EDUCATING THE FUTURE

After School Program

Pilot Evaluation Report

Early Childhood Education
Programa Depois Eskola

PILOT SITES

EPE Alfredo de Jesus Lemos Ermera
EPE Tomas Ximenes Ermera
EPE Lauhata Liquiçá
EPE Saboria Aileu



Reporting period: February – September 2026
Published November 2026

Contents

	The pilot at a glance	4
	Report Information	5
	Acronyms and Key Terms	6
1	Executive Summary	7
2	Background and Context	16
3	Program Design and Objectives	19
4	Target Participants	23
5	Program Governance, Safeguarding and Operating Policies	26
6	Implementation Overview	35
7	Activities Conducted	38
8	Evaluation Design and Methodology	41
9	Strength of Evidence: What the Pilot Can and Cannot Demonstrate	51
10	Participation and Attendance	55
11	Learning, Behavioural and Capacity Outcomes	59
12	Qualitative Findings: Parent and Facilitator Perspectives	68
13	Synthesis and Triangulation of Evidence	74
14	Monitoring, Data Quality and Classroom Observation	79
15	Challenges and Mitigation Measures	87
16	Lessons Learned	91
17	Coordination and Partnerships	94
18	Recommendations	97
19	Conclusion	102
20	References	105
21	Annexes	108



Mixed-methods evaluation of an eight-week, 16-session evidence-based pilot at four pre-schools in Timor-Leste, delivered in two rounds.

ROUND 1 · FEB – JUL 2026
EPE Alfredo de Jesus Lemos
Suku Lihu, Railaco, Ermera

ROUND 1 · FEB – JUL 2026
EPE Tomas Ximenes
Suku Lihu, Railaco, Ermera

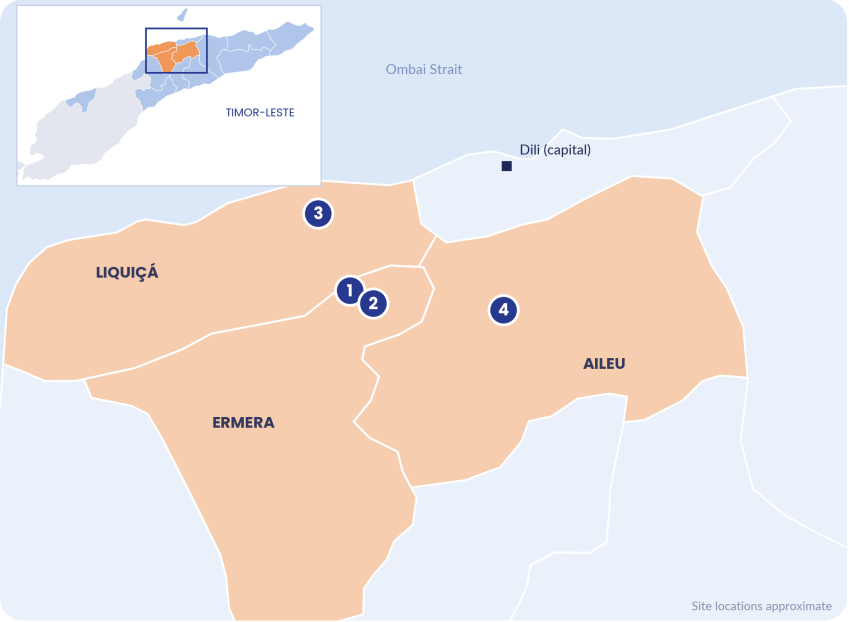
ROUND 2 · JUL – SEP 2026
EPE Lauhata
Liquiçá Municipality

ROUND 2 · JUL – SEP 2026
EPE Saboria
Aileu Municipality

All images in this report follow our child protection and consent process.

THE PILOT AT A GLANCE

Four pre-schools, three municipalities, one after-school model



1. **EPE Alfredo de Jesus Lemos**
   Railaco, Ermera · Round 1
   Feb – Jul 2026 · 41 children enrolled
2. **EPE Tomas Ximenes**
   Railaco, Ermera · Round 1
   Feb – Jul 2026 · 57 children enrolled
3. **EPE Lauhata**
   Liquiçá · Round 2
   Jul – Sep 2026 · 64 children assessed
4. **EPE Saboria**
   Aileu · Round 2
   Jul – Sep 2026 · 43 children assessed

Reach

205
children took part
111 girls (54%) · 94 boys

64
play-based sessions
16 per site over 8 weeks

113
parents & caregivers surveyed at baseline and endline

8
facilitator assessments
local Timorese facilitators

Key outcomes (all four sites, baseline → endline)

35% → 82%
learning-skill checks done independently

26% → 86%
health & hygiene practices seen often or always

23% → 73%
socio-emotional behaviours seen often or always

Pre/post measurement without a comparison group: results show measured change, not proven impact. See Section 9.

Report Information

Field	Detail
Program title	Early Childhood Education After School Program (Programa Depois Eskola)
Document title	After School Program Pilot Evaluation Report
Document type	Mixed-methods pilot evaluation: single-group pre/post measurement with qualitative enquiry at the two first-round sites; quantitative pre/post measurement only at the two second-round sites. Not an impact evaluation.
Project location	First-round sites: Ermera Municipality - Railaco Administrative Post, Suku Lihu (Aldeia Hi and neighbouring aldeias). Second-round sites: EPE Lauhata and EPE Saboria (administrative location not recorded in the datasets supplied; to be confirmed by ETF)
Implementation sites	First-round sites: EPE Alfredo de Jesus Lemos; EPE Tomas Ximenes. Second-round sites: EPE Lauhata; EPE Saboria
Reporting period	February - September 2026: first-round sites February - July 2026; second-round sites July - September 2026 (qualitative data collection August 2026, first-round sites only)
Implementing organisation	ETF - Educating The Future Australia (ETF Timor-Leste team)
Quantitative data collection	ETF Timor-Leste team: child baseline and endline assessments, facilitator pre/post knowledge tests, classroom observations, attendance registers, parent confidence surveys, enrolment records (first-round sites); child baseline and endline assessments, facilitator pre/post knowledge tests, attendance registers and parent confidence surveys only (second-round sites)
Qualitative data collection	Independent consultant team: parent and facilitator focus group discussions, key informant interviews, and child success-story interviews - at the two first-round sites only. No FGDs or KIIs were conducted at the second-round sites.
Analysis and drafting	Adolfo Silva, Lucia G. de Araujo & Natalina dos Santos | Consultants
Institutional counterpart	Ministry of Education, Timor-Leste (Municipal and central)
Date of this revision	October 2026 - revised to integrate data from EPE Lauhata and EPE Saboria

Acronyms and Key Terms

Tetun-language terms used in the assessment instruments are reproduced here as they appear in the original data-collection tools.

Term	Meaning
ASP	After School Program (Programa Depois Eskola)
AdJL	EPE Alfredo de Jesus Lemos (abbreviation used in tables)
CFM	Child Functioning Module (UNICEF / Washington Group disability identification tool)
DNEPE	Direção Nacional de Ensino Pré-Escolar - National Directorate of Pre-School Education
ECE / ECD	Early Childhood Education / Early Childhood Development
EPE	Edukasaun Pre-Eskolár - pre-school / early childhood education centre
ETF	Educating The Future - implementing organisation
FGD / KII	Focus Group Discussion / Key Informant Interview
IDELA	International Development and Early Learning Assessment (Save the Children)
LAFAEK	Nationally distributed Tetun language children's learning magazine used as a program material
LH	EPE Lauhata - second-round site (abbreviation used in tables)
MELQO	Measuring Early Learning Quality and Outcomes (UNESCO / UNICEF / World Bank)
MoE	Ministry of Education, Timor-Leste
First-round sites / Second-round sites	First-round sites: EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes (February-July 2026). Second-round sites: EPE Lauhata and EPE Saboria (July-September 2026) All four are pilot sites.
M&E	Monitoring and Evaluation
PSEAH	Prevention of Sexual Exploitation, Abuse and Harassment
SB	EPE Saboria - second-round site (abbreviation used in tables; spelled "Saboaria" in some file names)
ToT	Training of Trainers
TX	EPE Tomas Ximenes (abbreviation used in tables)
Liña Baze / Liña Ikus	Baseline / Endline (Tetun)
Sin	Able to perform the skill independently (Tetun rating)
Ho ajuda	Able to perform the skill with help (Tetun rating)
Lae	Not yet able to perform the skill (Tetun rating)
Sempre / Bebeik / Dala-ruma / Nunka	Always / Often / Sometimes / Never (Tetun frequency scale)
Suku / Aldeia	Village / hamlet - sub-municipal administrative units
Dona da casa / Agricultor	Homemaker / farmer (occupations recorded at enrolment)



01

SECTION

Executive Summary

1.1 Purpose and scope

The Early Childhood Education (ECE) After School Program (Programa Depois Eskola) was designed and piloted by Educating The Future Australia (ETF) between February and July 2026 at four public pre-schools (EPE) in Ermera Municipality, Railaco Administrative Post, Suku Lihu: EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes, and Aileu Municipality, Saboria, and Liquiçá Municipality in Lauhata. The program supported children aged approximately 3-9 years (with a core target of 3-7 years, and extended to children 9 years of age given the limited structured opportunities after school) through structured, play-based activities in literacy, numeracy, problem-solving, socio-emotional learning, and health and hygiene, delivered by trained local facilitators over **eight weeks**, with **two sessions per week at each site**, resulting in **16 sessions per EPE and 32 sessions across the two sites**. The pilot was designed to provide children with additional opportunities to strengthen foundational learning and developmental skills through engaging, child-centred activities beyond regular preschool hours. The implementation of this pilot was supported by the Swiss Embassy to Timor-Leste under their small-grants program.

Educating The Future is an Australian founded not-for-profit providing pre-school education in Timor-Leste to over 3000 children since 2015 and is a full member of the Australian Council for International Development (ACFID) since 2020. The organization has built and supports eight public pre-schools and runs Timor's first evidence-based After School Program improving educational, social-emotional and developmental outcomes for youth and communities. The organization also founded the Timor-Leste Australia Youth Dialogue providing a regional forum for youth-led policy action. Educating The Future is run completely by volunteers in Australia with paid staff in Timor-Leste. The Timorese organization is part of the Technical Working Group for Pre-School Education and is a recognized Ministry of Education partner with an active Memorandum of Understanding since 2019.

This report evaluates the pilot. It draws on two strands of primary evidence: quantitative measurement of children's skills, attendance, facilitator capacity and parental confidence, collected by the ETF Timor-Leste team; and qualitative interviews and focus group discussions with parents and facilitators, conducted by an independent consultant team in August 2026. A third strand, individual case studies, is published separately as a communications product and is treated here as illustrative material rather than as evidence.

This revised edition integrates data from the two second-round pilot pre-schools, EPE Lauhata and EPE Saboria, where the same 16-session model was delivered between July and September 2026. For these two sites the evaluation holds child baseline and endline assessments, facilitator pre/post knowledge tests, attendance registers and parent confidence surveys. No focus group discussions, key informant interviews or classroom observations were conducted there, so findings for the second-round sites rest on quantitative and documentary evidence only. Throughout the report, combined figures are recalculated from the underlying counts across all four sites, and findings that draw on the qualitative round or on classroom observation are identified as applying to the two first-round sites only. All four schools are pilot sites: this report refers to EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes (February-July 2026) as the first-round sites and to EPE Lauhata and EPE Saboria (July-September 2026) as the second-round sites.

1.2 Methods

Method

The pilot used a single group, pre-test/post-test design with no comparison group, no randomisation, no follow-up measurement, and incomplete matching of all individual children between rounds. The assessment instruments were developed internally by Educating The Future and have not been externally validated or psychometrically tested. All child assessments were conducted by the implementing team rather than by independent assessors. The two second-round sites used the same design and instruments, without the qualitative round.

This design can establish that children's measured performance changed over the pilot period, and it can describe the direction and approximate size of that change. It cannot establish how much of the change was caused by the program as opposed to ordinary schooling, maturation, repeated exposure to the same test, or assessor expectation. Readers, including the Ministry and prospective funders, should treat the numbers in this report as promising programmatic evidence rather than as impact estimates.

1.3 Scale of the pilot

A total of 205 children participated across the four sites: 98 formally enrolled at the two first-round sites (41 at EPE Alfredo de Jesus Lemos and 57 at EPE Tomas Ximenes), of whom 55 (56%) were girls[1], and 107 recorded in the baseline and endline assessment records of the two second-round sites (64 at EPE Lauhata and 43 at EPE Saboria), of whom 56 (52%) were girls. Across the four sites, 111 of the 205 children (54%) were girls. The second-round figure counts assessed children rather than enrolments, because no usable enrolment list exists for EPE Saboria and the EPE Lauhata list does not reconcile with its assessment records (Section 8.8). In total, 174 children completed baseline child assessments and 176 completed endline assessments (67 and 69 at the first-round sites; 107 at each round at the second-round sites), and 113 parents/caregivers participated in baseline and endline confidence surveys (34 at the first-round sites and 79 at the second-round sites). There was anecdotal evidence of children with disabilities participating in this program, however, was not recorded. Four facilitators (two per site) were trained at the first-round sites and assessed before and after training and observed throughout the pilot; at the second-round sites, four further facilitator assessments were recorded, drawing partly on the same people. The qualitative round, conducted at the two first-round sites only, added 25 further respondents across parent interviews, parent FGDs, facilitator FGDs and facilitator interviews. No incentives were provided for participation in the program and/or measurement of the pilot.

1.4 RESULTS

By the end, children did more on their own

35% → 82%

of skill checks done **independently**. A skill check is one child attempting one of ten skills, such as counting from 1 to 10.

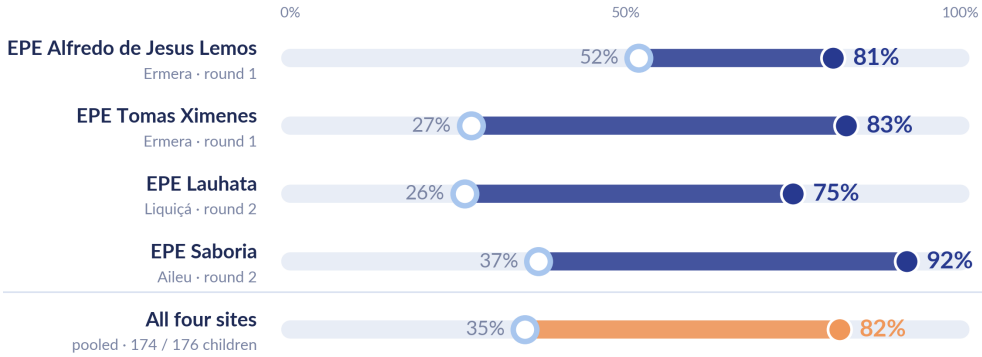
33% → 1%

of skill checks rated **"not yet able"**: for example, a child who could not count from 1 to 10 or retell a short story, even with help.

100 of 107

children at the second-round sites, assessed twice, did more skills on their own at endline.

SKILLS PERFORMED INDEPENDENTLY, BY SITE ○ baseline ● endline



ALL FOUR SITES

Early literacy Retells a simple story	20% → **82%**	Letter recognition 27% → 67%; letter sounds 21% → 73%.
Numeracy Counts 1–10 without help	22% → **68%**	"More" and "less" 40% → 91%; shapes 21% → 70%. Counting is the weakest skill.
Problem-solving Builds structures with blocks	48% → **97%**	Puzzles 42% → 91%; trying several ways to solve a problem 43% → 89%.
Health & hygiene Practices seen often or always	26% → **86%**	Handwashing, healthy-food recognition and movement skills.

FIRST-ROUND SITES EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes

Social & emotional Behaviours seen often or always	AdJL 35% → **90%** TX 31% → **48%**	The smaller gain at EPE Tomas Ximenes coincides with many children joining after baseline.
Parental confidence Mean score out of 5, EPE Alfredo de Jesus Lemos	3.3 → **5.0**	Not measurable at EPE Tomas Ximenes: the endline survey repeated the baseline records.
Facilitator capacity Knowledge test score out of 6	2–3 → **6**	Three facilitators with complete records reached full marks. Confidence applying it in class is still developing.

Source: child assessments, observation records, parent surveys and facilitator tests (Annexes B–D). Before-and-after comparison without a comparison group; see 1.2 Methods.

Measured performance rose substantially at all four sites. Across the ten learning-skill indicators covering literacy, numeracy and problem-solving, the share of skill checks performed independently rose from 52% at baseline to 81% at endline at EPE Alfredo de Jesus Lemos, from 27% to 83% at EPE Tomas Ximenes, from 26% to 75% at EPE Lauhata and from 37% to 92% at EPE Saboria. Recalculated from the underlying counts across all four sites (1,740 baseline and 1,759 endline skill checks), the share rose from 35% to 82%.

Ratings of "not yet able" (Lae), which accounted for 29% and 56% of baseline checks at the two first-round sites and 25% and 35% at the two second-round sites, fell from 33% to 1% of all checks. They were absent from the endline records at three of the four sites, where every skill check at endline was rated at least "with help"[2]. At EPE Lauhata, 18 endline checks (3%), recorded for 13 of its 64 children, were still rated "not yet able".

Health and hygiene practices - handwashing, healthy-food recognition and gross-motor coordination - were observed "often or always" in 100% of endline observations at both first-round sites, up from roughly 64-66% at baseline; at the second-round sites they rose from 0% to 63% (EPE Lauhata) and from 5% to 98% (EPE Saboria), and across all four sites from about 26% to 86%. Facilitator knowledge scores at the first-round sites rose from 2-3 out of 6 before training to 6 out of 6 after training for the three facilitators with complete records[3]; at the second-round sites scores rose by one to three points but no facilitator reached the maximum (post-training scores of 3 to 5 out of 6). Parental confidence rose at every site where it could be measured: at EPE Alfredo de Jesus Lemos from a mean of 3.3 to 5.0 on a five-point scale[4]. At EPE Lauhata and EPE Saboria the share of top ratings rose from 50% to 91% and from 36% to 85% respectively.

At the two first-round sites, the qualitative round confirmed the direction of these findings from the household side. All three parents interviewed individually judged the program beneficial, despite most children having attended for under three months at the time of interview. Parents independently described children studying without prompting, becoming more creative, teaching newly acquired skills to siblings and to parents themselves, and shifting away from passive screen time toward hands-on making, drawing and planting. Both parents and facilitators reported that children feel safe in the program, and no child protection concern was raised by any respondent. No equivalent qualitative evidence exists for the two second-round sites, where no FGDs or KIIs were conducted; their results rest on the assessment, attendance and survey records alone.

At the first-round sites, classroom observation carried out by ETF program staff supports the account of content delivery: every item in the four knowledge-development domains was rated as met at all four observed visits. The weaknesses the observer recorded - late starts, unsupervised transitions, older children dominating younger ones, and self-regulation are the same weaknesses the qualitative round surfaced four months later. No classroom observation records exist for the second-round sites.

The second-round results should be read with particular care. They come from the program's first fully matched cohorts - all 107 children were assessed at both rounds, and 100 of them (93%) performed more skills independently at endline than at baseline - but they rest on a single evidence strand, and the records contain patterns that require verification: an identical rating profile for 16 of 43 baseline children at EPE Saboria, a near-ceiling EPE Saboria endline, and an abrupt break between the first 40 and the last 24 EPE Lauhata endline records (Section 8.8). Recalculating the pooled result without the questionable records changes the endline figure by no more than three percentage points (79-83%), so the overall direction does not depend on them (Section 11.1).

1.5 Main challenges

Attendance was variable at all four sites, averaging 25 children per session at EPE Alfredo de Jesus Lemos and 22 at EPE Tomas Ximenes against enrolments of 41 and 57[5], and 28.5 at EPE Lauhata and 21.6 at EPE Saboria against 64 and 43 assessed children. The two first-round sites moved in opposite directions: EPE Tomas Ximenes climbed strongly in the second half of the cycle while EPE Alfredo de Jesus Lemos suffered a marked mid-May dip. At the first-round sites, the qualitative round identified the drivers: rain and difficult terrain including large river crossings, child tiredness, family circumstances and illness, and, most consistently cited of all, an unstable session schedule including a mid-cycle shift to morning delivery that excluded children available only in the afternoon. At the second-round sites, attendance was lower in the second half of the cycle than in the first, and EPE Saboria sessions paused for three weeks between sessions 4 and 5; the reasons are not documented, because no qualitative data were collected there.

Scheduling was the single most frequently raised operational problem across every respondent group at the first-round sites. Facilitators reported that afternoon start times left children tired and inattentive, while parents in some locations actively opposed afternoon sessions because of competing household commitments. This is not a minor logistical detail: it determines who can attend and how much they absorb when they do.

Facilitator capacity is the second constraint. Although post-training knowledge scores reached the maximum, facilitators themselves reported limited confidence in applying the guidelines, no prior experience managing groups of 3-7 year-olds, difficulty managing single rooms holding up to 39 children across a five-year age span, and incomplete Training of Trainers coverage. The four designated learning "corners"[6] were rarely implemented. The pilot achieved knowledge transfer but not skills transfer - a pattern well documented in the professional development literature and addressable through coaching rather than through more content. At the second-round sites, post-training knowledge scores of 3 to 5 out of 6 - including 3 out of 6 for the two facilitators who had scored the maximum after the first-round training - reinforce the conclusion that a single knowledge test is not a stable measure of capability.

1.6 Limitations of this evaluation

The limitations below are material to how the findings should be read and used. They are set out in full in Section 14 and summarised here because several bear directly on the headline numbers.

- **No counterfactual.** There is no comparison group and no randomisation, so no part of the measured gain can be attributed to the program with confidence. Children were also attending school throughout the period, and children of this age develop rapidly regardless of intervention.
- **Instruments not validated.** The child observation tool, facilitator checklist and parent survey were developed by consultant and may not been externally validated or checked for inter-rater reliability. No pilot-testing or cognitive-testing record was available to the evaluation.
- **Assessors were not independent.** Baseline and endline child assessments, and the facilitator pre/post tests, were administered by the ETF Timor-Leste team, who also delivered the program and trained the facilitators. This is common and often unavoidable in pilots of this size, but it creates a recognised risk of expectancy effects that independent assessment would remove. The same applies at the second-round sites.
- **First-round cohorts are not matched at the first two sites.** At the first-round sites, only around 20-22 children per site could be matched by name between baseline and endline. The comparison is therefore between two overlapping groups, not the same children measured twice. At EPE Tomas Ximenes the endline group contained

around 21 children who joined after baseline undertake. At the two second-round sites, by contrast, all 107 children (64 at EPE Lauhata and 43 at EPE Saboria) were assessed at both rounds and matched by name, so for these sites the comparison is of the same children measured twice.

- **Scales changed between rounds.** The EPE Alfredo de Jesus Lemos endline recorded socio-emotional items on a 1-4 numeric scale while the baseline used descriptive Tetun categories, requiring harmonisation during analysis and introducing interpretation risk.
- **Records are missing or duplicated.** One facilitator post-test was not recorded, and the EPE Tomas Ximenes parent endline survey duplicates the baseline entries, including the survey date. No endline change in parental confidence can be reported for that site. At the second-round sites, the EPE Saboria enrolment list is a copy of the EPE Lauhata list, nine EPE Lauhata attendance sheets are undated copies of an earlier sheet, and the EPE Saboria parent endline survey carries auto-filled dates running into November 2026 (Section 8.8).
- **Fidelity was not measured.** All 16 sessions were delivered (at each first-round site), but a designed component of the model - the learning corners - went largely unimplemented for a full cycle without detection. Delivery was complete in count but partial in content.
- **The near-elimination of "not yet able" ratings should be read with care.** A floor-to-ceiling movement on a three-point scale administered by the implementing team is exactly the pattern that measurement artefacts produce. It is consistent with real learning, and the qualitative and observational evidence supports that reading, but on its own it is not strong evidence. The same caution applies to the near-uniform EPE Saboria endline, where 22 of 43 children were rated independent on all ten skills.
- **Qualitative sample is small and purposive.** Twenty-five respondents, selected purposively, interviewed after the cycle ended. Findings indicate the range and strength of views held rather than their prevalence, and are subject to social-desirability effects among respondents who want the program to continue. No qualitative data were collected at the two second-round sites, so findings for EPE Lauhata and EPE Saboria rely on quantitative and documentary evidence only and cannot be corroborated or explained by respondents' accounts.
- **Second-round data quality.** The EPE Lauhata and EPE Saboria records contain patterns that require verification before the results are used externally: 16 of 43 EPE Saboria baseline records share an identical and developmentally implausible rating profile; the first 40 EPE Lauhata endline records rate every socio-emotional and health item "often or always" while the remaining 24 almost never do; and attendance for ten EPE Lauhata sessions can only be taken from hand-written session totals. These sites' results are reported in full, but with lower confidence than the first-round results.
- **Financial actuals are incomplete.** Expenditure is documented only for the program-materials category (US$3,349). Budget utilisation for personnel and operations - 63% of the revised US$12,418 pilot budget - cannot be confirmed, thus, all unit costs are stated on a budget rather than an expenditure basis. A further US$1,304 of the direct program cost line in the revised costing is not itemised and has no expenditure record. No financial data were provided for the second-round sites.

1.7 Conclusion of the Evaluation

The pilot delivered what it set out to deliver, and the three independent lines of evidence available - measured assessment, classroom observation and the accounts of parents and facilitators - point in the same direction. That convergence is the most persuasive feature of the evidence base, and it is stronger than any single strand. Children at the two first-round sites who began with very little exposure to structured early learning finished the cycle able to attempt every core skill, in two communities, with cohorts of different ages and starting points, using low-cost

materials and locally recruited facilitators working in Tetun. The two second-round sites show the same direction and a similar size of measured change in larger, fully matched cohorts, but that evidence comes from one strand only - implementer-administered assessment - and cannot be triangulated in the same way.

What the pilot has not established is how much of that change the program caused, whether it persists after the cycle ends, or how the model performs beyond the sites measured. The second-round sites extend the quantitative record but, without qualitative or observational evidence and with unresolved data-quality questions, they do not settle the question of transferability. These are not peripheral questions; they are the questions the Ministry and prospective funders will ask first, and the pilot design cannot answer them. The barriers the evaluation did identify scheduling, staffing ratios, communication, materials and data systems are operational rather than doubts about the model's value, and every respondent group affirmed that value.

The evaluation therefore supports a further, better-measured cycle rather than immediate scale-up on the strength of these numbers alone. The priority actions below are the conditions under which the next cycle could produce evidence robust enough to carry a scale-up argument.

1.8 Outlook

Priority actions to support the program's expansion

- Fix the schedule first. Agree a stable, jointly negotiated timetable per site with schools and parents before enrolment opens, and hold it for the full cycle.
- Resource the facilitator ratio. Move toward roughly 1:15 within the 3-7 band, and group children by age rather than teaching a five-year span in one room.
- Complete facilitator preparation. Full ToT coverage for every facilitator, refresher training on session content, plus child protection and classroom-management modules, and scheduled coaching visits rather than episodic support.
- Introduce a unique child ID at enrolment, used on every instrument, to enable genuine matched-cohort measurement.
- Strengthen the instruments. Version-lock the tools, adopt a single rating scale per instrument across rounds, add an observed-practice rubric for facilitators, and benchmark the child tool against an established instrument such as IDELA (International Development and Early Learning Assessment) or the MELQO (Measuring Early Learning Quality and Outcome) module.
- Introduce independence into measurement. Have at least the endline assessment conducted by assessors who did not deliver the program, and double score a sample to produce an inter-rater reliability figure.
- Build an aldeia-based parent communication system with phone numbers collected by hamlet, ending reliance on children to relay messages.
- Extend dosage from 16 sessions toward 24-32, or run a full six-month cycle, to consolidate the literacy and counting skills that remain "with help" for a substantial minority.
- Document the core policies. Formalise and evidence the safeguarding, disability inclusion and facilitator recruitment policies described in Section 5.
- **Close the evidence gap at the second-round sites.** Conduct parent FGDs and facilitator KIIs at EPE Lauhata and EPE Saboria, and verify the anomalies in their records, before their results are used externally; give every future site the full evaluation package.

Notes

1. Annex E – Enrolment/registration lists (first-round sites).
2. Annex B – Child baseline and Endline assesment.
3. Annex C – Facilitators pre/post training assessment record.
4. Annex D – Parents /caregiver survey dataset.
5. Annex A – Attendance registers (all four sites – 16 sessions each)
6. Learning Corners – refers to designated areas within a classroom that are set up to encourage active and independent learning. ASP has four learning areas/station: homework, reading, language and technology (see ASP curriculum).

02

SECTION

Background and Context

2.1 Why early learning matters, and what the evidence says

The case for investing in early childhood is among the better-evidenced propositions in development economics and developmental science. Heckman's work on skill formation established that the returns to human-capital investment are highest in the earliest years, because early skills make later skills cheaper to acquire - the dynamic complementarity that underpins the economic argument for early intervention (Heckman, 2006; Cunha & Heckman, 2007). The Lancet series on early childhood development has repeatedly documented both the scale of lost developmental potential in low-and middle-income countries and the effectiveness of interventions that provide stimulation and early learning opportunities (Grantham-McGregor et al., 2007; Engle et al., 2011; Black et al., 2017; Britto et al., 2017). A meta analysis of early childhood interventions across 23 countries found consistent positive effects on cognitive outcomes, with the larger effects associated with interventions that included a direct educational component rather than nutrition or cash alone (Nores & Barnett, 2010).

For the East Asia and Pacific region specifically, Rao and colleagues have shown that participation in early childhood education is associated with improved cognitive outcomes for young children in low- and middle-income countries, and have developed regionally validated measurement scales for the purpose (Rao et al., 2019). The relevance for this pilot is direct: the mechanism the After School Program relies on structured, adult-facilitated early learning opportunities for children who would otherwise have none is the mechanism this literature identifies as effective.

The evidence on out-of-school-time programs specifically is more qualified, and it is worth stating honestly. Meta-analyses of after-school programs find positive but modest average effects, with wide variation between programs (Lauer et al., 2006). The most useful finding for program design comes from Durlak, Weissberg and Pachan (2010), who found that after-school programs improved children's self-perceptions, school bonding, positive social behaviour and academic achievement - but that these gains were concentrated in programs using four practices, abbreviated as SAFE: a sequenced set of activities, active rather than passive learning, a focused allocation of time to skill development, and explicit definition of the skills being taught. Programs lacking these features showed no significant effects on most outcomes. This is a design standard the After School Program should be measured against, and Section 13 returns to it.

The play-based approach the program uses also has a specific evidence base. Guided play - child-directed activity within an adult-structured learning goal - produces learning gains equal to or greater than direct instruction on several early academic outcomes, including geometric knowledge and vocabulary, while sustaining engagement better (Weisberg, Hirsh-Pasek & Golinkoff, 2013; Fisher et al., 2013; Zosh et al., 2017). The pilot's own item-level results are consistent with this: the most game-like indicators (sorting, block construction, flexible problem-solving) reached 96-100% independent performance, while the most instruction-dependent (counting without help, letter recognition) remained the weakest. The second-round sites show the same broad ordering (Section 11.1).

2.2 The Timor-Leste context

Timor-Leste has expanded preschool access substantially but remains far short of its own targets. The country-led evaluation of the preschool sub-sector commissioned by the Ministry of Education supported by MFAT through HANDS Program found a gross preschool enrolment rate of 31,5% as of 2025, against national targets of 60% by 2028 and 80% by 2030, and identified underfunding, a shortage of qualified teachers and fragmented policy as the principal constraints (HANDS Report, 2025). HANDS has documented that access is weakest in rural areas, where distance and terrain are direct barriers to attendance.

The pilot communities sit at the difficult end of that distribution. Only 10 of 41 enrolled children (24%) at EPE Alfredo de Jesus Lemos and 8 of 57 (14%) at EPE Tomas Ximenes had any previous preschool experience according to enrolment records. Baseline assessment confirmed the consequence: at EPE Tomas Ximenes, 56% of all learning-skill checks were rated "not yet able", including more than two-thirds of children unable to recognise ten letters, count to ten unaided, or identify basic shapes without support. Baseline assessment at the two second-round sites was similarly low: 25% of learning-skill checks at EPE Lauhata and 35% at EPE Saboria were rated "not yet able". Prior preschool experience was not recorded for these sites.

There is also directly comparable programming in the same municipality. UNICEF has supported community-based alternative pre-schools in Ermera and Viqueque since 2015, established specifically in aldeias with no existing pre-school provision, and recent research has examined developmental outcomes for children attending them. ETF should engage with that body of work during the next cycle: it offers both a methodological reference point and, potentially, a comparison population for a more rigorous evaluation design.

2.3 Context of implementation

- **Geographic coverage:** two pilot EPE sites in Suku Lihu, Railaco Administrative Post, Ermera Municipality, drawing children from aldeias including Hi, Camalpun/Cmalpu, Camalrahei and Rai-udo. The two second-round sites, EPE Lauhata and EPE Saboria, delivered the model between July and September 2026; their suku and municipality are not recorded in the datasets supplied to the evaluation.
- **School and community conditions:** children largely walk to the program (30 of 39 with recorded transport at EPE Alfredo de Jesus Lemos; 37 of 57 at EPE Tomas Ximenes); a minority are transported by parents or caregivers. Parents in the FGD identified difficult terrain, including the crossing rivers, as a barrier during the wet season.
- **Socio-economic context:** the majority of mothers and caregivers are homemakers (dona da casa) and fathers predominantly farmers (agricultor), with a small number of civil servants, teachers and health workers, indicating largely subsistence-agriculture households. Household labour demands, particularly in the agricultural calendar, shape when parents can release children for activities.
- **Connectivity:** internet access among parents is limited, which constrains digital communication channels and has direct implications for how the program schedules, announces and adjusts sessions. Word of mouth was largely used to disseminate information about the program, including through teachers and children.
- **Language:** Tetun is the home language of nearly all enrolled children, with a substantial minority speaking Tetun and Mambae at home - supporting delivery of all activities and assessment tools in Tetun. Local language should be prioritized in sites where Tetum is not the dominant/preferred language, and this is emphasized in the program design and facilitator guidelines.
- **Stakeholders:** Foundation Educating The Future registered in Timor-Leste as implementer with Educating The Future Australia as the development partner; school directors and teachers of the four EPEs; five facilitators named in the test records across the four sites, with others named on some register sheets; parents and caregivers; suku and aldeia leadership; and the Ministry of Education as institutional counterpart.
- **Coverage of this section:** the contextual information above on transport, household occupation, connectivity and language is drawn from enrolment records and the qualitative round at the two first-round sites. Equivalent information was not supplied for EPE Lauhata or EPE Saboria and is not assumed.



03

SECTION

Program Design and Objectives

3.1 Program rationale

The After School Program was created to close the early learning gap in communities where the formal preschool network does not reach, by providing a safe, playful, structured learning environment after regular school hours, delivered by trained community facilitators using low-cost play-based methods, and by strengthening parents' confidence to support learning at home.

The qualitative round confirmed that families understood the program in exactly these terms. Parents explained their enrolment decisions in two ways: the site was close to home, and the program represented structured learning rather than the unstructured time that would otherwise fill the after-school hours. One family described the alternative as long idle periods at home during the dry season, or accompanying parents to agricultural work during the rains. This matters for expansion messaging: families were not primarily seeking childcare, they were seeking structured learning.

3.2 General objective

To improve foundational learning and holistic development of children aged 3-7 years through structured, play-based after-school activities delivered in their own communities.

3.3 Specific Objectives

- **Early literacy:** Nearly 70% of Grade 1 students failed national assessments in 2021, and few homes have books. The ASP extends preschool curriculum themes into the afternoon using Timorese stories, backed by evidence that ASPs improve literacy (Durlak et al., 2010).
- **Numeracy:** Almost all Grade 4 students failed assessments in 2020–2021, and stunting (over 47% in some areas) limits cognitive growth. Preschool Inspectors named numeracy a top priority, and ASP programs strengthen the executive functions that underpin early maths.
- **Problem-solving:** Over 70% of children aged 3–7 spend afternoons unsupervised, with little access to toys or enrichment. ASP through construction and discovery activities builds problem-solving and creativity.
- **Social-emotional development:** Without regular peer interaction, children can regress socially and emotionally. The ASP are shown to improve cooperation, self-control and confidence.
- **Parental engagement:** Parents face limited time and awareness of early learning. ASP Increase parental engagement and confidence in supporting home learning.
- **Facilitator capacity:** Most community facilitators have no minimal formal ECE qualification. Training is often shorterm with no follow up mentoring. ASP was evidently create an opportunity to build the local capacity to trive ECE with local materials and resources.
- **Overall fit with public preschool:** The ASP complements rather than duplicates preschool. It uses preschool facilities in the afternoon, follows the national curriculum themes, and aligns with the National Preschool Policy (2014) and the Preschool Network Plan (Diploma Ministerial N.º 64/2024).

3.4 Achievement of objectives at a glance

The table below records the evaluation's verdict against each objective. "Achieved" here means that the available measurement moved in the intended direction and that the change is confirmed by at least one other evidence strand. It does not mean that the program has been shown to have caused the change; Section 9 sets out the limits of attribution. Detailed evidence appears in Sections 10 to 12. For the second-round sites, where no second evidence strand exists, the table records measured change only and does not apply the "Achieved" label.

Table 3.1. Status against each specific objective.

Specific objective	Status	Principal evidence	Confidence
Early literacy skills	**Achieved (first-round sites); measured gains at LH and SB, uncorroborated**	Letter recognition, letter sounds and story retelling all rose at all four sites; story retelling reached 85% (AdJL), 93% (TX), 66% (LH) and 95% (SB) independent at endline.	Moderate at first-round sites - corroborated by observation and parent report. Low at second-round sites - assessment only
Numeracy competencies	**Largely achieved (first-round sites); measured gains at LH and SB, uncorroborated**	Shapes and more/less comparison rose sharply at all four sites; counting 1-10 remains the weakest indicator at TX (49% independent at endline) and among the weakest at LH (66%), against 91% at SB.	Moderate (first-round sites); low (second-round sites)
Problem-solving skills	**Achieved (first-round sites); measured gains at LH and SB, uncorroborated**	Sorting, block construction and flexible problem-solving reached 96-100% independent at endline at both first-round sites and 77-98% at the second-round sites.	Moderate, but note ceiling effects on a three-point scale
Social and emotional development	**Achieved at EPE Alfredo; partial at EPE Tomas. Measured gains at LH and SB, uncorroborated**	"Often or always" ratings rose 35% → 90% at AdJL, 31% → 48% at TX (where the endline cohort included many newly joined children), 13% → 64% at LH and 21% → 99.5% at SB; about 23% → 73% across all four sites.	Low to moderate - scale changed between rounds at AdJL; uniform rating blocks at LH and SB
Health, hygiene and physical development	**Achieved (first-round sites). Measured gains at SB and, partly, LH; uncorroborated**	100% of endline observations rated "often or always" at both first-round sites (from 64-66%) and 98% at SB (from 5%); 63% at LH (from 0%). Classroom observation corroborates the change at the first-round sites only.	Moderate to high (first-round sites); low (second-round sites)
Parental engagement and confidence	**Achieved at EPE Alfredo; not measurable at EPE Tomas. Self-reported gains at LH and SB, uncorroborated**	Mean confidence 3.3 → 5.0 at AdJL; share of top ratings 50% → 91% at LH and 36% → 85% at SB. TX endline survey duplicates baseline records. Qualitative evidence of engagement is strong at the first-round sites; none was collected at the second-round sites.	Low - one site unmeasured, self-report, only the scale end-points used at LH and SB, unverifiable SB endline dates
Facilitator capacity	**Partially achieved**	Knowledge scores reached 6/6 for three of four first-round facilitators, and classroom practice visibly changed. At the second-round sites post-training scores were 3-5 out of 6, including 3/6 for the two first-round facilitators retested at SB. Pilot facilitators report low confidence applying guidelines, gaps in ToT coverage, and unmet needs in group management.	Knowledge: moderate. Applied practice: not achieved

The single objective where measured performance and self-reported experience diverge is facilitator capacity. That divergence is examined in Section 13.2 and drives one of the report's central recommendations.



04

SECTION

Target Participants

4.1 Target participant numbers

Category	EPE Alfredo de Jesus Lemos	EPE Tomas Ximenes	EPE Lauhata	EPE Saboria	Total	Female	Male
Children enrolled (first round) / assessed (second round)†	41	57	64	43	**205**	111	94
Children assessed at baseline	39	28	64	43	**174**	93	81
Children assessed at endline	26	43	64	43	**176**	91	85
Children matched across both rounds	~20	~22	64	43	**~149**	-	-
Parents/caregivers surveyed	17	17	45	34	**113**	-	-
Facilitators trained / assessed‡	2	2	2	2	**5 people**	-	-

The first row counts enrolments at the first-round sites but children assessed at both rounds at the second-round sites, because the EPE Saboria enrolment sheet is a copy of the EPE Lauhata list and the EPE Lauhata list (49 entries, including apparent duplicates) does not reconcile with the 64 children assessed there. The 205 total is therefore a conservative count of participating children: further names appear in the second-round attendance registers but cannot be reliably reconciled. Sex for the second-round sites is as recorded in each round (it differs between rounds for three children). ‡ Eight facilitator assessments were recorded, two per site, covering five individually named facilitators: Juliandina and Mario (EPE Alfredo de Jesus Lemos) also delivered at EPE Saboria, and Vitri (EPE Tomas Ximenes) at EPE Lauhata with a new facilitator, Emigio. At the first-round sites three of the four facilitators were women; sex is not recorded in the second-round records.

The matched row is the most important line in this table and is easily overlooked. At the first-round sites, fewer than half the assessed children appear in both rounds, which is why this report compares group-level performance rather than individual progress for those sites and why EPE Tomas Ximenes shows a higher endline number of surveys completed. At the two second-round sites every assessed child appears in both rounds, so for these sites individual progress can be described directly (Section 11.1).

4.2 Qualitative respondents

The August 2026 qualitative round, conducted by the consultant team, engaged the following respondent groups. Data collection took place on 8 August at EPE Alfredo de Jesus Lemos and 11 August at EPE Tomas Ximenes. No qualitative data collection took place at EPE Lauhata or EPE Saboria: no focus group discussions or key informant interviews were conducted at either second-round site, and no respondent from those sites is included below.

Activity	Respondent group	Participants
Individual interviews	Parents / caregivers	3
Focus group discussion (two groups)	Parents / caregivers	16
Focus group discussion	Facilitators	4

Activity	Respondent group	Participants
Individual interviews	Facilitators	2
Total		**25**

The facilitator FGD engaged six participants, which exceeds the four facilitators formally assigned to the two first-round sites. The additional participants are understood to include facilitators associated with the wider ETF program. This should be documented in the qualitative annex before external dissemination, because it affects how facilitator findings should be attributed to the pilot.

4.3 Target participant profile

- **Age profile:** mean enrolment age was 5.3 years at EPE Alfredo de Jesus Lemos (range 2.8-7.8) and 4.9 years at EPE Tomas Ximenes (range 2.9-7.5). Assessment records include some children up to 9 years old, indicating that older siblings also participated; 38 of 41 and 55 of 57 enrolled children respectively fell within the 3-7 target band. As a structured play-based program that is free, many above the target age of 7 voiced their hope to participate in the program, often with their younger siblings. At the second-round sites, mean age recorded at the baseline assessment was 4.8 years at EPE Lauhata (range 3-7) and 4.7 years at EPE Saboria (range 3-6), so all assessed children there fell within the target band.
- **Sex:** girls were 59% and 54% of enrolments at the two first-round sites, 44% (28 of 64) of assessed children at EPE Lauhata and 65% (28 of 43) at EPE Saboria, as recorded at baseline.
- **Age-mix effects:** parents in the FGD reported that older children sometimes dominated younger ones, and that this discouraged some younger children from continuing. The ETF classroom observer recorded the same dynamic directly on 14 April. Parents recommended grouping children by age and prioritising children aged 10 and under, particularly those with weaker academic skills.
- **Inclusion and vulnerability:** enrolment screening recorded one child with an allergy and one child with a reported medical condition or special need at EPE Alfredo de Jesus Lemos; no allergies or medical conditions were reported at EPE Tomas Ximenes. All children were reported with complete immunisation status. Section 5.3 discusses why these figures almost certainly understate disability prevalence. No enrolment health or medical screening data were supplied for the second-round sites.
- **Civil documentation:** 31 of 41 children (76%) at EPE Alfredo de Jesus Lemos and 53 of 57 (93%) at EPE Tomas Ximenes had birth certificates on file - a useful proxy for household access to services, and a gap the program can help families address. Equivalent data were not supplied for the second-round sites.
- **Rural participation:** 100% of first-round participants live in rural aldeias of Suku Lihu; most walk to sessions. Residence and transport data were not supplied for the second-round sites.

05

SECTION

Program Governance, Safeguarding and Operating Policies

This section sets out the core policies and operating arrangements that surrounded delivery, and assesses the evidence available to the evaluation on each. All of the functions described here - facilitator recruitment, facilitator training, safeguarding practice, inclusion screening and site-level supervision - were the responsibility of the ETF Timor-Leste team. The consultant team had no role in delivery.

The section is included because a reader assessing whether this model should be expanded needs to know not only whether children learned, but whether the program was run safely and inclusively, and how the people delivering it were selected and prepared. On several of these questions the evaluation found practice to be sound but documentation to be thin. That distinction is maintained throughout: what follows separates what was done from what can be evidenced.

5.1 Delivery and governance responsibilities

Table 5.1. Who was responsible for what.

Function	Responsible party	Evidence available to the evaluation
Program design, curriculum and session guidelines	ETF (Australia and Timor-Leste teams)	Session guidelines and activity plan, program design and theory of change
Facilitator recruitment and selection	ETF Timor-Leste team	Facilitator names, sites and pairing; no written selection criteria, advertisement, shortlist or interview record was provided
Facilitator training and ToT	ETF Timor-Leste team	Training delivered April 2026; pre/post knowledge assessments for 4 first-round facilitators (1 post-test missing); pre/post records for 4 facilitator assessments at the second-round sites (July and August-September 2026); no training curriculum, attendance register or ToT completion record
Session delivery	Community facilitators, two per site	Attendance registers for all 32 pilot sessions with facilitator names and dates; registers for all 32 second-round sessions, nine of them undated (EPE Lauhata)
Safeguarding and child protection	ETF Timor-Leste team	Observed practice recorded in classroom observation forms; no written policy, code of conduct, signed undertaking, screening record or reporting procedure was provided
Disability and inclusion screening	ETF Timor-Leste team	Enrolment records with allergy and medical-condition fields (first-round sites only); no functional-difficulty screening instrument used
Supervision, mentoring and classroom observation	ETF Timor-Leste team	Four completed observation forms at the first-round sites (one visit per facilitator); none at the second-round sites; no supervision schedule or mentoring log
Baseline and endline child assessment	ETF Timor-Leste team	Assessment datasets: 67 baseline and 69 endline records (first-round sites); 107 baseline and 107 endline records (second-round sites)
Parent confidence survey	ETF Timor-Leste team	Survey datasets: 34 respondents at the first-round sites (TX endline duplicates baseline); 79 at the second-round sites (SB endline dates unverifiable)

Function	Responsible party	Evidence available to the evaluation
Qualitative evaluation (FGDs, KIIs, success stories)	Independent consultant team	Interview and FGD guides, respondent list, thematic notes - first-round sites only
Analysis and evaluation reporting	Independent consultant team	This report
Institutional approval and oversight	Ministry of Education (DNEPE, municipal directorate)	Approval meeting 14 January 2026; municipal introduction 27 February 2026; attendance at opening sessions

5.2 Safeguarding and child protection

ETF is a member of the Australian Council for International Development and is therefore bound by the ACFID Code of Conduct, which requires member organisations to maintain child safeguarding and PSEAH policies, to apply them to all personnel including locally recruited staff and volunteers, and to operate a complaints-handling mechanism. The evaluation proceeds on the basis that these organisational policies exist at ETF level. What it could not obtain was evidence of how they were operationalised at the two first-round sites.

What the evidence shows was done

- No child-protection concern was raised by any respondent in the qualitative round (first-round sites only; no respondents were interviewed at the second-round sites, so the absence of concerns there is unevidenced rather than confirmed). All parents interviewed stated that their children feel safe at the program, crediting attentive facilitators and children returning home content.
- The ETF classroom observation tool contains a five-item learning-environment safety section, which was rated as fully met at both EPE Tomas Ximenes visits and partially met at the two EPE Alfredo de Jesus Lemos visits.
- Facilitators controlled access to sharp or damaged materials after training, prepared first-aid materials, and placed soap at handwashing points. The observer recommended that garden forks and similar tools not be given to 3-4 year-olds - a hazard identified and acted on within the monitoring system.
- Children are identified by first name and nickname only in the case-study material, and the consultant team recommended withholding full family names pending consent verification.

Gaps identified

- **No site-level safeguarding record.** The evaluation was not provided with a facilitator code of conduct, a signed child-protection undertaking, a police or community reference check record, or a written incident-reporting procedure for the sites. Facilitators themselves requested child-protection training in the FGD, which indicates they had not received it.
- **A supervision gap was directly observed.** On 14 April at EPE Alfredo de Jesus Lemos, children were left playing in the playground unsupervised while both facilitators prepared the science activity, and the observer recorded older children aged 8-9 dominating younger ones during that period. With two adults and up to 39 children, any preparation task creates an unsupervised interval. This is a structural consequence of the staffing ratio, not a lapse by individuals.

- **One uncorroborated signal was not followed up.** A single parent in the FGD observed that some children are hesitant to participate, possibly because of the facilitator or teacher. The remark is uncorroborated and may be benign. It should nonetheless have triggered a documented follow-up, and no follow-up record exists.
- **Consent for publication is unverified.** Photographs and named child stories were collected. ETF should confirm that written parental consent for publication is on file for each named child and each identifiable image before any external dissemination, including of Document 2.

Recommended safeguarding measures for the next cycle

- Adopt a one-page facilitator code of conduct, signed before deployment, covering behaviour with children, one-adult-alone rules, physical contact, photography and confidentiality.
- Make child-protection training a compulsory module of the ToT, delivered before a facilitator takes a session - facilitators asked for this directly.
- Record a reference or community verification check for every facilitator at recruitment.
- Establish and publicise a simple reporting route for children, parents and facilitators, with a named ETF focal point and a documented response procedure.
- Never leave a session without an adult present. Where preparation is needed, stagger it or use a parent volunteer - the parent-volunteer capacity identified in Section 14.3 is directly relevant.
- Maintain a written consent register for photographs and named stories, with consent recorded at enrolment and refreshed before each publication.
- Log every safeguarding query or concern, including those that turn out to be nothing, so that the absence of incidents is evidenced rather than assumed.

5.3 Disability and inclusion

Enrolment screening recorded one allergy and one medical condition or special need at EPE Alfredo de Jesus Lemos and none at EPE Tomas Ximenes - approximately 2% of enrolled children at one site and 0% at the other. No screening data were supplied for the second-round sites, although anecdotal data suggests there were some children with disabilities participating at second-round sites. These figures should not be read as prevalence estimates. Global data indicate that a considerably higher proportion of children have functional difficulties in at least one domain, and difficulties in vision, hearing, communication, learning and behaviour are routinely under-identified where screening relies on a parent volunteering a diagnosis at registration, as it did here.

The practical risk is straightforward. Several findings in this evaluation - children hesitating to participate, younger children being dominated in mixed-age groups, and the persistence of “with help” ratings on counting and letter recognition for a minority of children - are consistent with undetected functional difficulty as one contributing explanation among several. The program has no way of distinguishing that possibility from the alternatives, because it does not screen for it.

The instrument designed for this purpose is the Child Functioning Module developed jointly by UNICEF and the Washington Group on Disability Statistics, which exists in versions for children aged 2-4 and 5-17, is administered to a caregiver in a few minutes, and asks about functioning in everyday domains rather than about diagnosis. It has been used at scale in national household surveys and is well suited to enrolment screening in this setting.

What the pilot did well on inclusion

- Delivery and assessment entirely in Tetun, with Mambae speaking children supported, which removed a language barrier that formal schooling does not always remove.
- Facilitators seated the youngest children at the front and supported them individually after training - an inclusion practice they adopted without being instructed to.
- The classroom observation tool rated “the facilitator involves all children” and “treats boys and girls equally” as met at all four visits.
- Girls were the majority of enrolments at both first-round sites (59% and 54%) and of assessed children at EPE Saboria (65%), though a minority at EPE Lauhata (44%); no systematic gender gap in attendance appears in the first-round registers.
- No fee was charged and sessions used existing school facilities, removing cost as a barrier to participation.

Gaps identified

- No functional-difficulty screening at enrolment, and therefore no disaggregation of any outcome by disability status.
- No physical-accessibility assessment of any venue, and no record of whether any child was unable to attend for reasons of mobility, distance or disability.
- No inclusion content identified in the facilitator training curriculum, and no facilitator raised inclusion as a topic in either the FGD or the interviews - a silence which is itself informative.
- Children out of the 3-7 band were enrolled and assessed without a written age policy, which is an inclusion decision made by default rather than by design.

Recommended inclusion measures for the next cycle

- Administer the UNICEF / Washington Group Child Functioning Module to caregivers at enrolment and record the result against the child ID.
- Disaggregate attendance and outcome reporting by sex, age band and functional-difficulty status from the first cycle in which the data exist.
- Add a short inclusion module to the ToT covering simple accommodations: seating, pace, visual and verbal instruction, peer pairing and how to notice a child who is struggling to see, hear or follow.
- Assess each venue for physical accessibility before the site agreement is signed, and record what was found.
- Set a written age policy and apply age-banded grouping, so that inclusion of older siblings is a deliberate choice with a plan attached rather than an unmanaged consequence.
- Use enrolment data to refer the roughly one in four children at EPE Alfredo de Jesus Lemos without a birth certificate to civil registration - lack of documentation is itself an exclusion risk for later schooling.

5.4 Facilitator recruitment, training and support

Facilitator recruitment and training were carried out by the ETF Timor-Leste team. Four facilitators were recruited locally, two per site: Juliandina and Mario Lugido Afonso at EPE Alfredo de Jesus Lemos, and Vitri and Liberty at EPE Tomas Ximenes. Three of the four facilitators were women, and all were residents of the communities where they served. Before the program commenced at each site, ETF delivered a Training of Facilitators (ToF) in early March for the facilitators at EPE Alfredo and in early April for those at EPE Tomas. Participants' knowledge was assessed immediately before and after the training. The training covered key areas, including relevant policies, child safeguarding, and child safety.

At the second-round sites, the facilitator records and attendance registers name Juliandina Pereira Soares and Mario Lugido Afonso as the facilitators at EPE Saboria - the same two facilitators who delivered at EPE Alfredo de Jesus Lemos - and Vitri Novi Selfi Lopes and Emigio Vasco Francisco Mau-Saco Maia at EPE Lauhata, as the core facilitators at EPE Lauhata. Other facilitators or ETF staff, including Cesar and Diana, are named alongside them on a number of register sheets at both sites - Diana as late as session 15 at EPE Saboria - so the two-person teams were not fixed. Pre-training knowledge tests were recorded on 13 July (EPE Lauhata) and 16 July 2026 (EPE Saboria), after the first sessions had begun, and post-training tests between 31 August and 7 September 2026, near the end of the cycle. The evaluation holds no record of the training itself for these sites.

Local recruitment is a genuine strength of the model and should be preserved. It produces facilitators who speak the children's home language, are known to families, are available in the community between cycles, and represent a durable local capacity asset rather than a project cost. The pilot's evidence supports this: parents cited the facilitators' attentiveness as the main reason they judged children safe, and facilitators introduced good practice on their own initiative.

What the evaluation could not establish

- The selection criteria. No written criteria, advertisement, applicant list or interview record was provided. It is therefore not possible to say what qualifications, experience or attributes ETF was recruiting for, or to replicate the selection in new sites.
- The training curriculum and its duration. Training is documented only through its outcome, the pre/post knowledge scores. Neither the content, the number of days, nor the attendance register was available.
- ToT coverage. Facilitators reported in the FGD that some of them had not attended the Training of Trainers. No attendance record exists against which to check who did and did not.
- Supervision frequency. Four classroom observations were conducted - one per facilitator, with both EPE Alfredo de Jesus Lemos visits in the same week and none there after session 9. There was no scheduled supervision cycle. No classroom observation was recorded at either second-round site.

The training gap, and what the literature says about it

The most consequential finding about facilitators is that knowledge transfer succeeded and skills transfer did not. Post-training scores reached 6 out of 6 while the same facilitators reported being unable to manage large mixed-age groups, keep to guideline timings, or run the learning corners. This is not a contradiction and not a failure of the training content; it is the standard finding of the professional-development literature. Joyce and Showers established decades ago that presentation and demonstration produce knowledge but little transfer to practice, and that transfer rises sharply only when coaching in the workplace is added. The second-round records add a cautionary data point. When Juliandina and Mario were retested before delivering at EPE Saboria in July, they scored 1 and 2 out of 6 - no higher than their first-round pre-training scores of 2 - and 3 out of 6 at the end of the second cycle.

Whether this reflects loss of knowledge, a different administration of the test, or the instability of a six-item binary measure cannot be determined from the records, but it confirms that a single post-training knowledge score is not a durable indicator of capability.

The contemporary meta-analytic evidence is consistent and unusually clear. Kraft, Blazar and Hogan (2018) pooled 60 causal studies of teacher coaching and found effects of 0.49 standard deviations on instructional practice and 0.18 standard deviations on student achievement - among the larger effects available from any professional-development approach - with coaching paired with group training outperforming either alone. Their second finding is equally relevant to ETF: effects attenuate as coaching programs scale, which means the coaching design must be built in deliberately from the start rather than added once a problem appears. Evidence from low-resource settings, including structured coaching within early-grade literacy programs, points the same way.

The operational conclusion for ETF is that facilitators do not need more content. They need supervised practice, a structured observation-and-feedback cycle, and an observed-practice rubric that measures what they do rather than what they know.

Recommended facilitator measures for the next cycle

- Write down the selection criteria and use a documented recruitment process, so that facilitator quality is reproducible across sites rather than dependent on individual judgement.
- Deliver full ToT to every facilitator before deployment, with an attendance register, and add compulsory modules on child protection, classroom and behaviour management for large mixed-age groups, and inclusion.
- Schedule fortnightly coaching visits with a short session-fidelity checklist and structured feedback, rather than episodic supervision.
- Develop an observed-practice rubric to sit alongside the knowledge test, and report both. A facilitator who scores 6/6 on knowledge and cannot run a session is not yet capacity built.
- Pair every new facilitator with an experienced first-round facilitator as mentor.
- Review remuneration and transport support against the actual workload, including preparation and travel, and budget transport for all facilitators rather than one.
- Move toward a 1:15 facilitator-to-child ratio. Almost every quality problem identified in this evaluation is downstream of two adults working with up to 39 children across five years of development.

5.5 Implementation coordination and monitoring structure

ETF's coordination and monitoring structure for the pilot is set out in the Diagrama Programa Depois Eskola, reproduced below and in Annex M. The diagram places the ETF team at the centre of three levels of stakeholders and assigns monitoring roles at each level.

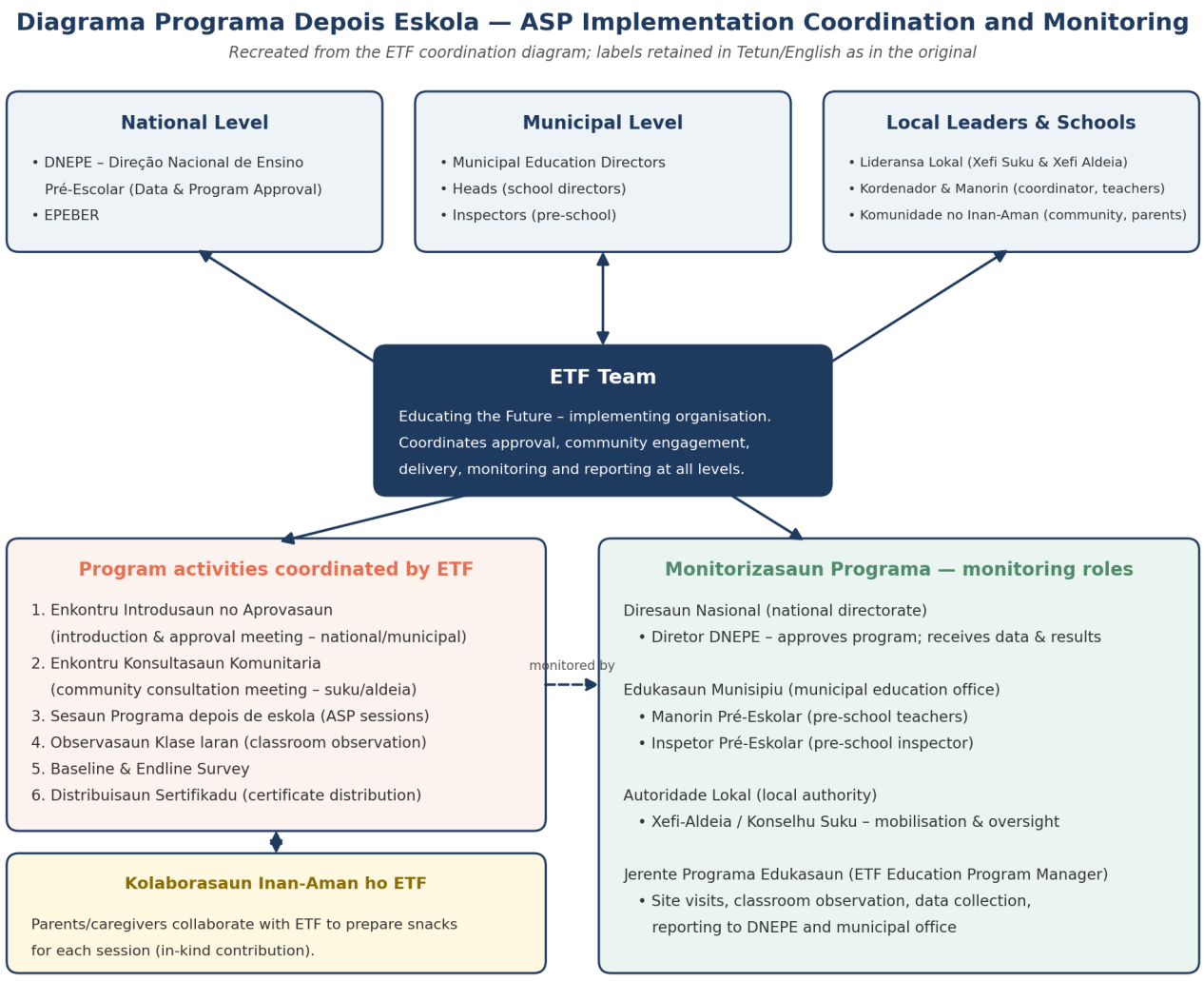


Figure 1. ASP implementation coordination and monitoring structure (recreated from the ETF diagram; see Annex M).

- **National level:** DNEPE and DGEPEBR, responsible for program approval and for receiving program data. In practice this was the meeting of 14 January 2026 and DNEPE's participation in the opening sessions.
- **Municipal level:** the Municipal Education Director, school heads and pre-school inspectors, engaged through the introductory meeting of 27 February 2026 and the opening sessions.
- **Local leaders and schools:** suku and aldeia leadership, the school coordinator and teachers, and the community and parents.
- **Program activities coordinated by ETF:** the introduction and approval meeting; community consultation; the after-school sessions; classroom observation; baseline and endline surveys; and certificate distribution. A separate strand records collaboration between parents and ETF in preparing snacks.
- **Monitoring roles:** the DNEPE Director at national level; pre-school teachers and the pre-school inspector at municipal level; the Xefi Aldeia and Konselhu Suku as local authority; and ETF's Education Program Manager.

The structure is sound on paper. Two observations follow from reading it against the evidence. First, the diagram places classroom observation and community consultation among ETF's core coordinated activities, but neither has a defined frequency; the pilot delivered one observation per facilitator and one consultation round at start-up.

Second, the monitoring roles assigned to the municipal inspector, teachers and local authorities are not evidenced in any monitoring record available to the evaluation. The structure should be attached to a simple visit log and reporting line so that it produces records as well as relationships.



06

SECTION

Implementation Overview

6.1 Program duration

- **Overall pilot period:** January-July 2026, with enrolment from 23 February 2026 at EPE Alfredo de Jesus Lemos and 4 March 2026 at EPE Tomas Ximenes. Qualitative evaluation data were collected in August 2026.
- **EPE Alfredo de Jesus Lemos:** baseline child assessment 6 March 2026; 16 sessions delivered from 10 March to 18 May 2026; endline child assessment 8-9 June 2026; qualitative data collection 8 August 2026.
- **EPE Tomas Ximenes:** baseline child assessment 8 April 2026; 16 sessions delivered from 15 April to 16 June 2026; endline child assessment 22 June 2026; qualitative data collection 11 August 2026.
- **EPE Lauhata (second-round site):** parent baseline survey 30 June 2026; baseline child assessment 2 July 2026; 16 sessions recorded, the first seven dated 7-27 July 2026 (sessions 1 and 2 both dated 7 July) and sessions 8-16 undated; endline child assessment and parent endline survey 23 September 2026; no qualitative data collection.
- **EPE Saboria (second-round site):** parent baseline survey 3 July 2026; baseline child assessment 8 July 2026; 16 sessions delivered from 9 July to 4 September 2026, with a three-week break between session 4 (17 July) and session 5 (6 August); endline child assessment 16 September 2026; parent endline survey dates recorded as 30 September to 2 November 2026 and unverifiable; no qualitative data collection.

> Note that the interval between the last session and the endline assessment differs between sites - three weeks at EPE Alfredo de Jesus Lemos and six days at EPE Tomas Ximenes. This is a minor but real source of non-comparability between the two sites' endline figures. At the second-round sites the interval was 12 days at EPE Saboria and cannot be determined at EPE Lauhata, where the final sessions are undated.

6.2 Locations

Both first-round sites are in Suku Lihu, Railaco Administrative Post, Ermera Municipality. Qualitative data collection identified Aldeia Hi as the principal catchment hamlet for both sites, which means the two "independent" sites in this pilot draw partly on the same community. The apparent replication across two sites is therefore weaker than it looks, and is discussed in Section 9.

The administrative location of EPE Lauhata and EPE Saboria is not recorded in the datasets supplied to the evaluation and should be confirmed by ETF before external dissemination. If, as their names suggest, they lie outside Suku Lihu, they are the first sites to provide measured results from beyond the pilot suku; this report does not assume their location.

6.3 Implementation structure

- **Facilitators:** two per site, recruited locally by the ETF Timor-Leste team, trained by ETF 9-20 of February 2026, and assessed pre- and post-training (Section 5.4). At the second-round sites the same two-per-site pattern applied, drawing partly on first-round facilitators (Section 5.4).
- **Coordination and supervision:** ETF program staff conducted training, monitoring visits, classroom observation and baseline/endline data collection. Facilitators confirmed in interviews that they receive support from the ETF team, while requesting more frequent and more in-depth mentoring than the pilot provided.

- **Venue: sessions were held in existing pre-school classrooms at both first-round sites; venue arrangements at the second-round sites are not documented.** Facilitators reported that children appeared less engaged when After-School Program sessions were conducted in the same classroom environment they were already familiar with from their regular school activities. They suggested that using alternative or adapted spaces could create greater variety, potentially helping to sustain children's interest and participation. Facilitators perceived that, for some children, conducting the program in the same classroom may have made the sessions feel similar to their regular school experience rather than providing a distinct opportunity for play, exploration, and creative learning. Future cycles could therefore consider using outdoor areas, community spaces, or reconfigured classroom environments where appropriate.

6.4 Session scheduling

Sessions were originally designed as an after-school activity. During implementation the schedule shifted at least once toward morning delivery. This change is recorded in the qualitative data as a significant disruption and is independently confirmed by the observation record: the June observation at EPE Tomas Ximenes took place in a 10:00-12:00 slot whereas all earlier observed sessions ran in the afternoon.

- Children who were most available or most alert in the afternoon were excluded by the move to mornings, according to parents.
- Facilitators reported the opposite problem: that the original after-school start time left children tired and with limited attention span, and that parents in some communities did not support afternoon sessions because of their own competing commitments.
- Late notice of schedule changes compounded the problem, since parents had limited means of receiving updates.

There is, in other words, no single timing that suits all children and parents. The evaluation's reading is that the pilot's error was less the choice of slot than the mid-cycle change and the late notice. Section 18 proposes that scheduling be settled per site through negotiation with schools and parents before enrolment opens. Session times are not recorded in the second-round registers, so whether the same scheduling problems arose there is unknown.

07

SECTION

Activities Conducted

7.1 Activity summary

Each of the four sites recorded 16 structured sessions (64 in total), each led by a core two-person facilitator team. The session-by-session activity log - dates, objectives, participants, facilitators and remarks - is maintained in the site attendance registers and can be annexed in full.

7.2 Types of activities

- **Literacy:** storytelling and story retelling (beginning-middle-end), reading circles, letter recognition and letter-sound games, and use of the Lafaek learning magazine, which parents specifically identified as a source of letter recognition and creative games at home.
- **Numeracy:** counting games (1-10), shape recognition (circle, square, triangle), and "more/less" comparison activities.
- **Problem-solving:** simple puzzles (3-6 pieces), sorting by colour and shape, building and repairing structures with blocks, and encouraging multiple solution strategies.
- **Social-emotional learning:** group games, sharing and turn-taking, naming and expressing feelings, empathy-building, respect, inclusion and self-regulation routines.
- **Health and physical development:** handwashing before eating and after using the toilet, healthy versus unhealthy food identification, and running, jumping and balance games; facilitators placed soap at the handwashing tap and integrated hygiene routines into each session.
- **Arts, creativity and science:** painting, drawing, singing, model-making including using recycle material, and simple science activities. Facilitators identified these as the most effective and best received activities, and parents specifically valued children's exposure to science.
- **Environmental and climate awareness:** watering plants and vegetables, waste disposal routines, and tidying the learning space - introduced by facilitators after training even where not formally scheduled. Parents reported children bringing these skills home, including raking soil and planting seedlings.
- **Parent engagement:** parent meetings at registration, baseline and endline parent surveys, and awareness-raising on supporting reading, counting, play and emotional development at home.

7.3 Activity effectiveness as reported by facilitators

Facilitators at the two first-round sites were asked which parts of the curriculum worked best in practice. Their responses point to a clear pattern and to one significant implementation gap.

Table 7.1. Facilitator assessment of program components.

Component	Facilitator assessment	Implication
Opening and closing sessions (1-5 and 16) - science, painting, games, creative activities	Identified as the most effective and best received by children	The most active, hands-on and novel content drives engagement. Distribute this content across the whole cycle rather than concentrating it at the start and end.
"All About Me", environmental care, arts and creativity, and science modules	Identified in interviews as the most valuable training content	These modules should anchor the ToT curriculum in future cycles.

Component	Facilitator assessment	Implication
The four learning "corners" (reading, language, homework and associated learning corner)	Rarely implemented in practice	A designed component of the model is not reaching children. Either the corners require space, materials and training that sites do not have, or the guidelines do not make their use practical. This must be diagnosed before any scale-up.
Time allocations set out in the guidelines	Difficult to adhere to	Session timings should be revised against observed practice rather than assumed, with built-in flexibility for group size and age mix.

The finding on the learning corners (kantu aprendizajen) deserves particular attention. It is the clearest instance in this evaluation of a gap between the program as designed and the program as delivered, and it was surfaced only by the qualitative round: the quantitative instruments contained no indicator that would have detected it. The implementation science literature treats this as a predictable failure rather than an unlucky one - Durlak and DuPre's review of over 500 studies found that implementation quality substantially moderates program outcomes, and that programs routinely assume fidelity in the absence of any measure of it.

08

SECTION

Evaluation Design and Methodology

This section states who conducted each part of the assessment, what instruments were used, whether those instruments were validated, how responses were scored and aggregated, and what was and was not done to check consistency. It is placed before the findings because the findings cannot be read properly without it.

8.1 Overall design

The pilot was evaluated using a mixed methods design combining routine monitoring data, structured pre/post measurement, and a qualitative round conducted after the cycle closed. The quantitative component is a single group, pre-test / post-test design. There is no comparison group, no randomisation, no baseline equivalence testing and no follow-up measurement. In the standard typology of Shadish, Cook and Campbell, this is the weakest of the quasi-experimental designs, and the threats to internal validity it leaves open - maturation, history, testing, instrumentation and selection - are all live in this case. Section 9 works through each of them. The two second-round sites were measured with the same quantitative instruments and the same single-group design, but without the qualitative round or classroom observation.

The qualitative component is an exploratory, thematic study designed to explain mechanisms, not to estimate prevalence. Together the two components can describe what changed and offer a credible account of why. They cannot isolate the program's contribution.

8.2 Who conducted each assessment

Responsibility for data collection was divided between the ETF Timor-Leste team and an independent consultant team. The division is stated here explicitly because it bears on how each finding should be weighted.

Table 8.1. Division of data-collection responsibilities.

Instrument / activity	Conducted by	Timing	Independence from delivery
Child baseline assessment (24 indicators, Tetun)	**ETF Timor-Leste team**	6 March 2026 (AdJL); 8 April 2026 (TX); 2 July 2026 (LH); 8 July 2026 (SB)	Not independent - assessors were the implementing team
Child endline assessment (24 indicators, Tetun)	**ETF Timor-Leste team**	8-9 June 2026 (AdJL); 22 June 2026 (TX); 23 September 2026 (LH); 16 September 2026 (SB)	Not independent
Facilitator pre-training knowledge test (6 domains)	**ETF Timor-Leste team**	April 2026, immediately before training (first-round sites); 13 and 16 July 2026 (LH, SB)	Not independent - ETF also designed and delivered the training
Facilitator post-training knowledge test (6 domains)	**ETF Timor-Leste team**	April 2026, immediately after training (first-round sites); 31 August - 7 September 2026 (LH, SB)	Not independent; one of four pilot records missing
Classroom observation (Feramentu Observasaun Sala Laran, 46 rated items)	**ETF Timor-Leste team**	14 and 17 April; 30 April; 11 June 2026 (first-round sites); none at LH or SB	Not independent; all four visits by the same observer
Attendance registers	**ETF Timor-Leste team / facilitators**	Every session	Recorded by facilitators, i.e. self-recorded

Instrument / activity	Conducted by	Timing	Independence from delivery
Parent/caregiver confidence survey	ETF Timor-Leste team	Baseline and endline (LH: 30 June and 23 September 2026; SB: 3 July 2026 and unverifiable endline dates)	Not independent
Enrolment and registration records	ETF Timor-Leste team	February-March 2026 (first-round sites); no usable enrolment record for LH or SB	n/a
Parent focus group discussions (2 groups, 16 participants) - first-round sites only	Consultant team	8 and 11 August 2026	Independent of delivery
Parent key informant interviews (3) - first-round sites only	Consultant team	8 and 11 August 2026	Independent of delivery
Facilitator focus group discussion (6 participants) - first-round sites only	Consultant team	August 2026	Independent of delivery
Facilitator key informant interviews (2) - first-round sites only	Consultant team	August 2026	Independent of delivery
Child success-story interviews (2 children and families) - first-round sites only	Consultant team	August 2026	Independent of delivery
Data cleaning, analysis, triangulation and reporting	Consultant team	August-September 2026	Independent of delivery

Why this matters

- **Every measure of child outcome in this report was collected by the same team that designed the program, trained the facilitators and had an interest in the result. This is entirely normal for a pilot of this size and implies no impropriety. It does mean that expectancy effects - the well-documented tendency of assessors who know the expected direction of change to score in that direction - cannot be ruled out, and that the near-elimination of "not yet able" ratings, which is the report's most striking single number, is exactly the kind of finding most vulnerable to it.**
- **The qualitative evidence is the part of this evaluation collected independently of delivery, and this is one reason the report gives it substantial weight. Its own limitations, chiefly small purposive sampling and social-desirability effects, are set out in Section 8.7.**
- **The remedy for the next cycle is inexpensive: have the endline child assessment administered by assessors who did not deliver the program, or at minimum have an independent assessor re-score a random sample so that an inter-rater reliability figure can be reported.**

8.3 Instruments and their validation status

All instruments were developed internally by ETF in Tetun. None has been externally validated. The evaluation was not provided with any record of pilot-testing, cognitive testing with children, expert review, factor analysis, internal-consistency statistics, test-retest reliability or inter-rater reliability for any instrument. This is stated as a plain fact rather than a criticism: internally developed observation checklists are the norm for small NGO pilots, and the tools are evidently fit for the practical purpose of guiding facilitators. The point is that they do not yet carry the psychometric properties that would allow the resulting numbers to be presented as measurement in a technical sense.

Table 8.2. Instruments, coverage and validation status.

Instrument	Structure	Coverage	Validation status
Child observation tool (Tetun)	24 indicators across three domains: 10 learning skills; 5 socio-emotional behaviours; health, hygiene and physical development items	First-round sites: 67 baseline and 69 endline records. Second-round sites: 107 baseline and 107 endline records, covering 18 rated items (10 learning, 5 socio-emotional, 3 health)	Internally developed. Not externally validated. No reliability testing, no piloting record, no scoring manual supplied to the evaluation.
Facilitator knowledge and practice check	6 domains, scored out of 6 (recorded at the second-round sites on a 1-5 confidence scale using only the values 1 and 5)	First-round sites: 4 facilitators, 3 with complete pre and post records. Second-round sites: 4 facilitator assessments, all complete	Internally developed. Measures knowledge, not applied practice. No observed-practice component.
Classroom observation tool (Feramentu Observasaun Sala Laran)	9 sections (A-I), 46 items rated yes/no, plus 2 descriptive items	4 observations, one per first-round facilitator; none at the second-round sites	Internally developed. Binary scale produces ceiling effects. No inter-rater check - a single observer conducted all four visits.
Parent confidence survey	EPE AdJL: 4 domains on a 1-5 scale. EPE TX: 8 domains, categorical. EPE LH and SB: the 8 TX domains on a 1-5 scale, recorded using only 1 and 5	17 respondents per first-round site; 45 (LH) and 34 (SB)	Internally developed. Three different formats used across the four sites, so results are not comparable between formats and are not pooled.
Attendance register	Presence marks per child per session	64 sessions (32 first-round, 32 second-round)	Not an assessment instrument. Hand-computed percentages used inconsistent denominators; this report recomputes first-round attendance from presence marks. At the second-round sites presence marks are unreliable for several sessions and written session totals are used (Section 8.8).
Qualitative guides (FGD and KII)	Semi-structured, by respondent group	25 respondents (first-round sites only)	Fit for exploratory purpose. Thematic analysis by a single team; no double-coding or inter-coder agreement check.

Benchmarking the child tool

Validated instruments exist for precisely this purpose and age group. The International Development and Early Learning Assessment (IDELA), developed by Save the Children between 2011 and 2015, covers emergent literacy, emergent numeracy, motor development and socio-emotional development for children aged 3.5 to 6 years; it has published evidence on internal consistency, inter-rater reliability, test-retest reliability and concurrent validity, takes roughly 35 minutes to administer, and has been shown to be usable with high fidelity in low-resource settings (Pisani, Borisova & Dowd, 2018). Subsequent work has demonstrated measurement invariance across five countries and predictive validity into early primary school. The MELQO package developed by UNESCO, UNICEF and the World Bank offers an alternative designed for national adaptation, and the East Asia-Pacific Early Child Development Scales were validated for this region specifically.

The evaluation does not recommend that ETF replace its own tool, which is short, practical, in Tetun, and usable by facilitators. It recommends that ETF administer a validated instrument alongside its own tool to a sample of children in the next cycle. That single step would establish whether the ETF tool tracks a validated measure, allow ETF to report results in units that the Ministry and donors already recognise, and convert the program's main evidential weakness into a strength at modest cost.

8.4 Scoring and how results were derived

The report was not supplied with a scoring manual, so the scoring rules below are those reconstructed and applied by the consultant team during analysis. They should be formalised in writing before the next cycle so that scoring is a documented procedure rather than an analytical reconstruction.

Learning skills (10 indicators)

- Each of the ten indicators was rated on a three-point ordinal scale: Sin (performs independently), Ho ajuda (performs with help), Lae (not yet able).
- The scale is ordinal, not interval. The distance between "not yet able" and "with help" is not assumed to equal the distance between "with help" and "independent", and no mean score is computed across the three categories.
- The headline figures (52% → 81% at AdJL, 27% → 83% at TX, 26% → 75% at LH and 37% → 92% at SB; 35% → 82% across all four sites) are the proportion of all skill checks rated Sin, where a skill check is one indicator for one child. The denominator at each round is ten indicators multiplied by the number of children assessed at that round - 390 checks at AdJL baseline, 260 at AdJL endline, 280 at TX baseline, 430 at TX endline, 640 at each LH round, 430 at SB baseline and 429 at SB endline (one blank rating).
- This is a check level, not a child level, statistic. It answers "what share of skills were mastered", not "what share of children mastered a given skill". Because the number of children differs between rounds, and because the children are not the same children, the two percentages describe two different groups. Item-level results, reported per indicator in Table 11.1, are the more interpretable figures.
- Missing or blank ratings were excluded from the denominator rather than imputed. Where a child's record was incomplete, that indicator was dropped for that child only.
- Pooled four-site figures are computed by summing skill checks (or observations) across sites and dividing by the summed denominators, never by averaging site percentages. Because the item-level pilot datasets were not re-supplied with the second-round files, first-round counts are reconstructed from the published site percentages and denominators. For independent ratings this reconstruction is exact: the item-level counts in Table 11.1 reproduce the published 52%, 81%, 27% and 83%. For "not yet able", socio-emotional and health ratings it is accurate to within about one percentage point, and pooled figures for those measures are stated as approximate.

Socio-emotional and health behaviours

- Rated on a four-point frequency scale: Sempre (always), Bebeik (often), Dala-ruma (sometimes), Nunka (never).
- Reported results collapse the scale to a binary: "often or always" (Sempre + Bebeik) versus the rest. Collapsing an ordinal scale in this way discards information and can exaggerate apparent change when responses cluster near the cut point; it was used for consistency with the original program reporting.
- A scale inconsistency required harmonisation. The EPE Alfredo de Jesus Lemos endline recorded socio-emotional items as numbers 1-4 while the baseline used the Tetun descriptive categories. The consultant team mapped 4 = always, 3 = often, 2 = sometimes, 1 = never. This mapping is a reasonable inference, but it is an inference, and

the AdJL socio-emotional result (35% → 90%) depends on it being correct. At the second-round sites both rounds recorded these items as numbers 1-4, so no harmonisation was needed; five non-numeric entries (ticks or the letter S) were treated as missing.

Facilitator knowledge

- Six domains, each scored as met or not met, producing a score out of 6. Reported as raw scores, not converted to percentages, given the small number of facilitators. At the second-round sites each domain was recorded as 1 or 5 on a 1-5 confidence scale; 5 is scored as met and 1 as not met, consistent with the pilot scoring.
- Liberty's post-training score was not recorded. No value was imputed; results are reported for three of four facilitators, and any statement about "all facilitators" in earlier drafts has been corrected.

Parent confidence

- At EPE AdJL, four domains on a 1-5 scale, reported as a mean across domains and respondents, plus the share of individual ratings at 4 or 5.
- At EPE TX, eight domains on a categorical scale, reported as the share of responses indicating strong confidence (Konfidente tebes).
- At EPE LH and SB, the eight TX domains on a 1-5 scale, on which only the values 1 and 5 were used. Results are reported as the mean and the share of ratings at 5; in practice the instrument functioned as a yes/no item.
- The sites used three different instrument formats. Parent results are not comparable between formats and are not pooled across sites.

Attendance

- Recomputed directly from presence marks in the registers, because the hand-written percentages in the registers used inconsistent denominators. At the second-round sites, the session totals written at the head of each register sheet are used instead, because the presence marks are demonstrably unreliable for several sessions (Section 8.8).
- Average attendance is the mean number of children present per session. Attendance as a share of enrolment uses the enrolment list as denominator, which is why EPE Tomas Ximenes shows a lower percentage despite similar absolute attendance: its enrolment list is larger. For the second-round sites, which lack a usable enrolment list, attendance is expressed as a share of children assessed.
- Per-child attendance was not summarised in the registers, so the evaluation cannot report how many children attended a given number of sessions. This prevents any dose-response analysis, which is the single most useful additional analysis a program like this can run on its own data.

8.5 Consistency and quality-assurance checks

The following checks were and were not performed. They are set out in full because the original draft did not state them, and because the credibility of the findings depends on them.

Table 8.3. Quality-assurance checks: what was done and what was not.

Check	Status	Detail
Inter rater reliability on child assessment	**Not performed**	No child was assessed independently by two assessors, so no agreement statistic can be reported.

Check	Status	Detail
Inter rater reliability on classroom observation	**Not performed**	All four observations were conducted by the same ETF observer. Consistency of application is therefore high by construction, but accuracy is unverifiable.
Test-retest reliability	**Not performed**	No child was re-assessed within a short interval.
Internal consistency (e.g. Cronbach's alpha)	**Not performed**	Not computed at the time and not computable retrospectively with confidence given missing records and scale changes.
Blinding of assessors to round or expected result	**Not performed**	Assessors knew which round they were administering and had delivered the intervening program.
Double data entry or entry verification	**Not evidenced**	No verification record was provided. Several transcription-type errors were found during analysis (see below), which suggests single entry without checking.
Double-coding of qualitative data	**Not performed**	Thematic coding by a single consultant team; no inter-coder agreement statistic.
Name-based matching of children across rounds	**Performed**	Fuzzy matching on name, sex and age, allowing for spelling variants. Approximately 20 of 26 endline children at AdJL and 22 of 43 at TX were matched. Matched subsets show the same direction of change as the full samples. At LH and SB all 64 and 43 children were matched across rounds.
Recomputation of attendance from source	**Performed**	All attendance figures recomputed from presence marks rather than taken from register totals (first-round sites). At the second-round sites, presence marks were checked against the written session totals and found inconsistent for 15 of 32 sessions (11 at EPE Lauhata, 4 at EPE Saboria); session totals are used.
Arithmetic reconciliation of the financial workbook	**Performed**	Balance column errors identified; correct net underspend on materials is US$0.47 rather than the US$5.57 shown.
Internal consistency checks across datasets	**Performed**	Cross-checks between enrolment lists, registers, assessment records and observation forms identified the duplicated TX parent survey, the missing facilitator post-test, the 4+16=19 discrepancy on the 14 April observation form, and several age inconsistencies between rounds. For the second-round sites, the same cross-checks identified the copied EPE Saboria enrolment list, copied EPE Lauhata attendance sheets, repeated rating profiles in the EPE Saboria baseline, the split in the EPE Lauhata endline ratings and auto-filled EPE Saboria parent survey dates.
Pattern checks on second-round assessment records	**Performed**	Records were screened for repeated identical rating profiles and abrupt changes in rating behaviour within a round. Findings are reported in Section 8.8 rather than corrected.
Triangulation across evidence strands	**Performed**	Every substantive finding was checked against at least two of: quantitative records, classroom observation, and qualitative accounts. Convergences and divergences are both reported (Section 13).

Errors found during the consistency checks were not corrected silently. The duplicated EPE Tomas Ximenes parent endline survey is reported as unusable rather than analysed; the missing facilitator post-test is reported as missing rather than imputed; the observation-form headcount discrepancy is flagged in Table 14.2; and the financial

workbook errors are set out in the business case. Where an error could not be resolved from source, the affected finding is withheld. The same rule was applied to the second-round records: anomalies are reported in Section 8.8 and the affected results are flagged rather than adjusted.

8.6 Analytical approach

- Quantitative results are reported descriptively. No significance testing was conducted, because with unmatched, non-random samples and a non-validated instrument, a p-value would convey a precision the design does not support.
- No effect sizes are reported, for the same reason and because the ordinal scale does not support a standardised mean difference.
- Qualitative responses were organised thematically by respondent group and activity type, covering satisfaction, observed changes, challenges, safety, training and support needs, and recommendations. Findings are presented by respondent group in Section 12 so that the distinct vantage points of parents and facilitators remain visible, followed by cross-cutting themes.
- Triangulation follows a simple rule: where strands agree, confidence is reported as higher; where they diverge, the divergence is reported rather than reconciled, because the divergence is usually itself the finding.

8.7 Limitations of the qualitative round

- The sample is small (25 respondents) and purposively selected; findings indicate the range and strength of views held rather than their prevalence across the whole parent body.
- Most children had attended for less than three months at the time of data collection, so respondents were assessing an incomplete exposure to the program.
- Interviews and FGDs were conducted after the cycle had ended, introducing recall effects, particularly for early-cycle events such as the mid-May attendance dip.
- The facilitator FGD engaged six participants against four facilitators formally assigned to the first-round sites; the composition of that group should be documented before external dissemination.
- Respondent groups were engaged separately by site, but findings are reported in aggregate across sites in the source material, which limits site-level attribution of specific views.
- Self-reported confidence and satisfaction are subject to social-desirability effects, particularly where respondents value the program and wish it to continue. Respondents were speaking to a team associated with the program's funder about whether the program should continue, which is not a neutral setting.
- Parents who never enrolled their children, and children who dropped out, were not interviewed. The qualitative round therefore samples on the dependent variable: it captures the views of those the program retained.
- The qualitative round covered the two first-round sites only. No FGDs or KIIs were conducted at EPE Lauhata or EPE Saboria, so no qualitative finding in this report can be attributed to, or verified for, those sites. Their results rely on the available quantitative data and other documented evidence (attendance registers and survey records).

8.8 Integration of the two second-round sites

This revision integrates the datasets supplied for EPE Lauhata and EPE Saboria: one workbook per site, each containing child baseline and endline assessments, attendance registers, facilitator pre/post knowledge tests and parent pre/post surveys. The program-summary sheet in each workbook was blank. The second-round assessment sheets record 18 rated items - the ten learning skills, five socio-emotional behaviours and three health and physical-development items - and results for these sites are reported on those items. The scoring rules in Section 8.4 were applied unchanged, and every combined figure in this report was recalculated from counts rather than by averaging site percentages.

Table 8.4. Evidence available, by site.

Evidence strand	EPE Alfredo de Jesus Lemos	EPE Tomas Ximenes	EPE Lauhata	EPE Saboria
Enrolment / registration records	Yes (41)	Yes (57)	Partial - list of 49 does not reconcile with records	None - sheet copied from LH
Child baseline / endline assessment	Yes (39 / 26)	Yes (28 / 43)	Yes (64 / 64, all matched)	Yes (43 / 43, all matched)
Attendance registers	16 sessions, presence marks	16 sessions, presence marks	16 sessions; 9 undated copied sheets	16 sessions, dated
Facilitator pre/post knowledge test	2 of 2 complete	1 of 2 complete	2 of 2 complete	2 of 2 complete
Parent confidence survey	Yes (17)	Baseline only usable (17)	Yes (45)	Yes (34); endline dates unverifiable
Classroom observation	2 visits	2 visits	None	None
Parent and facilitator FGDs and KIIs	Yes	Yes	None	None
Financial records	Materials only (pilot)	Materials only (pilot)	None	None

Data-quality findings for the second-round sites

The cross-checks described in Section 8.5 identified the following. None was corrected silently; each is reflected in how the affected results are reported.

- **Enrolment.** The EPE Saboria enrolment sheet reproduces the EPE Lauhata list name for name (49 entries), so no EPE Saboria enrolment record exists. The EPE Lauhata list (49 entries, including apparent duplicate entries for one child) does not reconcile with the 64 children assessed there. Participation at both sites is therefore counted from the assessment records.
- **Attendance.** At EPE Lauhata, the presence lists for sessions 8-16 are undated and identical to the session 1 list (39 names, all marked present), while the totals written on the same sheets range from 21 to 34; session 7 lists 19 names against a written total of 31. At EPE Saboria, the session 1 sheet contains about ten names carried over from the EPE Lauhata template. For both sites, attendance is taken from the written session totals - the only session-specific figures available - and should be treated as approximate.

- **Repeated rating profiles.** At the EPE Saboria baseline, 16 of 43 children (37%) share an identical profile: "not yet able" on all six literacy and numeracy items but independent on all four problem-solving items. A child who cannot yet attempt basic shapes or "more and less" but completes puzzles and builds independently is possible, but sixteen identical cases suggest rapid or templated recording. The pattern lowers the EPE Saboria baseline for literacy and numeracy and raises it for problem-solving.
- **Within-round break.** At the EPE Lauhata endline, the first 40 records rate every socio-emotional and health item "often or always", while the remaining 24 records almost never do (5% and 1% respectively) and contain all 18 of the site's endline "not yet able" ratings. Assessors are not identified in the records, so the evaluation cannot tell whether the break reflects different children, different assessors or different recording practice. The site's endline socio-emotional and health figures (64% and 63%) are averages over two very different blocks.
- **Ceiling effects.** At the EPE Saboria endline, 22 of 43 children were rated independent on all ten learning skills, 214 of 215 socio-emotional ratings were "often or always", and no "not yet able" rating was recorded. Results this close to the scale maximum leave almost no room for variation and are the kind most exposed to expectancy effects (Section 9.3).
- **Parent survey.** Both second-round sites used a 1-5 scale but recorded only the values 1 and 5, so the survey functioned as a yes/no item. The EPE Saboria endline survey dates run sequentially from 30 September to 2 November 2026, one date per respondent and several after the data were supplied; they are evidently auto-filled, and the actual endline survey date cannot be verified.
- **Facilitator tests.** Pre-tests were administered after the first sessions had been delivered, and post-tests seven to eight weeks later, near the end of the cycle, rather than immediately before and after training as at the first-round sites. Scores are therefore not directly comparable with the first-round scores.
- **Minor recording issues.** Five non-numeric entries in the socio-emotional and health fields, one blank learning rating and two invalid or blank entries in the EPE Saboria parent baseline survey were treated as missing; sex was recorded differently between rounds for three children; one child's age is missing at baseline.

What the second-round sites add, and what they cannot

The second-round sites substantially enlarge the quantitative record: they contribute 107 of the 176 children assessed at endline, about 61% of all skill checks, and the program's only fully matched cohorts. They show whether the measured pattern recurs with new cohorts and partly different facilitator teams. They cannot contribute to triangulation. Every finding in this report that draws on parents' or facilitators' accounts, on classroom observation, or on enrolment-record profiles relates to the two first-round sites only, and the text says so where it matters. Where a statement is made about all four sites, it is a quantitative statement.

09

SECTION

Strength of Evidence: What the Pilot Can and Cannot Demonstrate

9.1 What the pilot can demonstrate

- **That the model is deliverable.** All 16 planned sessions were delivered at each first-round site by locally recruited facilitators in two rural communities, in Tetun, using low-cost materials in existing school facilities, at a site-level delivery cost of roughly US$75 per enrolled child (revised costing) for the first-round sites. Sixteen sessions are also recorded at each second-round site, delivered partly by first-round facilitators. This is a genuine operational finding and it does not depend on the assessment data at all.
- **That children's measured performance changed substantially over the cycle.** Measured performance rose on every one of the ten learning indicators at all four sites, on health and hygiene practices, and on socio-emotional behaviours. The change is large, consistent in direction, and present in four cohorts with different ages, starting points and partly different facilitator teams, including two fully matched cohorts at the second-round sites.
- **That the program was acceptable to families and workable for facilitators.** This comes from the independently collected qualitative strand at the two first-round sites and does not depend on the assessment instrument. Every respondent group affirmed the model's value; the criticisms were operational.
- **That content was delivered in the classroom broadly as designed.** All 16 knowledge-development items were rated met at all four observed visits at the first-round sites, by an observer independent of the facilitators.
- **That specific, identifiable operational problems constrain the model.** Scheduling instability, a 1:20 staffing ratio, mixed-age grouping, weak parent communication, unused learning corners and a fragile data system. Each was identified by more than one source, and several were observed directly.
- **That facilitator knowledge can be raised quickly, and that knowledge alone is insufficient.** This is a robust finding because it is corroborated by the facilitators' own franker account and by the observation record, and it is consistent with a well-established literature. The second-round records add that knowledge gains are neither uniform nor necessarily retained.

9.2 What the pilot cannot demonstrate

- **Causal impact.** With no comparison group, the counterfactual is unknown. Children attended primary and pre-primary school throughout, and children aged 3-7 develop rapidly in any case. Some, possibly much, of the measured change would have occurred without the program.
- **The size of any program effect.** Even if the program caused all of the change, the magnitude cannot be expressed in units comparable to other programs, because the instrument is unvalidated and produces ordinal categories rather than a scaled score.
- **Individual progress.** At the first-round sites, fewer than half of assessed children appear in both rounds. Statements of the form "children improved from X to Y" describe two overlapping groups, not the same children twice. At the second-round sites individual progress can be described, but without a comparison group it still cannot be attributed to the program.
- **Persistence.** No measurement was taken after the cycle. The out-of-school-time literature is clear that short-cycle gains frequently fade without continuation, and nothing here speaks to that question.
- **Dose-response.** With attendance at 39-60% of enrolment at the first-round sites, about half of assessed children at the second-round sites, and no per-child totals, the evaluation cannot show that children who attended more gained more - which is the single most accessible internal test of whether a program is working.

- **Transferability.** Two sites in one suku, sharing a principal catchment aldeia, one language environment and one terrain type, is closer to one context than two. The facilitators' own view that Aileu would need a different delivery model implies transferability should not be assumed. The second-round sites widen the base, but their location is undocumented in the data supplied and, without qualitative evidence, how the model operated there is unknown.
- **Cost-effectiveness.** Expenditure is confirmed for about 27% of the revised budget. No cost-per-outcome statement is possible. No cost data were supplied for the second-round sites.
- **Equity of benefit.** Without disability screening and without per-child attendance, the evaluation cannot say whether the children who most needed the program were the ones who received it.
- **Qualitative findings for the second-round sites.** No FGDs or KIIs were conducted at EPE Lauhata or EPE Saboria. Nothing in this report about families' views, facilitators' experience, safety, scheduling or the reasons for attendance patterns can be verified for those sites.

9.3 Specific threats to validity, and how they apply here

Table 9.1. Standard threats to internal validity in a single group pre/post design, applied to this pilot.

Threat	What it means	How it applies here	Severity
Maturation	Participants change naturally over the period regardless of the intervention	Children aged 3-7 develop rapidly over three to four months. Several indicators - counting, shape recognition, story retelling - are exactly the skills that emerge with age. The interval between rounds was 10-12 weeks at the second-round sites.	**High**
History	Something other than the program changed during the period	All children were simultaneously attending school. The pilot cannot separate its contribution from ordinary schooling.	**High**
Testing	Taking the test the first time improves performance the second time	The same 24-indicator tool was administered twice, three to four months apart, with the same tasks. At the second-round sites the interval was 10-12 weeks.	**Moderate**
Instrumentation	The measure itself changes between rounds	The AdJL socio-emotional scale changed between rounds and had to be harmonised analytically. Two different parent instruments were used across sites. At the second-round sites, a within-round break in the LH endline ratings suggests the scale was applied inconsistently.	**Moderate to high**
Selection / attrition	Who is measured changes between rounds	At the first-round sites, fewer than half of children were matched. At AdJL the endline group is smaller (26 of 39) and may over-represent regular attenders; at TX it is larger, with ~21 late joiners. Not applicable at the second-round sites, where all children were assessed at both rounds.	**High**
Regression to the mean	Extreme baseline scores move toward the average on re-measurement	The TX baseline was extremely low (27% independent, 56% not yet able). Some upward movement is expected on statistical grounds alone. The LH (26%) and SB (37%) baselines were also low.	**Moderate**
Expectancy / assessor effects	Assessors who expect improvement record improvement	All assessments were conducted by the implementing team, unblinded, with no independent verification. The near-ceiling SB endline and the repeated SB baseline profiles are the patterns most consistent with this threat.	**Moderate to high**

Threat	What it means	How it applies here	Severity
Hawthorne effects	Participants behave differently because they are being observed and valued	A novel, well-resourced program with visiting staff, certificates and a closing ceremony. Both children and facilitators knew they were being assessed.	**Moderate**
Social desirability	Respondents give the answers they think are wanted	Parent confidence surveys and qualitative interviews were conducted in the context of a program respondents want continued.	**Moderate**
Ceiling effects	The measure cannot record further improvement	Four indicators reached 96-100%; parent confidence reached the scale maximum of 5.0; observation items are binary with three of four visits above 90%. At the SB endline, 22 of 43 children were independent on all ten skills and 99.5% of socio-emotional ratings were "often or always".	**Moderate**

9.4 Why the findings are nonetheless worth acting on

Listing threats is not the same as dismissing findings, and the evaluation does not dismiss them. Four considerations support taking the results seriously while declining to call them impact.

- **Three independent lines of evidence agree.** Assessment data, classroom observation by a different person (ETF Team), and interviews conducted months later by an independent team all describe the same program with the same strengths and the same weaknesses. Measurement artefacts do not usually reproduce themselves across independent instruments and independent observers. This holds for the first-round sites only; the second-round results agree with the pilot pattern in direction but have no independent corroboration of their own.
- **The internal structure of the results is coherent.** Gains are largest where practice was most intensive and hands-on, and smallest on the skills that require distributed retrieval practice. An inflated dataset would be expected to rise more uniformly. The same ordering recurs at EPE Lauhata; it is much weaker at EPE Saboria, where endline ratings are close to uniform.
- **The negative findings were not suppressed.** The same records that report large gains also record an unimplemented program component, a missing facilitator test, a duplicated parent survey, a supervision lapse and a site where socio-emotional gains were modest. A record that reports its own failures is more credible on its successes. The second-round records, which contain unresolved anomalies, are reported with those anomalies stated.
- **The underlying mechanism is well evidenced elsewhere.** Structured early learning opportunities for children who otherwise have none, delivered through active, play-based methods, is among the better-supported interventions in the development literature. The pilot is consistent with that literature; it is not asking to be believed on its own.

The proportionate conclusion is that the pilot provides sufficient evidence to justify a further, better-measured cycle, and insufficient evidence to justify presenting these figures to the Ministry or to funders as demonstrated impact. Section 18 sets out what would close that gap, and the accompanying business case makes the forward-looking argument on that basis. The second-round sites strengthen the case for that further cycle by showing the same direction of measured change in larger, fully matched cohorts, but they do not change its character: they add to the quantity of quantitative evidence, not to its independence or depth.

10

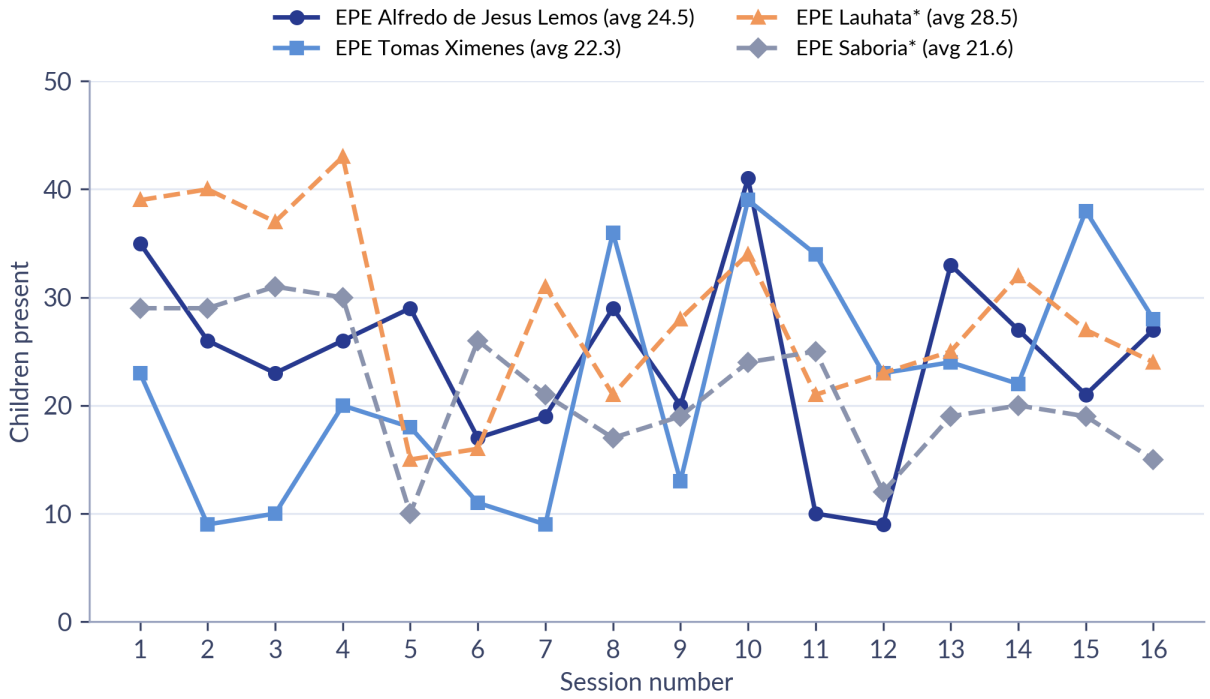
SECTION

Participation and Attendance

10.1 Attendance Summary

Location	Total sessions	Average attendance	Lowest / highest session	Enrolled or assessed (girls/boys)
EPE Alfredo de Jesus Lemos (Railaco, Lihu)	16	24.5	9 / 41	41 (24 / 17)
EPE Tomas Ximenes (Railaco, Lihu)	16	22.3	9 / 39	57 (31 / 26)
EPE Lauhata (second-round site)†	16	28.5	15 / 43	64 assessed (28 / 36)‡
EPE Saboria (second-round site)†	16	21.6	10 / 31	43 assessed (28 / 15)‡
First-round subtotal	**32**	**23.4**	-	**98 (55 / 43)**
All four sites	**64**	**24.2**	9 / 43	**205 (111 / 94)**

Attendance at the second-round sites is taken from the session totals written on each register sheet, because the presence marks are unreliable for several sessions (Section 8.8). ‡ No usable enrolment list exists for EPE Saboria and the EPE Lauhata list does not reconcile with its assessment records, so the children assessed at both rounds are shown, with sex as recorded at baseline.



* Second-round sites: session totals as recorded on each register sheet (see Section 8.8).

Figure 2. Children recorded present per session: first-round sites counted from the presence marks in the attendance registers; second-round sites from the session totals written on each register sheet.

10.2 Analysis

- **Attendance trends:** The two first-round sites show opposite trajectories. EPE Tomas Ximenes averaged only 17 children over its first eight sessions but 27.6 over the last eight - a 62% increase suggesting growing community trust and word of mouth as the program became established. EPE Alfredo de Jesus Lemos started strong (35 children at session 1; a peak of 41 at session 10) but dipped sharply to 10 and 9 children at sessions 11-12 in mid-May before recovering to 21-33 in the final sessions. The second-round sites both started higher than they finished: EPE Lauhata averaged 30.3 children over its first eight sessions and 26.8 over the last eight, and EPE Saboria 24.1 and 19.1, with single-session lows of 15 (EPE Lauhata, session 5) and 10 (EPE Saboria, session 5, the first session after a three-week break). The reasons are not documented, as no qualitative data were collected at these sites.
- **Dropout and cohort continuity:** 20 of 26 endline-assessed children at EPE Alfredo de Jesus Lemos and 22 of 43 at EPE Tomas Ximenes could be matched by name to the baseline cohort. At AdJL, 13 of 39 baseline children did not appear at endline; at TX, the endline cohort included around 21 children who joined after baseline. At the second-round sites the same 64 and 43 children were assessed at both rounds; no child assessed at baseline was missing at endline.
- **Participation consistency:** Average attendance equalled about 60% of enrolment at EPE Alfredo de Jesus Lemos and 39% at EPE Tomas Ximenes. At the second-round sites it equalled about 45% of children assessed at EPE Lauhata (58% of the 49 names on its enrolment list) and 50% at EPE Saboria. Percentages recorded manually in the registers used inconsistent denominators and should be recomputed against a fixed enrolment list in future rounds.
- **Gender participation:** Girls were the majority of enrolments at both first-round sites (59% and 54%) and the large majority of the EPE AdJL endline cohort (19 of 26), while the EPE TX endline cohort was majority boys (24 of 43). No systematic gender gap in attendance appears in the registers. At the second-round sites, girls were 44% of assessed children at EPE Lauhata and 65% at EPE Saboria; the written session totals show girls in the majority at 15 of 16 EPE Saboria sessions and boys at 12 of 16 EPE Lauhata sessions.
- **Dosage is unknown at child level:** The registers were not summarised per child, so the evaluation cannot say how many sessions any individual child attended. Given that the average child attended well under half the enrolled roster's worth of sessions, and that assessment happened regardless of attendance, some children counted in the endline results will have received very little of the program. This is the most important unmeasured variable in the dataset. The same applies at the second-round sites, where the presence lists are unreliable for several sessions.

10.3 Explaining attendance volatility

The attendance registers record what happened but not why. For the two first-round sites, the qualitative round supplies the missing explanation, and the causes reported map closely onto the observed pattern. No equivalent explanation is available for the second-round sites, and the drivers in Table 10.1 should not be assumed to apply there.

Table 10.1. Reported drivers of attendance, against the register evidence (first-round sites).

Reported driver	Source	Consistency with attendance data
Rain and wet-season weather	Parents (interviews and FGD); facilitators (FGD)	Consistent with the sharp mid-May dip at AdJL during the rainy tail-season. Cited independently by every respondent group.

Reported driver	Source	Consistency with attendance data
Difficult terrain, including crossing large rivers	Parents (FGD)	Explains why weather has an outsized effect: rain does not merely deter attendance, it can make the journey impassable.
Schedule shift to morning sessions	Parents (FGD); facilitators (FGD); confirmed by observation dates	Excluded children more available in the afternoon; a plausible contributor to sustained rather than one-off attendance loss.
Child tiredness and limited attention in after-school slots	Facilitators (interviews)	Suggests attendance and learning quality are affected by timing in opposite directions - the slot that maximises attendance may not maximise attention.
Older children dominating younger ones	Parents (FGD); directly observed by ETF on 14 April	Reported to have discouraged some younger children from continuing - a dropout mechanism invisible in the registers.
Facilitators occasionally arriving late	Parents (FGD); observation records late starts at both AdJL visits	Erodes the reliability that parents said drove their enrolment decision.
Family circumstances and occasional illness	Parents (interviews)	Consistent with normal episodic absence.
Missed invitations and communication failures	Parents (FGD)	Children were an unreliable channel for relaying information; limited internet access blocked digital alternatives.

Two conclusions follow. First, the largest attendance losses are addressable through operational decisions - timetable stability, communication systems and age grouping - rather than through demand generation; at one school, we noted increasing trends for enrolment which was justification of 'growing community trust and word of mouth as the program became established. Second, weather-related absence in this terrain cannot be eliminated and should be planned for, through make-up sessions or a longer cycle that absorbs lost days.

A third point deserves emphasis because it is a measurement issue as much as an operational one. Attendance in after-school programs is a known determinant of effect: programs with low or irregular participation produce weaker outcomes, and dosage effects are consistently observed in the out-of-school-time literature. With average attendance at 39-60% of enrolment at the first-round sites (about 45-50% of assessed children at the second-round sites) and no per-child attendance record, the pilot cannot examine whether children who attended more gained more. Establishing that relationship in the next cycle would do more to strengthen the evidence base than any other single analytical addition.

11

SECTION

Learning, Behavioural and Capacity Outcomes

The results in this section describe measured change between two rounds. They are presented as measured, with the caveats from Section 8 carried forward: the first-round samples are not matched, the instrument is not validated, and the assessors were the implementing team. Section 9 interprets what these results can and cannot support. Results for the second-round sites rest on assessment records alone, without the corroboration available for the first-round sites, and carry the data-quality caveats set out in Section 8.8.

11.1 Learning outcomes

Children were assessed at baseline (Liña Baze) and endline (Liña Ikus) on ten learning-skill indicators covering literacy (letter recognition, letter sounds, story retelling), numeracy (counting 1-10, shapes, more/less), and problem-solving (puzzles, sorting, block construction, flexible problem- solving). Each skill was rated "independent" (Sin), "with help" (Ho ajuda), or "not yet able" (Lae).

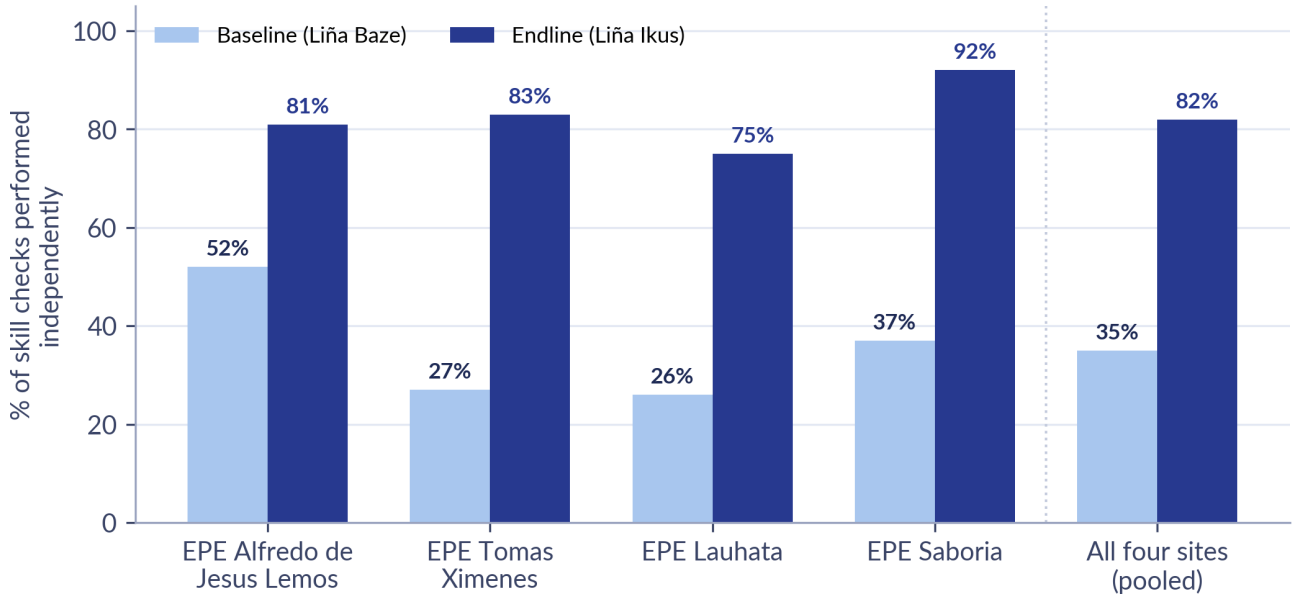
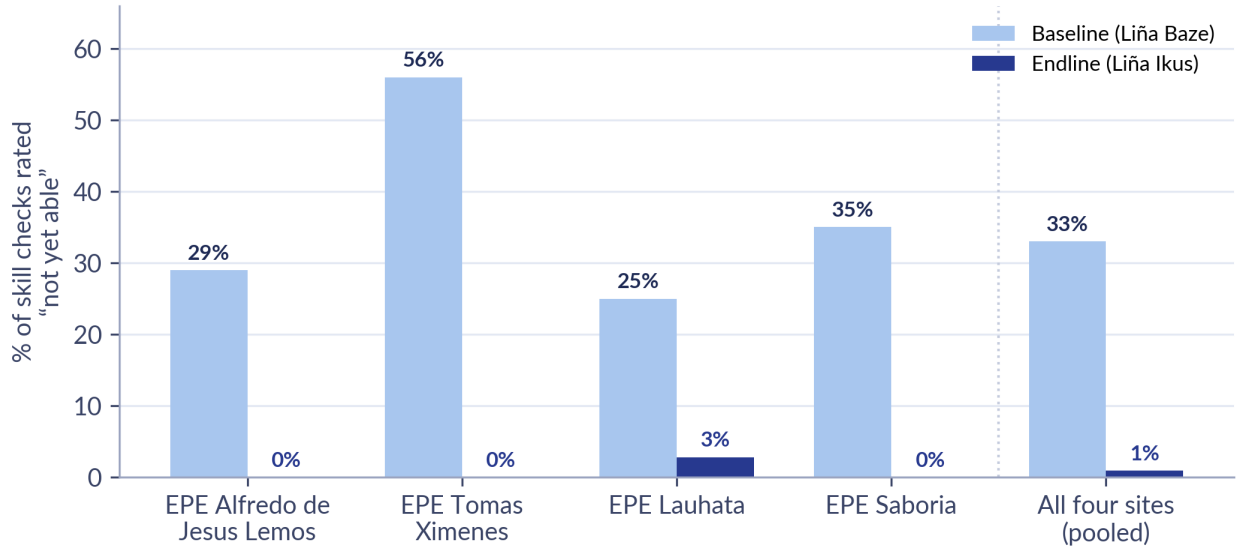


Figure 3. Share of all learning-skill checks performed independently, baseline versus endline, by site and pooled across all four sites.



First-round site baseline values for Lae are derived from published site percentages (±1 point).

Figure 4. Share of learning-skill checks rated "not yet able" (Lae) - eliminated at endline at three of the four sites.

Among the first-round sites, EPE Tomas Ximenes recorded the steepest change: independent performance tripled from 27% to 83% of skill checks, in a younger cohort (mean age 4.8) and from a much lower starting point. Item-level results show letter recognition rising from 4 of 28 children independent at baseline to 27 of 43 at endline, with the remaining 16 able with help; counting 1-10 rising from 2 of 28 to 21 of 43 independent; and understanding of "more/less" reaching 42 of 43 children independent. At EPE Alfredo de Jesus Lemos, every one of the 26 endline children could build with blocks and try multiple problem-solving strategies independently, and story retelling rose from 21 of 39 to 22 of 26.

At the second-round sites, independent performance rose from 26% to 75% of skill checks at EPE Lauhata and from 37% to 92% at EPE Saboria. Because every child was assessed at both rounds, change can also be described for individual children: 57 of 64 children at EPE Lauhata and all 43 at EPE Saboria performed more skills independently at endline than at baseline, while two EPE Lauhata children were unchanged and five performed fewer. The average number of the ten skills performed independently rose from 2.6 to 7.5 per child at EPE Lauhata and from 3.7 to 9.2 at EPE Saboria. At EPE Lauhata, however, 13 children still had at least one skill rated "not yet able" at endline, mostly in story retelling (7 ratings), counting (6) and shapes (3).

Pooled across all four sites, independent performance rose from 35% of 1,740 baseline skill checks to 82% of 1,759 endline checks, and "not yet able" ratings fell from 33% to 1%. The pooled figures are weighted towards the second-round sites, which contribute about 61% of all skill checks.

How sensitive are the learning results to the data-quality problems?

Because the second-round records contain anomalies (Section 8.8), the pooled learning result was recalculated under three alternative assumptions. Excluding the 16 EPE Saboria children with identical baseline profiles, the share of skill checks performed independently rises from 34% to 81%. Excluding EPE Saboria entirely, it rises from 34% to 79%. Using the first-round sites alone, it rises from 42% to 83%. The endline figure stays between 79% and 83% under every assumption, so the direction and approximate size of the measured change do not depend on the questionable records. What the anomalies do affect is confidence in the site-level figures for EPE Saboria and in the EPE Lauhata socio-emotional and health results.

The EPE Lauhata endline split is not explained by the children. The 24 children in the second block of records started from the same baseline as the first 40 (25% and 26% of skill checks independent) but finished far lower (45% against 92%), gaining on average 2.0 skills against 6.6. Since the two groups were alike at baseline, the difference at endline more probably reflects who assessed them and how, or how often those children attended, than any difference between the children themselves. The records cannot distinguish these explanations; a re-assessment of a sample from both blocks by a single independent assessor would (Section 18, action V1).

Results by sex and age (second-round sites)

Because the second-round cohorts are fully matched, their results can be broken down by sex and age. Girls and boys gained similarly: at EPE Lauhata, independent performance rose from 24% to 77% of skill checks for girls and from 28% to 73% for boys; at EPE Saboria, from 36% to 92% for girls and from 38% to 92% for boys. Younger children gained less at EPE Lauhata, where children aged 3-4 rose from 17% to 60% and those aged 5-7 from 31% to 83% (one child's age is missing). At EPE Saboria the two age groups rose from 34% to 90% and from 39% to 94%. The gap at EPE Lauhata supports the case for age-banded groups and extra support for the youngest children. The first-round datasets were not re-supplied with the second-round files, so the same breakdown cannot be produced for those sites.

Table 11.1. Percentage of children performing each skill independently: first-round sites, by round.

Skill indicator (assessed in Tetun)	AdJL baseline (n=39)	AdJL endline (n=26)	TX baseline (n=28)	TX endline (n=43)
Recognises at least 10 letters	44%	73%	14%	63%
Identifies basic letter sounds/names	41%	62%	18%	74%
Retells a simple story	54%	85%	11%	93%
Counts 1-10 without help	62%	69%	7%	49%
Identifies basic shapes	26%	69%	11%	77%
Understands "more" and "less"	59%	85%	14%	98%
Completes simple puzzles (3-6 pieces)	54%	73%	46%	86%
Sorts objects by colour or shape	62%	96%	82%	98%
Builds/repairs structures with blocks	62%	100%	32%	98%
Tries multiple ways to solve a problem	62%	100%	36%	98%

At endline, every skill check at both first-round sites and at EPE Saboria was rated at least "with help"; at EPE Lauhata, 18 endline checks (3%) were still rated "not yet able". The first-round columns describe different groups of children, not the same children measured twice.

Table 11.2. Percentage of children performing each skill independently: second-round sites and all four sites combined.

Skill indicator (assessed in Tetun)	LH baseline (n=64)	LH endline (n=64)	SB baseline (n=43)	SB endline (n=43)	All sites baseline (n=174)	All sites endline (n=176)
Recognises at least 10 letters	30%	53%	16%	88%	27%	67%
Identifies basic letter sounds/names	19%	66%	9%	88%	21%	73%
Retells a simple story	8%	66%	14%	95%	20%	82%
Counts 1-10 without help	19%	66%	2%	91%	22%	68%
Identifies basic shapes	23%	59%	21%	79%	21%	70%
Understands "more" and "less"	41%	86%	40%	98%	40%	91%
Completes simple puzzles (3-6 pieces)	16%	97%	67%	98%	42%	91%
Sorts objects by colour or shape	53%	83%	67%	98%*	63%	92%*
Builds/repairs structures with blocks	30%	95%	72%	98%	48%	97%
Tries multiple ways to solve a problem	23%	77%	60%	91%	43%	89%

At the second-round sites the same children were assessed at both rounds. All-sites columns are calculated from summed counts across the four sites; first-round counts are reconstructed exactly from Table 10.1. * EPE Saboria endline sorting: n=42 (one blank rating); all-sites n=175.

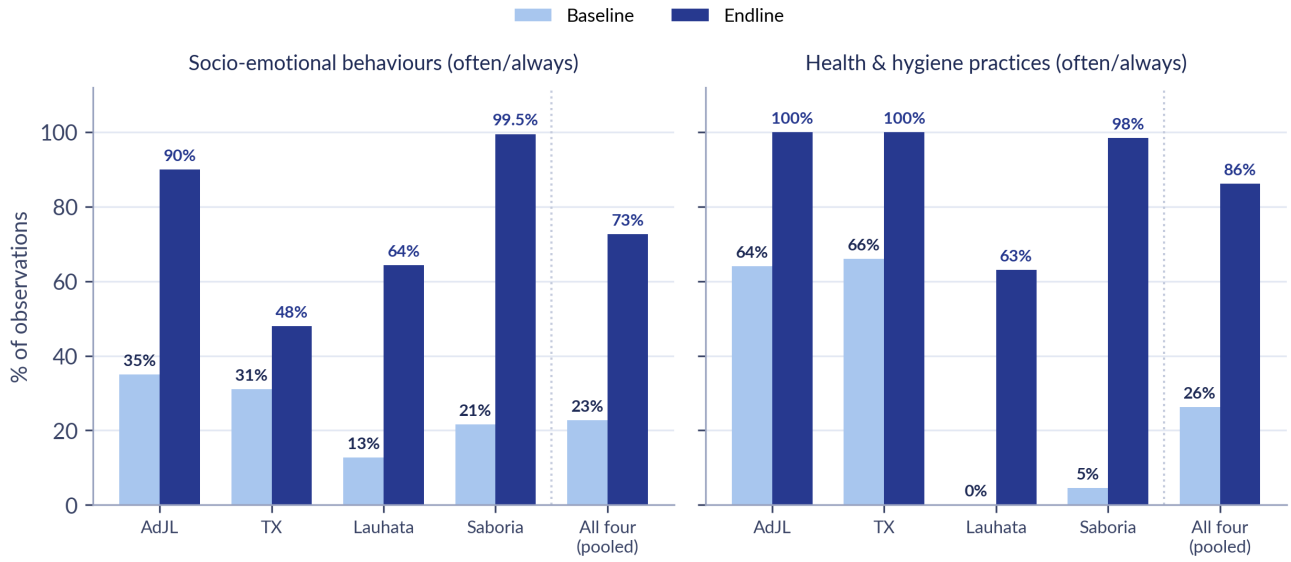
Reading the pattern

Three features of these results are worth drawing out, and two of them are cautionary.

- **The pattern across items is internally coherent.** The largest gains are on the most game-like, materially supported indicators - sorting, block construction, flexible problem-solving, more/less - which are also the activities facilitators identified as most effective and which children practised most. The smallest gains are on counting without help and letter recognition, which depend on repeated retrieval practice over time. A fabricated or uniformly inflated dataset would not normally show this structure, and the differentiation is mild evidence that the assessment is tracking something real. The same broad ordering recurs at the second-round sites: problem-solving items reach 77-98% independent at endline, while letter recognition (53%) and shapes (59%) are weakest at EPE Lauhata and shapes (79%) at EPE Saboria. Pooled across all four sites, letter recognition (67%), counting (68%) and shapes (70%) are the three weakest endline indicators. The EPE Saboria endline is a partial exception: with nine of ten indicators at 88% or above it is close to uniform, and it adds little to the differentiation argument.

- **Two indicators warrant continued attention.** Counting 1-10 without help is the weakest endline result at EPE Tomas Ximenes (49%), and letter recognition the second weakest (63%). These are precisely the skills that benefit most from distributed practice over time, which supports extending cycle length beyond 16 sessions. Across all four sites the same two indicators, with shape identification, remain the weakest at endline (67-70% independent).
- **Ceiling effects limit what the top-end figures mean.** Four indicators reach 96-100% at endline at the first-round sites; at EPE Saboria four reach 98%, and 22 of 43 children were rated independent on every skill. On a three-point scale with no gradation above "independent", the instrument cannot distinguish a child who has just mastered a skill from one who mastered it long ago, and cannot register further progress. A more granular scale, or a validated instrument administered alongside, would resolve this.

11.2 Behavioural outcomes



Site values are as published; pooled values combine observation counts, with first-round site counts derived from published percentages (±1 point).

Figure 5. Socio-emotional behaviours and health and hygiene practices observed "often or always," by site and pooled across all four sites.

- At EPE Alfredo de Jesus Lemos, socio-emotional behaviours (sharing, turn-taking, empathy, appropriate expression of feelings, self-control) were observed "often or always" in 90% of endline observations, up from 35% at baseline, with no "never" ratings at endline. This figure depends on the scale harmonisation described in Section 8.4 and should be treated as the least robust of the headline numbers.
- At EPE Tomas Ximenes, "often or always" ratings rose from 31% to 48%, with "never" ratings absent in both rounds. The more modest gain is consistent with a younger cohort and a large influx of newly joined children at endline; the direction of change is nonetheless positive across all five behaviours.
- At EPE Lauhata, "often or always" ratings rose from 13% to 64%, and "never" ratings fell from 25% to 8%. As described in Section 8.8, the endline figure averages two very different blocks of records: 100% for the first 40 children and 5% for the remaining 24.

- At EPE Saboria, ratings rose from 21% to 99.5% (214 of 215 ratings), and “never” ratings fell from 54% to under 1%. A result this close to the scale maximum is the least informative kind of result the instrument can produce.
- Pooled across all four sites, socio-emotional behaviours were observed “often or always” in about 23% of baseline and 73% of endline ratings.
- Facilitators at the first-round sites reported improved confidence, calmer classrooms, and greater willingness of three-year-olds to participate in group activities following the introduction of small-group play techniques after training.

At the first-round sites, parents confirmed the behavioural findings independently, reporting improved listening and behaviour, greater discipline, friendship-building through games, and a marked shift from passive phone use toward creative and hands-on activity. Several parents reported children teaching newly learned skills back to siblings and to the parents themselves - an outcome the assessment instrument does not capture but which, if accurate, indicates genuine consolidation of learning. There is a counterweight worth noting: the classroom observer rated “children control their frustration or sadness” as not met at both EPE Alfredo de Jesus Lemos visits, the site that later recorded the largest socio-emotional gain. Section 14.3 discusses this. No parent or facilitator accounts are available for the second-round sites, so their behavioural results cannot be corroborated in the same way.

11.3 Health and physical development outcomes

- Handwashing before eating and after toilet use, identification of healthy versus unhealthy foods, and running, jumping and balance coordination were rated “often or always” for 100% of endline observations at both first-round sites, up from 64% at EPE Alfredo de Jesus Lemos and 66% at EPE Tomas Ximenes at baseline.
- At the second-round sites, the same practices rose from 0% to 63% “often or always” at EPE Lauhata and from 5% to 98% at EPE Saboria. Baseline ratings at both second-round sites were far lower than at the first-round sites - almost all “never” or “sometimes” - and whether this reflects genuinely lower practice or stricter rating cannot be determined. Pooled across all four sites, the share rose from about 26% to 86%.
- Post-training, facilitators at both first-round sites positioned soap at handwashing taps, controlled access to sharp or damaged materials, and embedded hygiene routines into every session.
- Parents at the first-round sites reported children applying environmental and practical skills at home, including raking soil and planting seedlings - evidence that the environmental-care component transferred beyond the session venue.

This is the domain where the evidence is strongest, because it is the only one where an independent-of-assessment record corroborates the change: the classroom observation forms document the physical infrastructure (soap present, water available, functioning toilet) and the behaviour (children washing hands after playing with soil at all four visits). Physical evidence of a practice is a harder measure than a rating of it. Note also that the observer qualified two hygiene items at EPE Tomas Ximenes as “yes, but only some children” at both visits, which sits awkwardly with a 100% endline rating and illustrates why the binary observation scale should be graded. That corroboration exists for the first-round sites only. At the second-round sites the health results rest on assessment ratings alone, and the EPE Lauhata endline (63%) shows that the 100% pilot result was not universal.

11.4 Facilitator outcomes

Facilitator	Site	Pre-training score (of 6)	Post-training score (of 6)
Facilitator 1	EPE Alfredo de Jesus Lemos	2	6
Facilitator 2	EPE Alfredo de Jesus Lemos	2	6
Facilitator 1	EPE Tomas Ximenes	2	6
Facilitator 2	EPE Tomas Ximenes	3	*Not recorded*
Facilitator 1	EPE Lauhata	3	5
Facilitator 2	EPE Lauhata	1	4
Facilitator 1	EPE Saboria	1	3
Facilitator 2	EPE Saboria	2	3

Second-round scores: domains recorded as 5 are scored as met and those recorded as 1 as not met. Pre-tests were taken on 13 July (EPE Lauhata) and 16 July 2026 (EPE Saboria), after the first sessions; post-tests between 31 August and 7 September 2026. The two EPE Saboria facilitators also delivered the pilot at EPE Alfredo de Jesus Lemos.

Pre-training, facilitators understood child development stages based on baseline and endline assessments but did not systematically use play-based activities, did not organise materials for children's independent access, and were not managing first-aid or hygiene materials. Post-training assessments document concrete changes: use of songs and games to calm and group children, inclusive seating for the youngest, soap placed at handwashing points, first-aid materials prepared, and environmental-care routines introduced on facilitators' own initiative.

These scores measure knowledge, not applied confidence, and the qualitative round shows the two diverge. Facilitators' own account of their capability is set out in Section 12.4 and the divergence is analysed in Section 13.2. A test administered by the trainer immediately after training, with three of four scoring the maximum, measures recall of just-delivered content; it is not a measure of capability.

The second-round scores are lower and more uneven. Post-training, no second-round facilitator reached 6 out of 6: the two EPE Lauhata facilitators scored 5 and 4, with neither meeting the safe and inclusive learning environment domain, and the two EPE Saboria facilitators scored 3 each, with neither meeting the health, hygiene and first-aid domain. One EPE Saboria facilitator was rated as meeting the environmental awareness domain before the cycle but not after it. Because the two EPE Saboria facilitators had scored 6 out of 6 after the first-round training, these results suggest either that knowledge was not retained between cycles or that the test does not produce stable scores. In either case they reinforce the conclusion that knowledge scores are not evidence of capability.

11.5 Parent outcomes

- At EPE Alfredo de Jesus Lemos, 17 parents and caregivers rated their confidence to support their child in four areas (reading and storytelling; counting and basic maths; educational games; managing emotions and behaviour) on a 1-5 scale. Mean confidence rose from 3.3 at baseline to 5.0 at endline; the share of ratings at 4 or 5 rose

from 57% to 100%. The weakest baseline area, playing educational games (mean 2.6), reached the maximum score at endline. While these results represent a perfect increase, qualitative data revealed a story beyond their confidence of the knowledge.

- At EPE Tomas Ximenes, 17 parents completed an eight-domain categorical survey. At baseline, 77% of responses indicated strong confidence (Konfidente tebes), with educational games (53%) and homework support (65%) the weakest domains. The endline records for this site duplicate the baseline entries, including the survey date, so no endline change can be reported for this site.
- At EPE Lauhata, 45 parents and caregivers completed the eight-domain survey on a 1-5 scale at baseline (30 June) and endline (23 September 2026). Mean confidence rose from 3.0 to 4.6, and the share of ratings at the top of the scale from 50% to 91%. The weakest baseline domains were art and creativity (16%) and educational games (18%); at endline the weakest was art and creativity (76%).
- At EPE Saboria, 34 parents completed the same survey. Mean confidence rose from 2.4 to 4.4 and the share of top ratings from 36% to 85%. Educational games and art and creativity were the weakest baseline domains (3% each). Homework support was the only domain at either second-round site to fall, from 68% to 62% top ratings. The endline survey dates are auto-filled and cannot be verified (Section 8.8).

> A mean of exactly 5.0 - every respondent selecting the top option on every domain - is a result that should prompt scrutiny rather than celebration. It is possible, and the qualitative evidence of parental enthusiasm is genuine. It is also the signature of a self-report instrument administered by the program team to a small, engaged group of parents at the end of a program they want continued, and it leaves no headroom to detect further improvement. The qualitative round provides partial compensation for the missing EPE Tomas Ximenes survey: parents at that site described concrete changes in their children and expressed clear support for the program. The second-round surveys show a related limitation: respondents used only the lowest and highest points of the five-point scale, so the instrument recorded whether parents felt confident rather than how confident. The direction of change is consistent across the three sites with usable endline data, but no qualitative evidence is available to test it at the second-round sites.



12

SECTION

Qualitative Findings: Parent and Facilitator Perspectives

This section presents the findings of the qualitative round conducted by the consultant team in August 2026. They are reported as the perceptions of respondents. Where a perception is corroborated by, or conflicts with, the quantitative or observational record, that is noted in Section 13. The qualitative round was conducted at EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes only. No FGDs or KIIs were conducted at EPE Lauhata or EPE Saboria, and none of the findings in this section can be attributed to, or assumed to hold for, those sites.

What the qualitative round adds to the evaluation

- It identifies scheduling as the single most consequential operational decision in the model - a finding entirely absent from the quantitative dataset.
- It reveals that a designed component, the learning corners, was rarely delivered due to unexpected time on other program components which no quantitative indicator would have detected.
- It supplies the causal explanation for attendance volatility that the registers alone could not provide.
- It documents an unmeasured outcome: children transferring learning to siblings, parents and the home environment.
- It confirms that demand-side belief in the program is already strong, so effort should concentrate on supply-side reliability rather than persuasion.
- It is the only strand of evidence in this evaluation collected independently of the delivery team.
- It covers the two first-round sites only; there is no equivalent for EPE Lauhata or EPE Saboria, whose results therefore cannot be explained or corroborated from respondents' accounts.

12.1 Parent interviews

Satisfaction with the program

Despite children having attended for less than three months at the time of the evaluation, all three parents interviewed agreed the program is beneficial. Children enjoy the sessions, are progressing in their learning, and facilitators were described as highly supportive. While the program was not compared to formal pre-school or primary school, some participants mentioned that the ASP provided children with more opportunities for independent learning and participation.

Information and enrolment decisions

Parents learned about the program through the school, the ETF team, or local authorities. Two factors drove enrolment: proximity to home, and the perception of the program as a meaningful learning opportunity rather than unstructured play at home. The second point matters for expansion messaging - families were not primarily seeking childcare; they were seeking structured learning.

Observed changes and most-valued aspects

Parents observed children studying independently, becoming more creative (for example through games in the Lafaek magazine), and in some cases teaching skills back to their parents and siblings. Parents most valued children's exposure to science, friendship-building through games, and the discipline fostered by extracurricular activities.

Challenges and safety

Weather, family circumstances and occasional illness were cited as barriers to attendance. All parents agreed their children feel safe in the program, crediting attentive facilitators and children returning home safe and content. No parent raised a child-protection concern.

Recommendations from parent interviews

- A more consistent and predictable schedule.
- Weekly information updates to parents.
- Continuity of the program, avoiding pauses between cycles.
- Broader outreach so that more children can participate.
- Community sensitisation on the value of extracurricular learning.
- Additional learning materials and toys, and dedicated storage for them.
- Fence protection at implementation sites, which may be a recommendation for the learning facility itself rather than the program.

12.2 Parent focus group discussion

Benefits and observed changes

Parents reported children gaining literacy, numeracy, drawing and social skills, along with a shift from passive activities (phone use) to creative and hands-on activities (crafts from leaves, drawing). Parents also noted improved listening and behaviour, and new practical skills such as raking soil and planting seedlings.

Participation barriers

- Rain, and difficult terrain including the crossing of rivers.
- Child tiredness.
- Older children dominating younger ones, which discouraged some younger children from continuing.
- A schedule shift to mornings, which excluded children who were more active in the afternoon.
- Facilitators occasionally arriving late.
- A shortage of facilitators relative to the number of children.

Communication barriers

- Limited internet access among parents.
- Incomplete certificate distribution.
- Missed invitations to program events.
- Children proving an unreliable channel for relaying information from facilitators to households.

Parents proposed a concrete remedy: collecting phone numbers organised by aldeia, so that information can be distributed through a structured local channel rather than through children.

Suggestions from the parent FGD

- A consistent schedule, coordinated with school teachers.
- Continuing the program for at least six months; if limited to three months, running daily sessions instead.
- Prioritising children aged 10 and under, especially those with weaker academic skills.
- Grouping children by age.
- Scheduling activities after school hours.

One respondent noted that some children are hesitant to participate, potentially because of the facilitator or teacher. This observation came from a single respondent and is not corroborated elsewhere in the dataset, but it warrants documented follow-up during monitoring visits given its child-protection and engagement implications (Section 5.2).

> "Bring children to the After School Program because they have an opportunity to learn something new."
>
> - Mother of a participating child, EPE Tomas Ximenes

12.3 Facilitator focus group discussion

Implementation experience

As the program is new, facilitators reported not yet having full confidence to implement the guidelines, but viewed the experience positively as an opportunity to develop skills in guiding and managing children. Many had no prior experience managing groups of children aged 3-7 and instead drew on informal, home-based experience such as calming crying children and building connection with them.

Most effective activities

The opening and closing sessions (1-5 and 16), which include science, painting, games and creative activities, were identified as most effective and best received by children. By contrast, the learning "corners" (reading, language and homework) were rarely implemented in practice - a designed element of the model that did not reach children.

Challenges

- Poor attendance during rain.
- The schedule change to morning sessions.
- Limited skills for managing large groups of children.
- Low engagement when using standard school classrooms, attributed to a familiar and unstimulating environment.
- Lack of Training of Trainers attendance for some facilitators.
- Difficulty adhering to the time allocations set out in the guidelines.

Training

Existing training on the guidelines was seen as clear and effective. Facilitators nonetheless requested further training aligned with the Ministry's curriculum, along with training on child protection and on classroom and behaviour management strategies.

Suggestions from the facilitator FGD

- Holding activities at alternative locations rather than only in schools.
- Avoiding overlap between program sessions and formal schooling.
- Providing transport funds in the future implementation and transport for learning materials.
- Strengthening community engagement.
- Conducting a needs survey before implementation, rather than defaulting to sites where ETF has already built a school.
- For Aileu specifically, running the program as a continuation of the school day rather than an afternoon activity, since parents there do not support afternoon sessions.
- Ongoing sensitisation and consultation with parents.

12.4 Facilitator interviews

Confidence in leading the program

Facilitators interviewed individually reported a high level of confidence in their ability to lead the program. This contrasts with the collective account given in the FGD, where facilitators described not yet having full confidence to implement the guidelines. The divergence is examined in Section 13.2.

Most valuable training content

Facilitators identified the "All About Me", environmental care, arts and creativity, and science modules as the most important components of their training.

Implementation challenges

- An after-school start time was difficult, as children were often tired and had limited attention.
- Parents were reported to disagree with afternoon scheduling.
- Managing a large group of 39 children spanning a 3-7 age range in a single room was cited as a further significant challenge.

Support needs and mentoring

Facilitators requested additional training on child management and refresher training on program content. They confirmed they receive support from the ETF team but would benefit from more frequent and in-depth mentoring. They also noted that while the ETF program is well regarded, parent awareness and understanding remain limited.

12.5 Cross-cutting themes

- Both parents and facilitators consistently viewed the program as beneficial for children's academic, creative and social development. No respondent group questioned the value of the model itself.

- Scheduling is the most frequently cited operational challenge across all respondent groups.
- Facilitator capacity - training, classroom management skills and ToT coverage - is an emerging gap, especially given large class sizes and wide age ranges.
- Communication between the program, facilitators and parents needs strengthening, particularly given limited internet access and unreliable child-mediated communication.
- Both groups independently recommended greater community sensitisation and consultation with parents.

13

SECTION

Synthesis and Triangulation of Evidence

13.1 Where the evidence converges

Table 13.1. Findings corroborated across multiple evidence strands (corroboration available for the first-round sites only).

Finding	Quantitative evidence	Classroom observation	Qualitative evidence
Children's foundational learning improved	Independent performance 52%→81% (AdJL), 27%→83% (TX); "not yet able" absent at endline at both first-round sites. Second-round sites, assessment only: 26%→75% (LH), 37%→92% (SB)	All 16 knowledge-development items met at all four visits	Parents report children studying independently and progressing; facilitators report classroom gains
Play-based, creative activity is the engagement driver	Most game-like indicators (sorting, blocks, flexible problem-solving) reached 96-100% at the first-round sites and 77-98% at the second-round sites	Songs, games and local materials in use at every visit	Facilitators name science, painting, games and creative sessions as most effective
Hygiene practice changed	64-66% → 100% "often or always" (first-round sites); 0% → 63% (LH) and 5% → 98% (SB), uncorroborated	Soap, water and functioning toilet present; handwashing after soil play at all four visits	Parents report children applying practical and environmental skills at home
Children transfer learning to the home	Not measured	Not measured	Parents report children teaching siblings and parents; new practical skills such as planting
Weather and terrain suppress attendance	Mid-May dip to 9-10 children at AdJL during the rainy tail-season	Late arrivals recorded at both AdJL visits	Rain and river crossings cited independently by parents and facilitators
The program is a safe environment	Enrolment health screening; no incidents recorded	Safety section fully met at both TX visits; soap and first aid prepared; sharp-tool hazard identified and acted on	All parents agreed children feel safe; facilitators described as attentive; no child-protection concern raised
Mixed-age grouping causes problems	Not measured	Observer recorded 8-9 year-olds dominating younger children on 14 April	Parents report the same dynamic and say it discouraged some younger children from continuing

13.2 Where the evidence diverges

Divergence 1: Facilitator capability

The quantitative record shows facilitator knowledge scores rising from 2-3 out of 6 to a maximum of 6 through pre/post test by ETF Timor team. Facilitators interviewed individually described high confidence in leading the program. Yet in the group discussion the same cadre reported not yet having full confidence to implement the guidelines, no prior experience with 3-7 year-olds, limited skills for managing large groups, difficulty keeping to prescribed time allocations, and incomplete ToT coverage. At the second-round sites, post-training scores of 3-5 out of 6, including 3 out of 6 for two facilitators who had previously scored the maximum, widen the gap between the knowledge measure and any claim of capability.

These are not contradictory so much as measuring different things. The post-training test measures recall of just-delivered content; individual interviews invite positive self-presentation; the group setting produced a franker account of applied practice. The operational conclusion is that knowledge transfer succeeded and skills transfer did not - facilitators know what to do but are not yet equipped to do it with 39 children of mixed ages in one room. As Section 5.4 sets out, this is the expected result of a training without coaching model, and the meta analytic evidence points to a specific remedy: structured coaching, which produces effects of around 0.49 standard deviations on instructional practice and works best when paired with group training rather than replacing it (Kraft, Blazar & Hogan, 2018).

Divergence 2: Session timing

Parents in the FGD recommended scheduling after school hours and objected to the shift to mornings. Facilitators reported the opposite: that the after-school slot left children tired and inattentive, and that parents did not support afternoon sessions because of competing commitments. Facilitators recommended, for future implementation, running the program as a continuation of the school day.

The most likely reading is that optimal timing is genuinely site-specific, varying with the school day, the agricultural calendar and household labour patterns in each community. A single centrally imposed timetable will be wrong in some locations by construction. The implication is that timing must be negotiated per site and then held stable: the pilot's error was less the choice of slot than the mid-cycle change and the late notice.

Divergence 3: Site-level socio-emotional gains

EPE Alfredo de Jesus Lemos recorded socio-emotional gains from 35% to 90% "often or always", while EPE Tomas Ximenes moved only from 31% to 48%, despite the latter posting the larger learning gains. Three explanations are available and the evaluation cannot choose between them. The first is compositional: the TX endline cohort included roughly 21 children who joined after baseline and had therefore received less exposure. The second is the mixed-age dynamic, which would depress precisely the sharing, turn-taking and self-regulation behaviours this domain measures. The third is methodological: the EPE AdJL figure depends on the scale harmonisation described in Section 8.4, and is the one headline number in this report that rests on an analytical inference rather than a direct reading. The second-round sites reproduce the spread rather than resolve it: EPE Saboria recorded the largest socio-emotional gain of any site (21% to 99.5%) and EPE Lauhata a moderate one (13% to 64%), with the EPE Lauhata figure itself split between two blocks of records. That the widest variation between sites occurs in the domain most dependent on assessor judgement is consistent with the methodological explanation.

Divergence 4: Socio-emotional ratings versus what the observer saw

This divergence was not identified in the original draft and is the most awkward in the dataset. "Children control their frustration or sadness" was rated as not met at both EPE Alfredo de Jesus Lemos visits, and "children share materials or games with others" as not met at the first - at the site that subsequently recorded a rise from 35% to 90% on exactly these behaviours. Both items were met at EPE Tomas Ximenes, the site that recorded the smaller gain. The observations were made in week 5 of a 16-session cycle and the endline in June, so the difference may reflect genuine change over the cycle. It may equally reflect the limits of a single binary observation, or a difference between how an external observer and a facilitator-assessor rate the same behaviour. The honest position is that the socio-emotional finding is the least secure in this report.

Divergence 5: Health practice across sites

Both first-round sites reached 100% "often or always" on health and hygiene, and classroom observation corroborated the practice. EPE Lauhata reached 63%, from a baseline of 0%, while EPE Saboria reached 98%. With no observation or qualitative data at EPE Lauhata, the evaluation cannot say whether the lower result reflects weaker delivery of the health component, stricter rating, or the split in the endline records, but it shows that the first-round result should not be generalised.

13.3 The pilot against an evidence-based design standard

It is useful to assess the program against the design features that distinguish effective after-school programs from ineffective ones. Durlak, Weissberg and Pachan's meta-analysis found that gains in personal and social skills were concentrated in programs meeting four criteria, abbreviated SAFE. The pilot's position against each is set out below. This is an assessment of design conformity, not of outcome, and it draws on first-round evidence only.

Table 13.2. The pilot against the SAFE design criteria (Durlak, Weissberg & Pachan, 2010).

Criterion	What it requires	Pilot position
Sequenced	A connected, coordinated set of activities building toward the skill, rather than isolated sessions	Partially met. A 16-session sequence exists in the guidelines, but facilitators reported the most effective content concentrated in sessions 1-5 and 16, and prescribed timings were not followed. Sequence as designed was not sequence as delivered.
Active	Active forms of learning that require children to practise the skill, not passive receipt	Met. This is the pilot's clearest strength. Play-based, hands-on activity was consistently delivered and consistently identified as effective by facilitators, parents and the observer.
Focused	Time explicitly and sufficiently allocated to skill development	Not met. Guideline time allocations were reported as impossible to keep, one observed session overran by an hour, and the learning corners - the component with dedicated skill time - were largely unused.
Explicit	The specific skills being developed are named and targeted, not left implicit	Partially met. Learning and health skills are explicit in the assessment tool and the activities. Socio-emotional skills are assessed but were not observed being explicitly taught, which may help explain why this is the weakest outcome domain.

The pattern is coherent: the program is strong where it is active and weak where it requires structure, timing discipline and explicit skill instruction. That maps precisely onto the facilitator practice gap identified in Section 13.2, and it suggests that the highest-return improvements are the coaching and fidelity measures rather than new curriculum content.

13.4 What remains unknown

- Whether learning gains persist after the program cycle ends. No follow-up measurement was conducted.
- Whether gains would have occurred without the program, through ordinary schooling and maturation. The design cannot separate program effect from these.
- Whether children who attended more gained more. Per-child attendance was not summarised, so no dose-response relationship can be examined.
- Whether the measured gains would survive assessment by an independent assessor using a validated instrument.

- Why the learning corners were not implemented - whether the constraint is space, materials, time, training or guideline design.
- Whether the single reported observation of children hesitating to participate reflects an isolated case or a wider issue.
- How the model performs outside Suku Lihu. Both first-round sites share one suku, one language environment, one terrain type and, in part, one catchment aldeia. The second-round sites extend the quantitative record, but their location is not documented in the data supplied and no qualitative or observational evidence exists for them.
- How the model worked at the second-round sites - what drove attendance, how sessions were run, whether the learning corners were used, and how parents and facilitators experienced the program at EPE Lauhata and EPE Saboria. No FGDs, KIIs or classroom observations were conducted there.
- Whether the anomalies in the second-round records (Section 8.8) reflect recording practice or genuine patterns among children.
- Whether children with disabilities participated and benefited, since functional difficulty was never screened for.
- What the model actually cost, since expenditure is confirmed for only about 27% of the revised budget. No financial data were provided for the second-round sites.

14

SECTION

Monitoring, Data Quality and Classroom Observation

14.1 Monitoring instruments and coverage

Table 14.1. Evaluation instruments and coverage.

Instrument	Coverage	Timing	Purpose
Attendance registers	64 sessions across four sites (32 first-round, 32 second-round)	Every session	Child presence marks; participation and dosage
Child observation tool (24 indicators, Tetun)	174 baseline, 176 endline records (first round 67 / 69; second round 107 / 107)	Baseline and endline	Learning skills, socio-emotional skills, health and physical development
Facilitator knowledge and practice check (6 domains)	8 facilitator assessments, 7 complete (4 first-round, 4 second-round)	Pre- and post-training	Facilitator capacity gain
Classroom observation tool (46 rated items)	4 observations, one per first-round facilitator; none at the second-round sites	April-June 2026	Delivery quality and learning-environment safety
Parent confidence survey	113 respondents: 17 per first-round site (4 domains at AdJL; 8 at TX); 45 (LH) and 34 (SB), 8 domains	Baseline and endline	Parental confidence to support home learning
Enrolment and registration records	98 children (first-round sites); no usable records for the second-round sites	At enrolment	Demographics, prior preschool exposure, health, documentation, transport
Qualitative interviews and FGDs	25 respondents (first-round sites only)	August 2026	Perceived benefits, implementation barriers, safety, training needs, recommendations
Case-study interviews	2 children, their families plus facilitators	August 2026	Individual change narratives (published separately as Document 2)

14.2 Key findings on implementation quality

- **Progress toward objectives:** consistent measured gains on all three child-outcome domains at all four sites, although the EPE Lauhata socio-emotional and health endline figures remain well below those of the other sites; facilitator capacity objective met on the knowledge measure for three of four first-round facilitators, and for none of the four second-round assessments, but not on applied practice; parental-confidence objective measurable at three of the four sites.
- **Delivery completeness:** all 32 planned first-round sessions were delivered by the assigned facilitator pairs, and 16 sessions are recorded at each second-round site, although nine EPE Lauhata sessions are undated.
- **Fidelity gap:** the four designated learning corners were rarely implemented, and prescribed session time allocations were reported as difficult to follow. Delivery was complete in count but partial in content - a distinction routine monitoring did not capture. Fidelity at the second-round sites was neither observed nor reported.
- **Practice change:** hygiene and inclusion practices visibly improved after training, including several routines introduced by facilitators on their own initiative.

14.3 Classroom observation findings

Instrument and coverage

ETF program staff observed each of the four first-round facilitators once during delivery, using the Tetun-language classroom observation tool (Feramentu Observasaun Sala Laran). The tool has nine sections:

1. observation details;
2. learning-environment safety (5 items);
3. facilitator practice - inclusive, child-centred teaching (3 items) and learning materials (5 items);
4. knowledge development - problem-solving and critical thinking, literacy and language, digital literacy, and science (16 items);
5. social-emotional development (6 items);
6. physical development and health (8 items);
7. life skills and environmental awareness (3 items);
8. parent and community involvement (4 items); and
9. general observations. Forty-six items are rated yes/no, with the observer free to qualify a rating.

All four observations were conducted by the same ETF observer, so no inter-rater check is possible. No classroom observation was conducted at EPE Lauhata or EPE Saboria, so everything in this section relates to the two first-round sites.

Table 14.2. Classroom observations conducted.

Site	Date	Facilitator observed	Session	Time	Children present (boys/girls)	Share of enrolment
EPE Alfredo de Jesus Lemos	14 April 2026	Facilitator 2	Week 5 / Session 9	15:00-18:00	19 (4 / 16)*	46% of 41
EPE Alfredo de Jesus Lemos	17 April 2026	Facilitator 1	Week 5 / science review	15:00-17:00	31 (13 / 18)	76% of 41
EPE Tomas Ximenes	30 April 2026	Facilitator 2	Session 3	15:00-17:39	10 (6 / 4)	18% of 57
EPE Tomas Ximenes	11 June 2026	Facilitator 1	Session 13	10:00-12:00	24 (12 / 12)	42% of 57

* The form records 4 boys and 16 girls but a total of 19; the discrepancy could not be resolved from the source.

Two features of this coverage are themselves findings. First, the observation dates confirm the schedule shift discussed in Section 6.4: the June observation at EPE Tomas Ximenes took place in a morning slot, whereas all earlier sessions were observed in the afternoon. Second, the 14 April session at EPE Alfredo de Jesus Lemos ran for three hours, longer than the two-hour sessions observed elsewhere - consistent with facilitators' reports that guideline time allocations were difficult to keep.

Results by section

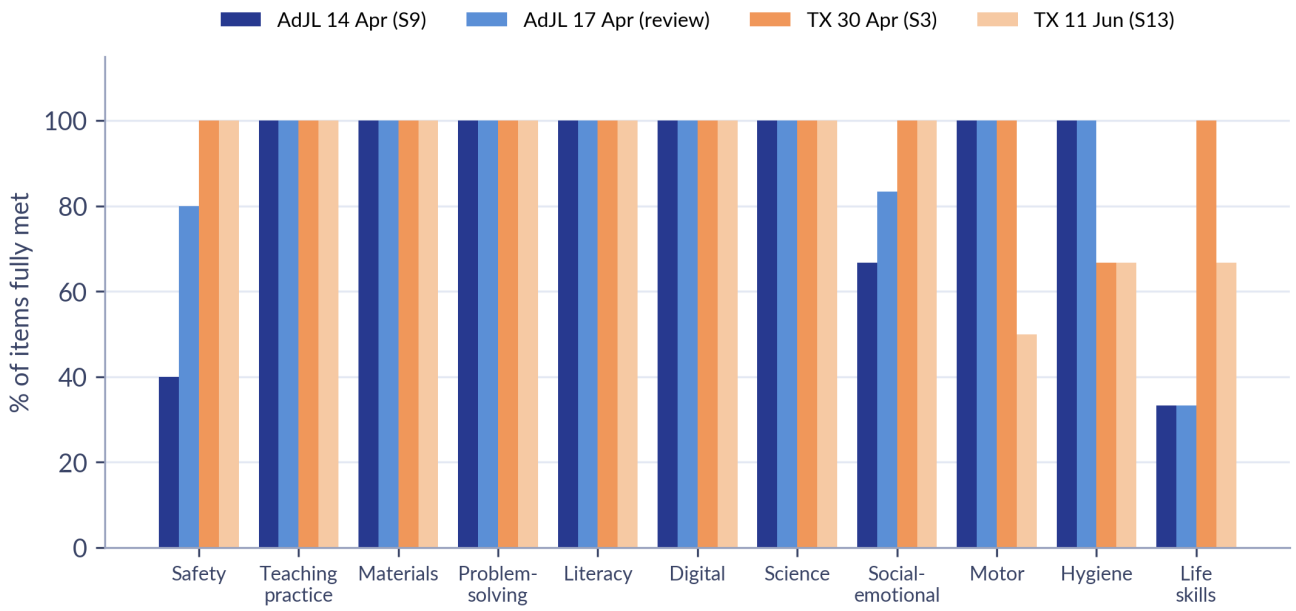


Figure 6. Share of classroom observation items fully met, by section and visit.

Table 14.3. Classroom observation results: items met, partially met and not met (46 rated items per visit).

Section (items)	AdJL 14 Apr	AdJL 17 Apr	TX 30 Apr	TX 11 Jun
B. Safety of learning environment (5)	2 met, 1 partial, 2 not met	4 met, 1 not met	5 met	5 met
C. Inclusive, child-centred teaching (3)	3 met	3 met	3 met	3 met
C. Learning materials (5)	5 met	5 met	5 met	5 met
D1. Problem-solving and critical thinking (4)	4 met	4 met	4 met	4 met
D2. Literacy and language (5)	5 met	5 met	5 met	5 met
D3. Digital literacy (3)	3 met	3 met	3 met	3 met
D4. Science (4)	4 met	4 met	4 met	4 met
E. Social-emotional development (6)	4 met, 2 not met	5 met, 1 not met	6 met	6 met
F1. Motor skills (2)	2 met	2 met	2 met	1 met, 1 not met
F2. Hygiene and nutrition (6)	6 met	6 met	4 met, 2 partial	4 met, 2 partial
G. Life skills and environmental care (3)	1 met, 2 partial	1 met, 2 partial	3 met	2 met, 1 not recorded
All rated items (46)	**39 met (85%)**	**42 met (91%)**	**44 met (96%)**	**42 met (91%)**

At every visit the facilitator was also rated as "often" (Dalabarak) using simple, age-appropriate language, and as responding to rule-breaking by speaking calmly. These two items are descriptive rather than yes/no and are excluded from the counts.

What was consistently strong

- **Content delivery.** Every item in the four knowledge-development domains - children using construction materials purposefully, completing puzzles, asking "why" and "how" questions, trying different approaches, learning new Tetun, English and Portuguese words, being read to and reading with the facilitator, following instructions for simple digital games, predicting what happens next, and showing interest in science materials - was rated as met at all four visits. This is the strongest independent corroboration in the evaluation that the content reached children.
- **Facilitator practice and materials.** At every visit facilitators involved all children, treated boys and girls equally, encouraged and praised children's work, prepared materials before the session (fully at three visits, partially at one), used songs and games, and made local materials available - the observer recorded empty water bottles reused as planters and community-supplied soil for planting.
- **Hygiene infrastructure.** A functioning toilet, clean water and soap were present at both first-round sites at every visit, and children washed hands after playing with soil at all four. At EPE Tomas Ximenes, handwashing before eating and after toilet use were rated "yes, but only some children" at both visits.
- **Environmental care.** Children disposed of waste correctly and showed care for the environment at every visit where it was recorded, though at EPE Alfredo de Jesus Lemos the observer noted they needed frequent reminding - the routine exists but is not yet self-directed.

Where practice fell short

- **Session start and transitions at EPE Alfredo de Jesus Lemos.** At both visits the session did not start on time. On 14 April children arrived late and the room was only full by 16:00, which the observer attributed to the school holiday period; on 17 April children again arrived late, some saying they were going to help their parents in the fields. On 14 April children were also left playing in the playground without supervision while both facilitators prepared the science activity, and the observer noted older children of 8-9 dominating the younger ones during this time. This is a direct observation of the mixed-age dynamic parents later reported, and a child-safety gap in its own right (Section 5.2).
- **Materials placement.** On 14 April soap and towels were not positioned where children could reach them; by 17 April all materials were in place, and the recommendation was repeated that soap, towels and first-aid materials be kept within children's reach.
- **Self-regulation.** "Children control their frustration or sadness" was rated not met at both EPE Alfredo de Jesus Lemos visits, and "children share materials or games with others" at the first. Both were met at EPE Tomas Ximenes. See Section 13.2, Divergence 4.
- **Gross motor activity.** At the June visit to EPE Tomas Ximenes no running, jumping, climbing or balance activity was observed - the only content-related item not met at that site. The observer's recommendations at that visit focus on song, picture and letter work, suggesting the physical component was simply not scheduled that day.

Parent and community involvement

Parents were present at only one of the four observed sessions (EPE Tomas Ximenes, 11 June), where they accompanied their children, waited for them at the centre, helped organise games and arranged bottles and soil for activities. Community contributions in kind were recorded at three visits. Given the parent-confidence results and

the practical requests parents made in August, the low physical presence of parents at sessions is an untapped resource - particularly for the supervision gap noted above, where a single parent volunteer would have prevented the unsupervised interval on 14 April.

Observer recommendations

The observer's recommendations fall into five groups, and they anticipate almost every operational finding of the August qualitative round. That they were made in April and the same problems were still being reported in August is itself a finding about the supervision system.

- **Time and pace:** control the minutes allocated to each content block so the session finishes on time; control break time and keep children close so they can be supervised; substitute a shorter song when time is short.
- **Demonstrate first, then instruct:** introduce and demonstrate each activity before children attempt it; give clear instructions; introduce pictures one at a time rather than covering the board; attend to upper- and lower-case letters on song sheets.
- **Rules and routines:** introduce classroom and outdoor rules repeatedly; set rotation rules at the activity corners to stop children running between them (recorded at the June EPE Tomas Ximenes visit - the one observed instance of corners in use); let all children finish an activity before anyone goes to break.
- **Age-differentiated attention and safety:** give maximum attention to the 3-5 year-olds; control the children aged 7 and above; avoid giving sharp tools such as garden forks to 3-4 year-olds; plant seedlings in individual bags so each child takes responsibility for their own plant.
- **Documenting children's work:** collect children's drawings and activity sheets into portfolios to present to parents at the closing session, and display them in the room.

Implications

- The observation record is the strongest available evidence that the model's content reached children as designed at the level of activities. Where it is weakest is exactly where the qualitative round later located the practice gap: managing time, transitions, mixed ages and self-regulation in a large group.
- The tool did not, and could not, detect the non-use of the learning corners, because it contains no item on corner use, guideline timings or group structure. The single reference to corners is the June recommendation to set rotation rules, which implies corners were in use at EPE Tomas Ximenes by session 13.
- One visit per facilitator, with both EPE Alfredo de Jesus Lemos visits in the same week and none there after session 9, cannot show change over the cycle. Two visits per facilitator, one in each half, should be the minimum.
- A binary rating produces a ceiling effect: three of four visits score above 90% and the differences between them rest on a handful of items. A three-point scale (not observed / partly / consistently) would preserve the tool's simplicity while making it sensitive to improvement.
- Observation was used as a record rather than as a feedback loop. The recommendations were written and then, on the evidence of the August findings, not systematically acted on. Coaching converts observation into practice change; recording alone does not.

14.4 Data quality and limitations

The limitations below constrain the strength of the conclusions and should be addressed before any scale-up. None of them undermines the direction of the reported changes, but several prevent claims being made with the rigour the Ministry and donors will expect.

Table 14.4. Data limitations and required remedies.

Limitation	Effect on findings	Remedy
No comparison group, no randomisation, no baseline equivalence test	No part of the measured change can be attributed to the program. Maturation, schooling and testing effects are uncontrolled.	Comparison group at the next cycle - a waitlisted site, a delayed-start cohort, or non-participating children at the same schools
Instruments internally developed and not validated	Results cannot be expressed in recognised units or compared with other programs; psychometric properties unknown	Administer a validated instrument (IDELA or MELQO) alongside the ETF tool to a sample; document scoring rules in a manual
Assessors were the implementing team	Expectancy effects cannot be excluded, particularly on the near-elimination of "not yet able" ratings	Independent endline assessment, or independent re-scoring of a random sample with a reported agreement statistic
Cohort tracking incomplete at the first-round sites: 39→26 at AdJL; 28→43 at TX; ~42 children matched. (All 107 second-round children were matched.)	First-round results compare group-level performance rather than a matched cohort	Unique child ID assigned at enrolment and carried across every instrument
Per-child attendance not summarised	No dose-response analysis possible; children with minimal exposure counted in endline results	Attendance totalled per child ID at cycle end; outcomes reported by attendance band
Instrument inconsistency: AdJL endline used a 1-4 numeric scale for socio-emotional items while baseline used descriptive categories	Required harmonisation during analysis, introducing interpretation risk into the headline socio-emotional figure	Version-locked instruments; one rating scale per tool for both rounds
Missing record: Liberty's post-training assessment	Facilitator capacity gain reported for three of four facilitators only	Completion checklist signed off per site before data submission
Duplicated record: TX parent endline identical to baseline, same date (26/03/2026)	No endline change reportable for the TX parent survey; a program objective is unmeasured at one site	Endline surveys administered on a distinct, verified date with supervisor sign-off
Three different parent instrument formats used across the four sites	Parent results are not comparable between formats and cannot be pooled	One parent instrument across all sites
Register arithmetic: hand-written percentages used inconsistent denominators	Register figures unreliable; this report recomputes attendance directly from presence marks	Fixed enrolment list as denominator; percentages computed centrally
Age-band adherence: some children under 3 or over 7 enrolled and assessed	Mixed-age effects confound outcome comparison and, per parents and the observer, suppress younger children's participation	Explicit written age policy; age-banded grouping within sites
No fidelity measurement	Non-implementation of the learning corners went undetected for a full cycle	Short facilitator session-fidelity checklist completed at every coaching visit
No disability screening	Outcomes cannot be disaggregated by functional difficulty; inclusion cannot be evidenced	UNICEF / Washington Group Child Functioning Module at enrolment

Limitation	Effect on findings	Remedy
No follow-up measurement	Persistence of gains cannot be established	Follow-up assessment one term after cycle end
Financial actuals recorded for about 27% of the budget only; none for the second-round sites	No cost-effectiveness or expenditure-based cost-per-child statement possible	Expenditure consolidated and reported alongside outcomes
No qualitative data, classroom observation or enrolment profile for the second-round sites (no FGDs or KIIs conducted)	Second-round results cannot be triangulated, explained or linked to delivery practice; contextual and qualitative findings apply to the first-round sites only	Include every site in the qualitative round and the observation schedule, and budget for both per site
Copied or templated records at the second-round sites: SB enrolment list copied from LH; nine undated, copied LH attendance sheets; repeated rating profiles at the SB baseline; auto-filled SB parent survey dates	Participation counted from assessment records; attendance approximate; SB baseline literacy and numeracy may be understated and problem-solving overstated	Site-specific templates; dated, signed register sheets; supervisor check of every dataset before submission
Within-round break in the LH endline ratings (100% vs 5% "often or always" between record blocks)	LH socio-emotional and health endline figures average two incompatible blocks of records	Record assessor identity on every record; double-score a sample
Parent survey at the second-round sites recorded using only 1 and 5 on a 1-5 scale	Measures whether, not how much, parents feel confident; the ceiling is reached quickly	One parent instrument with defined anchor points for every value

15

SECTION

Challenges and Mitigation Measures

Table 15.1. Challenges encountered during the pilot and at the second-round sites.

Challenge	Impact	Action taken	Result / status
Variable and volatile attendance (9-10 children at AdJL sessions 11-12; slow start at TX; single sessions of 15 at LH and 10 at SB)	Reduced dosage for absent children; harder session planning	Facilitators used engaging games and songs; community word of mouth; sessions continued on schedule	TX attendance rose 62% in the second half; AdJL recovered to 21-33 children in final sessions; LH and SB attendance lower in the second half (causes undocumented)
Low prior preschool exposure for those students who also attend formal education (14-24% of enrolees) and very low baseline skills at TX	More than half of baseline skill checks "not yet able" at TX	Play-based, small-group methods; youngest children seated in front and supported individually	"Not yet able" ratings absent from endline records at both first-round sites and SB; 3% of checks at LH
Facilitator capacity gaps pre-training (scores 2-3 of 6 at first-round sites; 1-3 of 6 at second-round sites)	Limited use of play-based pedagogy, materials management and hygiene practice	ETF facilitator training (April 2026) with pre/post assessment	Post-training scores of 6/6 for three first-round facilitators and documented behaviour change in classrooms; 3-5 of 6 at the second-round sites, with no classroom observation
Unstable session schedule, including a mid-cycle shift to morning delivery	Excluded children available only in the afternoon; late notice compounded by weak communication channels	Sessions continued; no formal renegotiation with parents recorded	**Unresolved. Cited by every respondent group as the most significant operational problem**
Large, age-diverse groups (up to 39 children aged 3-7 in one room) with only two facilitators	Facilitator overload; older children dominating younger ones; unsupervised intervals; some younger children discouraged from continuing	Facilitators grouped children by age where possible and supported the youngest	Partially mitigated. Staffing ratio and age-banding recommended
Learning corners rarely implemented; guideline time allocations hard to follow	A designed component did not reach children; delivery complete in count but partial in content	Not identified during implementation; surfaced only in the qualitative round	Requires diagnosis before any scale-up
Weak parent communication channels	Missed invitations; incomplete certificate distribution; parents unaware of changes	Information passed via school, ETF team and local authorities	Partially mitigated. Aldeia-based phone list recommended by parents
Unstimulating venue: standard school classrooms	Reduced child engagement due to environment familiarity	Facilitators varied activities within the available space	Alternative or adapted venues recommended by facilitators
Cohort churn between baseline and endline	Only ~20-22 matched children per site; prevents matched-cohort comparison	Names recorded at each round; fuzzy matching applied during analysis	Partially mitigated; unique child IDs required. Not an issue at the second-round sites (all 107 children matched)

Challenge	Impact	Action taken	Result / status
Inconsistent data recording (scale changes, spelling variants, ad-hoc denominators, missing and duplicated records; at the second-round sites, copied enrolment and register sheets, repeated rating profiles and auto-filled dates)	Reduced comparability and completeness of several indicators	Data cleaned and standardised during analysis; gaps flagged transparently	Direction of findings robust; second-round results reported with lower confidence; data-quality protocol required
Some participants outside the 3-7 target band (up to age 9)	Mixed-age groups require differentiated activities	Older children integrated into group work; facilitators grouped by age	Managed; written age policy recommended
Incomplete ToT coverage among facilitators	Uneven preparation across the facilitator cadre	Site-level induction by ETF staff	Full ToT coverage required before any expansion
No safeguarding, inclusion or recruitment documentation at site level	Good practice cannot be evidenced; facilitators had not received child-protection training	Practice was sound where observed, but undocumented	Policies and records required (Section 5)
No FGDs or KIIs conducted at the two second-round sites (EPE Lauhata, EPE Saboria)	No qualitative data for these sites; their findings rely on quantitative data and other documented evidence and cannot be corroborated or explained by respondents	Second-round results reported separately and flagged; qualitative findings attributed to the first-round sites only	Unresolved. Qualitative follow-up at both sites recommended (Section 15.1)
No classroom observation, enrolment profile or financial records for the second-round sites	Delivery quality, fidelity, child profile and cost at the second-round sites unknown	Gaps stated wherever relevant in this report	Unresolved

15.1 Limitation: no qualitative data at the second-round sites

Unfortunately, focus group discussions (FGDs) and key informant interviews (KIIs) were not conducted at the two second-round pilot preschools, EPE Lauhata and EPE Saboria. Consequently, qualitative data from these sites are unavailable, and the findings for these preschools rely on the available quantitative data - child baseline and endline assessments, attendance registers, facilitator pre/post knowledge tests and parent confidence surveys - and other documented evidence. The implications for interpretation are specific:

- **Results cannot be triangulated.** The convergence of assessment, classroom observation and respondent accounts that gives the first-round findings their credibility (Section 13) does not exist for the second-round sites. Their results rest on implementer-administered assessment alone.
- **Patterns cannot be explained.** The causes of attendance variation, the three-week break in EPE Saboria sessions, the low EPE Lauhata health ratings and the anomalies in the records (Section 8.8) cannot be explored with the people involved.
- **First-round qualitative findings are not transferred.** Findings on scheduling, mixed-age grouping, parent communication, safety, learning corners and facilitator confidence are reported for the first-round sites only and are not assumed to apply to EPE Lauhata or EPE Saboria, even where the same facilitators delivered at both.

- **Unmeasured outcomes stay unmeasured.** Outcomes the pilot captured only through parents' accounts - children teaching siblings, shifting from screens to making things, applying skills at home - are unknown for the second-round sites.
- **Absence of reported concerns is not evidence of their absence.** No safeguarding or child-protection concern was reported from the second-round sites, but no parent, facilitator or child there was asked.

For these reasons the second-round results are presented as quantitative findings with lower confidence than those for the first-round sites, and they are not used on their own to support any conclusion in this report. A short qualitative follow-up at both sites - for example, one parent FGD and two facilitator KIIs per site - would close much of the gap at modest cost and should be completed before the second-round results are used externally (Recommendation 21).

15.2 Recurrent constraints

- **Weather and terrain:** rain suppresses attendance directly and, where routes involve crossing rivers, can make attendance impossible. This is structural rather than solvable, and should be planned around through make-up sessions and longer cycles.
- **Scheduling:** no timing suits all constituencies; the pilot changed timing mid-cycle without an effective way to inform parents. Both the choice and the stability of the timetable matter.
- **Materials and storage:** facilitators improved children's access to materials after training, but parents requested more learning materials and toys, and dedicated weather-protected storage. Transport of materials to remote sites was raised as a separate constraint.
- **Site safety and infrastructure:** parents requested fence protection at implementation sites - a low-cost intervention with direct child-safety and parental-confidence benefits.
- **Staffing ratio:** two facilitators for up to 39 mixed-age children is the binding constraint on quality, and was raised independently by parents, facilitators and the classroom observer.
- **Parental awareness:** facilitators observed that although the program is well regarded, parent understanding of it remains limited - pointing to sensitisation as an ongoing activity rather than a launch event.

16

SECTION

Lessons Learned

16.1 What worked

- **Play-based learning drives engagement and the largest measured gains.** Sorting, block building and flexible problem-solving - the most game-like indicators - reached 96-100% independent performance at the first-round sites and 77-98% at the second-round sites. Facilitators independently identified the most creative and hands-on sessions as the most effective, and this is the one SAFE criterion the program clearly meets.
- **Facilitator training is high-leverage, but incomplete on its own.** A single training cycle moved first-round facilitators from 2-3 to 6 out of 6; at the second-round sites post-training scores were lower (3-5 out of 6), including for two of the same facilitators. However, this evaluation recognizes the potential desirability bias and it is only limited to confidence, rather than skills. It was observed that facilitators also were able to support hygiene stations, environmental care routines, inclusive seating - that no one instructed them to adopt. The limits of training without coaching are the subject of 16.2.
- **Community trust builds over time.** EPE Tomas Ximenes attendance grew 62% between the first and second halves of the cycle as families saw the program in action. The second-round sites did not follow this pattern: attendance at both was lower in the second half of the cycle, for reasons that are not documented.
- **Local-language delivery works (first-round sites).** Teaching and assessing entirely in Tetun, with Mambae-speaking children supported, enabled full participation including of three-year-olds.
- **Local recruitment produces a durable asset.** At the first-round sites, facilitators live in the communities they serve, are known to families, and remain available between cycles. Parents cited their attentiveness as the reason they judged children safe.
- **Learning travels home.** Children taught skills back to siblings and parents, shifted from phone use to making and drawing, and applied practical skills such as planting. This is among the strongest indicators of genuine consolidation, and it is currently unmeasured. This is reported by first-round parents; it is not known for the second-round sites.

16.2 What did not work, and why

- **A centrally set timetable cannot fit every community.** The pilot set a time, changed it mid-cycle, and had no reliable way to tell parents. Attendance paid the price. Timing is a site-level negotiation, not a program-level default.
- **Knowledge training does not produce classroom competence on its own.** Facilitators scored full marks and still could not manage 39 mixed-age children, keep to session timings, or run the learning corners. The professional-development literature predicts exactly this, and identifies coaching as the missing element.
- **Delivering all sessions is not the same as delivering the model.** All 32 pilot sessions ran, yet a designed component was largely absent. Counting sessions is not measuring fidelity, and implementation quality is a known moderator of outcomes.
- **Children are not a communication channel.** Relying on children to relay messages produced missed invitations and undistributed certificates. With limited household internet, an aldeia-based phone list is the practical alternative.
- **Mixed-age groups have a cost.** Older children dominated younger ones and discouraged some from continuing - a dropout mechanism attendance registers cannot detect, observed directly in April and reported by parents in August.

- **Observation without a feedback loop changes nothing.** The April observer identified almost every operational problem the August qualitative round found. The recommendations were recorded but not converted into practice change, because no coaching cycle existed to carry them.
- **Site selection by convenience is a false economy.** Facilitators recommended a needs survey before implementation rather than defaulting to sites where ETF has already built a school. Convenience of infrastructure is not evidence of need or of scheduling fit.

16.3 Lessons for measurement

- **Simple data systems matter more than sophisticated analysis.** Most limitations of this evaluation stem from recording practice, not program quality. Unique IDs, one fixed scale per tool, per-child attendance totals and a completion checklist would resolve them at negligible cost. The second-round records - copied sheets, templated ratings and auto-filled dates - confirm the point.
- **Design the evaluation before the cycle, not after it.** A comparison group costs almost nothing if identified at the start - a waitlisted site, a delayed-start cohort - and cannot be constructed afterwards at any price.
- **Independence is cheap and worth buying.** Having someone other than the delivery team conduct the endline, or re-score a sample, would remove the single most serious objection a reviewer can make to these results.
- **Benchmark against a validated tool.** Running IDELA or MELQO alongside the ETF instrument for one cycle would let ETF report in units the Ministry and donors already understand, and would test whether the in-house tool tracks a validated measure.
- **Qualitative enquiry is not optional.** The three most consequential findings - scheduling, the facilitator practice gap, and non-implementation of the learning corners - were invisible in the quantitative data. The second-round sites show the cost of omitting it: results that can be measured but neither explained nor corroborated.
- **Ask the same question in two settings.** Facilitators reported high confidence individually and low confidence in a group. Both were honest answers; the difference between them is itself a finding.
- **Measure what families notice.** The outcomes parents valued most - independent study, creativity, teaching siblings, discipline, safety - are only partly represented in the current instrument. A short parent-observed behaviour module would capture them at low cost.
- **Give every site the full evaluation package.** Extending the program to new sites with only the quantitative instruments produced numbers without context. The evaluation plan for the next cycle should specify qualitative and observation coverage for every site.



17

SECTION

Coordination and Partnerships

17.1 Key partners

- **Foundation Educating The Future Timor-Leste**, the local implementing partner, which delivered the program in the four communities through its Timor-Leste team and locally recruited facilitators.
- **Educating The Future Australia**, which worked with the consultants and the local organisation to fund and design the program, and provided support during implementation, financial management, governance, safeguarding and evaluation review. Independent consultants supported the qualitative evaluation.
- EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes school directors and teachers, who hosted sessions and supported facilitators. EPE Lauhata and EPE Saboria are recorded as the session venues for the second-round sites, but the role of their school staff is not documented.
- Parents' groups and caregivers, engaged through registration meetings, surveys, interviews and focus group discussions.
- Suku Lihu and aldeia leadership, supporting community mobilisation and, in some cases, informing families about the program.
- Ministry of Education (municipal and central), as institutional counterpart.
- Community facilitators, recruited locally, who constitute both the delivery mechanism and a durable local capacity asset.

17.2 Coordination mechanisms used

- Introduction and approval meeting with the National Director of Pre-School Education (DNEPE) and the Director-General on 14 January 2026, at which the pilot was approved; introductory meeting with the Ermera Municipal Education Director on 27 February 2026.
- Engagement meetings with the two host pre-schools and parents on 1 and 8 October 2025 to brief school authorities and agree session hours, registration dates and parents' contribution of meals.
- Registration and orientation meetings with parents at program start (February-March 2026).
- Public opening sessions at each site (5 March 2026 at AdJL; 26 March 2026 at TX) attended by DNEPE and the municipal education office, and a joint closing session on 25 June 2026 at which certificates were distributed.
- Facilitator training and post-training verification visits (April-June 2026), and four classroom observation visits by the ETF program team.
- Baseline and endline data-collection visits by the ETF Timor-Leste team at all four sites.
- Ongoing communication between facilitators, school staff and ETF coordinators via the site registers and phone contact.
- Qualitative evaluation visits by the consultant team engaging parents and facilitators at both first-round sites (August 2026); no qualitative visits took place at the second-round sites.
- Coordination records for the second-round sites (approval, community engagement, opening and closing sessions) were not supplied to the evaluation.

17.3 Coordination weaknesses identified

- **School-program scheduling coordination.** Parents recommended that the schedule be set in coordination with school teachers, and facilitators recommended avoiding overlap with formal schooling. This implies the timetable was not jointly agreed with the host school - an omission that is manageable at two sites and unmanageable at twenty.
- **Program-parent communication.** There was no reliable channel from the program to households. Information flowed through children, or not at all. An aldeia-based contact list with a named parent focal point per hamlet is the mechanism parents themselves proposed.
- **Mentoring and supervision.** Facilitators confirmed ETF support but requested more frequent and in-depth mentoring. Supervision was episodic rather than scheduled, which is why fidelity problems such as the unused learning corners persisted undetected for a full cycle.
- **Monitoring roles at municipal and national level.** The only monitoring records held by the evaluation are ETF's own observation visits; no inspector, teacher or local-authority monitoring visit is documented. The structure exists on paper and should be activated with a simple visit log.

18

SECTION

Recommendations

These recommendations follow from the evaluation findings. They are about strengthening the model and the evidence base for it, not about expansion; the case for expansion, its sequencing and its cost are set out in the accompanying business case. Priority 1 items should be completed before any further cycle. Priority 2 items should be implemented during the next cycle. Priority 3 items are medium-term.

18.1 For the implementer (ETF)

Priority 1 - complete before the next cycle

Table 18.1. Priority 1 recommendations. Items V1-V5 were added following the review of results (Section 11.1).

#	Recommendation	Rationale
1	Fix the schedule per site before enrolment opens, through joint negotiation with the host school and parents, and hold it for the full cycle. Document it in a written site agreement.	Scheduling was the most frequently cited operational problem across every respondent group, and mid-cycle change directly cost attendance.
2	Move toward a facilitator-to-child ratio of approximately 1:15 within the 3-7 band, and cap enrolment per facilitator pair accordingly.	Two facilitators managing up to 39 children of mixed ages was raised independently by parents, facilitators and the classroom observer, and directly produced a supervision lapse.
3	Group children into age bands within each site rather than teaching a 3-7 span in one room, and adopt a written age policy.	Older children dominated younger ones and discouraged some from continuing; observed directly on 14 April and reported by parents in August.
4	Complete ToT coverage for every facilitator before deployment, with an attendance register, and add compulsory modules on child protection, classroom and behaviour management, and inclusion.	Some facilitators never attended ToT; facilitators explicitly requested all three additional modules.
5	Adopt a facilitator code of conduct, a recruitment and reference-check procedure, and a documented safeguarding reporting route with a named ETF focal point.	No site-level safeguarding documentation exists; facilitators had not received child-protection training; one uncorroborated engagement signal went unfollowed.
6	Establish an aldeia-based parent communication system: phone numbers collected by hamlet, a named parent focal point per aldeia, and weekly updates.	Missed invitations, undistributed certificates and unannounced schedule changes all trace to this gap; parents proposed the remedy themselves.
7	Diagnose why the four learning corners were not implemented - space, materials, time, training or guideline design - and either resource them properly or remove them from the model.	A designed component of the program did not reach children for an entire cycle without detection.
8	Version-lock the assessment instruments, adopt one rating scale per tool across both rounds, write a scoring manual, and use a single parent instrument across all sites.	Scale changes between rounds, two incompatible parent instruments, a missing facilitator post-test and a duplicated parent survey all reduced the evidence base.
V1	Have an assessor independent of delivery re-assess a random sample of at least ten children per second-round site, including children from both EPE Lauhata endline blocks, and report agreement with the original ratings.	Tests the credibility of the second-round results and gives a first reading on persistence, at low cost.

#	Recommendation	Rationale
V2	Check second-round data entry against the original paper forms, record which assessor rated each child, and annotate or correct the datasets.	Copied sheets, repeated profiles and a within-round break cannot otherwise be resolved (Section 8.8).
V3	Analyse and report the ASP 3-Month Learning Outcome Follow-up (Annex P) if the data exist.	It would answer the persistence question directly.
V4	Reconcile second-round enrolment lists with the schools, and reconstruct per-child attendance for EPE Saboria from its dated registers; report outcomes by attendance band.	Gives the program its first dose-response evidence and a reliable participation count.
V5	Retest all facilitators with the same knowledge instrument under the same conditions, alongside an observed-practice rubric.	Second-round scores fell for facilitators who had previously scored the maximum.

Priority 2 - implement during the next cycle

Table 18.2. Priority 2 recommendations.

#	Recommendation	Rationale
9	Introduce scheduled fortnightly coaching visits with a short session-fidelity checklist and structured feedback, replacing episodic supervision.	Coaching, not more content, is what converts facilitator knowledge into practice; meta-analytic effects on instruction are around 0.49 SD (Kraft et al., 2018). The April observer identified the same problems still being reported in August.
10	Develop an observed-practice rubric to sit alongside the facilitator knowledge test, and report both.	A 6/6 knowledge score coexisted with a self-reported inability to run a session. Knowledge is not capability.
11	Build a comparison condition into the next cycle - a waitlisted site, a delayed-start cohort, or assessed non-participants at the same schools.	Without a counterfactual, no measured change can be attributed to the program. This is the single largest weakness in the evidence base and it must be designed in advance.
12	Have the endline assessment conducted, or a random sample re-scored, by assessors independent of delivery, and report an agreement statistic.	All outcome data were collected by the implementing team. Independence removes the most serious objection available to a reviewer.
13	Administer a validated instrument (IDELA or the MELQO module) alongside the ETF tool to a sample of children.	Allows results to be reported in recognised units and tests whether the in-house tool tracks a validated measure.
14	Screen for functional difficulty at enrolment using the UNICEF / Washington Group Child Functioning Module, and disaggregate outcomes accordingly.	Recorded disability prevalence of 0-2% is implausibly low; inclusion is currently asserted rather than evidenced.
15	Redistribute the most engaging content across the full cycle rather than concentrating it in sessions 1-5 and 16, and revise guideline time allocations against observed practice.	Facilitators identified these sessions as most effective and the prescribed timings as impractical. Sequencing and time focus are two of the four SAFE criteria the pilot does not currently meet.
16	Extend dosage: move from 16 sessions toward 24-32, or run a six-month cycle; where a cycle must be short, increase frequency instead.	Counting and letter recognition remained “with help” for a substantial minority; parents explicitly requested six months or daily sessions.

#	Recommendation	Rationale
17	Ensure continuity between cycles and expand access so more children can participate.	Requested by parents in both interviews and the FGD; pauses erode community trust that takes a full cycle to build.
18	Invest in learning materials and toys, secure weather-protected storage, transport for materials, and site safety works including fencing.	All four were raised directly by parents or facilitators; storage and fencing are low-cost, high-visibility investments in parental confidence.
19	Use alternative or adapted activity spaces rather than standard classrooms where possible, and avoid overlap with formal schooling.	Facilitators attributed low engagement to the familiarity of the everyday classroom environment.
20	Verify written parental consent for every named child and identifiable image before publishing any success story or photograph.	Applies directly to Document 2 and to all future success-story collection.
21	Conduct a short qualitative follow-up (parent FGDs and facilitator KIIs) at EPE Lauhata and EPE Saboria, and verify the anomalies in their datasets with the site teams, before second-round results are used externally. Include every site in the qualitative round and observation schedule of future cycles.	No qualitative or observation data exist for the second-round sites, and several of their records show copied or templated entries (Sections 8.8 and 15.1).

Priority 3 - medium term

Table 18.3. Priority 3 recommendations.

#	Recommendation	Rationale
22	Add a short parent-observed behaviour module to the endline instrument, capturing independent study, teaching siblings, creative activity and home application of skills.	These were the outcomes parents valued most and the current instrument does not measure them.
23	Conduct a follow-up assessment one term after cycle end to test whether gains persist.	Persistence is currently unknown and is the question donors will ask first.
25	Engage with the UNICEF-supported community-based alternative pre-school programming in Ermera	A comparable model operating in the same municipality offers a methodological reference point and a potential comparison population.
26	Follow up and document the single report of children hesitating to participate because of a facilitator or teacher.	Uncorroborated but material given its child-protection and engagement implications.

18.2 For government (Ministry of Education)

- Recognise the after-school model as a complementary pathway to preschool in communities where only 14-24% of children have preschool experience, noting that the pilot evidence is promising but not yet causal.
- Align the facilitator training curriculum with the Ministry's own early childhood curriculum - a request raised directly by facilitators, and a prerequisite for eventual institutionalisation.
- Support provision of learning materials (letter and number cards, puzzles, blocks, Tetun storybooks, Lafaek magazines) and basic hygiene supplies to program sites.
- Facilitate use of school facilities after hours, and encourage EPE directors to coordinate timetables jointly with the program rather than in parallel.

- Leverage program enrolment data to help the roughly one in four children at EPE Alfredo de Jesus Lemos without birth certificates obtain civil documentation.
- Support an independent, comparison-group evaluation at the end of the next cycle, to generate evidence of a standard suitable for national policy discussion.
- Activate the monitoring roles already assigned to the municipal pre-school inspector and DNEPE in the coordination structure: at least one documented inspector visit per site per cycle, with findings shared with ETF.

18.3 For communities and parents

- Participate in setting the session timetable, so that the agreed time reflects household and agricultural realities in each aldeia and can then be held stable.
- Nominate a parent focal point per aldeia to receive and distribute program information, replacing reliance on children as messengers.
- Sustain and expand parent participation; parent sessions should prioritise the weakest baseline domains - educational games and homework support at EPE Tomas Ximenes, and educational games and art and creativity at the second-round sites.
- Support consistent attendance, including community arrangements for walking groups in poor weather and on difficult river routes.
- Encourage volunteer engagement to assist facilitators with large mixed-age groups - a single parent volunteer per session would close the supervision gap observed in April.
- Continue peer-to-peer advocacy. Parents recommending the program to other parents proved one of the most effective mobilisation mechanisms in the pilot.

19

SECTION

Conclusion

The After School Program pilot at EPE Alfredo de Jesus Lemos and EPE Tomas Ximenes did what it set out to do. Sixteen sessions were delivered as planned per targeted EPE, by four locally recruited facilitators working in Tetun with low-cost play-based materials inside existing school buildings, in two rural communities where only 14-24% of enrolled children had any prior preschool experience. Measured performance rose on every learning indicator at both sites. By endline, no assessed child at either site was recorded as unable to attempt any of the ten core skills. Facilitators moved from partial to full knowledge scores and visibly changed their classroom practice. Parents at the first-round site with usable survey data reported universal confidence in supporting their children's learning.

The two second-round pilot pre-schools, EPE Lauhata and EPE Saboria, extend this picture quantitatively. The same model, delivered there between July and September 2026, was followed by the same direction of measured change in two larger, fully matched cohorts: independent performance rose from 26% to 75% and from 37% to 92% of skill checks, and 100 of the 107 children assessed performed more skills independently at endline than at baseline. Across all four sites, independent performance rose from 35% to 82% of skill checks. The second-round sites also temper the pilot's headline: "not yet able" ratings persisted for 13 children at EPE Lauhata, health practices there reached 63% rather than 100%, and facilitator post-training scores were 3-5 out of 6 rather than the maximum.

At the two first-round sites, the qualitative round conducted independently in August corroborates this from the household side and adds what the numbers could not. Parents describe children who study without being asked, who make things instead of watching screens, who teach what they have learned to younger siblings and sometimes to their parents. Facilitators describe a program they believe in and a set of practical constraints - too many children, too wide an age range, an unstable timetable, an unfamiliar skill set - that they are asking for help to overcome. Classroom observation, conducted by a third party again, records the same strengths and the same weaknesses. That three independent lines of evidence describe the same program is the most persuasive feature of this evidence base. No such corroboration exists for the second-round sites, where no FGDs or KIIs were conducted and no classroom was observed: their results rest on one strand of evidence, and several of their records show signs of copied or templated entry that must be verified.

19.1 What this evaluation does not establish

It would be a disservice to ETF, and to the Ministry, to leave the conclusion there. The pilot used a single-group pre/post design with no comparison group, no randomisation, no follow-up measurement and incomplete matching of children between rounds. The assessment instruments were developed internally and have not been validated. All child assessments were conducted by the team that delivered the program. At the first-round sites fewer than half of assessed children appear in both rounds (at the second-round sites all do, but there the evidence is quantitative only), and per-child attendance was never summarised, so the evaluation cannot even test whether children who attended more gained more.

The consequence is specific. This evaluation cannot say how much of the measured change the program caused, rather than ordinary schooling, normal development over three to four months, familiarity with a test taken twice, or the expectations of assessors who had every reason to hope for improvement. It cannot say whether the gains persist. It cannot express the change in units comparable to any other program. It cannot say whether the model

would work outside Suku Lihu, where both first-round sites sit and where they partly share a catchment; the second-round sites add measured results from two further pre-schools, but without the qualitative and observational evidence needed to understand how the model worked there. It cannot say whether children with disabilities participated, because they were never identified. And it cannot say what the model cost, because expenditure is confirmed for only about a quarter of the budget.

The most striking single figure in this report the near-disappearance of “not yet able” ratings (from 33% to 1% of checks across the four sites) is also the figure most vulnerable to these limitations. A floor-to-ceiling movement on a short, unvalidated, three-point scale administered twice by the implementing team is exactly the pattern that measurement artefacts produce. The evaluation’s judgement, on the balance of the evidence and particularly on the strength of the independent corroboration, is that real learning occurred. That judgement should be stated as a judgement, not presented as a demonstrated result.

19.2 What follows from this

What limits the model is operational, not conceptual. Scheduling, staffing ratios, age grouping, communication with parents, materials and storage, safeguarding and inclusion documentation, fidelity monitoring and the data system are all fixable, and every one of them was identified by the people delivering or receiving the program. No respondent group questioned whether the After School Program should continue. What they described was how it should be run.

The program’s significance for Timor-Leste lies in what it plausibly demonstrates: that trained community facilitators, working in the children’s own language with inexpensive materials in buildings that already exist, can deliver structured early learning in rural communities the formal preschool network does not reach, in a country where gross preschool enrolment stood at about 31.5% in 2025 against a 2030 target of 80%. The mechanism is well supported by the international evidence. What is not yet supported is the specific claim that this program produced these results.

The proportionate conclusion is therefore a conditional one. The evidence justifies a further cycle, better measured, with the corrective package in Section 18 in place - and it justifies ETF seeking the resources to run that cycle. It does not yet justify presenting these figures as evidence of impact, and ETF’s credibility with the Ministry and with funders is better served by saying so. A second cycle with a comparison condition, unique child IDs, independent endline assessment, a validated benchmark instrument, and qualitative and observation coverage at every site - which the second-round sites lacked - would cost little more than the first and would produce evidence of an entirely different order. That is the recommendation on which this evaluation ends, and it is the premise on which the accompanying business case is built.

20

SECTION

References

American Institutes for Research & Universidade Nacional Timor Lorosa'e (2022). Country-Led Evaluation of Preschool Education in Timor-Leste (2012-2020). Commissioned by the Ministry of Education, Youth and Sport, supported by UNICEF.

Australian Council for International Development (ACFID). ACFID Code of Conduct - quality principles on safeguarding, child protection and PSEAH.

Black, M. M., Walker, S. P., Fernald, L. C. H., et al. (2017). Early childhood development coming of age: science through the life course. The Lancet, 389(10064), 77-90.

Britto, P. R., Lye, S. J., Proulx, K., et al. (2017). Nurturing care: promoting early childhood development. The Lancet, 389(10064), 91-102.

Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. American Economic Review, 97(2), 31-47.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? Clinical Psychology Review, 18(1), 23-45.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. American Journal of Community Psychology, 41(3-4), 327-350.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. American Journal of Community Psychology, 45(3-4), 294-309.

Engle, P. L., Fernald, L. C. H., Alderman, H., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. The Lancet, 378(9799), 1339-1353.

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: supporting preschoolers' acquisition of geometric knowledge through guided play. Child Development, 84(6), 1872-1878.

Government of Timor-Leste & UNICEF (2020). Situation analysis of children in Timor-Leste. Dili: UNICEF Timor-Leste.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. The Lancet, 369(9555), 60-70.

HANDS Report, 2025. HANDS Annual implementation report. Dili

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. Science, 312(5782), 1900-1902.

Joyce, B., & Showers, B. (2002). Student achievement through staff development (3rd ed.). Alexandria, VA: ASCD.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: a meta-analysis of the causal evidence. Review of Educational Research, 88(4), 547-588.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: a meta-analysis of effects for at-risk students. Review of Educational Research, 76(2), 275-313.

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. Economics of Education Review, 29(2), 271-282.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). International Journal of Educational Research, 91, 1-15.

Rao, N., Sun, J., Richards, B., Weber, A. M., Sincovich, A., Darmstadt, G. L., & Ip, P. (2019). Assessing diversity in early childhood development in the East Asia-Pacific. Child Indicators Research, 12, 235-254.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

UNESCO, UNICEF & World Bank (2017). Overview: MELQO - Measuring Early Learning Quality and Outcomes. Paris: UNESCO.

UNICEF & Washington Group on Disability Statistics (2016). Module on Child Functioning. New York: UNICEF.

Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided play: where curricular goals meet a playful pedagogy. Mind, Brain, and Education, 7(2), 104-112.

Yoshikawa, H., Rojas, N. M., Kabay, S., Pisani, L., & Dowd, A. J. (2019). Measuring early learning and development across cultures: invariance of the IDELA across five countries. Developmental Psychology, 55(1), 23-37.

Zosh, J. M., Hopkins, E. J., Jensen, H., et al. (2017). Learning through play: a review of the evidence. Billund: LEGO Foundation.

21

SECTION

Annexes

Annex	Contents	Status
A	Attendance registers, all four sites (16 sessions each)	Available
B	Child baseline/endline assessment datasets (first-round sites: 67 baseline and 69 endline records, 24 indicators; second-round sites: 107 baseline and 107 endline records, 18 rated items)	Available
C	Facilitator pre/post-training assessment records (8 assessments across four sites; 1 pilot post-test missing)	Available
D	Parent/caregiver survey datasets (113 respondents: 34 first-round, 79 second-round; TX endline duplicates baseline; SB endline dates unverifiable)	Available
E	Enrolment/registration lists (98 children, first-round sites) with demographic, health, documentation and transport fields; the EPE Lauhata list does not reconcile with its records and the EPE Saboria list is a copy of it	Available
F	Qualitative evaluation materials (first-round sites only): interview and FGD guides, respondent list, thematic coding notes, and confirmation of facilitator FGD composition	Available
G	Stories of Change: Case Studies from the After School Program Pilot	Published separately as Document 2
H	Written parental consent records for named children and identifiable images	Available
I	Photos of activities, training materials, and monitoring checklists	Available
J	Financial records and work plans	Available (analysed in Document 3)
K	Classroom observation records: four completed observation forms with observer recommendations	Available (analysed in Section 14.3)
L	ASP implementation coordination and monitoring diagram (Diagrama Programa Depois Eskola)	Available (Figure 1)
M	Proposed instruments for the next cycle: unique-ID protocol, scoring manual, site agreement template, needs-assessment instrument, session-fidelity checklist, observed-practice rubric, Child Functioning Module	Availaible
N	Core policy documents: child safeguarding policy, facilitator code of conduct, recruitment procedure, inclusion policy, incident reporting procedure	Available
O	Question for local leader, counterpart, and ministry	Available
P	ASP 3-Month Learning Outcome Follow-up	Available
Q	Second-round datasets: EPE Lauhata and EPE Saboria workbooks (child assessments, attendance registers, facilitator tests, parent surveys)	Available (analysed in Sections 8.8, 10 and 11)

EDUCATING THE FUTURE

November 2026

Obrigadu barak!

Thank you to the children, families, facilitators, schools and Ministry of Education partners who made this pilot possible, and to every supporter who helps us build dreams alongside Timorese communities.

Australia

Educating The Future Australia Limited
ABN 31 632 970 819
people@educatingthefuture.org
ahegedus@educatingthefuture.org
www.educatingthefuture.org

Timor-Leste

Foundation Educating The Future Timor-Leste
Registered in Dili
Avenida Run Presidente Nicolau Lobato, Fatuhada 04,
Dom Aleixo, Comoro
education@educatingthefuture.org

Member of the Australian Council for International Development since 2020

The implementation of this pilot was supported by the Swiss Embassy to Timor-Leste under their small-grants program.