson preparation, and demonstration, TLRs, positive environment, coaching and mentoring, and teacher champion. A refresher training of what has been done throughout the year and an action plan for PLCs.”

A teacher from Oshiyie then explained the role of a district employee,

“The [part] I like most is the coaching and mentoring because mentors from the district and RtP come to teach us how to improve our teaching. I like how they teach us starters that will help predict the lesson and how to use games.”

Additionally, teachers attributed the support received as a reason why they were confident that the LtP methodology will continue once the program ends. A teacher from Greater Accra South spoke about his relationship with one RtP staff member, noting specifically how he was trained to write and prepare “proper” lesson notes. In thinking ahead the teacher shared,

“If the RtP program ends now, I will continue to teach learners through play using RtP’s methodologies and materials I have received...[and use] the Reflect-Connect-Apply approach of teaching lessons. RtP gave teachers pen drives, which have various play activities that can be used when teaching lessons.”

School Leadership

This section focuses on the guidance and support provided by HTs while implementing P3. Some items, specific to the intervention, were only asked of HTs in treatment schools and, therefore, impacts were not measured.

At endline, **virtually all HTs in treatment groups reported that P3 helped to improve teaching practices at their schools (Table 31)**, with 100% of HTs in the direct implementation schools citing success in improving the teacher practices.

Table 31. Percent of HTs Who Felt P3 was Successful in Improving Teacher Practice at Their School

	Endline
Indirect	92%
Direct	100%
GALOP	91%

When asked the ways in which the program was successful (**Table 32**), the most common responses from HTs included improved teaching, improved student outcomes, and (in direct schools) the provision of more teaching and learning materials.

Table 32. In What Ways Has the LtP Program Been Successful? (Endline)			
Response	Indirect	Direct	GALOP
Teaching has improved	84%	91%	87%
More teaching and learning materials	50%	83%	42%
Student outcomes have improved	80%	75%	82%
Teachers are being supported more	38%	70%	40%
Students are more active in the classroom	48%	65%	69%
Student attendance has improved	50%	60%	50%
Teacher attendance has improved	50%	50%	40%
Parents are more involved in school	29%	44%	28%
I am getting more support from supervisors	31%	23%	19%

When asked what role they played in helping P3 improve learning (**Table 33**), most HTs across all groups cited monitoring teachers in their classrooms, followed by providing information to teachers on program expectations and coaching and mentoring teachers. Between 73% (GALOP) and 90% (direct) of HTs reported being trained by the program to perform their role supporting LtP.

Table 33. What Role Have HTs Played in Helping P3 Improve Learning (Endline)			
Response	Indirect	Direct	GALOP
Monitor teachers in their classrooms	92%	91%	77%
Provide information to teachers on program expectations	63%	74%	61%
Coach/mentor teachers	75%	68%	61%
Facilitate professional development activities (training learning circles etc.)	46%	63%	58%
Provide information about the school's progress to program staff	46%	47%	37%
Provide information about the school's progress to Ministry staff	29%	12%	16%

As **Table 34** and **Table 35** show that most HTs rely on student results and regular interaction and discussions with teachers to determine if a teacher is not performing as expected. The use of coaching data and feedback as a measure of teacher performance increased across all groups. If the teacher is not performing as expected the majority of HTs reported providing extra support themselves to the teachers and requested additional coaching support for the teacher. The proportion of comparison teachers who said they provided follow-up support to teachers increased significantly compared to direct implementation and GALOP. However, it is important to note that the direct and GALOP groups already had a very high percentage of HTs providing support.

Table 34. How Do HTs Know If a Teacher in Their School Is Not Performing as Expected					
Response	Timepoint	Comparison	Indirect	Direct	GALOP
Student results	Baseline	78%	67%	73%	80%
	Endline	82%	79%	81%	74%

Table 34. How Do HTs Know If a Teacher in Their School Is Not Performing as Expected

Response	Timepoint	Comparison	Indirect	Direct	GALOP
Regular interactions/discussions with teachers	Baseline	57%	76%	73%	67%
	Endline	76%	73%	77%	67%
Coaching data	Baseline	2%	4%	11%	11%
	Endline	29%	40%	32%	39%
Coaching feedback	Baseline	4%	11%	9%	7%
	Endline	22%	38%	35%	33%
Supporting the teacher	Baseline	2%	7%	20%	11%
	Endline	20%	42%	28%	35%

Table 35. What Action Do School Directors Take if a Teacher is Not Performing as Expected

Response	Timepoint	Comparison	Indirect	Direct	GALOP
Provide additional support to the teacher yourself	Baseline	70%	80%	93%	87%
	Endline	93%	92%	97%*	87%*
Request additional coaching support for the teacher	Baseline	30%	35%	33%	38%
	Endline	60%	65%	65%	74%
Ask another teacher to mentor the teacher	Baseline	11%	17%	18%	27%
	Endline	42%	60%	61%	57%
Observe the teacher (more frequently)	Baseline	65%	54%	49%	36%
	Endline	69%	58%	74%	63%

* $p > 0.05$

During the qualitative interviews, HTs' perceptions of their support role also became evident. HTs frequently described their role as vetting lesson preparation notes, observing classrooms, and mentoring and coaching teachers. Outside of the classroom, HTs indicated their support by providing materials, such as manila cards, to support LtP's implementation. One HT from Weija Gbawe explained their role as follows:

"I observe the teacher by using lesson observation sheets (LOS) which are designed by both GES and RtP. I sit with the teachers and point out their shortcomings and address them. I also do individual coaching based on the scores from the LOS. I also do follow-ups to see if the teacher is integrating what I addressed."

HTs also said they began to incorporate the child-centered methodological approach in how they assessed teachers. One HT explained, *"I organize end-of-term analysis where performance of [the] teachers are monitored. I give learners questionnaires where I ask learners how they...like or relate with teachers."*

Home Environments

This section presents feedback gathered from parents and caregivers in quantitative and qualitative interviews. When interpreting the findings presented here, it is important to note the convenience sample of parents and caregivers willing to come to the school for an interview (i.e., one per school for the qualitative interview and two per school for the quantitative interview). Responses are likely biased toward more engaged and invested parents and caregivers.

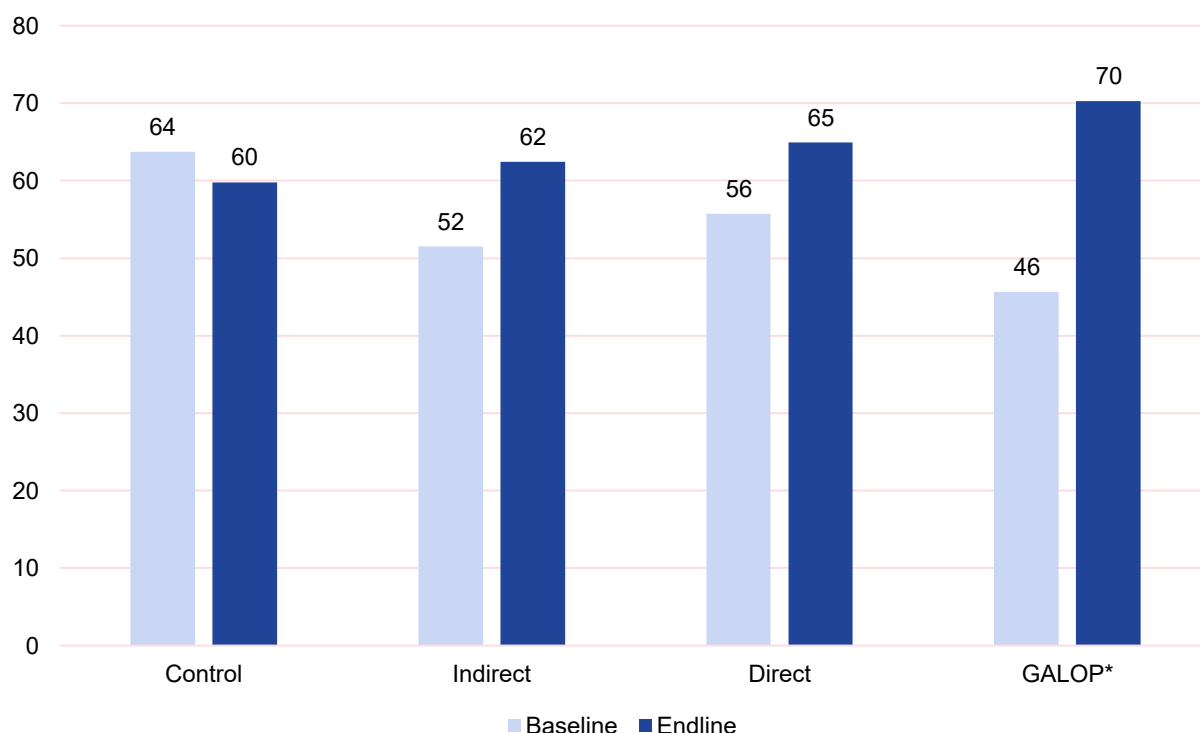
Table 36 shows the percentage of caregivers who reported engaging with the school in the month prior to the study, through different activities. **In direct implementation schools, the percentage of caregivers in who met with their child's teacher increased significantly at endline, when compared to the comparison group** (where the percent of caregivers meeting with teachers decreased). **Further, the percentage of caregivers in direct and indirect implementation schools who volunteered in class also increased significantly from baseline, when compared to the comparison group.**

Table 36. Activities that Caregivers Participated in within the Month Prior to the Study			
Action	Treatment Group	Baseline	Endline
Attended a PTA meeting	Comparison	55%	45%
	Indirect	49%	43%
	Direct	59%	57%
	GALOP	59%	56%
Met with child's teacher	Comparison	69%	49%
	Indirect	56%	55%
	Direct	50%	59%*
	GALOP	61%	57%
Volunteered in class	Comparison	19%	15%
	Indirect	4%	16%*
	Direct	6%	28%*
	GALOP	13%	24%
Attended a school event	Comparison	15%	22%
	Indirect	7%	20%
	Direct	18%	36%
	GALOP	19%	32%

* $p > 0.05$

Figure 11 shows the percentage of caregivers who discussed how their child was doing with the teacher. **The percentage of caregivers in GALOP schools who reported discussing how their child was doing in school with teachers increased significantly from baseline, when compared to the comparison group.**

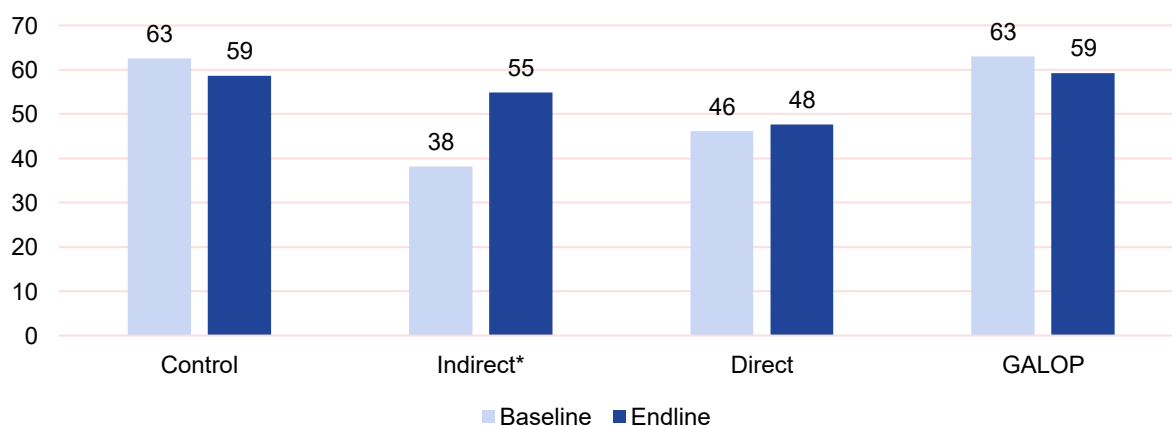
Figure 11. Percentage of Caregivers Who Discussed with the Teacher How Their Child Was Doing: Baseline vs. Endline



* $p > 0.05$

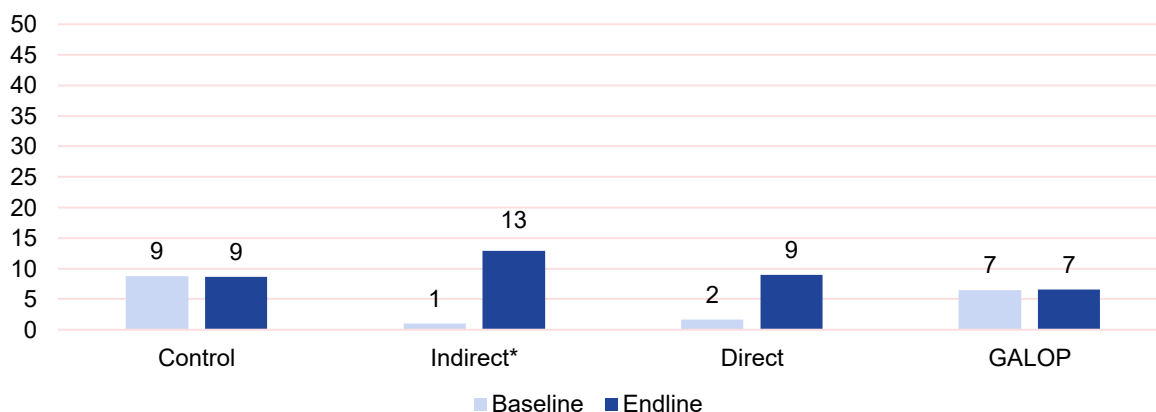
Figure 12 shows the percentage of caregivers who said they helped their child with homework in the 3 days prior to the study, while **Figure 13** shows the percentage of caregivers who reported having craft supplies at home. **The percentage of caregivers in indirect implementation schools who reported affirmatively to both of these questions increased significantly from baseline, when compared to comparison.** It is worth noting that frequencies for these items were also much lower in the indirect group, at baseline. No other significant changes were detected in the caregiver interview data.

Figure 12. Percentage of Caregivers Who Reported Helping Their Child with Homework in the 3 Days Prior to the Study



* $p > 0.05$

Figure 13. Percentage of Caregivers Who Reported Having Craft Supplies at Home



* $p > 0.05$

During qualitative interviews, caregivers stated that they were aware of the LtP approach and were observant of their children's engagement in the learning process by using activities or materials at home. A parent from Greater Accra South described LtP as follows:

"I think play makes learners easily grasp what is been taught when learning. I also believe some learners get more understanding on lessons when lessons are delivered using play-based methods."

Another parent from Weija Gbawe explained their observation at home as,

"She plays with her brother in the upper primary most of the time, so they both benefit from each other's playful activities. Sometimes they will be playing with blocks. They will arrange it and ask each other to pick and then record some numbers. I don't know what work they are doing but I believe they are learning something they learned in class."

The majority of caregivers indicated that they were in communication with the school either through the PTA or the school's WhatsApp, which helped them be aware of their child's progress in school. However, caregivers suggested that instead of sending messages home with their children or through the WhatsApp platform, it would be beneficial for the school to call the parents. Caregivers noted that they considered the information they received from the school as important and that they participated in different capacities, such as developing levies through the PTA, developing learning materials, or providing feedback. A caregiver from Weija Gbawe shared,

"It helps to allow us—the parents—[to] also contribute to what the school needs to improve situations....[Through] WhatsApp, meetings, and sometimes they invite us...They discuss with us enrolment issues. That is the number of learners to be enrolled and how that will affect teaching and learning....It is very important to get involved in such a decision process."

HTs and teachers said they expected for caregivers to be involved in the learning process of their children. They also reported receiving support from caregivers and the PTA in different forms. For example, according to one teacher, *“During a lesson on ‘some selected individuals’ in our community, a parent came and served as a resource person. The parent, who was a police officer, came in her police uniform and spoke to learners about the police as a profession learners can take up in the future.”*

Government Support of LtP

In-country qualitative researchers also conducted interviews with an array of government actors at the district and national levels. Responses were coded, analyzed, reviewed, and triangulated between different government levels.

National-Level Interviews

In interviews held with MOE officials at the national level, two key themes emerged: (1) officials’ relationship with and perceptions of LtP and (2) sustainability.

Relationship with and Perceptions of LtP

At the national level, employees had varying understanding of the LtP initiative. Some officers had stronger understandings of the LtP initiative and, therefore, were able to define LtP in the following ways:

- “Learning through play means children researching and finding out information themselves, so the learning stays with them.”
- “It is the process where learners are taken through play-based activity, including group work, to come out with learning outcomes of the lesson.”
- “Using [an] activity base in learning.”

For other national-level employees, definitions did not necessarily align as well to LtP’s methodology. These individuals were more disconnected from the initiative and remarked that they did not participate in a great deal of training. Additionally, in reflecting on the LtP initiative, these individuals often spoke in vague terms and would reference RtP, and not the specific initiatives associated with LtP. For example, one officer explained, *“Learning through play is a methodology that promotes the learner’s ability to learn concepts more effectively.”*

In reflecting on the initiative, one director expressed positive sentiments and support for play-based methodologies, but acknowledged how his position was one to steer administrative and managerial decisions, explaining, *“Looking at the impact of the play-based method we have decided to bring all [nongovernmental organizations] together to share experience[s] and consolidate a common guideline. For example, I was in one of the districts, and during a math lesson, I realized that play-based learning is very effective. I represent management, so whatever we do reflects management.”*

Sustainability

National employees shared a level of optimism that aspects of the LtP initiative will be sustained after P3’s end because play-based methodologies have been infused into the national curriculum, training, and policy documents. The Deputy Director of Early Childhood Education spoke at length about her work with RtP and where she has advocated to develop a kindergarten curriculum. Although she has only been in the position for 1 year, she received prior training on LtP methodologies in her previous position, and RtP has supported her work in creating early childhood training manuals. She explained, *“RtP has*



helped us to develop a national [kindergarten] insert training manual. RtP has also trained us, the Early Childhood Education (ECE) unit, on how to implement this manual.”



Another director spoke about using technology to create sustainable spaces for sharing, noting, “For me, I want to lead a team to video some of these initiatives and put them online for easy access.” **Other national employees believed aspects of the initiative will inevitably end due to the lack of funding to support materials, teacher training, and monitoring.** One officer spoke directly about the relationship between materials and monetary resources, stating, “Funds won’t be available, and textbooks and manila cards might also not be available again because there will be no funding.” Conversely, another officer believed sustainability was inevitable because “The standards-based curriculum we developed already prioritized learner-centered approaches to teaching. So, teachers are supposed to be facilitators,” although he did share his concerns about how this will occur when funding ends. He commented, “It will be difficult to implement the RtP activities when funding ends. As we speak, the [science, technology, engineering, and math] STEM guide that RtP supported us to develop...lack[s] the funding to even print to train the teachers. So, not much can happen when their program ends.”

District-Level Interviews

Interviews were conducted with officials in two districts, Greater Accra South and Weija-Gbawe, each of which include a mix of direct and indirect implementation schools. In interviews held with district officials, five key themes emerged: (1) LtP perceptions, (2) observable shifts in teachers’ instructional practices, (3) collaboration and capacity building, (4) reflections on training, and (5) sustainability.

LtP Perceptions

Across all participants holding positions at the district level (i.e., SISOs, Trainers, and DEDs), there was a shared understanding and definition of LtP, which included child-centered learning, where students were happy as they engaged in group activities that were meaningful and interactive. Two examples of how LtP was defined by district-level officials are below.

- “Learning through play is any lesson that is meaningful, engaging, interactive, and joyful.” (DED)
- “Using play-based pedagogies that will bring fun, make learners happy, make learners interactive, and make learners grasp what is being taught in class.” (DTO)

Observable Shifts in Teachers’ Instructional Practices

When speaking about shifts in teachers’ pedagogy, district-level employees described distinct changes in teachers’ instructional practice. Prior to LtP, teachers were dependent on teacher-focused, lecture-style pedagogy, and few to no materials were observed across instruction. One SISO explained, “*teacher(s) struggled to execute lessons because there were not enough resources to be used. Also, teacher(s) kept using the lecturing method without involving learners in lesson delivery.*” However, in the year prior to the study, **those supporting and observing teachers (i.e., SISOs and training officers) noted distinct shifts in how and the frequency in which teachers used games, cooperative learning strategies such as think-pair-share, and manipulatives to enhance learning.** Another SISO shared, “*I observe teachers every day and 98% of the teachers are using games—except new teachers. It’s part of their daily work.*” He went on to share how he observed teachers using culturally specific materials, such as local games that are meaningful, teaching-learning resources, “TUMATU,” i.e., lines drawn on the floor that students have to

jump to teach math, the “think, pair, share” methodology, and the use of manilla cards with numbers written on them.

A Training Officer described an observed lesson he believed embraced aspects of LtP that engaged students in cooperative and interactive processes in the following way:

“One of the teachers I observed last year used an energizer to get the class settled. And he instructed one of the learners, called Kofi, to lead the class with the energizer. So, if Kofi says “sit,” you have to sit, but if Kofi says, “stand,” you have to stand. The activity was interesting and got the learners to pay attention. Getting to the middle, what the teacher did was, he grouped the learners into smaller groups and gave them a task. That is, sorting out nouns using manila cards. The learners had to sort the cards with words written on them. Then the teacher will ask questions why they chose those words and which of the words are nouns. All the learners were participating in the group activity. At the end of the lesson, the teacher used questions and answers to summarize the lesson. After that, he announced the next lesson.”



District-level employees did identify challenges teachers faced in using the LtP pedagogy, including classroom management, lack of resources, and space. One SISO reflected on these challenges in the following way:

“Teachers had challenges on how to comparison learners in the course of using play activities while teaching. This is because class size was bigger than the size of the classroom. The teacher did not...manage time well; the teacher got carried away during play time while teaching.”

Collaboration and Capacity Building

All district-level employees described how collaboration with HT and RtP employees were important aspects in carrying out their work with the LtP program. SISO and Training Officers explained how frequent and consistent visits to the schools, which occurred approximately monthly or bi-monthly, supported LtP’s implementation. During these observations and school visits, district-level employees reflected on how they worked with HTs and teachers to align observed teaching practices with RtP’s and the LtP methodology. One SISO explained his relationship with RtP schools in the following way:

“For the RtP schools (schools involved with the RtP[-implemented] project) I meet with HT and teachers and have discussions with them on the strengths and weaknesses when using the RtP methodologies. I also meet with teachers every month to find ways of how best to use RtP methodologies in delivering lessons.”

A DED advocated that RTP’s program be implemented in his district because he believed it was a good program that sought to improve education, sharing, *“I appealed to RtP for my district to be added to their new project, which is the LtP/RtP P3 project.”* Similarly, a Training Officer reflected on how his director advocated that RtP move into their district, noting that the district *“... was not part of RtP, but during a ceremony, in the Ga South District, our director happened to be part of the program, and, having heard about RtP activities, he lobbied for RtP to select our district for implementation.”* **The Directors of Education spoke highly about their collaborative partners in the schools, namely how the training officers and HTs were the biggest advocates and supporters in developing RtP’s LtP program.**

Reflection on Training

District-level employees reflected on the importance of training received from RtP on the LtP pedagogy where this training included the following:



1. Coaching and Mentoring. Employees discussed how they learned how to coach and mentor teachers through collaborative means. *“If I go to monitor a class, I am able to support the teachers and I now provide one-on-one coaching with the teachers after observations. I am also now able to have interactive sessions with the teachers to discuss their shortcomings and this has helped them to improve. I do frequent follow-ups to see if they are still having challenges.”* (SISO)
2. Material Development. Employees learned how to support teachers in using the local environment to create manipulatives to support learning. *“We received training on how to use no-cost or low-cost materials to create teaching and learning resources. RtP has also given learning and teaching resources like pen drives (which [have] songs for play), cardboards, manila cards, markers, and teaching manuals.”* (SISO)
3. Instructional Practice. Employees reflected on how they received training on pedagogical approaches to support students. *“The activities that align with RtP are lesson note preparation; energizers (warm-ups); starters (must be related to the topic); asking questions (Reflect-Connect-Apply) [where] the learner should be able to reflect on the activity and then connect to the topic being taught and apply it to answer the question; performance indicator shared with learners; think per share (pedagogy) [where] two learners think about the lesson and share with the class; group activities; group presentation; assessment; summary; exercise; and evaluation. At the end, announce the next lesson.”* (Training Officer)

Sustainability

All district-level employees expressed optimism in how LtP will be sustained once funding ends. **Participants described systematic structures of teacher monitoring and development through PLCs, which will support sustainability.** A Training Officer attributed PLCs, training, and mentoring programs established by RtP, as well as buy-in from national officials, as reasons why the program will continue, stating:

“Support from RtP has built officers in terms of coaching and mentoring to provide technical support. I believe it will continue because, the Deputy Minister of Education was here and from the discussion that we had, he was so happy, and, per our discussion, I believe RtP will continue.”

Additionally, employees reflected on how school-based organizations like the PTA and school management committees will provide funding. District-level employees identified how they are in contact with national-level employees and there have been discussions and plans on how to move forward. District-level employees were confident the MOE will continue to fund and support the RtP initiative. For example, one Director of Education was confident that all aspects of RTP’s program would be sustained, sharing,

“All parts of the RtP initiative [will be sustained]. We are leveraging LtP in non-RtP schools to improve learning outcomes. We have planned [for RtP’s] exit [since] year 1. We also have functional municipal executives to support the activities of LtP. We are also taking community engagement seriously, as well as the RtP pedagogies. We will continue the monitoring and coaching and PLCs in the schools.”



4. CONCLUSION

4.1 Measurement Considerations

This study incorporates the use of well-established measures (such as the EGRA and EGMA) alongside constructs that have been less commonly measured in low- and middle-income countries. For example, we developed new play attitude scales, as well as a new SEL tool (which went through a rigorous development and piloting process, prior to administration). These tools were reviewed after baseline and revised accordingly, while ensuring that we could maintain comparability across time points.

Although the intent of this study was to collect baseline data on all programs prior to the start of any school-level implementation, this approach was not always possible, due to COVID-19-related delays. Therefore, all impact estimates at endline comparison for baseline differences.

Additionally, while our student-level estimates of impact are strengthened by the student-level longitudinal design of this study, there was attrition at the student level (30%). There was no evidence of differential attrition, but the absenteeism may still have had an impact on the estimates. Several methods were considered for mitigating attrition bias. However, recent research has shown the limitations of imputation and inverse probability weighting, which are the two most common approaches for dealing with attrition (Anderson et al., 2021; Cumming & Goldstein, 2021; Metten et al., 2022). The most recent of these articles (Metten et al., 2022) found that inverse probability weighting does not actually reduce the bias more so than estimates that simply account for baseline covariates, while the others noted that imputation can actually lead to more biased estimates with less precision. Therefore, we ultimately elected to account for attrition by estimating regression models that included baseline scores as comparisons. These regressions will produce slightly biased estimates when differential attrition is higher, but the estimates should simply be seen as upper or lower bounds (depending on whether the differential attrition favors treatment or comparison groups).

It must also be noted that identifying directly comparable comparison schools for P3 was difficult. The challenge was the limited availability of at-risk schools in the selected regions of Ghana, given that GALOP engages those schools directly. Therefore, it is possible that some of the negative impacts on EGRA and EGMA outcomes found at endline resulted from comparison schools consisting of higher performing schools across the board, beginning at baseline. It is otherwise unclear why teacher practices would improve in treatment schools but that the strongest gains in student performance would still occur in comparison schools.

4.2 Implications and Recommendations

LTP in the Classroom

Teachers continued to incorporate many basic best teaching practices in their lessons (e.g., student monitoring and support, material preparation, good instructional pace, making connections and sharing ideas, inclusionary practices). However, there was little change in the use of playful practices in the classroom from baseline to endline, across all four groups. There were, however, some notable shifts, particularly in direct implementation schools.

Teaching Approach. In direct implementation schools, there was a significant increase in the proportion of time that students spent reading during observed lessons. Teachers spent the largest proportion of time explaining concepts, modeling activities, and monitoring students' understanding, while students spent the largest proportion of time listening and

writing. ***We recommend that teachers build on the additional reading time, in order to continue to engage students with print-based materials. Additionally, the MOE could learn from the direct implementation approach to ensure that students have more opportunities to read in class.***

Material Availability. In both direct and indirect treatment schools, the percentage of teachers who said they received play-based materials increased significantly. The availability of play materials, however, was far more prevalent in direct schools (73%) than in indirect schools (28%). The percentage of observed lesson time during which books were used also increased significantly in direct implementation schools. Overall, despite a focus on materials under P3, there was considerable variation in their availability and use across classrooms. Without access to materials, teachers' implementation of programs will continue to be limited. ***We recommend careful examination of the conditions driving greater availability of materials in direct schools compared to other implementation groups, to help inform planning that ensures classrooms are provided with both traditional materials (e.g., teachers' guides) and play-based materials.***

Student Engagement. There was some evidence of increased student engagement under P3, with teachers noting that active engagement and a child-centered approach was at the heart of the program. ***We recommend building on the program's successes and continuing to increase children's opportunities for meaningful engagement through problem solving, reflection, and critical-thinking activities.***

Group Work. Although the majority of teachers (between 62 and 75%) noted the importance of group work for student learning, neither the use of nor attitudes towards small and large group work changed significantly from baseline to endline, across classes. This may be in part attributed to differences between instruction in grades 1 and 3. ***We recommend providing teachers with explicit guidance, informed by consultative discussions with teachers, on how to increase the use of group work and/or learning corners to improve social interaction and iteration among students. This is particularly valuable in overcrowded classrooms, in which students are at risk of being left behind by whole class instructional approaches.***

Student Agency. Although student agency is a major component of the discussion around LtP, nearly all class time was dedicated to activities initiated and directed by teachers. Opportunities for students to select from choices, choose roles, set goals, make rules, or choose when to move around were virtually non-existent. However, almost one-half of lessons included a chance for students to develop and share ideas with classmates of their teacher. Whether student-initiated or teacher-initiated and student-directed, ***we recommend building on teachers' existing practice of asking students to share ideas, to encourage students to share ideas about the direction of their learning. This, in turn, provides students with increased levels of autonomy over how they engage in an activity or, at a minimum, with improved opportunities for asking questions, expressing opinions, and practicing skills.***

LTP Enabling Environment

Learning Environment. While students increasingly reported incidences of physical and verbal aggression by their teachers at endline in Indirect, GALOP and comparison schools, this figure did not increase in direct implementation schools – showing a significant impact on the use of yelling and physical punishment. Although these are systemic issues, ***we recommend further investigating how direct implementation schools were able to abate this growing issue and provide a more supportive environment for children.***

Teacher Support Staff. In direct implementation and GALOP schools, significantly more teachers said coaching helped them improve – and most often cited coaches’ friendliness, confidence-building, and use of play as influential factors on their teaching. However, while a variety of actors are engaged to support teachers in P3, but they do not always appear to work together. For example, head teachers claimed that they give additional support to struggling teachers themselves, and teachers who received coaching most often said it came from the head teacher - despite the availability of coaches/mentors as well as teacher meetings. ***We recommend ensuring that support staff understand one another’s roles and that they work together to offer complementary support to teachers.***

Ministry Stakeholders. P3 has made strides in sensitizing ministry officials on the benefits of incorporating LtP in classrooms. However, it was also reported in qualitative interviews that MOE staff at the national level didn’t always understand the nuances of the P3 programming itself. ***We recommend continuing to leverage relationships with officials to clearly communicate expectations around learning through play, including continued sensitization of playful and interactive classrooms for teachers.*** In particular, the high level of engagement with, and involvement of, district level staff should be continued, promoted and leverages as the level of buy-in and understanding fostered at this level of the system is a unique strength of P3.

Caregivers. At endline, P3 showed a significant increase in parental engagement with teachers and classrooms. ***We recommend continuing to engage caregivers in school-based activities (through PTAs/SMCs) with an explicit focus on children’s learning. Additionally, further guidance could be provided to households on how to develop low-cost or homemade toys, crafts, and play material in order to engage in LtP practices at home.***

Perceptions of LTP

Attitudes Toward Play. Caregivers, teachers and education stakeholders generally supported LtP in theory but were less supportive of a student-centered, “free” classroom environment. This finding may have stemmed in part from a misconception that student agency and more student-centered approaches will lead to a less disciplined and more chaotic classroom. ***We recommend continuing the work of orienting key stakeholders on the benefits of learning through play in the classroom and understanding their concerns and objections in order to suggest solutions for improving attitudes toward specific playful practices.***

Student Holistic Outcomes

SEL. There was minimal impact on SEL scores for P3 students at endline, with observed confidence actually going down for students in all groups. This means that despite improvements in some teacher practices, there were not significant gains in SEL, aside from relationships. ***We recommend introducing more agency and group work for children in classrooms, which would be expected to increase their confidence, collaboration, conflict resolution, and relationships.***

Literacy and Numeracy. The EGRA and EGMA results represent endline estimates for students in Grade 4. As expected, large gains were found from baseline to midline across all treatment groups. Additionally, the overall impacts for direct implementation schools were typically positive relative to indirect and GALOP schools. Interestingly, the most consistent impacts were found for students in direct schools relative to those in indirect schools. These findings are very promising given the short timeline of the program. ***We recommend continuing to focus on supports for teachers and caregivers, along with materials, in***

order to promote the best opportunity for continued improvement in literacy and numeracy outcomes.

Conclusions

Overall, direct implementation schools in P3 led to some meaningful shifts at the classroom, school and community level. This is particularly impressive given the relatively short timeframe of the program, combined with the complexities of teacher behavior change. For example, while there remains a continued overreliance on the use of oral materials in classrooms, there was a considerable shift toward a greater use of books in direct implementation schools—which provided students with more time for engaging with text.

Additionally, teachers in direct implementation schools reported higher levels of feeling valued, than their comparison school counterparts. The program also produced improvements in parental engagement with teachers and classrooms—and led to high levels of ‘pro-play’ attitudes for school-level stakeholders in direct schools. These may not have been the large-scale ‘playful pedagogy’ shifts that the program was hoping for, but they are essential first steps toward greater active engagement for students, creating a foundation for additional changes and improvements in the future.

Some of these changes were able to be achieved due to Right to Play’s adaptive approach to program implementation. During the program’s life cycle, a number of concerns and challenges were identified (e.g., teacher commitment to implementing new practices in the classroom; teacher ability to create learning materials; lack of structured lesson planning). As a response, RtP developed targeted approaches to address these issues. However, there are additional challenges that students, teachers and parents face at a system/structural level (e.g., teacher transfer, lack of infrastructure, high student-teacher ratios, etc.). These challenges cannot be overcome by programs alone but also require the support of government staff, along with funding and policy changes.

The Ministry of Education in Ghana has made it clear that they are interested in moving toward greater use of child-centered approaches, like those touted in Right to Play’s P3 activity. This shift can be seen by their interest in expanding the reach of LtP approaches from several hundred direct and indirect implementation schools to nearly 10,000 schools under the GALOP program. In addition, there appears to be strong support for LtP by MOE staff at the national and district levels.

However, there remains a mixed understanding among central ministry staff regarding what ‘learning through play’ actually means, as well as the specifics of how RTP’s intervention is intended to work. Therefore, while there is broad ‘buy-in’ for the P3 activity and LtP more broadly, it will be important to continue to orient and sensitize national, district, community, and school level stakeholders to the intricacies of LtP programming in schools. This can also help to garner additional support and funding, which will be required to ensure that effective practices can be integrated and sustained into a system beyond the life cycle of programs such as P3.

ANNEX A. ADDITIONAL TABLES

Figure A-1. Teachers' Responses to "I Feel Valued by Parents," Baseline vs. Endline

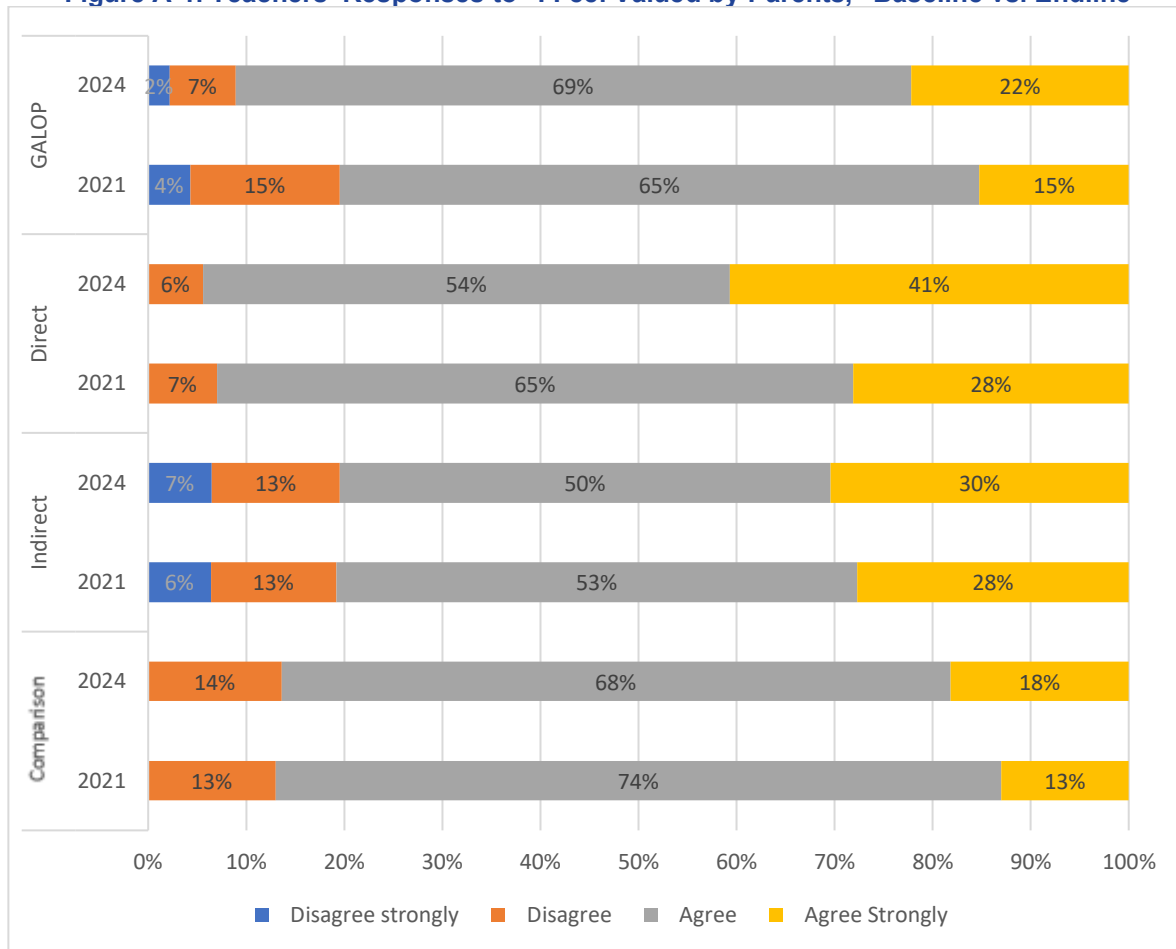


Figure A-2. Teachers' Responses to "I Find My Job Rewarding," Baseline vs. Endline

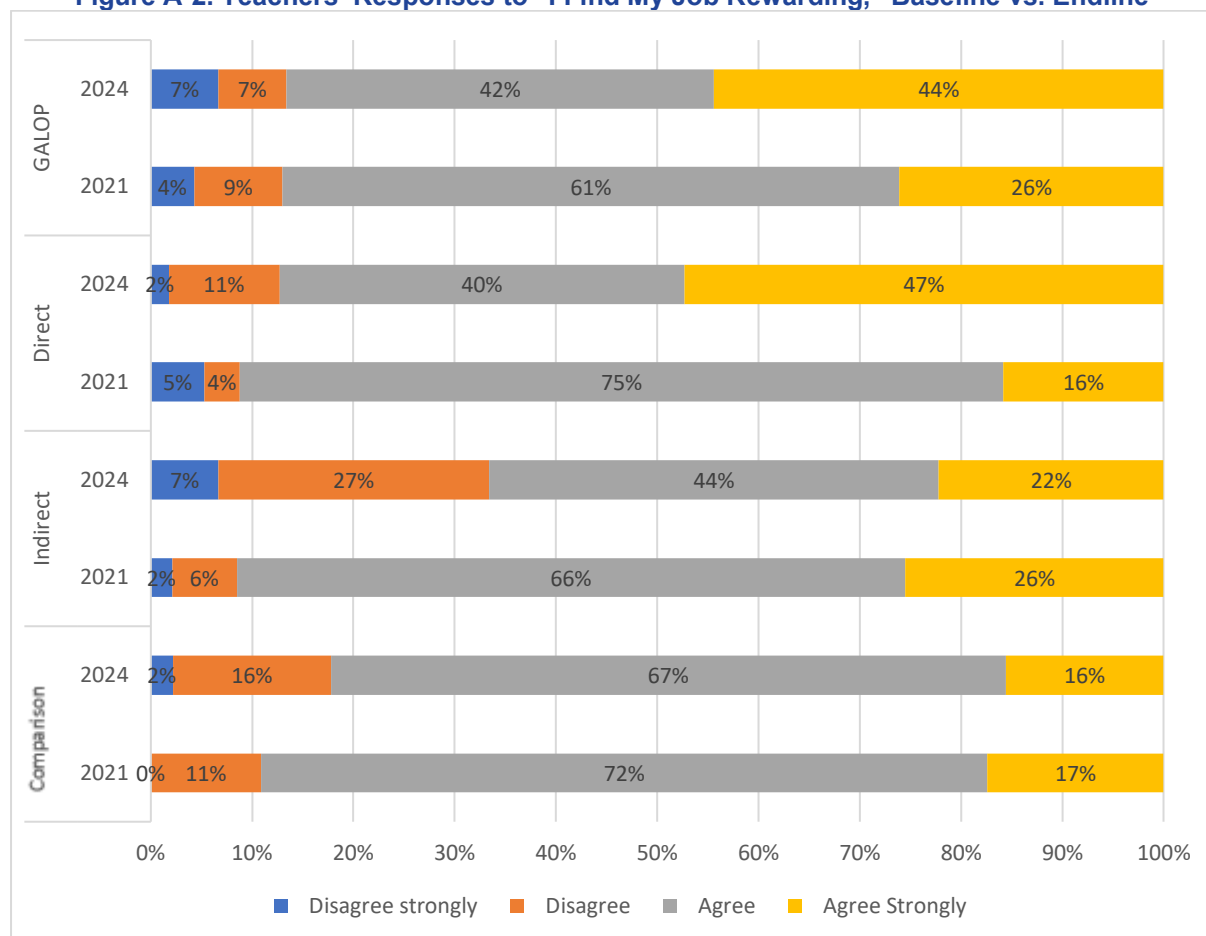


Figure A-3. Teachers’ Responses to “I Have Support I Need from the School Administration to Effectively Implement LtP,” Baseline vs. Endline

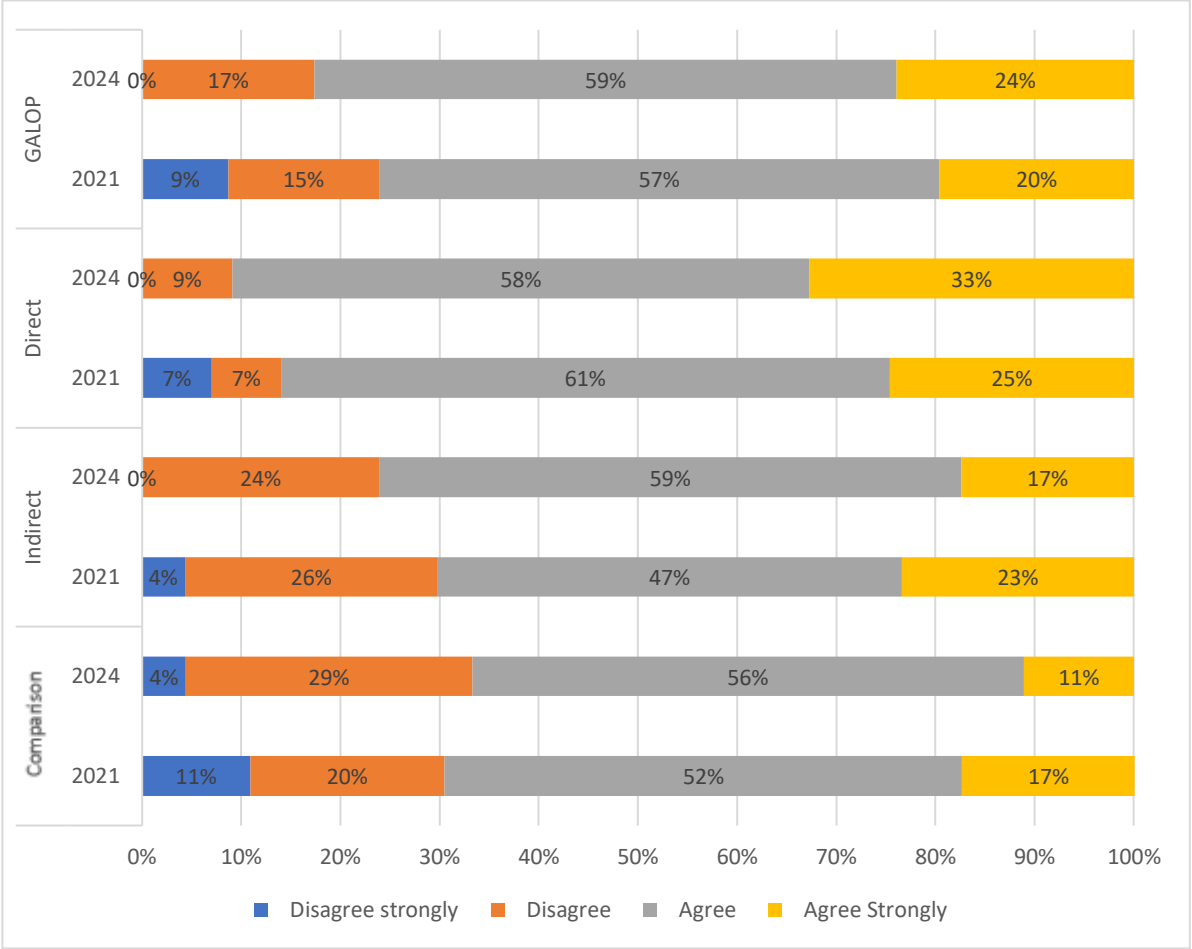


Figure A-4. Teachers' Responses to "I Regularly Use Play-Based Teaching Methods in the Classroom," Baseline vs. Endline

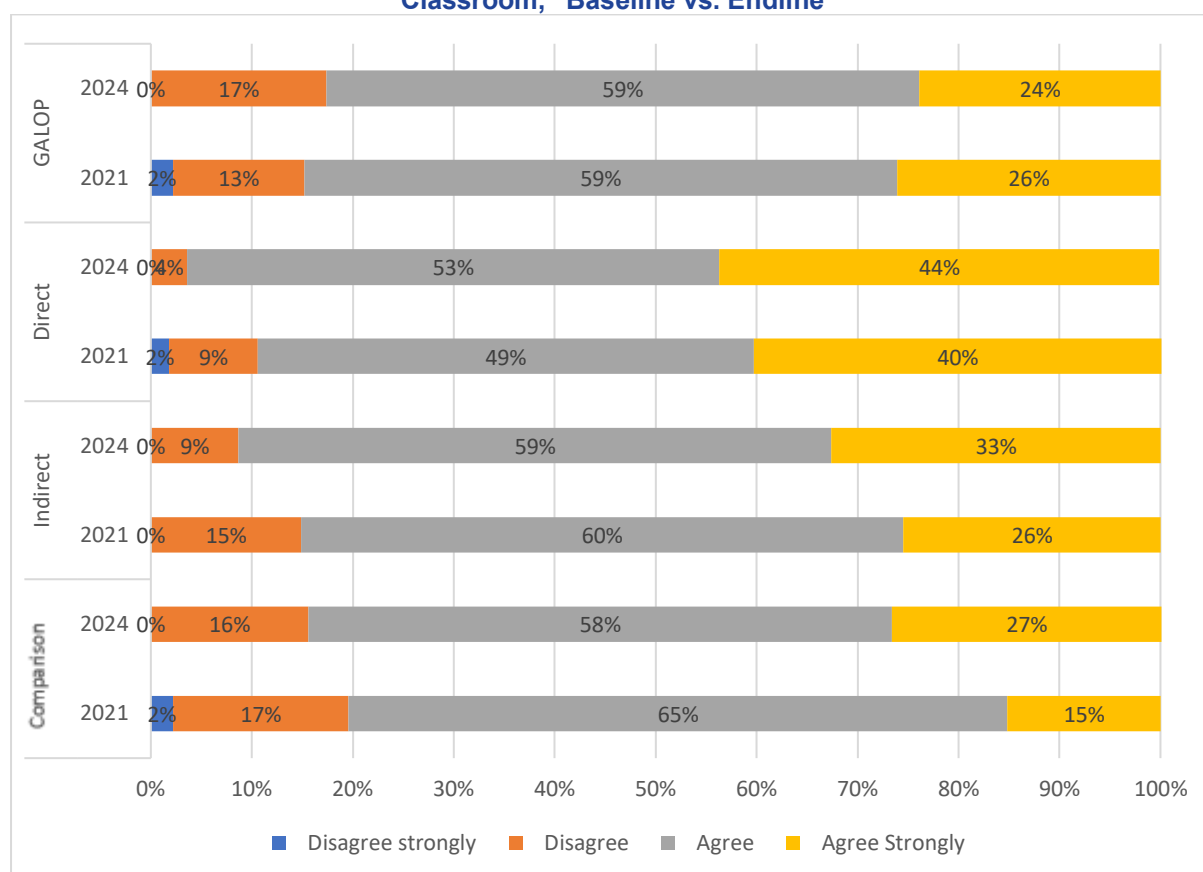
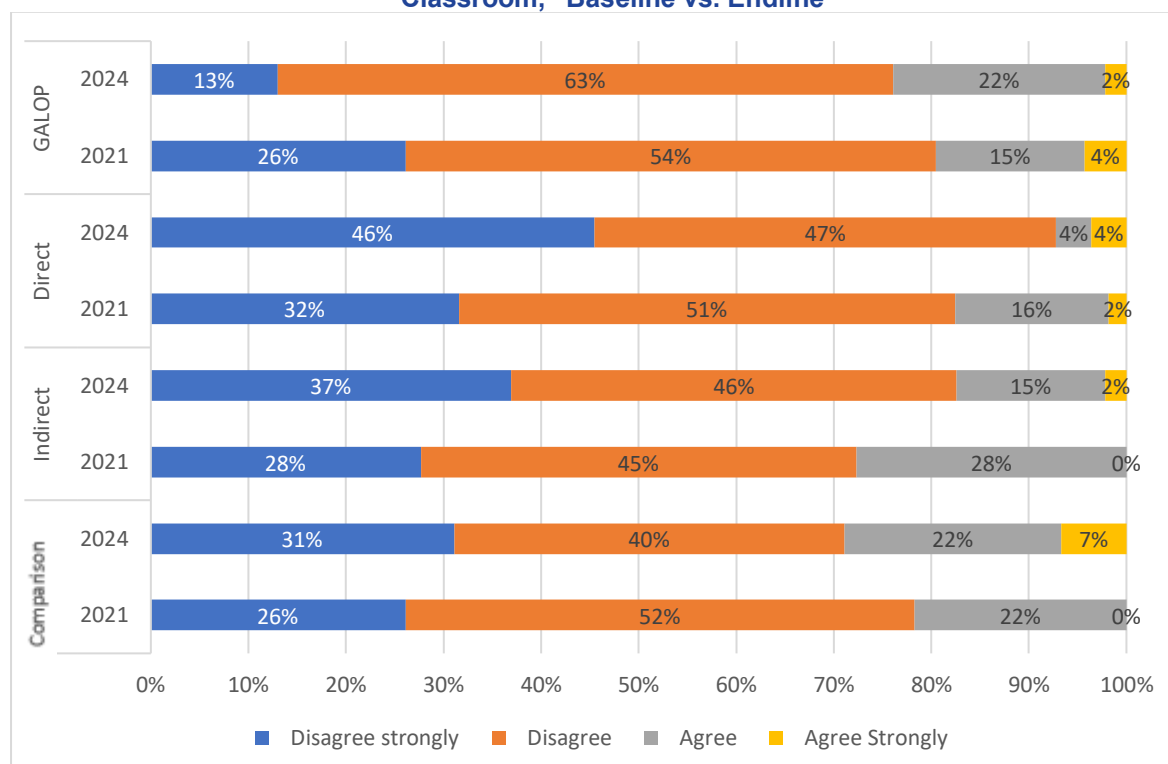


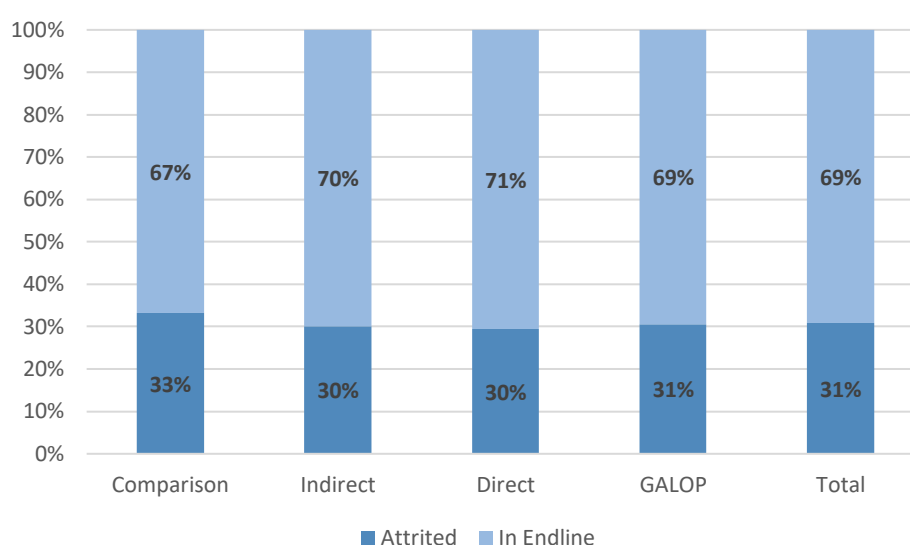
Figure A-5. Teachers' Responses to "LtP Methods are Too Difficult to Implement in the Classroom," Baseline vs. Endline



ANNEX B. ATTRITION ANALYSIS

The first step in examining attrition was to estimate the proportion of students who attrited (i.e., were not available at midline) along with those who were in the midline data. This was done by treatment group as well as overall (see Figure B-1 below). Overall, attrition was consistent across the three treatment groups, with slightly (but not significantly) higher attrition in the comparison group. The overall attrition rate for the Ghana study is 31%. It is worth noting that attrition was slightly higher among older children – the mean age of children who were assessed at endline is 7.8 years, while the mean age of children who were not assessed at endline is 8.1 years. This has no impact on the comparability of treatment and comparison groups.

Figure B-1. Percentage of Attritors by Treatment Group in Ghana



Following these initial attrition analyses, we examined the baseline EGRA and EGMA outcomes for each treatment group by attrition status. As shown in Tables A-1 and A-2, there was a significant difference in performance between attrited students and those who were tested at endline was found in the indirect implementation group for three subtasks: letter sound fluency, quantitative comparison, and basic addition. Significant differences were found in the direct implementation group for level 2 addition and subtraction, and in the comparison group for level 2 addition. Follow-up regression analyses were run to estimate the effects of treatments status and attrition on these EGRA and EGMA outcomes (comparing comparison schools with the three treatment groups). In these models, no significant effects were detected when treatment group and attrition status interact. This analysis did suggest slightly higher performance among students who attrited in indirect schools - this may therefore represent lower-bound estimates observed in the indirect group.

In Ghana, the average age of students missing from the endline sample was higher than that of students who were in the sample (8.1 vs 7.8 years). Additionally, overage students were more likely to be missing at endline (63% vs 55%). This finding, consistent with other countries, points to evidence that older students and overage students may be more likely to be absent or drop out of school at rates higher than their 'of-age' counterparts.

Table B-1. Baseline EGRA Scores by Attrition Status by Treatment Group in Ghana

	Treatment Group	Missed Endline n=917	Assessed At Endline n=2,059	p-value
Letter Sounds Fluency (correct letters per minute)	Comparison	7.4 [±1.7]	7.4 [±1.2]	p=.986
	Indirect	6.9 [±1.5]	4.7 [±.9]	p=.009*
	Direct	6.6 [±1.4]	6.2 [±.9]	p=.660
	GALOP	6.2 [±1.7]	5.6 [±1.0]	p=.535
Invented Words Fluency (correct wpm)	Comparison	.6 [±.4]	.3 [±.2]	p=.157
	Indirect	.4 [±.2]	.4 [±.2]	p=.817
	Direct	6.6 [±1.4]	6.2 [±.9]	p=.660
	GALOP	6.2 [±1.7]	5.6 [±1.0]	p=.535
Oral Reading Fluency (correct wpm)	Comparison	.5 [±.5]	.4 [±.3]	p=.485
	Indirect	.3 [±.2]	.3 [±.3]	p=.780
	Direct	6.6 [±1.4]	6.2 [±.9]	p=.660
	GALOP	6.2 [±1.7]	5.6 [±1.0]	p=.535
Reading Comprehension Score (% correct items out of 5)	Comparison	.3 [±.5]	.2 [±.4]	p=.891
	Indirect	.1 [±.2]	.2 [±.3]	p=.377
	Direct	6.6 [±1.4]	6.2 [±.9]	p=.660
	GALOP	6.2 [±1.7]	5.6 [±1.0]	p=.535

*p<0.05

Table B-2. Baseline EGMA Scores by Attrition Status, by Treatment Group in Ghana

	Treatment Group	Missed Endline n=917	Assessed At Endline n=2,059	p-value
Quantitative Comparison % Correct	Comparison	48 [±4]	48.1 [±2.7]	p=.965
	Indirect	43.4 [±3.7]	37.8 [±2.4]	p=.014*
	Direct	43.4 [±3.4]	41.3 [±2.1]	p=.285
	GALOP	39.8 [±3.9]	39.4 [±2.5]	p=.854
Missing Number % Correct	Comparison	16.5 [±1.7]	17.1 [±1.2]	p=.554
	Indirect	14.7 [±1.7]	13.2 [±.0]	p=.158
	Direct	15.9 [±1.8]	14.4 [±.9]	p=.155
	GALOP	14.8 [±1.9]	13.1 [±1.1]	p=.135
Basic Addition Fluency Correct Items per Minute	Comparison	3.1 [±.5]	2.8 [±.3]	p=.238
	Indirect	2.9 [±.5]	1.7 [±.3]	p<0.001*
	Direct	2.4 [±.5]	2.3 [±.3]	p=.585
	GALOP	2.7 [±.5]	2.4 [±.3]	p=.347
Basic Subtraction Fluency Correct Items per Minute	Comparison	1.2 [±.4]	.9 [±.2]	p=.195
	Indirect	1.2 [±.4]	.8 [±.2]	p=.061
	Direct	1.4 [±.4]	1.1 [±.2]	p=.177
	GALOP	1 [±.3]	1 [±.2]	p=.889
Addition Level 2 % Correct	Comparison	9.2 [±2.7]	5.7 [±1.4]	p=.022*
	Indirect	4.5 [±1.9]	3.4 [±1.1]	p=.342
	Direct	8.7 [±2.8]	5.5 [±1.3]	p=.040*
	GALOP	6.3 [±2.4]	4.4 [±1.3]	p=.175

Table B-2. Baseline EGMA Scores by Attrition Status, by Treatment Group in Ghana

	Treatment Group	Missed Endline n=917	Assessed At Endline n=2,059	p-value
Subtraction Level 2 % Correct	Comparison	2.3 [±1.2]	1.4 [±.7]	p=.192
	Indirect	1.8 [±1.1]	.6 [±.5]	p=.050
	Direct	3.3 [±1.6]	1.1 [±.6]	p=.011*
	GALOP	1.6 [±1.0]	1.6 [±.7]	p=.973

*p<0.05

ANNEX C. INSTRUMENTS

C.1 Overview

Classroom Observation

To understand the learning environment, classroom observation tools were used twice: during the lesson and following its conclusion. The tool used during the observation was a timed instrument that captured what happened every 2 minutes at seven levels of classroom practice (facilitator, student, source [state of play], materials, focus [the organization and the nature of the social interaction], engaged, climate [tone and initiator]). The tool used after the observation was a binary checklist with 29 items organized under six categories (integrated pedagogies, application, responsiveness, resources, efficiency, student autonomy). The post-observation tool also included a materials inventory with four categories (teacher materials, student materials, textbook ratio, library) and a means to capture observer statements.

The constructs and the items that were included in the classroom observation tool were informed by the descriptions of play with an emphasis to include items that could be directly observed. The timed tool is considered low inference as the observers did not make evaluative statements. The post-observation checklist required more inferencing as it captured the presence of a desired behavior at least once during the observation. The materials checklist noted the presence of materials to support mathematics and literacy development. The observer statement was a means to capture an opinion of the observer and potentially was to be used to help further validate understanding of the environment.

Classroom observations were conducted following school-level introductions and after securing teacher consent. Observers sat in a part of the classroom that was unobtrusive so as not to distract students or the instruction. They coded on an A4-sized paper version of the timed instrument so that all seven levels and their codes would be visible simultaneously. Following the observation, the codes written on paper for each two-minute interval were entered into Tangerine®.⁵ The responses for the post-observation tools were entered directly into Tangerine. Item definitions and examples used in training were readily available to observers during observations.

Child Assessment: SEL and Learning Environment Tool

The social and emotional learning (SEL) tool consists of six subtests. It was developed after a review of the SEL literature and is based on existing tools (see below for details). We conducted two rounds of piloting with the SEL tool assessing learning environments and attitudes toward play in Nairobi and Kitui counties in Kenya. The first pilot was with 150 children, 50 teachers, and 100 parents in six schools. Results pointed to the need for more tool adaptation. We made the adjustments then conducted a second round of piloting in Kitui with 75 children, 50 teachers, and 50 parents. After these pilots, we used their results to shorten the subtests by dropping items that performed less well. Below is a summary of what we learned about each assessment.

Child behavior during assessment: In this subtest, the assessor rates students on six behaviors during the SEL assessment. It has been used in previous studies; many items were adapted from the Preschool Self-Regulation Assessment. It had high reliability during

⁵ Tangerine is an RTI-developed open-source data collection software optimized for phones, used for observations, interviews, and assessments in education systems.

the pilots in Kenya and Vietnam and is a helpful measure to triangulate with children's self-ratings of confidence, for example.

Relationships: This subtest assesses whether children have friends they can ask for help in difficult situations or when studying. It was adapted from the International SEL Assessment. It was quick and easy for children to understand and had reasonable reliability.

Solving conflict: This subtest assesses students' response in a conflict situation and their hostile attribution bias (the tendency to attribute a negative motive to another child in the scenario). This subtest was adapted from the New York University/International Rescue Committee Social-Emotional Response and Information Scenarios test (with items added from the International SEL Assessment). A key adaptation from the original version was that we used an open-ended question about children's response to conflict instead of a series of closed-ended questions. The tool had excellent psychometrics and was enjoyable for children.

Confidence paired statements: This subtest involves reading a series of paired statements to students, one of which involves a confident child and the other a non-confident child. The student is asked which child is most similar to them. The subtest was developed by RTI for use in Tanzania. The statements had good reliability in Kenya, and a reduced version had reasonable reliability in Vietnam. We piloted 20 items and selected the best 7.

Creativity: This subtest requires students to suggest alternative uses for ordinary objects (e.g., a spoon). It was based on the Alternative Uses Test, a well-established test created in 1967. The data on this assessment had a good distribution.

Child assessment of learning environment: This scale involved students rating their learning environment on ten items. This scale took items from other assessments, including RTI's school-related gender-based violence classroom climate assessment and a 2012 tool developed for United Nations Children's Fund (UNICEF) child-friendly schools. Children's assessment of their learning environment in school had good reliability around two sets of items: positive aspects of the environment (warm, encouraging teachers) and negative aspects (verbally and physically aggressive teachers).

EGRA and EGMA

For measuring literacy and numeracy, versions of the EGRA and EGMA were adapted. For the EGRA, four subtasks were selected: letter sound identification, nonword reading, oral passage reading, and reading comprehension. The letter sound and nonword subtasks for all evaluations were drawn from existing adaptations of the EGRA in each context. A new oral reading passage and accompanying reading comprehension questions were developed, in lieu of using existing passages. The decision to use a universal passage was spurred by the need for comparability across evaluations (because EGRA passages can vary greatly in their level of difficulty). The newly developed passages and comprehension questions were purposefully designed to be context-agnostic by focusing on issues common to all students. However, these passages and questions underwent translation, back-translation, and review in an effort to ensure that all language was contextually appropriate. All subtasks were administered in local languages of instruction.

For numeracy, we opted to use four of the core EGMA subtasks: level 1 addition, level 1 subtraction, quantity discrimination, and missing number. Instead of creating new items, we used the items from the EGMA Toolkit, which have been rigorously vetted and piloted.

Teacher and Caregiver “Attitudes to LtP” Scale

This tool was based on a method developed by RTI to assess attitudes by asking respondents to endorse one or the other of a pair of contrasting statements. Statements are based on activities and attitudes that facilitate play in classrooms. Initially the same set of statements was piloted with parents and teachers, but responses were substantially different, which led staff to develop two separate scales. The parent scale focuses on attitudes related to child agency and teacher behaviors. The teacher scale focuses on child agency and aspects of the environment and teacher behavior that may facilitate play. Both scales had good reliability after poorly performing items were removed. Although initially developed for teachers, the teacher attitudes to play scale was also conducted with HTs and a nonrepresentative sample of trainers, coaches, district officials, and teacher meeting facilitators. English versions of both scales are in **Annex B**.

Stakeholder Interviews

Stakeholder interviews focus on support provided by the program and supports that these stakeholders offer to teachers and students.

These instruments were designed to collect background information while also measuring attitudes toward play and information about key program supports to a play-enabling environment. These survey instruments were informed by prior RTI work but were significantly adapted to ensure their alignment with and appropriateness for the Play Accelerator evaluation.

C.2 Attitudes Scale: Teachers

Directions: In the following set of questions, I will read you two statements. I want you to tell me which of the two statements you agree with most or which statement you prefer.

For example, let me read two statements to you: Statement 1: I like tea. Statement 2: I like coffee. Which do you agree with? If they agree with both, say “Which do you agree with more? Perhaps you like both tea and coffee, but we want to know which statement you agree with most.”

When answering the following questions, I want you to think of children in Grades 1 through 3.

Table C-1. Attitudes Scale (Teachers)

Statement 1	Statement 2
Children learn best from the teacher.	Children learn best from their peers.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
It is important for the class to be orderly.	It is important for children to be free to talk.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Children learn literacy/language best in small groups.	Children learn literacy/language best as a whole class.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
It is better if children tell the teacher when they don't understand.	If children don't understand, they should try to find the answer independently before asking the teacher.

Table C-1. Attitudes Scale (Teachers)

Statement 1	Statement 2
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
If children answer their own questions they will be more likely to understand.	If the teacher answers children's questions they will be more likely to get the correct answer.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Group work undermines the class's sense of togetherness.	The class does not have to sit together to have a sense of togetherness.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Parents welcome more play-based approaches to learning in the classroom	Parents prefer that I stick to traditional teaching approaches in the classroom
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Children learn best by exploring new things on their own.	Children learn best when new things are explained by the teacher.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Play is an important part of learning for children in primary school.	Play is only important for learning for preschool-aged children.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
Children can choose their own learning activities in the classroom.	Teachers are responsible for choosing the learning activities for his/her students.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
It is more important that the teacher is engaged.	It is more important that students are engaged.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	
A good teacher oversees the whole class at all times.	A good teacher is able to work with one group of students while other students work independently.
1. Agree with 1 much more than 2 2. Agree with 1 and 2 equally 3. Agree with 2 much more than 1	

C.3 Attitudes Scale: Caregivers

Directions: In the following set of questions, I will read you two statements. I want you to tell me which of the two statements you agree with most or which statement you prefer.

For example, let me read two statements to you: Statement 1: I like tea. Statement 2: I like coffee. Which do you agree with? If they agree with both, say "Which do you agree with

more. Perhaps you like both tea and coffee, but we want to know which statement you agree with most.”

When answering the following questions, I want you to think of children in Grades 1 through 3.

Table C-2. Attitudes Scale (Caregivers)	
Statement 1	Statement 2
It is more important to teach creativity than academic subjects.	It is more important to teach academic subjects than creativity.
Teachers should create a friendly classroom environment.	Teachers should create a well-disciplined classroom environment.
Teachers should use play to promote learning in the classroom.	Play should only happen outside of the classroom.
Children learn best when they are having fun.	Children learn best when they are serious.
I would be happy if teachers focus on building my child's confidence and curiosity in the classroom.	I'd prefer teachers focus on academic skills in the classroom.
It is important for the class to be orderly.	It is important for children to be free to talk.

C.4 Timed Classroom Observation

Directions: Using a timing device, at two-minute intervals capture a snapshot of what is happening at each level. The observer codes what is happening at the exact two-minute interval not what happened previously. Only one item per level per interval can be selected.

Table C-3. Timed Classroom Observation Sheet (Minute 2–26)

Lesson Minute	2	4	6	8	10	12	14	16	18	20	22	24	26
Level: Facilitator (teacher)													
Explains concept (X)													
Models (L)													
Affirms (F)													
Question Closed-Ended (QC)													
Question Open-Ended (QO)													
Monitors (R)													
Suggest (S)													
Analyzes (Z)													
Management (M)													
Nothing (N)													
Level: Student													
Reading (R)													
Listening (L)													
Writing (W)													
Orally (O)													
Turn-taking (T)													
Physical (P)													
Manipulate (M)													
Getting (G)													
Not Engaged (N)													
Level: Source (State of Play)													
Following directions (F)													
Planning (P)													
Self-directed (S)													
Level: Materials													

Table C-3. Timed Classroom Observation Sheet (Minute 2–26)

Lesson Minute	2	4	6	8	10	12	14	16	18	20	22	24	26
Book (B)													
Chalkboard (CB)													
Written (W)													
Oral (O)													
Illustration (I)													
Manipulatives (M)													
Physical Space (P)													
Crafts (C)													
Repurposed (R)													
Stations (S)													
Toy (T)													
Nothing (N)													
Level: Focus (pick 2: organization and the nature of the social interaction)													
Whole Class (W)													
Large Group (L)													
Small Group (S)													
Individual (I)													
Corners (C)													
Competition (T)													
Cooperation (P)													
Conflict (F)													
Level: Engaged													
Most (M)													
Half (H)													
Few (F)													
Level: Climate (pick 2: tone and initiator)													
Perseverance (R)													
Positive Verbal (PV)													
Positive Physical (PP)													
Negative Verbal (NV)													
Negative Physical (NP)													

Table C-3. Timed Classroom Observation Sheet (Minute 2–26)

Lesson Minute	2	4	6	8	10	12	14	16	18	20	22	24	26
Ordinary (O)													
Teacher Initiated (T)													
Child Initiated (C)													
Not Applicable (N)													

C.5 Post-Observation Checklist

Directions: This checklist is completed immediately after observing the lesson. Tick *yes* if you observed the behavior during the observation; tick *no* if you did not observe it. Some advanced observers may tick items as they are seen during the observation.

Table C-4. Post-observation Checklist

Integrated Pedagogies		
1	The teacher asked questions or raised topics that got students to respond to each other not to the teacher (i.e., active).	yes no
2	Student(s) persisted with the activity after the rest of the class has started doing something else (i.e., engage).	yes no
3	Student(s) started the task again or adjusted when they thought of a better idea/approach (i.e., iterative).	yes no
4	Student(s) had a chance to make connections between the new experiences/content to their prior knowledge or experiences (i.e., meaningful).	yes no
5	Students problem-solved or had a chance to think critically or reflect about a topic.	yes no
<i>Application: Students have generalized games.</i>		
6	Songs or energizers start or divide activities.	yes no
7	Students played games with obvious rules.	yes no
8	Students had physical play.	yes no
19	Students used objects to represent something else.	yes no
10	Students had symbolic play.	yes no
<i>Responsiveness: The teacher adapts to student behaviors.</i>		
11	The teacher gave students hints, suggestions, or redirection when needed.	yes no
12	The teacher circulated, leaned in, redirected to support students as needed.	yes no
13	The teacher asked the class or a group for a solution to a student question or problem.	yes no
14	The teacher included students who did not volunteer to answer (i.e., inclusion).	yes no

Table C-4. Post-observation Checklist		
Integrated Pedagogies		
<i>Resources: Materials that are needed for implementation are used.</i>		
15	Students used a resource that was dedicated for their use (e.g., book, workbook, slate, tablet).	yes no
16	Students used something to create something else (e.g., art supplies, items in nature, objects repurposed).	yes no
17	Students used a resource to solve a problem (e.g., manipulatives, math charts, dictionary).	yes no
18	Instruction used an item designed for interaction between teacher and students or between students (e.g., read aloud, pocket chart, counters, puzzle).	yes no
19	Students used physical space for learning activities (e.g., counting large steps for measurement, sit in a circle for sharing).	yes no
20	Students used stations/corner work/learning centers.	yes no
<i>Efficiency: Instructional time is maximized.</i>		
21	The teacher had materials ready at the start of an activity or the class.	yes no
22	The pace of the instruction was lively.	yes no
<i>Student-Directed (Ownership/Autonomy)</i>		
23	Students developed and shared ideas with either classmates or the teacher.	yes no
24	Students had a chance to select from choices (state of play: exploring).	yes no
25	Students chose collaborators or roles.	yes no
26	Students made or changed rules for an activity.	yes no
27	Students chose how long to do an activity (state of play: owning).	yes no
28	Students set goals or helped to plan how to use time.	yes no
29	Students chose when to move around.	yes no