Comparative Analysis of Teaching and Learning Outcomes for ALMA

May 2025

Submitted to Abt Associates / Australia Timor-Leste Partnership for Human Development

Under the terms of agreement #PHD-2025-0001

Authors:
Dr. Lotte Renault, CARE Timor-Leste
Camilo Boavida, CARE Timor-Leste

Table of figures

Table of Tables

Acronyms

ALMA Apoiu ba Lideransa, Mentoria no Aprendizajen / Support to Leadership, Mentoring and Learning
DFAT Department of Foreign Affairs and Trade/ Australian Government
EGRA Early Grade Reading Assessment
HATUTAN II Hahán ne'ebé Atu Fó Tulun ba Nutrisaun no Edukasaun/ Food to Support Nutrition and Education
GTL Grupu Traballu Líder/ Leader Working Group

GTP Grupu Traballu Profesór/ Teacher Working Group
MoE Ministry of Education of Timor-Leste
PTA Parent-Teacher Association

Executive Summary

Introduction to ALMA

The ALMA program aims to support the Ministry of Education (MoE) of Timor-Leste to strengthen the capacity of school leaders to support teachers to roll out the new national basic education curriculum launched in 2015. Since its inception in 2016, ALMA reached 1,084 schools across all municipalities of Timor-Leste. ALMA's approach is based on the set-up of an integrated mentorship system, where local mentors support school administrators and teachers through school-based sessions and cluster-based peer working groups. The mentorship process is informed by classroom observations, with school leaders entering the results into an electronic system - eskola.tl – and visualizing patterns over time through an automated dashboard. The mentorship process has since been integrated in MoE's systems and staffing structure.

Study Purpose and Methodology

The present study aimed to assess ALMA's impact on student acquisition of foundational literacy skills and, in parallel, validate the use of the classroom observation tool by school leaders. In the absence of a baseline for ALMA-supported schools or a counterfactual group established at baseline, the study was initially designed to compare the literacy outcomes of (i) 1,036 grade 3 students attending 70 schools supported by ALMA since 2016-2018 ('long-term' group) with (ii) the results of 1,029 students attending 70 schools supported by ALMA from 2024 onwards ('recently joined' group). The design assumed that the 'long-term supported' schools would have a greater impact from ALMA's implementation, serving as a treatment group, while the 'recently joined' schools would serve as a comparison group. The study also conducted 249 classroom observations in the 140 schools, comparing the results with the online classroom observation data generated by school leaders during 2025 mentorship sessions.

Validating the Design

Findings indicated that the that long-term supported schools generally have poorer conditions than those that joined the program in 2024, thus affecting comparisons between the two groups. An attempt to match schools with similar characteristics across both study groups found that even when comparing matched samples, the recently joined schools continued to outperform long-term supported schools. It is likely that the frequency of mentorship and participation in peer learning groups is higher in locations where ALMA was recently rolled out. Conversely, secondary data suggest that there is a decline in the

frequency of mentorship to teachers in the long-term supported schools. The results do not support the validity of the original study design as a means of measuring ALMA's impact, as the recently joined schools seem to have equal or higher access to mentorship and therefore cannot serve as a comparison group. Additionally, ALMA staff noted that the initial phases of the program focused mainly on developing school leadership capacity, with support to teacher working groups on literacy not integrated until 2018, thus further affecting the ability to compare the two groups.

In parallel, this study found that the exposure to ALMA's mentorship component, regardless of the phase of the program, is strongly associated with higher reading outcomes among students. Therefore, this study has focused primarily on exploring the impact of mentorship on learning outcomes and the pathways through which this is occurring.

ALMA's Mentorship Interventions Are Associated with Higher Literacy Outcomes

The findings have shown a robust association between ALMA's mentorship approach and improved literacy outcomes. Nearly a third of the school administrators (31%) self-reported ALMA mentors as a key source of assistance to them in supporting teachers. The self-reported mentorship of school administrators was associated with systematically higher reading fluency results, measured through three EGRA tasks – invented word reading, familiar word reading, and passage reading. The benefits of mentorship seem to apply to both boys and girls; however, girls systematically outperform boys. Self-reported mentorship is also associated with a significantly lower proportion of non-readers – a reduction of 10 percentage points. Most importantly, self-reported mentorship of school administrators is associated an increase of nearly 8 percentage points in reading comprehension scores. It is also associated with a significantly lower proportion of students scoring zero in reading comprehension – a difference of 11 percentage points. Overall, the impact of the mentorship of school administrators on reading outcomes tends to be higher in long-term schools, which may indicate a maturation effect where the use of improved practices resulting from mentorship is consolidated over time. Taken together, these results suggest that the mentorship of school administrators is an effective model to boost literacy outcomes in early grades. Furthermore, the mentorship of administrators seems to play an important role in supporting low-performing students, reducing the proportion of non-readers.

The study also considered that the impact of mentorship of administrators might be the result of having a sub-set of more experienced or qualified administrators, or simply better managers. However, administrators who self-reported mentorship do not have a significantly different background in terms of experience or academic background in relation to administrators who did not report a mentor support. Administrators who self-reported mentorship are also not significantly better than other administrators in terms of managing teacher attendance or strengthening participatory governance. The results support the hypothesis that ALMA's mentorship is indeed the key factor contributing to improved learning outcomes in the schools where administrators self-reported receiving it.

The findings show that teacher self-reported mentorship is not immediately associated with higher literacy outcomes. However, teacher mentorship is associated with higher utilisation of formative assessments – a difference of 13 percentage points. The use of formative assessments is associated with improved reading outcomes - increased passage reading fluency and reading comprehension scores. Teacher participation in Teacher Working Groups / GTPs, another ALMA-supported approach, is associated with the application of improved teaching practices and with increased reading fluency. The findings suggest that ALMA's support to teachers is having a positive impact on improving teaching practices; in turn, the use of improved teaching practices correlates positively with passage reading fluency and reading comprehension.

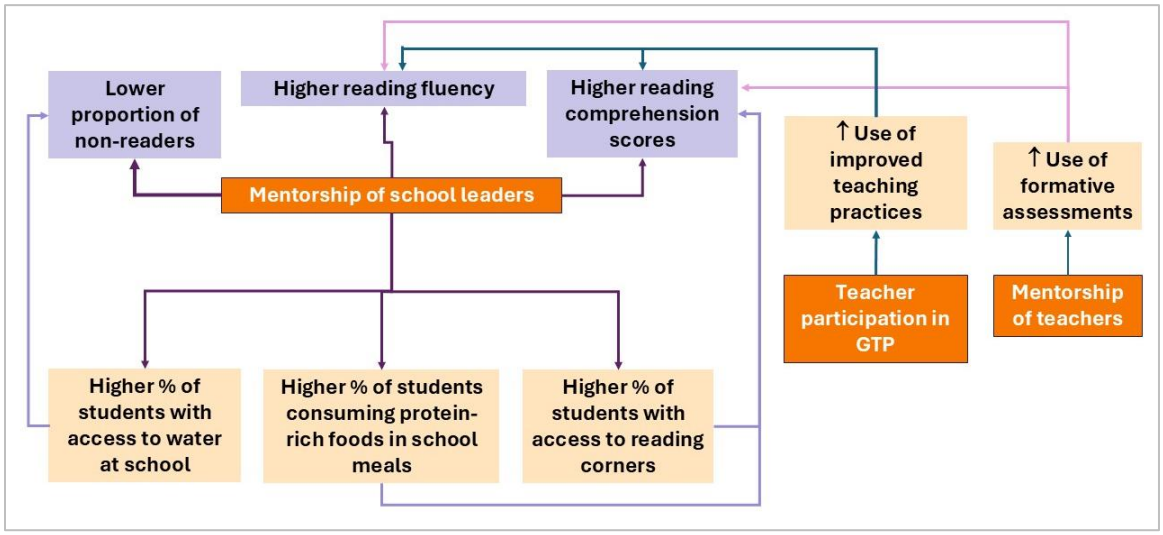


Figure 1: Diagram summarizing how ALMA-supported interventions are associated with improved literacy outcome sand management of resources

A sub-set of the assessed schools is extremely remote and accessible on foot only. The proportion of school leaders self-reporting mentorship in these schools is lower, potentially due to the challenges involved in accessing remote locations during the rainy season, particularly in the case of female mentors. Students assessed in these schools have significantly lower literacy outcomes. These schools represent a special case, which may require a different approach to mentorship and GTPs.

Validating ALMA's Classroom Observation Tool and Practices

Findings indicate that mentors tend to overrate classroom practices during observations, with a significantly higher proportion rating teachers as applying 'very good' practices in relation to external enumerators. However, the classroom observation tool shows high internal reliability, which enables its use as a scale. The 'teaching practices score' derived from the classroom observation tool increases with GTP participation, and most importantly, correlates with improved reading outcomes. Taken together, these findings validate the use of the classroom observation tool as a sensitive instrument able to capture changes in teaching practice, which are positively associated with higher literacy outcomes.

Attendance

There is no association between mentorship of school leaders or teachers and student attendance. Teacher and student attendance rates are generally low – 82% for teachers, 75% for grade 1-6 students. Grade 1 has the lowest attendance rate – 68% (66% girls). As expected, there is a positive correlation between teacher and student attendance, particularly in grades 1-3.

Other Factors Affecting Learning Outcomes

Literacy outcomes are positively associated with the availability of reading corners at grade 2 classes; the availability of water at school; and student consumption of protein-rich foods through school meals. Specifically, the availability of water at school has a significant impact in reducing the proportion of non-readers among female students, supporting assumptions that girls are disproportionately impacted by water shortages. Interestingly, those factors are positively affected by mentorship of school administrators. Schools whose administrators self-reported receiving mentorship had a higher availability of reading corners and protein-rich foods in meals, indicating an association between ALMA's school administrator mentorship approach and increased capacity for resource management and higher accountability for the use of improved practices, ultimately contributing to improved learning outcomes.

Conclusions and Recommendations

The results generally support the assumption that ALMA's approach to mentorship of school administrators is associated with improved reading fluency and reading comprehension. However, there are disparities in the results, with the proportion of administrators self-reporting mentorship declining in extremely remote schools and among men. The findings also demonstrate that ALMA's approach to teacher mentorship is associated with an increased use of formative assessments, while the participation in GTPs is associated with a higher use of improved classroom practices. The use of improved classroom practices and formative assessments are predictors of improved literacy outcomes – reading fluency and reading comprehension.

The findings show that the strongest results of ALMA's implementation are seen in schools whose administrators self-reported mentorship. There is a need to further explore if self-reported mentorship is associated with fidelity of implementation, and to understand which factors are associated with ALMA's stronger results in these schools. Additionally, there is a need to further understand the gender gaps observed in student literacy results, teaching practices, and administrator self-reported mentorship. In all cases, the gender gaps favour girls and women. Last but not least, there is a need to further refine ALMA's mentorship model to better serve extremely remote schools accessible on foot only.

Background

1.1 Introduction

The Apoiu ba Lideransa, Manorin no Aprendizajen (Support to Leadership, Teachers, and Learning) program[1], known as ALMA, was designed to support the Ministry of Education (MoE) of Timor-Leste to strengthen school capacity to roll out the basic education curriculum. ALMA is part of the Partnership for Human Development (PHD), a 10-year multi-sectoral collaboration between the Australian and Timorese governments towards achieving and maintaining healthier, educated, and more prosperous communities.

ALMA's approach includes five components:

- Leaders of Learning program, through which ALMA provides training to school administrators and emerging female leaders (manorin feto potensiál) to develop skills for school management and counselling teachers on the use of improved teaching practices and assessments.
- Reflective practise based around a classroom observation tool, where ALMA trains mentors and school administrators to use structured classroom observations to inform teacher mentorship sessions.
- School and cluster-based peer professional learning groups for teachers and school leaders, including Teacher Working Groups (Grupu Traballu ba Profesór/ GTP) and Leader Working Groups (Grupu Traballu Líder/ GLP). These periodic sessions enable peer-to-peer learning exchanges and group troubleshooting for the implementation of the new curriculum.
- Mentor support to leaders and teachers from local education mentors: ALMA supports national and municipal-level mentors to (i) provide on-site mentorship sessions to school administrators and teachers informed by a structured classroom observation tool, and (ii) facilitate GTL and GTP sessions.
- Educational technology to enable efficient information sharing and monitoring: ALMA provided tablets to school administrators and mentors with pre-loaded content (curriculum and associated materials) and eskola.tl, a one-stop solution to collect, store, analyse, and display education data, including from classroom observations and student assessments.

These components aim to create conditions for an integrated, evidence-driven mentorship cycle, strengthening the capacities of school administrators and teachers, and building a localised ecosystem for technical support to the implementation of the national basic education curriculum. ALMA started being implemented in 2016, following the introduction of the new curriculum in 2015. The program used a phased approach, completing its roll-out to all municipalities in 2024, reaching 1,084 schools. Prior to the COVID-19 crisis, ALMA

[1] Previously called *Apoiu Lideransa Liuhusi Mentoria no Aprendizajen* (Support to Leadership, Mentorship, and Learning).

had covered 402 schools across nine municipalities. ALMA's mentorship process has been integrated into MoE's systems and staffing structure.

ALMA did not conduct a baseline study or established a counterfactual group at inception. A previous multi-year study (2018-2021)[2] assessed changes in teaching practices and teacher knowledge associated with ALMA's implementation. Findings from this study indicated that ALMA contributed to improve school leaders' capacity to support teachers, including through the use of technology to document classroom observations and GTPs. The study found that teachers were increasingly using more efficient classroom management practices, student-centred activities, and to a certain extent, gender-responsive classroom practices. Teachers identified mentors as a source of support. However, the study observed that GTP application tended to decline after the program had been rolled out in a municipality and highlighted the need for continued mentorship support to teachers. The study did not include student learning assessments and recommended conducting follow-up research to determine ALMA's impact on learning outcomes. The present study was designed to respond to this recommendation.

1.2 Purpose

The present study seeks to:

1. Better understand the impact of the ALMA program on student literacy skills, supporting PHD's reporting against the new Tier 2 Indicators and providing the MoE with further evidence of the ALMA program's impact.
2. Assess teacher competency in effectively delivering the Basic Education Curriculum, through the use of classroom observations that align with the ALMA classroom observation tools and methodologies. This approach aims to validate or refute the ongoing teacher observation findings that are collected through the eskola.tl platform.
3. Through the administration of an EGRA and a validation of the classroom observations, answer the question: "Does the longer-term and accurate implementation of ALMA lead to improvements in literacy outcomes for students?".

1.3 Research Questions

Specifically, the study seeks to respond to the following questions:

1. Has ALMA improved literacy learning outcomes for students entering Grade 3?
- Is there an observable link between teacher adherence to curriculum-prescribed teaching methodology[3] and improved learning outcomes in long-term ALMA-supported schools?

[2] Australian Council for Educational Research (2022) Teacher Development Multi-Year Study Series. Timor-Leste, Final Report, pg.11; pg.47

[3] As assessed in the classroom observation used by ALMA mentors.

- Are there significant differences between learning outcomes in rural and urban[4] schools?
- Are there significant differences between the learning outcomes for girls and boys in long-term and recently joined ALMA-supported schools?
- Are there any important external factors that are demonstrating significant influence over learning outcomes?

2. Are the classroom observations conducted by school leaders accurate representations of actual teaching practice?
3. Attendance

- Has ALMA improved student attendance for Cycles 1 and 2?
- Are there significant differences in the impact on Cycle 1-2 student attendance by location or sex?

Methodology

2.1 Tools

The study used the following tools:

- Early Grade Reading Assessment/ EGRA: The study used an adapted version of EGRA currently being applied by the HATUTAN II project in Timor-Leste. This version was reviewed and approved by the MoE. The assessment contains the following sub-tasks: Letter recognition (timed); non-word reading (timed); familiar word reading (timed); passage reading (timed); reading comprehension questions (not timed, two levels)[5]. The EGRA tool is provided in Annex 2.
- Short student survey inclusive of background questions (age, sex, mother tongue, access to reading materials, food consumption) and a working memory test.[6] Both tools were previously used by HATUTAN II with large student samples in 2019-2025. The short survey and working memory test were applied with students participating in the EGRA.
- Classroom observation: Applied the classroom observation tool currently used by ALMA mentors and school leaders, with the addition of a short teacher survey collecting background information (sex, highest teaching qualification, exposure to training, participation in teacher working groups[7], years of teaching experience).
- School survey: Applied with the school administrator (director or coordinator). The tool is an adapted version of a longer instrument currently used at scale by

[4] The MoE categorizes school location as urban, rural, remote, very remote, and extremely remote.
[5] The use of two levels of reading comprehension passages is intended to prevent a ceiling effect when applying the tool with grade 4 students.
[6] Working memory refers to the ability to actively sustain information for a period of time, which is a critical predictor of reading comprehension. The pictorial working memory test consists of presenting to students 19 images representing items familiar to children, requesting the respondent to remember them, and subsequently asking the respondent to recall the items.
[7] *Grupu Traballu ba Profesór*/GTP – cluster-level coaching and peer-to-peer support platforms.

HATUTAN II, previously approved by the MoE. It assesses school conditions, availability of reading materials, management practices, exposure to teacher professional development (training and coaching), functionality and practices of parent-teacher associations, recordkeeping, total enrolment by grade and sex, and teacher and student attendance (using headcounts).

All tools were designed in Tetum and subsequently back-translated into English. Instructions for students to perform EGRA tasks were provided in the students' mother tongue for those who are not native Tetum speakers. However, the EGRA tasks themselves were in Tetum[8]. This approach ensured that students could understand what they were expected to do at each EGRA task and excluded the possibility of poor results due to students not understanding the instructions.

2.2 Sampling

School sample: The sampling frame for "long-term" schools included all those targeted by ALMA in 2016-2018. The sampling frame for "recently joined" schools included those targeted in 2024 in the municipalities of Ainaro, Baucau, and Viqueque. The sampling frames excluded schools offering Cycle 3 (grades 7-9) only and those targeted by another program providing school-based coaching (HATUTAN II) to prevent bias. The school sample was selected proportional to the distribution of schools by municipality in each sampling frame. A total of 70 randomly selected long-term and 70 recently joined schools were identified for each sample.

Student sample: The study aimed to assess randomly selected grade 3 students in treatment and comparison schools. The student sample size was calculated as follows:

a. 1,361 grade 3 students[9] from treatment (long-term exposure) sample schools
b. 1,361 grade 3 students[10] from comparison (recently joined) sample schools.

The sampled students were randomly selected among those listed in the enrolment records for each grade. In schools where grade 3 classes had over 20 students present on the day of the survey, the team leader for the data collection team used a random number generator app (Random UX) to sample students from enrolment records. The random selection ensured that the proportion of male and female students in the sample would mirror the overall distribution of male and female grade 3 students in those schools. In schools where 20 or less grade 3 students were present on the day of the survey, a 'take all' approach was used.

Classroom observation sample: The study intended to conduct two classroom observations with teachers in grades 2-3 in each of the sampled schools. The observations were

[8] With the exception of invented word reading
[9] Considering 95% confidence level, 80% power, an estimated effect size of 0.22, a design effect of 2, potential 5% non-response or data loss, and sex disaggregation. The base sample size is calculated considering $n=2[Z\alpha+Z\beta/ES]^2$
[10] As above

conducted exclusively during Tetum classes, allowing the study to cross-reference teaching practices and literacy outcomes.

Realized sample

All sampled schools were assessed by the data collection teams. However, there was a shortfall in the number of classroom observations due to (i) teacher absences and (ii) single observations conducted in multigrade classes. There was also a shortfall in the number of students participating in learning assessments due to absenteeism and smaller-sized classes.

Table 1: Intended and actual samples.

Category	Long-term: intended	Long-term: actual	Recently joined: intended	Recently joined: actual	Shortfall
Students[11]	1,361	1,029	1,361	1,036	24%
Schools	70	70	70	70	–
Teachers (classroom observations)[12]	140	126	140	123	10%

2.3 Data collection

Training

CARE trained a total of 62 data collectors (enumerators), out of whom 56 were selected for deployment. The training included an overview of ALMA; research ethics; child protection; gender and inclusion; sampling; overview of the tools; data quality control; mock practices; and two days of field practice, with feedback on performance provided to the enumerators. The enumerators were selected based on the results of three post-tests and performance during the field practice. A total of 14 data collection teams were formed, with 42 enumerators and 14 team leaders.

Procedures

The teams used tablets provided by ALMA for data collection. The data was collected using ONA, an electronic data collection platform, with ODK-based xlsx tools. Classroom observation data was collected initially on paper-based tools and then entered in ONA's electronic tool. The use of paper-based tools for classroom observations enabled enumerators to simultaneously record a variety of behaviours and practises used by teachers and prevented issues with missed data due to the time needed to scroll through the electronic survey.

[11] 240 students were assessed in schools included in the initial sample, but not in the revised sample.
[12] 15 teachers were assessed in schools included in the initial sample, but not in the revised sample.

The data collection process was managed by two CARE staff - Camilo Boavida, Assessment Coordinator, and Antonio Soares, Assistant Field Coordinator. Both staff conducted daily spot-checks for data quality control. At the beginning of each data collection day, the team leaders provided details on their data collection plans and at the end of the day, debriefed on the actual results achieved and provided details about each school location via WhatsApp. CARE's Director of Research conducted field spot-checks during the first week of data collection and daily data quality control checks on electronic data, providing real-time feedback to data collection teams to address emerging issues.

Interpretation of Findings

The consultants conducted a presentation of preliminary findings (April 09) with DFAT and ALMA staff, and a workshop for interpretation of findings (April 28) with the presence of DFAT, MOE, and ALMA staff. These sessions aimed to gather inputs to validate the interpretation of the results, considering differences in implementation by phase or geography. Additionally, these sessions also included consultations on potential factors that might explain the results observed, thus informing additional analyses. The present report incorporates these inputs and additional analyses associated with them.

2.4 Limitations

Design. The original design assumed that the two study groups – schools receiving long-term support from ALMA and recently joined schools – would have comparable characteristics. However, the survey results show that the long-term group includes a larger proportion of highly vulnerable schools, including 16% reachable only on foot compared to 11% in the recently joined group. The differences in the distribution of schools by municipality across ALMA's phases also affect comparability. The proportion of schools in Baucau is nearly twice as high in the recently joined group – 40% compared to 21% in the long-term support group. Students in Baucau have significantly higher reading fluency and comprehension scores, ultimately biasing the results. While the study includes a sub-analysis using matched samples from both study groups, the substantial differences between the two groups mean that a matching procedure in itself generates bias, removing a large proportion of vulnerable schools from the analysis. Additionally, the recently joined schools, originally thought to be a valid comparison group due to the limited timeline of ALMA's support, have actually had similar levels of exposure to school administrator and teacher mentorship.

Furthermore, ALMA's original design focused heavily on strengthening school administrators' leadership capacity. According to input gathered during the interpretation workshop, in its initial phases, the program did not provide specific mentorship to administrators or teachers on literacy skills development. The support to GTPs, with a particular focus on foundational literacy and numeracy, was incorporated by the program in 2018. As a result, long-term supported schools targeted in 2016-2017 (or even in early 2018) may have received very limited support on literacy skills development, while the

recently joined schools have received a far more focused intervention, aiming at improving foundational literacy skills.

Therefore, the long-term x recently joined design does not allow for a meaningful analysis of ALMA's impact. However, during the exploratory analysis of the data, it became evident that school administrators' self-reported exposure to mentorship – one of ALMA's intervention components - was systematically associated with improved reading outcomes. As a result, the design was modified to explore the associations between mentorship of administrators and teachers and reading outcomes.

Sample size achieved. The study was conducted during an exceptionally severe rainy season. The average rainfall in March 2025, when the data was collected, exceeded 156mm, compared to 114mm in March 2024 and 99mm in March 2023[13]. Heavy storms, landslides, full rivers, collapsed roads and tracks, and flooding affected access to schools, resulting in teacher and student absenteeism. The situation had a negative impact on the number of students available for learning assessments and teachers present in school for observations. Additionally, three classroom observations were excluded from the final analysis due to the odd timing of entries, despite having apparently valid data.

Timeline. The study was conducted at the end of the first school trimester and therefore in a compressed timeline. Therefore, if the targeted teachers were not present at an assessed school for two consecutive days, one or more classroom observations were not conducted at the school, as the data collection team was unable to remain at the school for an extended period of time. Similarly, it was not possible for the data collection teams to return to the site later on to conduct 'mop up' observations due to the compressed timeline and the time required to access sites, often amounting to a day of travel in the case of locations reachable only on foot (overall, 14% of the sites were not accessible to vehicles due to collapsed roads and river crossings).

Definition of remoteness. The study considered an alternative definition of 'extremely remote' schools – those accessible on foot only. The MoE classification system rates schools as 'urban', 'rural', 'remote', 'very remote', and 'extremely remote'. However, some of the schools accessible only on foot were classified as 'urban' or 'rural', despite requiring data collection teams to walk up to 5km to reach the site. The alternative classification was used as the inability to access schools using a vehicle has a direct impact on the frequency of mentorship provided. Additionally, the findings show that schools accessible only on foot have a clear disparity in student learning levels, as noted in the following sections.

[13] See Timor-Leste Climate Explorer

Figure 2: Example of river crossing required to reach an extremely remote school

Figure 3: Collapsed Road with landslide on the way to an extremely remote school

Disability inclusion. Due to the limited use of sign language and Braille outside the capital Dili, students with complete loss of vision or hearing functionality could not be assessed through EGRA. In two cases, enumerators were informed by teachers that sampled students had severe hearing disabilities. The study incorporated some adaptations for disability inclusion, namely (i) the use of large fonts and high contrast print in student EGRA forms, catering to those with mild loss of vision functionality, who lacked access to corrective lenses; and (ii) the repetition of instructions where applicable, allowing students with low attentiveness and mild cognitive disabilities to engage in the tasks. The enumerators were trained to prevent discrimination of students with disabilities, including (i) non-discrimination in sampling; (ii) use of games with students with special needs who could not participate in the assessment, thus preventing them from feeling excluded; and (iii) the

use of adaptive strategies such as the repetition of instructions in case students seemed disengaged or did not respond, and applying games/ songs as part of the introduction (catering to those with mental health issues, particularly severe anxiety).

Key Findings

This section includes:

- Profiles of assessed grade 3 students, teachers, and schools.
- An analysis of literacy outcomes among grade 3 students, showing the results by each EGRA subtask, and the individual and school-level factors associated with the acquisition of literacy skills.
- A comparative analysis of the utilisation of ALMA's classroom observation tool and an analysis of teaching practices.
- An analysis of student and teacher attendance patterns.

3.1 Student Characteristics

Sex. Overall, 55% of the assessed students are male and 45% female. The proportion of male and female students replicates the proportion of male and female students enrolled in grade 3 in assessed schools (54% and 46%, respectively) as expected from a randomly selected sample. The proportion of girls among assessed students is similar across both study groups (47% long-term, 44% recently joined).

Age. The average age of grade 3 students is 8.7 years, without statistically significant differences by sex (8.8 years boys, 8.5 years girls) or study group (8.7 both). Overall, 18% of the students are 10 years or older, reflecting a pattern of repetition and /or late enrolment; 7% are ages 6-7[14]; and 2% did not know their age.

Language. Overall, 28% of the assessed students have a mother tongue other than Tetum.[15] The proportion of students whose mother tongue is not Tetum is highest in the recently joined group (35%), compared to 21% in the long-term group, largely due to the high prevalence of second language speakers in Baucau and Viqueque. The most common second languages are Makassae (28%) and Mambae (12%).

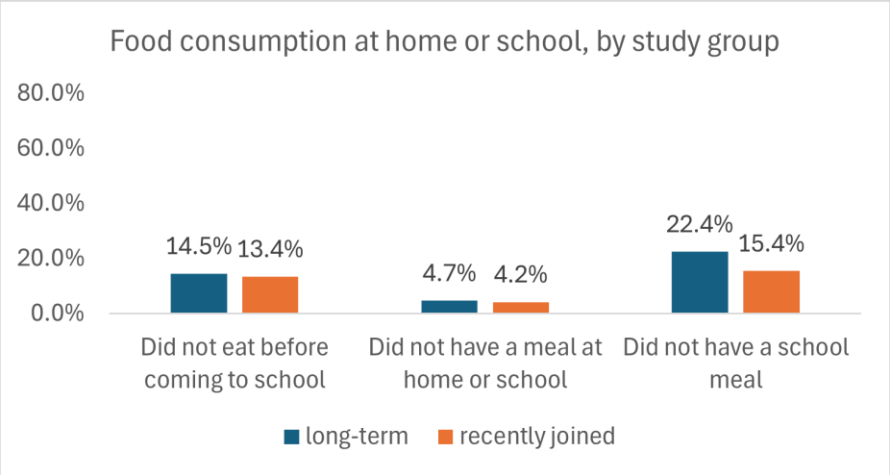
Food consumption. Among assessed students, 15% did not eat anything before coming to school; out of those, 32% attend schools which did not provide a meal on the day of the assessment. Overall, 5% of the students in long-term schools and 4% of those in recently joined schools did not eat at home or at school on the day of the assessment.

Overall, 81% of the assessed students attended schools that provided meals on the day of the survey (as observed by the enumerators). However, the proportion of students who did

[14] It is possible that the age was not informed correctly as some students may be innumerate.
[15] Including Tetum-Prasa and Tetum Terik.

not have access to a school meal is higher among the long-term group (22%) compared to the recently joined group (15%).

Figure 4: Food consumption patterns among assessed students, disaggregated by study group. Sources: Student survey (own consumption, self-reported) and school survey (observations of school feeding by enumerators).



Proportion of students attending schools receiving mentorship. The study asked grade 2-3 teachers to report when they received mentorship for the last time. Teachers who self-reported access to mentorship during the current school year were considered as having a mentor's support.

The study has also asked school administrators – i.e. directors or coordinators – who supported them to mentor teachers. School administrators who self-reported receiving support from a mentor or a female teacher coached for leadership (manorin feto potensiál) were considered as having access to mentorship. This measurement reflects school administrators' recall of support and therefore may not capture all exposure to mentorship. However, it is likely to identify the most effective cases of mentorship (i.e. recognized by administrators as a critical source of support).

The proportion of students whose teachers self-reported access to mentorship does not differ in long-term or recently joined schools. Similarly, the proportion of students whose school administrators reported access to a mentor does not differ between long-term or recently joined schools (see table below).

Table 2: Proportion of assessed students whose teachers and school administrators who self-reported access to mentorship, disaggregated by study group.

Study group	School administrator self-reported mentor support	School administrator did not self-report mentor support	Teacher received mentorship	Teacher did not receive mentorship
Long-term	31.7%	68.3%	83.3%	16.7%
Recently joined	30.3%	69.7%	85.0%	15.0%

The proportion of students attending schools where the administrator self-reported mentor support differs considerably by municipality in each study group. However, the sample is not designed for disaggregation by municipality. Therefore, the results should not be considered as indicative of the mentors' actual performance in the municipality as a whole, but rather only in the assessed schools in these municipalities.

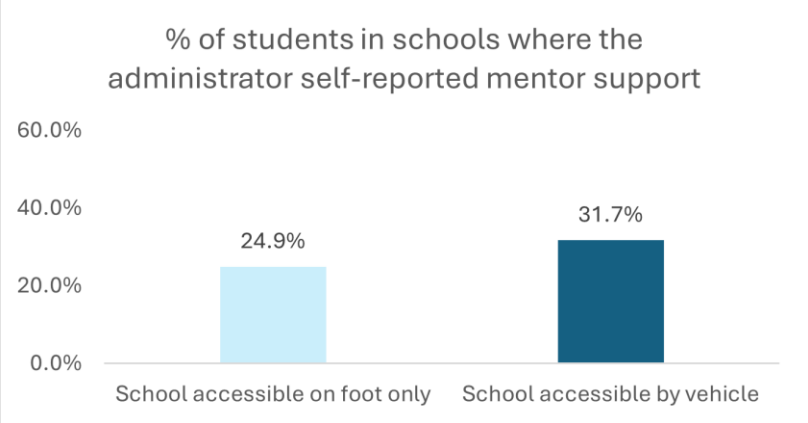
Table 3: Proportion of students in schools where the administrator self-reported receiving mentor support in each municipality, within the study groups

Study group	Municipality	% of students in schools where administrators self-reported mentor support
Long-term	Aileu	31%
Long-term	Baucau	27%
Long-term	Bobonaro	34%
Long-term	Liquica	69%
Long-term	Manatuto	44%
Long-term	Manufahi	10%
Long-term	Viqueque	0%
Recently joined	Ainaro	40%
Recently joined	Baucau	32%
Recently joined	Viqueque	20%

The proportion of students attending schools where the administrator self-reported mentor support is significantly lower[16] in extremely remote locations (i.e. accessible on foot only) – 25% compared to 32% in the schools accessible by vehicle.

[16] χ^2=3.896, p<0.028

Figure 5: Proportion of students attending schools where administrators self-reported mentorship, disaggregated by remoteness.



3.2 Teacher Characteristics

Observed teachers

Slightly over half of the observed grade 2-3 teachers are male – 52% compared to 48% female. There are substantial differences between male and female teachers:

- Female teachers are generally younger - 61% under the age of 50, compared 30% of the male teachers.
- Female teachers have less years of experience – 15.9 compared to 19.2 years among male teachers.
- The proportion of male teachers with any teaching qualification is higher – 88% compared to 82% of the female teachers. However, the proportion of female teachers with a higher education teaching degree[17] is larger – 21.5% compared to 15% of the male teachers. Overall, 67% of the teachers (60% female, 74% male) have bacharelato (the minimum teaching qualification in Timor-Leste) as their highest qualification.
- Overall, 9% of the observed teachers have taught for five years or less. However, 14.3% of the female teachers have taught for five years or less, compared to 4.6% of the male teachers.

Over half of the teachers (63%) received mentorship during the 2025 school year, and 72% attended GTP sessions. Overall, 21% are teaching multigrade classes.

All teachers in assessed schools

- The majority of the primary teachers are male – 61%. Recently joined schools have a slightly higher proportion of female teachers - 42% compared to 36% in long-term supported schools – but the difference is not statistically significant.

[17] *Licenciatura* (bachelor's degree) in education

- Schools where the administrator has self-reported receiving a mentor's support have a significantly higher proportion of female teachers – 46% compared to 36%[18].
- 51% of the grade 1-3 teachers have attended training on reading skills development, and 9% received training on multilingual education.

3.3 School Characteristics

- In average, the school administrator (director/ coordinator) has six years of experience as an administrator. However, 40% of the administrators are yet to complete a year in this position. The pattern is similar in long-term and recently joined schools.
- 90% of the school coordinators have a minimum teaching qualification (89% long-term schools, 90% recently joined), with 74% having bacharelato (75% long-term, 74% recently joined) as their highest qualification.
- Long-term supported schools tend to have poorer conditions than those that recently joined the program (see table below).

Table 4: Characteristics of assessed schools, disaggregated by study group

Characteristics	Long-term	Recently joined
Filial school	89%	83%
Multigrade classes	22%	18%
Accessible on foot only	16%	11%
Parent-Teacher Association (PTA) met in 2025 (active)[191]	26%	51%
Provided school meals	77%	81%
Students did not eat before coming to school	15%	13%
Has water (at least part of the time)	90%	84%
Has electricity	89%	90%
Has toilets for students	61%	60%
Has toilets for female students	19%	30%
Handwashing station	31%	50%
Handwashing station with soap	16%	20%
Reading corner in Grade 2 classes	59%	54%
Reading corner in Grade 3 classes	56%	54%

[1] Active PTA defined as having met at least once in 2025.

The results described above formed the basis for shifting the study design from a comparison between long-term and recently supported schools to an analysis of the impact

[18] $F_{1,140}=4.722$, $p<0.031$

[19] Regardless of the number of members.

of mentorship of administrators and teachers. For transparency, the results of the comparison of long-term and recently supported schools are provided in Annex 5.

3.4 EGRA Results

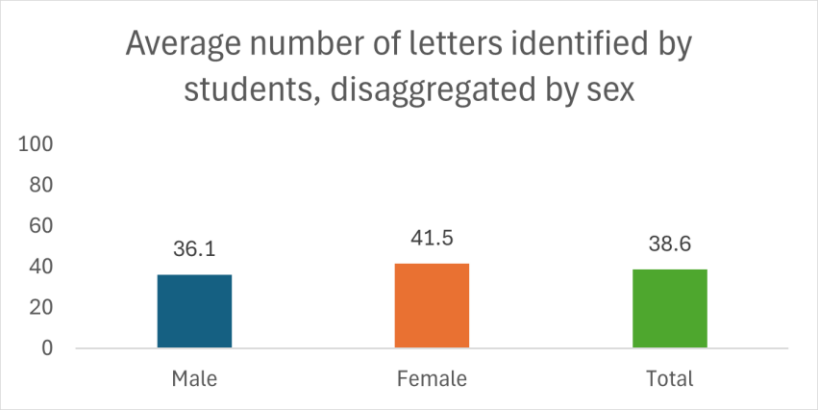
In this section, we describe the overall and disaggregated results in each EGRA task. Due to the modification in the original design, the results do not include disaggregation by study group (long-term or recently joined schools). These results are provided in Annex 5.

3.4.1 Letter Recognition

The letter identification task assesses student ability to identify letters in the Tetum alphabet in a timed manner. It counts the total number of letters recognized by students within a period of one minute, out of a maximum of 100 letters.

Average scores. Nearly all students (94.4%) could identify letters. In average, students could identify nearly 39 letters per minute. The average number of letters identified per minute is significantly higher[20] among female students, compared to boys.

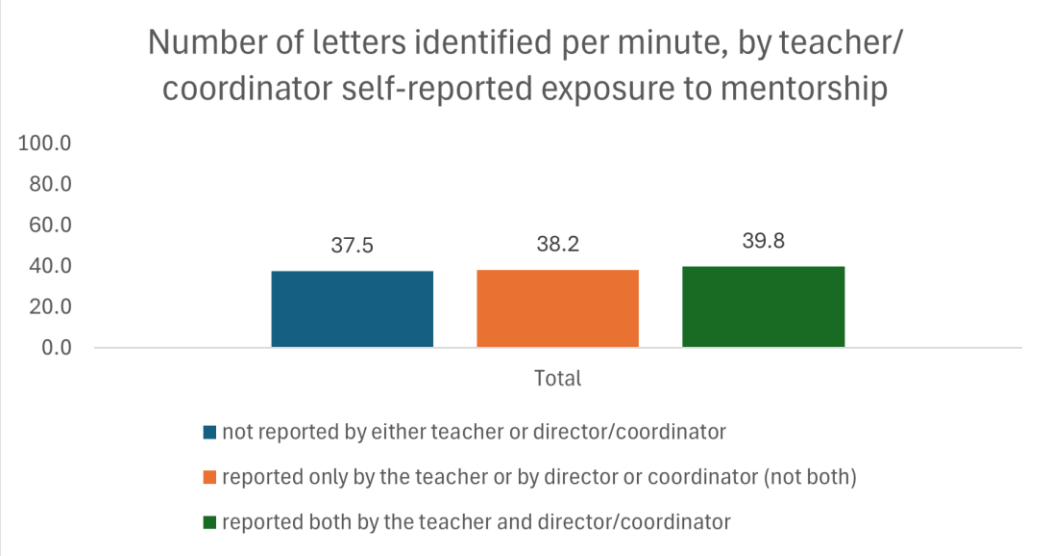
Figure 6: Average number of letters identified per minute (out of 100 maximum), disaggregated by sex



Mentorship. There is no statistically significant difference in the number of letters identified by students whose teachers or school directors/ coordinators self-reported receiving support from an ALMA mentor. However, there is a trend towards improved performance among students attending schools where both teachers and directors/ coordinators self-reported receiving support from an ALMA mentor.

[20] $F_{1,2065}=30.956$, $p<0.000$

Figure 7: Average number of letters identified per minute by students, disaggregated by level of exposure to mentorship (self-reported by teachers and/or school directors or coordinators)



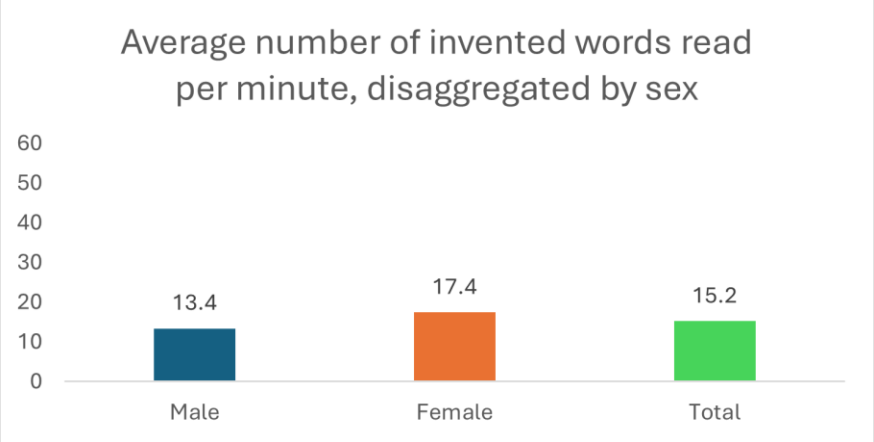
3.4.2 Invented Word Reading

"Invented words" are words with a similar structure to Tetum words, which do not exist in that language or in the students' mother tongues. The invented word reading task assessed reading fluency by counting the number of invented words read by students within a period of one minute, to a maximum of 60 words. The task seeks to assess reading fluency while removing the potential bias caused by sight recognition of familiar words, thus theoretically leveling the playing field for all students, regardless of language background.

Average reading fluency. In average, students could read 15 invented words per minute, out of a maximum of 60. Overall, girls are reading invented words at a higher speed, compared to boys - 17.4 words per minute/wpm compared to 13.4 wpm. The difference between boys and girls in invented word reading fluency is statistically significant.[21]

[21] $F_{1,2305}=45.678$, $p<0.000$

Figure 8: Average number of invented words read per minute (wpm), out of a maximum of 60, disaggregated by sex



Non-readers. Over a quarter of the students (27%) were unable to read any invented words. The proportion of male students unable to read any invented words is substantially higher (31%) compared to female students (22%). It is also higher in extremely remote schools (i.e. accessible on foot only) and among students whose mother tongue is not Tetum, as shown in the table below.

Table 5: Proportion of students unable to read any invented words, disaggregated by subgroups (sex, mother tongue, remoteness)

Subgroups	% of zero scores
Male	31%
Female	22%
Mother tongue: Tetum	26%
Mother tongue: Other	29%
School accessible on foot only	33%
School accessible by road	26%
Total	**27%**

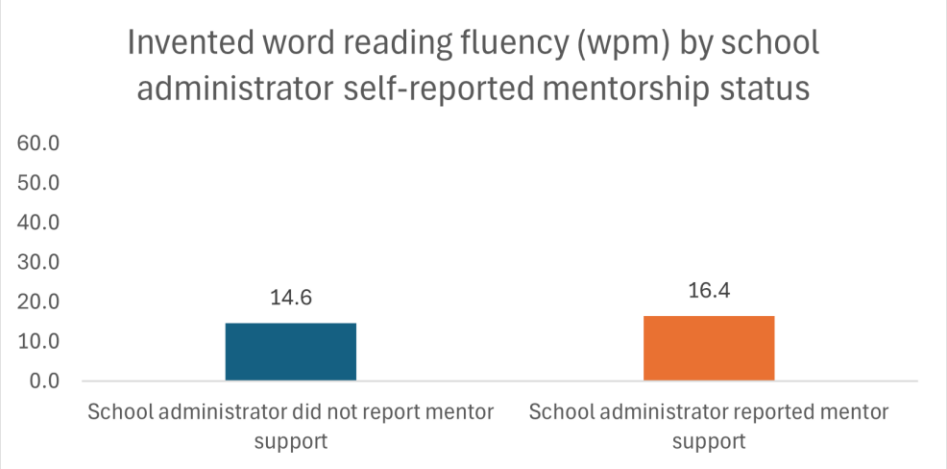
Mentorship. The exposure to mentorship identifies where ALMA is being implemented as intended – and therefore enables the study to assess if ALMA is having an impact on reading skills development when implemented as intended.

The average invented word reading fluency is significantly higher[22] among students attending schools whose directors/ coordinators self-reported receiving support from an ALMA mentor. The effect is similar across study groups (see figure below). There is no

[22] $F_{1,2065}=7.938$, $p<0.005$

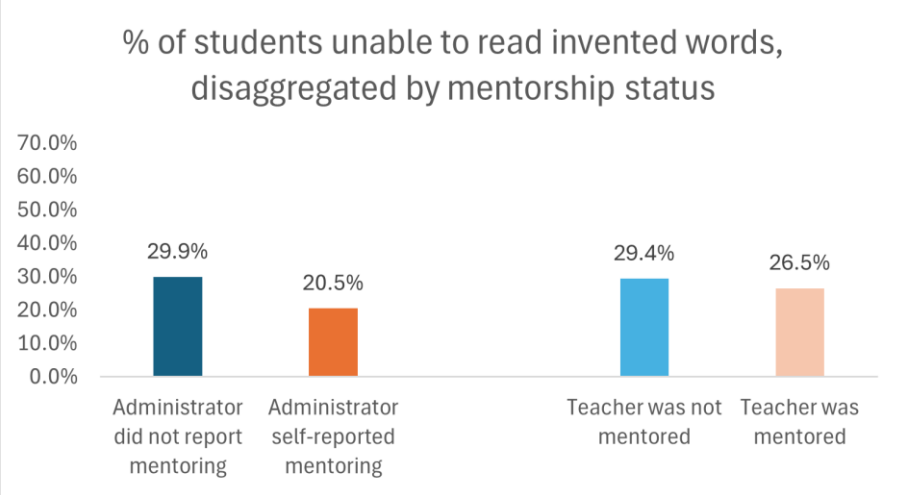
statistically significant difference in invented word reading fluency between students whose teachers self-reported mentor support and those who did not, regardless of study group.

Figure 9: Average invented word reading fluency (words per minute/wpm), disaggregated by self-reported exposure to mentorship by school administrators (directors or coordinators)



Mentorship seems to have had a strong impact in reducing the proportion of students scoring zero. This finding is further explored in the subsequent sections. In this sub-task, the proportion of students unable to read any words increases by nine percentage points where the administrator did not report receiving a mentor's support. The proportion of students unable to read any words increases by three percentage points when teachers did not receive mentorship.

Figure 10: Proportion of students unable to read any invented words, disaggregated by self-reported teacher or administrator exposure to mentorship



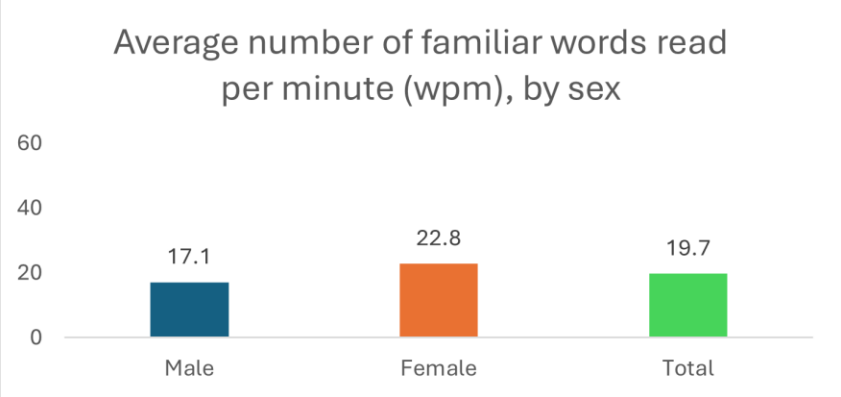
3.4.3 Familiar Word Reading

The familiar word reading task assesses the number of Tetum words read by the student in a period of one minute, out of a maximum of 60 words. "Familiar words" are part of the

regular vocabulary used by young children on a daily basis, including family members, body parts, animals, household items, etc. The familiar words are presented to student in a table arranged by level of difficulty – i.e. length, structural complexity, phonemes common across languages – starting with the least complex, shorter words.

Average familiar word reading fluency. In average, students read about 20 familiar words per minute. As expected, students are able to read familiar words faster than invented words, potentially due to sight recognition. Girls are reading familiar words at a faster pace than boys – 23 wpm compared to 17 wpm. The difference is statistically significant.[23]

Figure 11: Average number of familiar words read per minute, out of a maximum of 60, disaggregated by student sex



Non-readers. Overall, 27% of the students could not read any familiar words. The proportion of non-readers is much higher among boys; students whose mother tongue is not Tetum; and those who attend extremely remote schools.

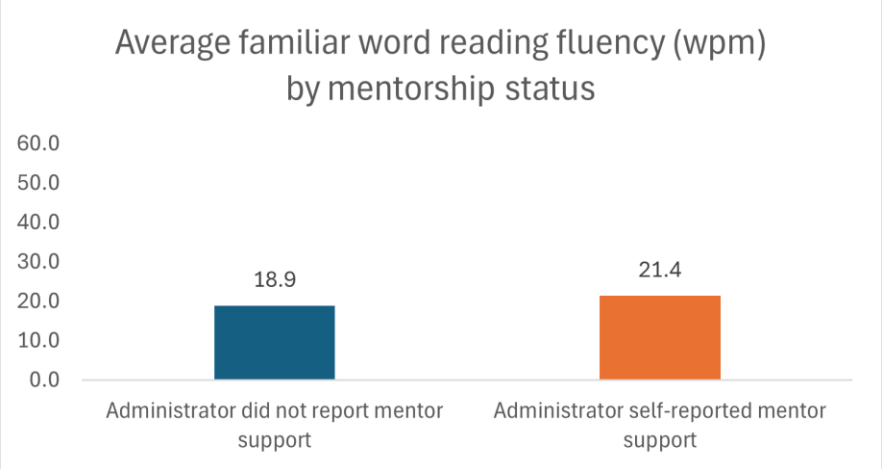
Table 6: Proportion of students scoring zero (unable to read any familiar words), disaggregated by subgroup

Subgroups	% of zero scores
Male	31%
Female	22%
Mother tongue: Tetum	26%
Mother tongue: Other	29%
School accessible on foot only	33%
School accessible by road	26%
Total	**27%**

[23] $F_{1,2065}=63.943$, $p<0.000$

Mentoring. The average familiar word reading fluency is significantly higher in schools where the administrator identified mentors as a source of support[24].

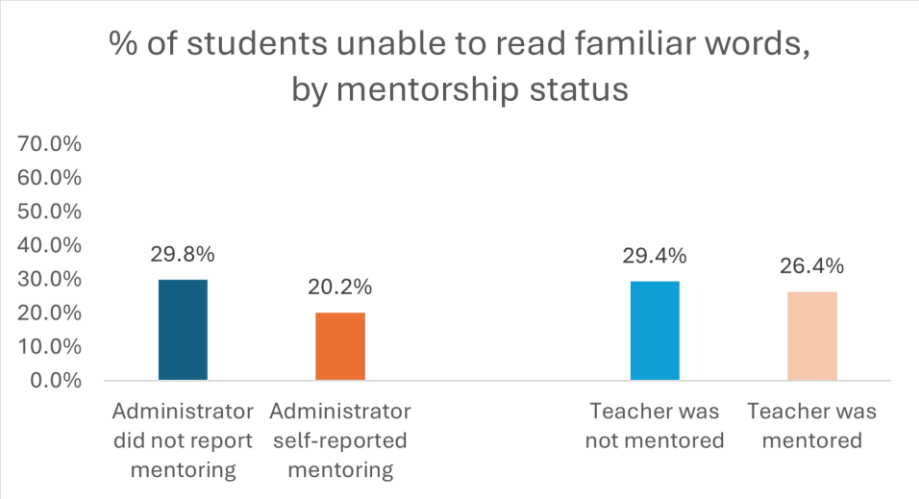
Figure 12: Average familiar word reading fluency (out of a maximum of 60 wpm), disaggregated by administrator self-reported access to mentorship



Mentorship and non-readers. The access to mentorship is associated with a reduced proportion of non-readers. In schools where the administrator self-reported access to mentorship, the proportion of non-readers is nearly 10 percentage points lower than in schools where the administrator did not report receiving a mentor's support. In schools where the teacher self-reported being mentored during this school year, the proportion of non-readers is three percentage points lower than in schools where teachers did not report being mentored. The results mirror those observed for invented word reading. Therefore, the combined findings indicate that ALMA has an impact in reducing the proportion of zero scores, which may be associated to school administrator actions supporting the most vulnerable learners.

[24] $F_{1,2065}=10.153$, $p<0.001$

Figure 13: Proportion of students scoring zero in familiar word reading in schools where (i) the administrator reported/did not report mentor support and (ii) the teacher received/did not receive mentorship.



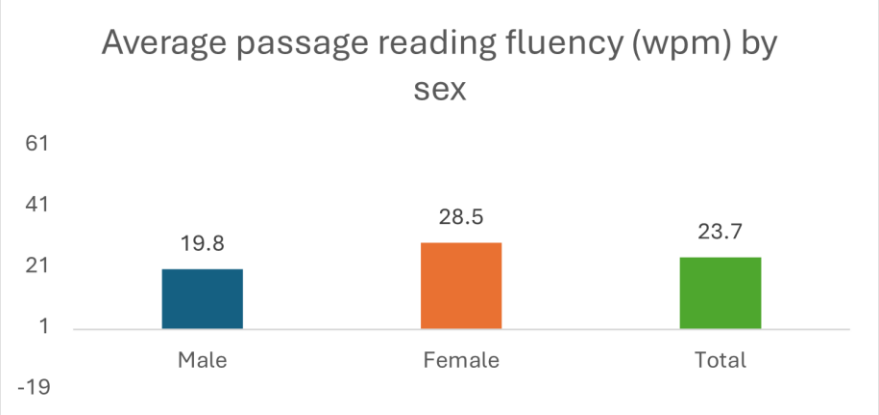
3.4.4 Passage Reading Fluency

The passage reading task measures the student's ability to read a short passage (61 words) in Tetum. The enumerator assessed the number of words read by the student within a period of one minute. The task allows the study to determine the proportion of students able to read text in the language of instruction; and to assess if the students can read at a minimum speed required to extract information from the text. Students who read very slowly are more likely to forget the information at the beginning of a text, hindering their capacity to read with comprehension.

Average passage reading fluency. Grade 3 students are reading an average of 24 wpm in the passage reading task. Girls have significantly higher passage reading scores compared to boys.[25] There is no difference in reading fluency if the student is attending a multigrade class or not.

[25] $F_{1,2064}=75.583$, $p<0.000$

Figure 14: Average passage reading fluency (in wpm), out of a maximum of 61wpm, disaggregated by student sex



Non-readers. Overall, 30% of the grade 3 students were unable to read any words in the passage reading task, compared to 27% in the invented word and familiar word reading tasks. The proportion of non-readers is significantly different by sex,[26] with a higher proportion of boys unable to read any words. There is also a significant difference by location, with a high proportion of non-readers among students from schools accessible on foot only.[27] There is no statistically significant difference between the proportion of non-readers by mother tongue, but the proportion of non-readers is higher among those whose mother tongue is not Tetum, in absolute terms.

Table 7: Proportion of students unable to read any words in a passage, disaggregated by subgroups

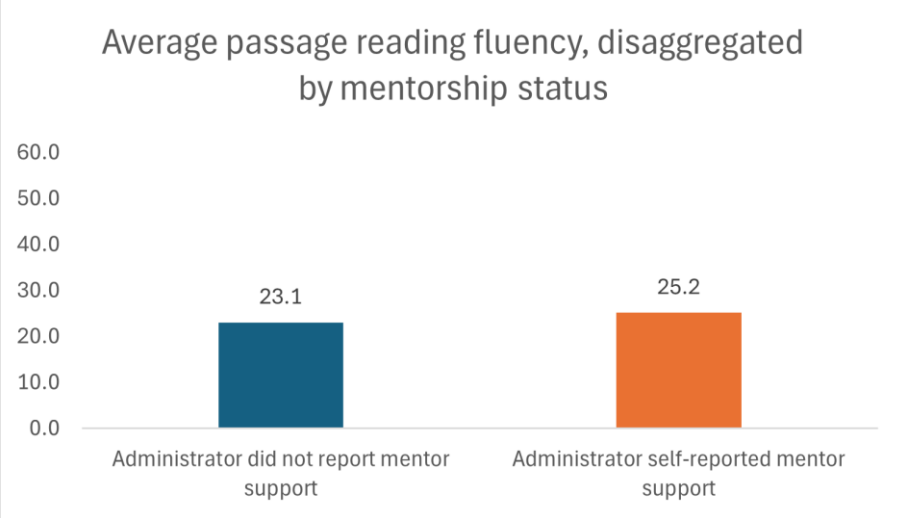
Subgroups	% zero scores
Male	35.5%
Female	24.1%
Mother tongue: Tetum	29.6%
Mother tongue: Other	32.2%
School accessible on foot only	36.8%
School accessible via road	29.6%
Total	**30.3%**

Mentorship. The average passage reading fluency is higher among students attending schools whose administrator (director/ coordinator) reported receiving support from a mentor. The difference is nearly significant.[28] There is no statistically significant difference of teacher self-reported exposure to mentorship on passage reading fluency.

[26] χ^2=30.761, p<0.000
[27] χ^2=4.385, p<0.023
[28] Significant at the 10% level; $F_{1,2064}$=3.689, p<0.055

The finding mirrors the results observed for familiar word reading and seem to indicate a pattern where the mentorship provided by ALMA to school administrators had a positive impact on reading fluency.

Figure 15: Average passage reading fluency (wpm), disaggregated by school administrator self-reported mentorship status

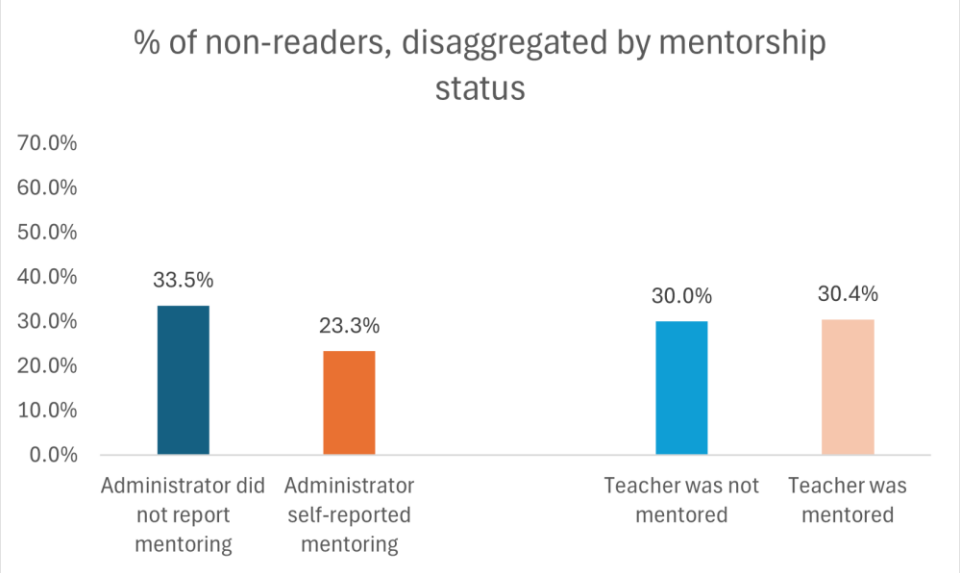


Mentorship and non-readers. The self-reported provision of mentorship to school administrators is associated with a significantly lower proportion of non-readers – a difference of 10 percentage points compared to schools where administrators did not self-reported mentorship. The difference is statistically significant. [29]There is no similar effect when mentorship is reported by teachers.

The finding is very similar to what was observed in the invented word and familiar word reading tasks and seems to indicate a strong positive impact of ALMA's mentorship in reducing the proportion of non-readers when provided to school administrators.

[29] χ^2=21.999, p<0.000

Figure 16: Proportion of non-readers (students unable to read a single word in a passage), disaggregated by mentorship status of administrators and teachers.



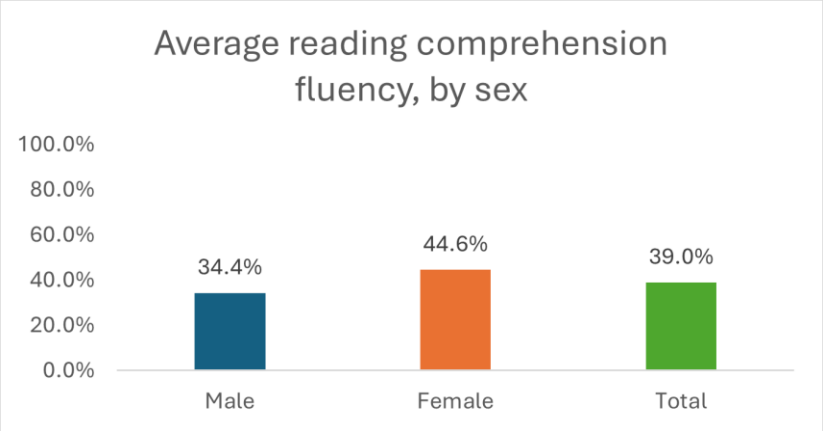
3.4.5 Reading Comprehension

The reading comprehension task involves asking the student to respond five questions about a short passage (the same applied in the passage reading task). The task is not timed as children reading in a second language tend to take longer to respond. By not timing the task, the study prevented potential over-reporting of students unable to read with comprehension.

Average reading comprehension score. In average, students scored 39% in this task. There is a large and significant difference by sex, with girls having a higher performance than boys (45% compared to 34%).[30] There is no difference between students attending multigrade or single-grade classes.

[30] $F_{1,2063}=39.824$, $p<0.000$

Figure 17: Average reading comprehension fluency score, disaggregated by student sex



Zero scores. Overall, 37% of the students were unable to respond reading comprehension tasks. Out of those, 82% were students unable to read any words. [31]The proportion of zero scores was considerably higher among boys, students whose mother tongue is not Tetum, and those attending extremely remote schools (i.e. accessible on foot only).

Table 8: Proportion of students scoring zero in the reading comprehension task, disaggregated by subgroups of students

Subgroups	% zero scores
Male	42.5%
Female	30.2%
Mother tongue: Tetum	35.5%
Mother tongue: Other	40.6%
School accessible on foot only	45.8%
School accessible by vehicle	35.9%
Total	**37%**

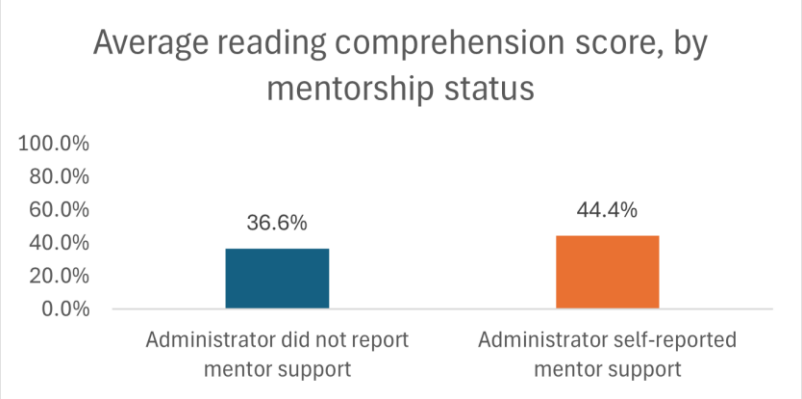
Mentorship. Students attending schools where the administrator (director/ coordinator) self-reported receiving mentor support had significantly higher reading comprehension scores.[32] The difference reaches nearly eight percentage points. There is no difference between the performance of students in schools where grade 2-3 teachers reported receiving mentorship and those who did not receive it.

The finding shows a pattern of higher student reading performance in schools where the administrator self-reported receiving mentorship, mirroring the results observed in the

[31] The assessment was terminated early for students unable to read words. The analysis allocates a zero score in reading comprehension to non-readers.
[32] $F_{1,2063}=19.478$, $p<0.000$

word/ passage reading tasks. Taken together, these results suggest a strong association between ALMA's mentorship on administrators and improved reading results.

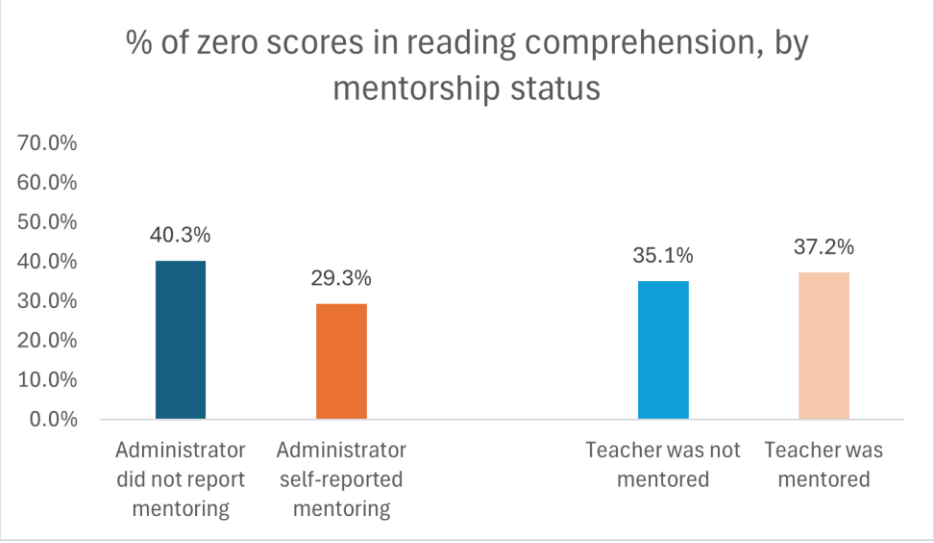
Figure 18: Average reading comprehension scores, disaggregated by school administrator self-reported mentorship status



Zero scores and mentorship. The proportion of zero scores in reading comprehension declines by 11 percentage points in schools where the administrator self-reported having mentor support. There is no difference in the proportion of zero scores in schools where grade 2-3 teachers reported receiving mentorship and those who did not.

Again, the result reinforces the trend of improved reading performance among students attending schools where administrators self-reported support from an ALMA mentor.

Figure 19: Proportion of zero scores in reading comprehension in schools where the administrator or teacher reported receiving mentorship



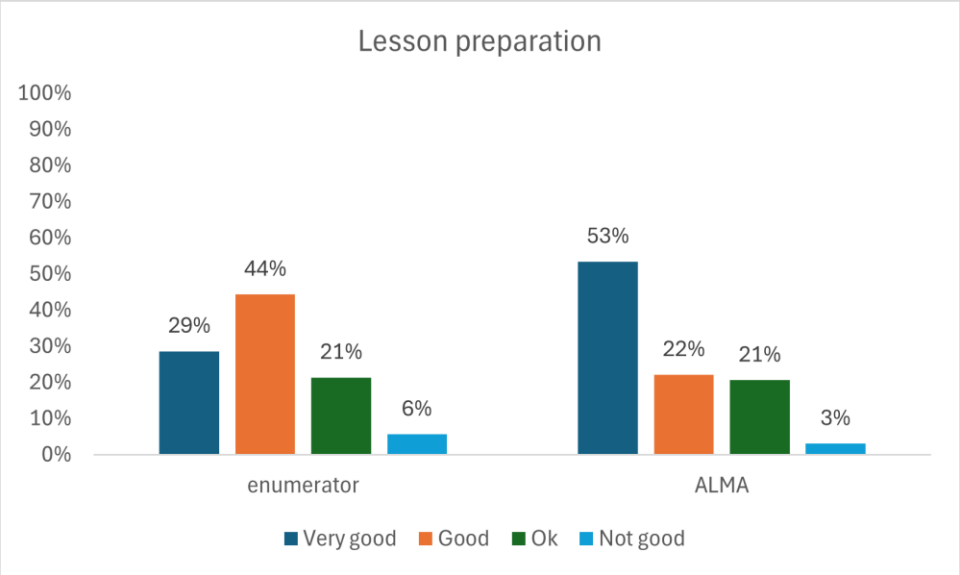
3.5 Comparative Analysis of Classroom Observations

The classroom observation tool assesses five domains – lesson preparation, instruction, implementation of the lesson plan, classroom management, and socio-emotional skills. The

tool is used by ALMA mentors to inform the technical support provided to school leaders and teachers. This component of the study compares the results from (i) observations conducted by ALMA mentors and (ii) observations conducted by external enumerators to assess consistency and analyse patterns. To ensure comparability, only observations of Tetum classes were included in the analysis.[33]

Overall, ALMA mentors tend to assess the majority of the teachers as using 'very good' practices- from 40% to 63%, across all observations in each of the domains. In contrast, the external enumerators tend to rate teaching practices as 'good' – 36% to 53% - with a smaller proportion rated as 'very good' – 10 to 29%. The proportion of teachers rated as 'ok' is similar in ALMA-led and enumerator-led observations. The proportion of teachers whose practices were rated as 'not good' is very small in both cases, although systematically higher in enumerator-led observations.

Figure 20: Domain 1- lesson preparation -results of classroom observations conducted by ALMA mentors and external enumerators



[33] The classroom observation dataset provided by ALMA included 385 valid observations, out of which 131 were for Tetum classes. The results were compared with 249 observations conducted by external enumerators.

Figure 21: Domain 2 - instruction strategies - results of classroom observations conducted by ALMA mentors and external enumerators

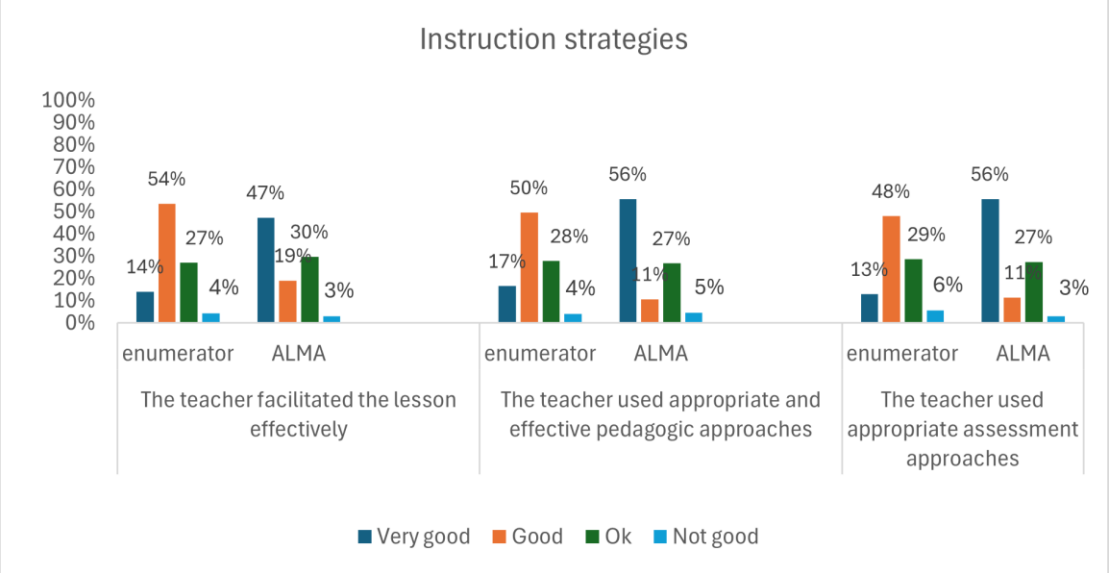
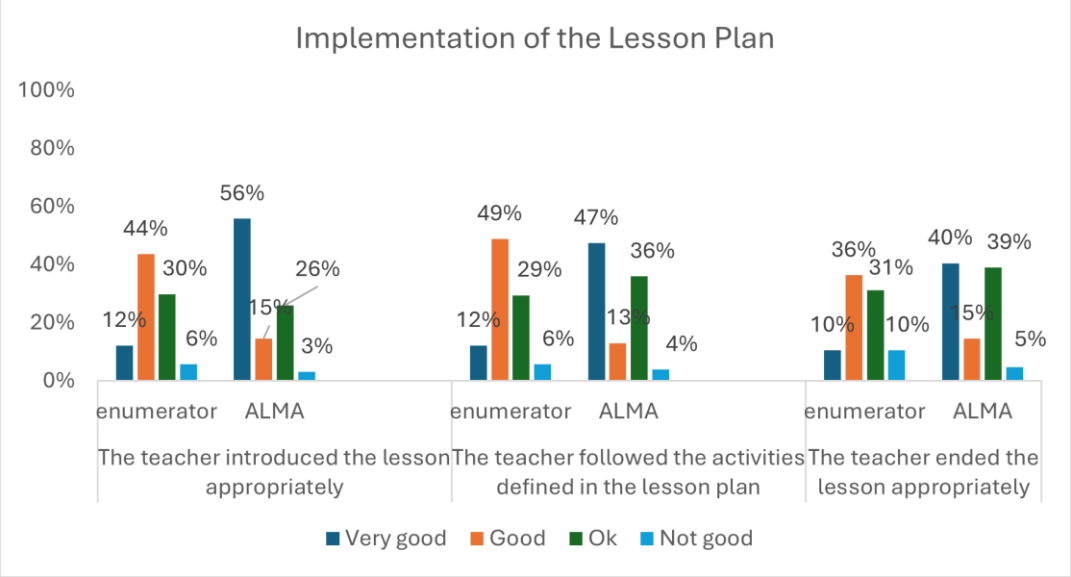


Figure 22: Domain 3 - implementation of the lesson plan - results of classroom observations conducted by external enumerators and ALMA mentors



Multi-panel bar chart showing ratings of classroom management for enumerators and ALMA across two aspects: creating an inclusive classroom environment and promoting positive attitudes. For both aspects, ALMA is more frequently rated ‘very good’ than enumerators. Enumerators are more often rated ‘good’, while ratings of ‘not good’ are low for both groups.

Figure 23: Domain 4 - classroom management - results of classroom observations conducted by external enumerators and ALMA mentors

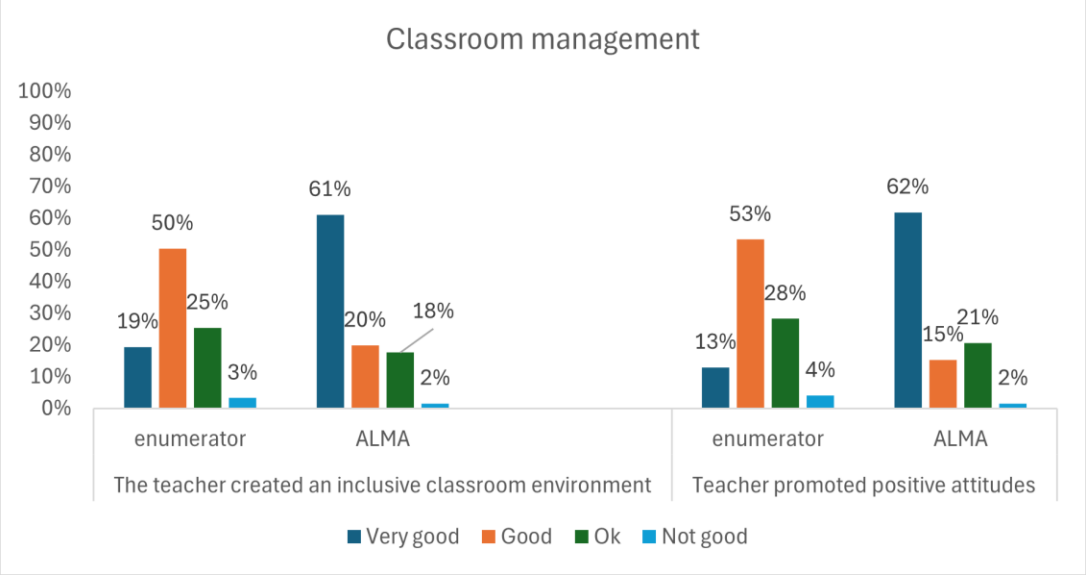
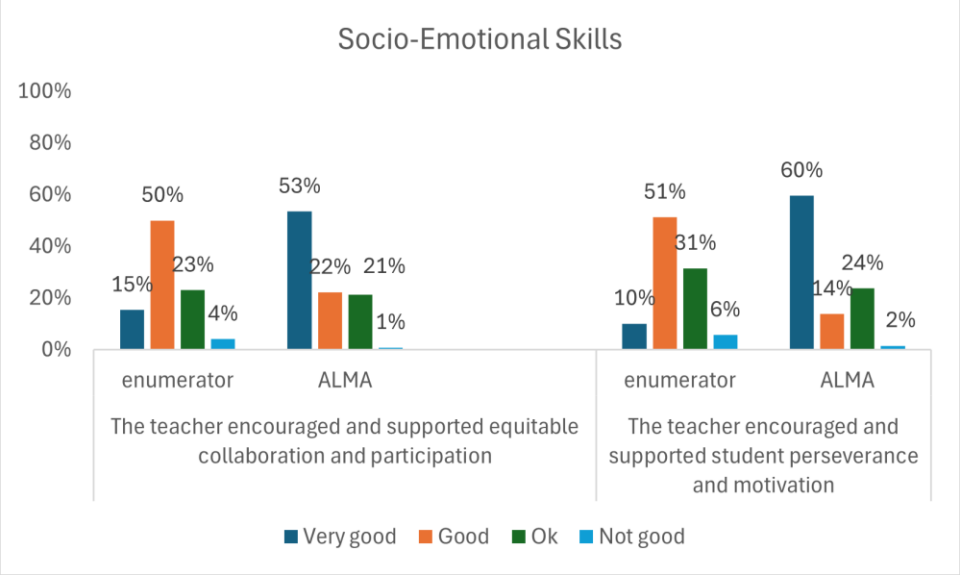


Figure 24: Domain 5 - socio-emotional skills - results of classroom observations conducted by external enumerators and ALMA mentors



Out of the 131 observations of Tetum classes conducted by ALMA mentors, 102 (78%) were from Baucau schools, compared to only 29% of those conducted by external enumerators. To eliminate potential bias in interpreting the results, a secondary comparison was conducted, looking exclusively at Baucau observations from both ALMA and external enumerators. The results are consistent with previously observed trends: ALMA enumerators consistently rated most teachers as 'very good' across all domains and practices. The proportion of teachers whose practices were rated by external enumerators as 'not good' is higher than in the previous analysis, while ALMA mentors systematically rated very few teachers (1-3% only) as showing 'not good' practices.

Table 9: Comparison of classroom observation results obtained by ALMA mentors and external enumerators, considering only Baucau schools

Domain 1: Lesson preparation

Source	Very good	Good	OK	Not good
External enumerator	38%	39%	19%	4%
ALMA mentor	54%	25%	18%	3%

Domain 2: Teacher facilitation and pedagogic practice: Teacher facilitated the lesson effectively

Source	Very good	Good	OK	Not good
External enumerator	15%	49%	32%	4%
ALMA mentor	50%	19%	27%	3%

Teacher used appropriate and effective pedagogic approaches

Source	Very good	Good	OK	Not good
External enumerator	10%	47%	32%	8%
ALMA mentor	56%	11%	30%	2%

Teacher used appropriate assessment approaches

Source	Very good	Good	OK	Not good
External enumerator	14%	38%	32%	10%
ALMA mentor	57%	11%	29%	2%

Domain 3: Lesson structure and implementation: Teacher introduced the lesson appropriately

Source	Very good	Good	OK	Not good
External enumerator	15%	29%	35%	10%
ALMA mentor	54%	16%	27%	3%

Teacher followed the activities defined in the lesson plan

Source	Very good	Good	OK	Not good
External enumerator	18%	36%	33%	10%
ALMA mentor	45%	14%	38%	3%

Teacher ended the lesson appropriately

Source	Very good	Good	OK	Not good
External enumerator	13%	26%	39%	14%
ALMA mentor	40%	15%	43%	2%

Domain 4: Classroom environment and attitudes
Teacher created an inclusive classroom environment

Source	Very good	Good	OK	Not good
External enumerator	24%	47%	25%	3%
ALMA mentor	62%	21%	17%	1%

Teacher promoted positive attitudes

Source	Very good	Good	OK	Not good
External enumerator	11%	44%	39%	4%
ALMA mentor	63%	15%	22%	1%

Domain 5: Student participation and motivation
Teacher encouraged and supported equitable collaboration and participation

Source	Very good	Good	OK	Not good
External enumerator	13%	57%	15%	6%
ALMA mentor	52%	24%	22%	1%

Teacher encouraged and supported student perseverance and motivation

Source	Very good	Good	OK	Not good
External enumerator	10%	43%	38%	7%
ALMA mentor	64%	13%	22%	1%

The results suggest a bias of ALMA mentors towards overly positive perceptions of classroom practices, and a reluctance from both mentors and enumerators in rating practices as 'not good'. However, it is interesting to note the similarities in rating practices as 'ok' – which may suggest that both ALMA mentors and external enumerators may be rating poor practice as 'ok' to avoid a 'not good' response.

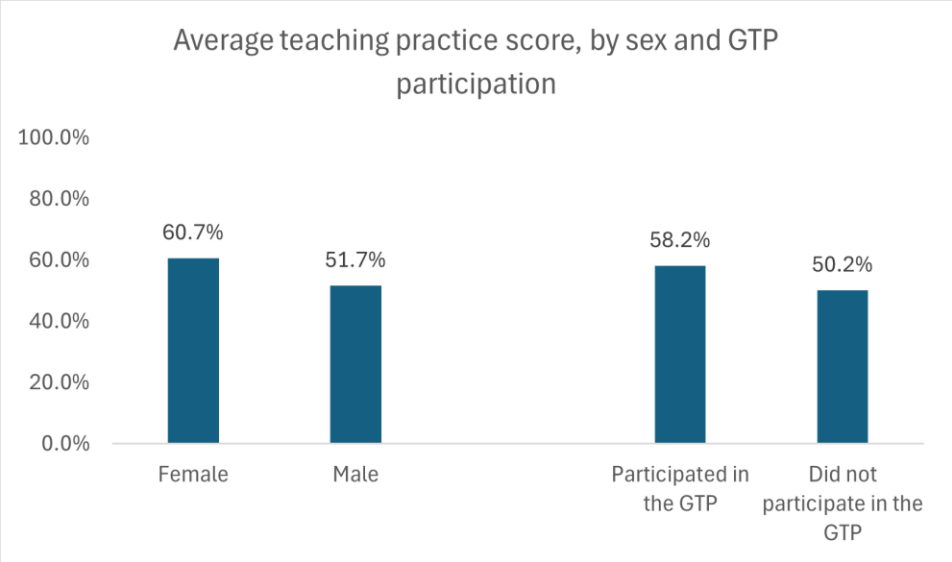
Reliability. The results for the 11 items (teaching practices/ behaviours) in the classroom observation tool tend to have high internal consistency (statistical reliability - with a

Cronbach's alpha of 0.884). Therefore, the results can be summed together to form a scale. In the following sections, we will use the scale to explore patterns in teaching practices.

Patterns in the enumerator-led observations. The participation in the GTP is associated with a significantly higher teaching practice score[34]. The average teaching practice score is 58% among teachers who participated in the GTP during the 2025 school year (January – March), compared to 50% among those who did not. Teachers were asked about what they have learnt from their participation in the GTP. There is a positive correlation between the number of items recalled by teachers as learnt from the GTP and the average teaching practice score: the higher the number of items recalled by teachers as learnt in the GTP, the higher the average teaching score.[35]

Female teachers have a significantly higher use of improved teaching practices, reaching an average score of 61% compared to 52% among men.[36]

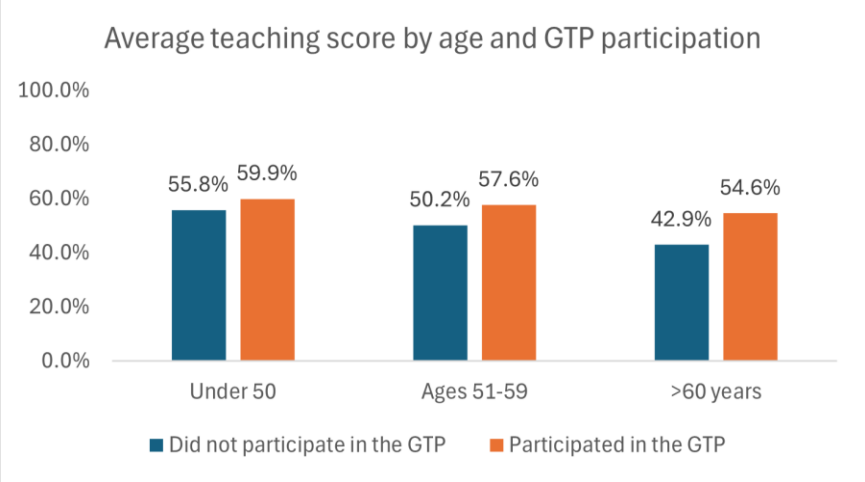
Figure 25: Average teaching practice score, disaggregated by sex and GTP participation. Differences are statistically significant.



Younger teachers (under 50 years of age) have significantly higher average teaching scores – 59% compared to 56% among those aged 51-59, and 50% among those over the age of 60. However, teachers who participated in the GTP in 2025 have higher teaching practice scores regardless of the age range. While there is no statistically significant interaction of age and GTP participation, possibly due to the smaller sample sizes, the trend suggests that teachers who systematically participate in the GTP are benefitting from the practice, and this is likely to be a stronger determinant of teaching performance than age.

[34] $F_{1,249}$=8.634, p<0.004
[35] r =.127, p<0.046
[36] $F_{1,249}$=14.059, p<0.000

Figure 26: Average teaching practice score, disaggregated by age and GTP participation.



Differences by age and GTP participation are statistically significant, but there is no statistically significant interaction age x GTP participation.

There is no statistically significant difference in the average teaching practice score between (i) multigrade and single-grade classes, (ii) teachers who received mentorship during this year, or (iii) administrator self-reported mentorship. There is also no statistically significant difference by qualification – although teachers who have a teacher training college or university degree in education have a higher average result (59% compared to an average of 56%), their total representation in the sample is probably too small for results to be statistically significant. There is also no correlation between the total number of years of teaching practice and the average teaching practice score.

3.7 Factors Affecting Reading Outcomes

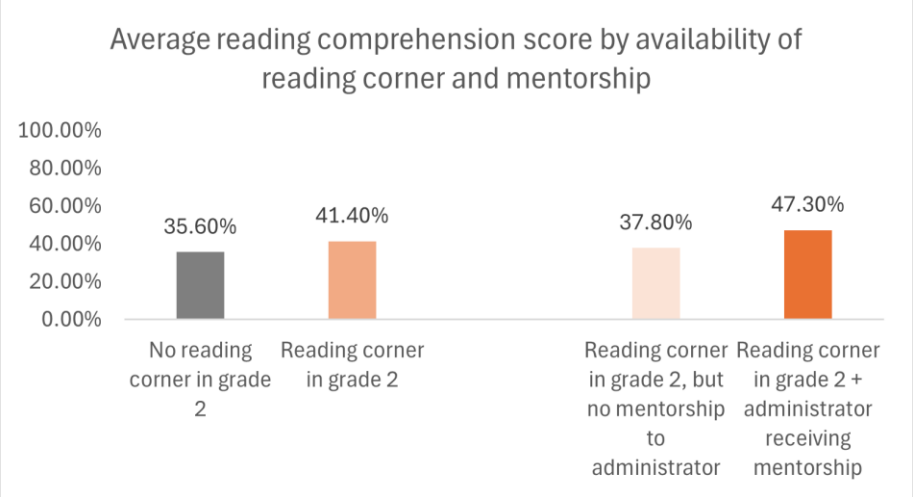
Access to Reading Materials at School

The average reading comprehension score among Grade 3 students is significantly higher in schools where there is a reading corner in the Grade 2 classroom (where the student theoretically attended class last year). The availability of a reading corner in grade 2 is significantly higher in schools where the administrator self-reported mentorship –69.8% compared to 50.5% in schools where the administrator did not self-report receiving a mentor's support.[37]

The combination of (i) having a reading corner in grade 2 and (ii) administrators self-reporting support from a mentor result in an increase of nearly 12 percentage points in the average reading comprehension score.

[37] χ^2=4.491, p<0.026

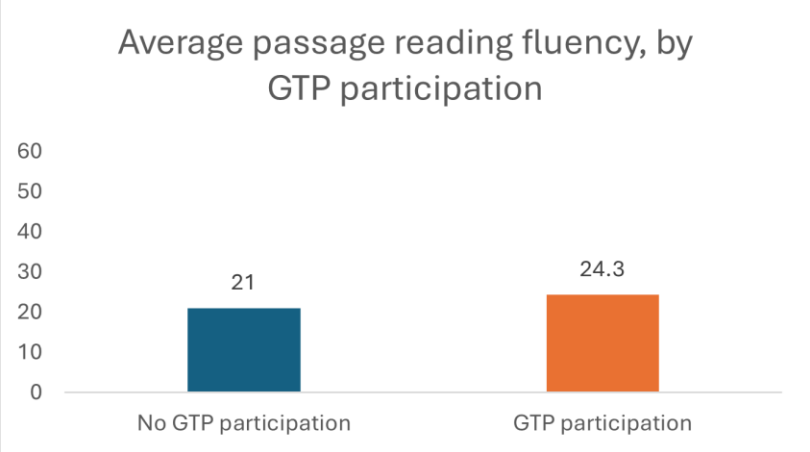
Figure 27: Average reading comprehension score (lower level) disaggregated by presence of reading corner in Grade 2 classes, with and without mentorship reported by administrators.



GTP Attendance

Teacher participation at the GTP during the 2025 school year (January-March) is associated with significantly higher passage reading scores.[38]

Figure 28: Average passage reading fluency score (wpm), disaggregated by teacher participation at the GTP



There is no statistically significant association between the participation in the GTP and reading comprehension scores. As noted above, the participation in the GTP is associated with a higher utilization of improved teaching practices. The result suggests that the participation in the GTP is particularly beneficial to strengthen reading practices associated with lower-level reading skills (i.e. decoding) rather than comprehension.

[38] $F_{1,2064}=6.524$, $p<0.011$

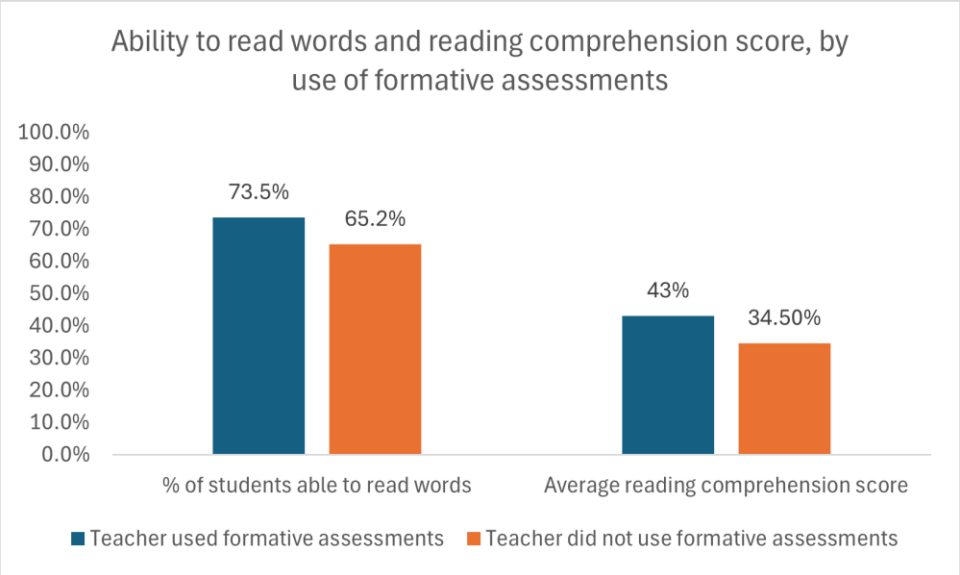
Use of Formative Assessments

The proportion of teachers using formative assessments[39] is considerably higher among those who self-reported receiving mentorship during the current school year - 55% compared to 42% among those who did not receive mentorship in January-March 2025. The use of formative assessments by teachers is associated with:

-A higher proportion of students able to read words in a passage – 74% vs 65% among students whose teachers did not use formative assessments.[40]

-Higher student reading comprehension scores – 43% vs 35% among those whose teachers did not use them.[41]

Figure 29: Proportion of students able to read words in a passage and average reading comprehension score, disaggregated by teacher use of formative assessments



Teaching practices

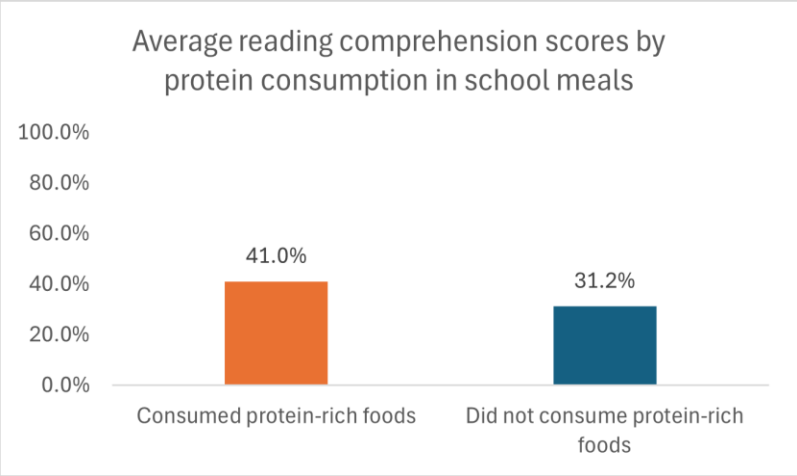
There is a positive correlation between the use of improved teaching practices - measured through the teaching practices score, based on the results from enumerator-led classroom observations – and passage reading fluency.[42] Similarly, there is a positive correlation between the use of improved teaching practices and the reading comprehension score.[43] Therefore, higher teaching practice scores are associated with higher passage reading scores and reading comprehension scores.

[39] Based on observed evidence of use. When the teacher affirmed using formative assessments, the enumerator requested to teacher to show evidence of use.
[40] χ^2=16.830, p<0.000
[41] $F_{1,2063}$=27.164, p<0.000
[42] r=.092, p<0.000
[43] r=.109, p<0.000

Protein Consumption, Cognitive Development, and Mentorship

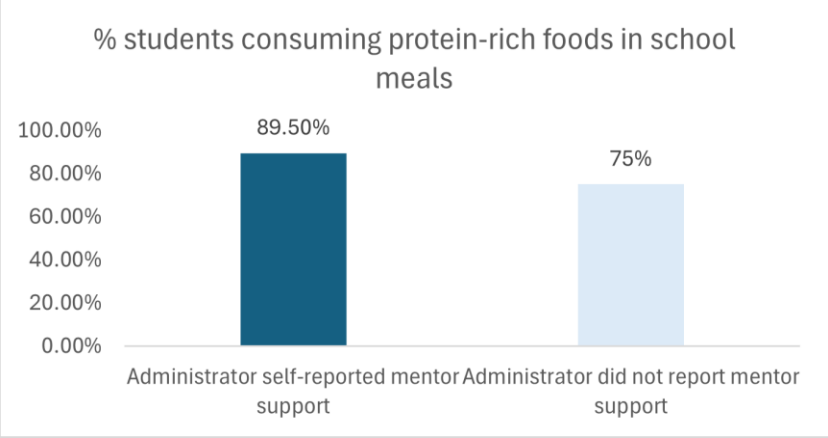
Students who did not consume protein-rich foods at school (20% of the total) have significantly lower reading comprehension scores – a decline of 10 percentage points (31% compared to 41% among those who consumed protein-rich foods).[44]

Figure 30: Average reading comprehension score (lower level) among students who self-reported consuming protein-rich foods in school meals and those who did not consume them.



Interestingly, the proportion of students who reported consuming protein-rich foods at school is significantly higher[45] in schools where the administrators self-reported receiving mentorship: 89.5% compared to 75% in schools where the administrators did not report mentorship support.

Figure 31: Proportion of students reporting consumption of protein-rich foods by administrator's self-reported mentor support

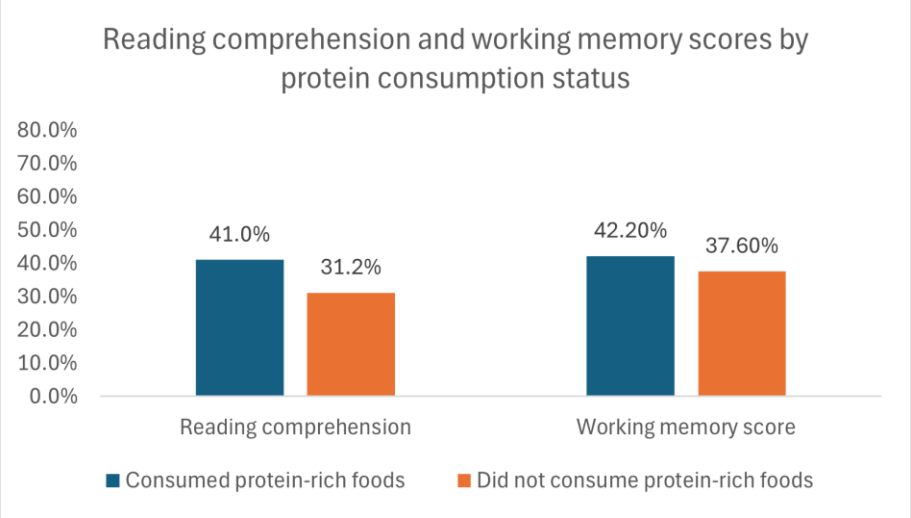


The consumption of protein-rich foods is associated with significantly higher working memory scores (the ability to sustain information for a pre-determined period of time). As

[44] $F_{1,2063}=24.176$, $p<0.000$
[45] $\chi^2=57.116$, $p<0.000$

expected, there is a strong correlation between working memory scores and reading comprehension scores.[46]

Figure 32: Average reading comprehension and working memory scores, disaggregated by student self-reported consumption of protein in school meals.



Availability of water at school

Overall, 11% of the students are attending schools where there is no water. The proportion of female non-readers is significantly higher[47] in schools without water – a difference of 13 percentage points. The proportion of male non-readers is nearly-significantly higher[48] in schools without water – a difference of 8 percentage points. The proportion of students with access to water at school is higher among those whose administrators self-reported receiving a mentor's support – 92% compared to 88%. The difference is statistically significant.[49]

[46] $p<0.000$; $r=0.397$
[47] [47] $\chi^2=9.555$, $p<0.002$
[48] $\chi^2=2.783$, $p<0.060$
[49] $\chi^2=4.977$, $p<0.014$

Figure 33: Proportion of students unable to read a single word in the passage reading task, disaggregated by sex and water availability at school

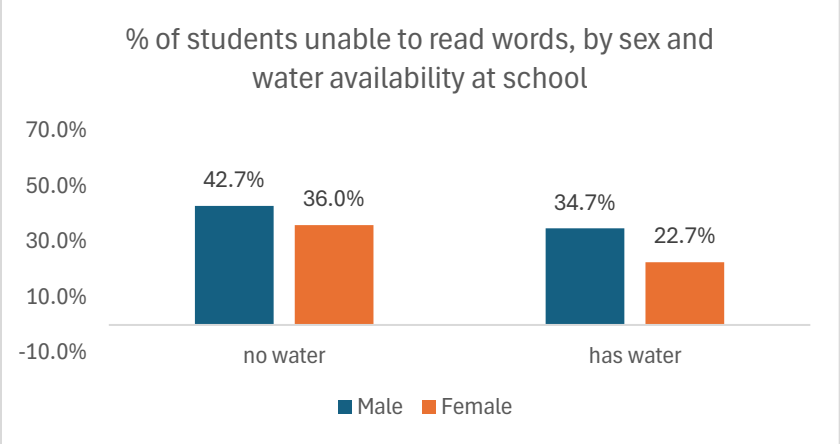
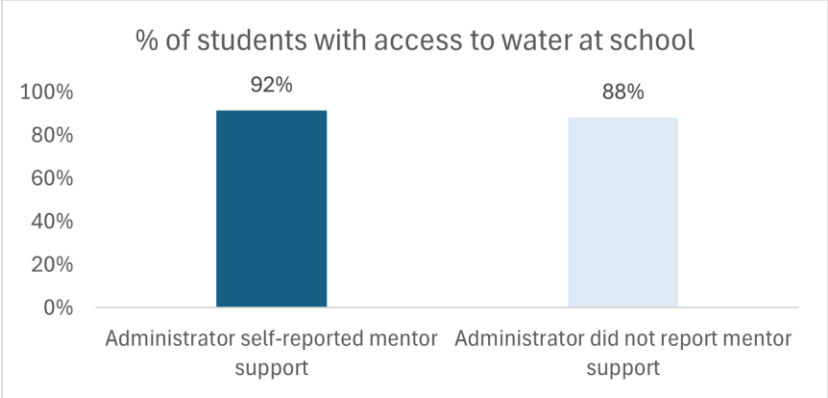


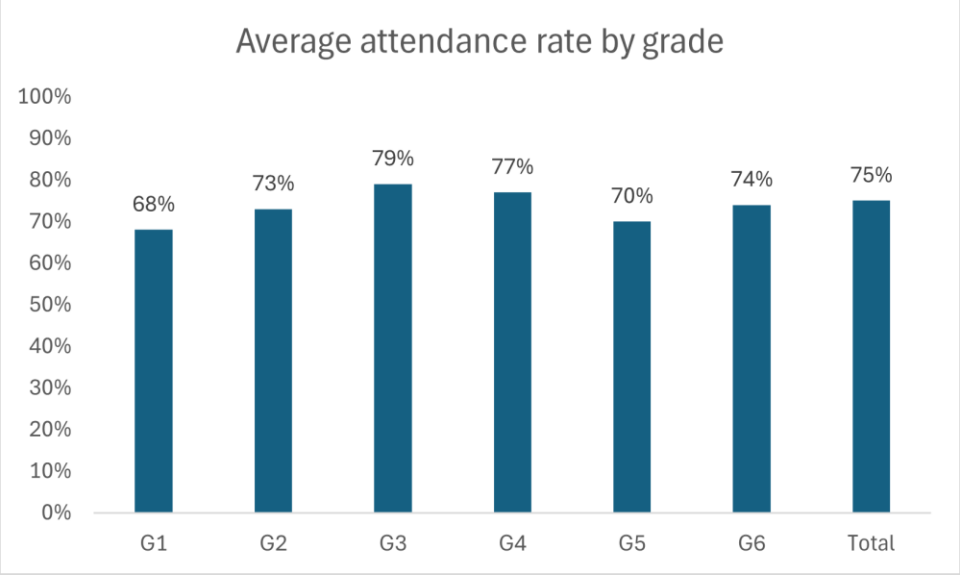
Figure 34: Proportion of students with access to water at school, disaggregated by administrator self-reported mentor support



3.8 Attendance

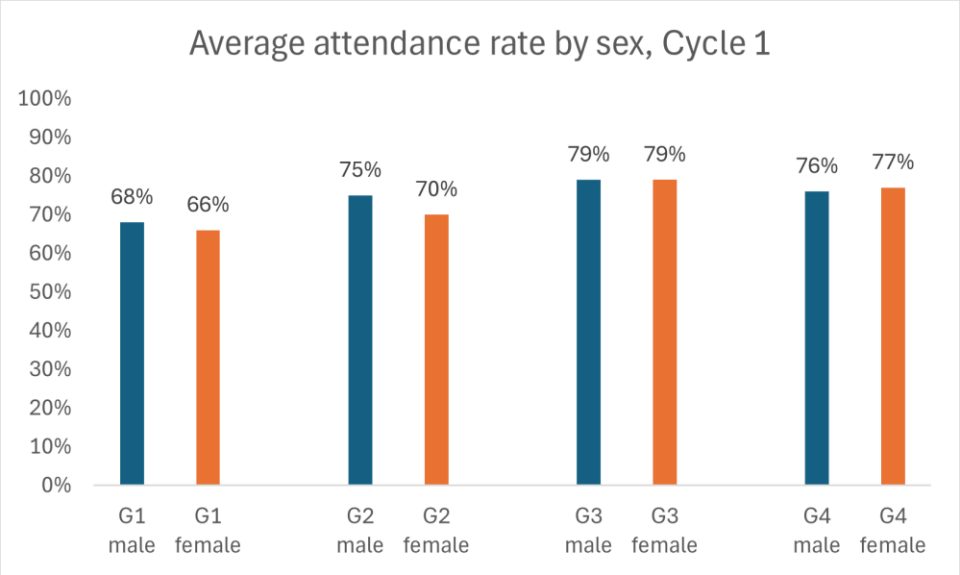
Student attendance. The average student attendance rate is low (75%), and tend to be lower in grades 1-2. The average student attendance rate in grades 1-3 is considerably lower in schools accessible on foot only - 69% compared to 74% in schools accessible by vehicle – but the difference is not statistically significant (possibly due to the smaller sample).

Figure 35: Average attendance rates by grade (Cycles 1-2)



When disaggregated by sex, female students have lower attendance rates in grades 1-2, compared to boys. Attendance rates are similar for both sexes in grades 3-4

Figure 36: Average attendance rates of Cycle 1 students (grades 1-4), disaggregated by sex and study group



Teacher attendance. The average teacher attendance rate is 82%, without significant differences by mentorship status of school administrator or school characteristics (remoteness, water availability, meal provision). It is possible and likely that teacher attendance is mostly influenced by individual teacher characteristics (food insecurity, health conditions) and/or by external factors (weather conditions). As expected, there is a

positive correlation between teacher and student attendance, particularly in early grades (1-3).[50]

The high absenteeism rates – with nearly a quarter of the students and a fifth of the teachers not physically present at school – are likely to be eroding ALMA's impact on learning outcomes. To ultimately effect change in literacy skills development, students need to have minimum exposure time to the improved practices implemented by teachers.

3.9 Are Administrators Self-Reporting Mentor Support Simply "Better Administrators"?

This study found a systematic association between the non-prompted recall of mentor support by school administrators and improved reading outcomes. This pattern calls into question if the findings are actually associated with mentorship or are simply a reflection of the work of a better administrator.

A better school administrator – be it more skilled, professional, and/or motivated – is likely to be more receptive of a mentor's support and to invest time and effort in learning from a mentor. Conversely, less skilled school administrators may have realized greater benefits from mentorship, as they started from a lower performance point.

To explore this question, the study:

- Compared the profiles of administrators who self-reported a mentor's support and those who did not do so; and
- Conducted a comparative analysis of patterns associated with improved management performance in schools where the administrator self-reported mentor support and in those where administrators did not do so.

Experience and academic background of school administrators

The proportion of administrators with a university degree in education[51] is higher among those who self-reported mentor support, but still very small in both groups (18% and 11%, respectively). The average number of years as an administrator is not significantly different across both groups.

Table 10: Comparative analysis of school administrator experience and academic background

Characteristic	Administrators who self-reported mentor support	Administrators who did not self-report mentor support
Average number of years of experience as an administrator	6.3 years	5.5 years
Proportion with a university degree in education	18%	11%

[50] r=.394, p<0.000

[51] Defined as having completed *licenciatura* (equivalent to a bachelor's degree) in education, as opposed to the short-term *bacharelato*, which is the minimum teaching qualification in Timor-Leste

Comparative analysis of management performance

A competent school administrator is expected to enforce teacher attendance; ensure participatory governance processes; coach teachers as part of a supportive supervisory role; and foster teacher participation in learning circles. The following table provides an overview of the status of these key performance indicators in schools where administrators self-reported or did not report mentor support. The only statistically significant difference between the two groups is the higher proportion of teachers attending GTPs in 2025 in schools where administrators self-reported mentor support – which may be a reflection of their own recognition of the value of mentorship.

Table 11: Comparison of School Management Practices by Administrator Reported Mentorship

Indicator	Schools where administrators self-reported mentor support	Schools where administrators did not self-report mentor support	Statistically significant difference
Teacher attendance	83%	82%	No
Administrator coached teachers weekly or monthly (reported by administrator)	77%	77%	No
Administrator coached teachers during the 2025 school year (reported by teacher)	66%	62%	No
Highly active Parent-Teacher Association (PTA)[52]	12%	20%	No
Teacher attended GTP during the 2025 school year (reported by teacher)	80%	69%	Yes

The comparative analyses above indicate that administrators who self-reported mentor support have very limited differences in terms of background and management performance in relation to those who did not self-report mentor support. This analysis supports the hypothesis that mentorship is the catalytical component contributing to the observed changes in reading outcomes, particularly benefiting vulnerable students who are struggling to acquire foundational skills.

[52] Defined as having more than two members and having met at least once during the previous month.

3.10 A Reflection on Contextual Shifts

To interpret these findings, it is necessary to consider that during the eight years of ALMA's implementation, Timor-Leste has faced multiple shocks, which had a negative impact on education outcomes. The COVID-19 crisis (2020-2021) triggered recurrent school closures and led to the temporary implementation of a system of multiple school shifts, reducing exposure to mentorship and other support systems due to mobility restrictions. Despite agile pivoting by the MoE leading to the implementation of a remote learning system, there is evidence that these shocks triggered major learning losses among students living in rural and remote areas of the country. A large-scale study conducted in rural and remote areas found learning losses ranging between 3-6 percentage points for passage reading fluency.[53] These shocks were further aggravated by Cyclone Seroja, which led to unprecedented flooding and landslides in April 2021, destroying over 30,000 homesteads and damaging agricultural production. Following the economic crisis triggered by COVID-19, the country experienced a sharp increase of food inflation (from 1.6% to 8.4%) in 2022-2024, with a particular impact on the price of rice - a key staple - due to international supply dynamics and erratic rainfall patterns[54] induced by the La Niña phenomenon. The combination of natural disasters and economic shocks resulted in an estimated 300,000 individuals facing acute food insecurity in 2022-2023.[55]

The results of the present study should be considered against the backdrop of the impact of these shocks on school administrators, teachers, students, and the mentors themselves. While the 'long-term' schools started receiving ALMA's support prior to the COVID-19 crisis, it is likely that the mentorship process was negatively affected in 2020-2021 by the COVID-19 school closures and mobility restrictions. The previous ALMA multi-year study noted that following the completion of the municipal roll-out of each phase, there was a decline over time in conducting activities associated with the program, most notably GTPs. It is likely that this pattern was exacerbated in 2020-2021, which may have led part of the school administrators and teachers to disengage from mentorship activities.

Weather-related shocks are associated with a pattern of absenteeism, clearly visible in this study, particularly among students in early grades. It is likely that the extreme weather patterns observed during this period have reduced the frequency of mentorship/ GTP sessions due to teacher absenteeism and mentors' inability to reach schools. High student absenteeism due to severe weather is likely to have further reduced the impact of mentorship on learning outcomes.

In parallel, economic and climate shocks have contributed to food insecurity and poor dietary diversity, affecting both students and teachers. As shown in this study, 14% of the students did not eat before coming to school. The proportion of non-readers is significantly higher among those who did not eat before coming to school – a difference of 3.5 percentage

[53] Consilient (2021) *Midline Evaluation: McGovern-Dole Food for Education Program, HATUTAN*, pgs.46-51
[54] World Bank (2024) *Timor-Leste Economic Report: Leveraging WTO Accession for Economic Transformation*, pg.11
[55] IPC, Timor-Leste Acute Food Insecurity Snapshot, November 2022-September 2023

points.[56] While no such data is available for teachers, school administrators, or mentors, it is likely that the quality of mentorship and the ability to engage with mentorship activities have also been affected by food insecurity and further worsened by disaster-driven absenteeism.

Conclusions

Has ALMA improved literacy learning outcomes for students entering Grade 3?

The findings indicate that schools where administrators self-reported ALMA mentorship support had significantly higher literacy outcomes across the board. The impact of administrator mentorship on reading outcomes is particularly strong in the long-term supported schools (see Annex 5), possibly indicating the need for a 'maturation effect' to fully manifest itself. Schools where administrators self-reported mentor support correspond to 31% of the total. It is likely that ALMA's mentorship package was implemented with the highest fidelity/ quality in those schools, thus triggering self-recall of mentor support.[57]

Administrators who self-reported mentorship do not differ significantly from their peers in terms of experience or academic background. Additionally, schools where administrators who self-reported mentorship do not differ from other schools in terms of teacher attendance, PTA functionality, or frequency of teacher coaching. The similar profiles of administrators who self-reported mentorship and those who did not suggest that the differences in reading outcomes are not simply associated with more qualified administrators or administrators who had a higher management performance overall. These results corroborate the hypothesis that mentorship is indeed the critical component contributing to make administrators more effective in recognizing and addressing the needs of vulnerable students, both through better teaching and by ensuring the presence of functional reading corners and the inclusion of protein-rich foods in school meals.

ALMA's mentorship to administrators – and to a certain extent, to teachers – is strongly associated with a reduction in the proportion of non-readers. The results suggest that the mentorship process enhanced administrator and teacher capacity to support low-performing students. This assumption is supported by the increased proportion of teachers observed using formative assessments among those self-reporting mentorship – and by the strong association between the use of formative assessments and the increased proportion of students able to read words. Taken together, these results suggest that the mentorship of school leaders is an effective model to boost the literacy outcomes of Cycle 1 students.

[56] $p<0.046$

[57] This may have included greater frequency of mentorship, more frequent administrator exposure (for instance, due to lower absenteeism among these administrators), and/or more skilled/ motivated mentors. In the absence of longitudinal data on fidelity of implementation and mentoring practices, it is not possible to determine if those assumptions are accurate or not.

Is there an observable link between teacher adherence to curriculum-prescribed teaching methodology and improved learning outcomes in long-term ALMA-supported schools?

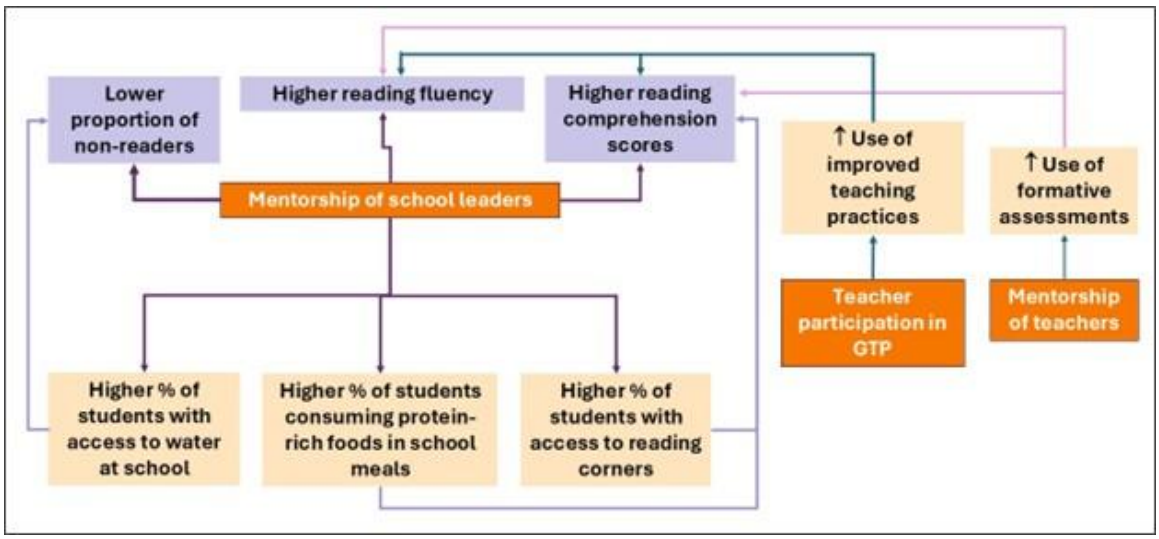
The findings suggest that ALMA's support to teachers is having a positive impact on improving teaching practices; in turn, the use of improved teaching practices – measured by ALMA's classroom observation tool - correlates positively with passage reading fluency and reading comprehension.

There is no association between mentorship (self-reported by teachers or administrators) and teaching practice scores. However, teacher mentorship is associated with an increased use of formative assessments, which is in turn associated with higher reading fluency and reading comprehension scores. Therefore, it is likely that the exposure to mentorship is improving teacher capacity to identify and support students who are struggling to read through the use of formative assessments, contributing to improved reading outcomes.

The participation in the GTP, another ALMA-supported practice, is associated with an increased utilisation of improved teaching practices (measured through classroom observations). In turn, the use of improved teaching practices is associated with higher reading fluency and reading comprehension. The findings support the assumption that the participation in the GTP sessions is beneficial to boost the use of improved teaching practices, contributing to improving literacy outcomes.

The following diagram illustrates how ALMA-supported interventions are contributing to improved literacy outcomes:

Figure 37: Diagram summarizing how ALMA-supported mentorship interventions are associated with improved literacy outcomes and management of school resources.



Factors Affecting Mentorship

Remoteness. Mentorship provision is conditioned by accessibility. Extremely remote schools are significantly less likely to have an administrator self-reporting receiving mentor

support. While the results demonstrate that staff in these schools could benefit greatly from exposure to mentorship, it may require a modified approach to enable regular support. It is unrealistic to expect mentors covering a large school cluster to be able to walk regularly to remote schools, which may require a full day of travel to reach the site and expose mentors to considerable risk, particularly during the rainy season and in the case of women traveling by themselves.

Gender. The proportion of female administrators self-reporting a mentor's support is considerably higher – 48% compared to 27% of the men. However, only 19% of the school administrators are female, and the small sample size may bias results. Nonetheless, further investigation is required to clarify this pattern. It is possible that female administrators have historically had less exposure to training opportunities, and thus may place greater value on mentorship, as well as benefiting from it.

Are the classroom observations conducted by school leaders accurate representations of actual teaching practice?

The tool used by ALMA to assess teaching practices has high internal reliability. This enables the results of the classroom observation tool to be used as a scale, forming a teaching practice score. The teaching practice score has a positive correlation with both passage reading fluency and reading comprehension. Taken together, the results support the assumption that ALMA's classroom observation tool is a valid tool for measurement of practices associated with improved reading outcomes. However, it should also be noted that school leaders seem to have a bias towards rating teaching practices as 'very good', when compared to the results from observations conducted by external enumerators. This does not affect the validity of the tool itself but rather indicates a need to reinforce the training of leaders on quality control procedures when conducting observations.

Has ALMA improved student attendance for Cycles 1 and 2?

There is no association between ALMA-supported interventions - including school administrator mentorship, teacher mentorship, or GTP participation - and student or teacher attendance. Absenteeism remains very high, with over a quarter of the students not being present in class during headcounts (highest in grade 1 and among girls).

As expected, there is a positive correlation between teacher and student attendance. However, teacher attendance seems to be primarily linked to external factors. As noted previously, contextual factors – including the severe rainy season and food insecurity – are likely to affect teacher and student attendance, potentially to a level where school-level interventions have a limited impact in mitigating absenteeism. In the case of teachers, sick leave is also likely to have a considerable impact on attendance – which may be disproportionately high among the aging workforce and those facing higher levels of food insecurity. Since the study did not collect individual student or teacher attendance data, it is not possible to conclude to what extent individual-level factors – age, illness, food security, sex – are affecting attendance rates. However, during data collection, teacher illness was reportedly a main cause of absenteeism.

Recommendations

Fidelity of implementation. The findings show a strong relationship between ALMA's school administrator mentorship model and improved literacy outcomes. Similarly, there is a strong relationship between the participation in the GTP, improved teaching practices, and passage reading fluency. However, only a third of the school administrators recalled a mentor's support without prompting. Teacher self-reported mentorship (with prompting) is not strongly associated with the use of improved teaching practices or literacy outcomes. There is a likelihood that part of ALMA-supported schools had lower fidelity of implementation – i.e. reduced frequency of mentorship and/or GTPs, or even mentorship not implemented as intended – resulting in limited impact on school administrators and teachers. It is also possible that fidelity of implementation varied across geographies, influenced by a higher proportion of remote schools, ultimately affecting results. There is a need to track fidelity of implementation in mentorship implementation to better understanding the full magnitude of ALMA's impact when implemented as intended.

Conversely, the fact that 32% of the administrators in long-term supported schools self-reported mentor support is remarkable, particularly given the time lapsed since the initial roll-out. This provides a valuable opportunity for a future case study to explore which factors have contributed to the high effectiveness of these mentors. The results could inform the MoE's planning and further investment in mentorship models, as well as DFAT's investment in similar models in Timor-Leste and in other settings.

Impact on non-readers. ALMA's mentorship approach seems to be particularly beneficial in improving conditions for students who are struggling to decode words. It seems to increase school administrators' ability to improve learning conditions for vulnerable students – including both teaching as well as school conditions – and to improve teacher use of formative assessments, enabling them to identify non-readers. It is important to explore why this is occurring, and how this effect could be further refined and scaled up. It is also important to explore the connections between mentorship of school administrators and increased accountability and professionalism in the use of school resources – and potentially also towards identifying and supporting struggling students.

Remoteness. There is a need to develop an alternative mentorship model for extremely remote schools, perhaps exploring the use of technology and supporting mentors to conduct 'residences' of a week or longer in schools accessible on foot only. While such approaches are undoubtedly challenging to implement, the major disparities in learning outcomes associated with remoteness clearly have an impact on the overall results of the program.

Ancillary approaches. While mentorship is having an undeniable impact on administrators, teachers, and students, it is also clear that student learning outcomes are heavily affected by school conditions – the availability of reading corners, water, and school meals containing protein-rich foods. The impact of mentorship is likely to be exacerbated when rolled out in partnership with other initiatives improving school infrastructure, access to meals, and reading materials.

Absenteeism remains a major underlying barrier to learning and is likely to erode part of the potential impact of approaches such as mentorship and GTPs/GTLs. There is a need to reinforce teacher attendance – including through institutional management, participatory accountability, and multi-sectoral interventions addressing external factors related to absenteeism. Other studies indicate that illness and natural disasters are among the key drivers of absenteeism in early grades.[58] Therefore, multi-sectoral interventions may be particularly critical to boost attendance in early grades, thus maximising the impact of mentorship and improved teaching practices on learning outcomes.

Gender. There is a significant gender gap in learning outcomes, with boys performing systematically below girls across all EGRA tasks. Additionally, female school administrators seem to be benefitting more from mentorship, while female teachers have better teaching practices than men. There is a need to further explore how mentorship could be made more effective to address gender gaps in learning – and to understand the patterns that make mentorship more effective for female teachers and school administrators.

[58] Consilient (2023) *HATUTAN Endline Evaluation*, pg.69

Annex (report)

Annex 4 – Regression Tables

Passage Reading Fluency

Variables	Beta	t	p	Statistically significant *** p<0.000 ** p<0.05 * p<0.1
Sex	.188	8.738	.000	***
Mother tongue: Tetum	.049	2.258	.024	**
Ate protein-rich foods at school	.066	3.060	.002	**
Extremely remote (walking access only)	-.057	-2.388	.017	**
Reading corner in grade 2	.069	2.954	.003	**
School leader self-reported mentorship	.043	1.946	.052	*
Teacher used formative assessments	.125	5.692	.000	***

Non-readers (based on the passage reading fluency task)

Variables	Beta	S.E.	p	Statistically significant *** p<0.000 ** p<0.05 * p<0.1
Sex	.550	0.099	.000	***
Ate protein-rich foods at school	.394	0.116	.001	**
School administrator self-reported mentorship	.514	.110	.000	***
Teacher used formative assessments	.398	0.097	.000	***
Extremely remote school (walking access only)	-.323	0.157	.040	**
School has access to water	.381	.148	.01	**

Reading comprehension

Variables	Beta	t	p	Statistically significant *** p<0.000 ** p<0.05 * p<0.1
Sex	.137	6.316	.000	***
Mother tongue: Tetum	.075	3.467	.001	**
Ate protein-rich foods at school	.106	4.881	.000	***

Variables	**Beta**	**t**	**p**	**Statistically significant** ***** p<0.000** **** p<0.05** *** p<0.1**
Extremely remote school (walking access only)	-.075	-3.135	.002	**
Reading corner in grade 2	.052	2.248	.025	**
School leader self-reported mentorship	.097	4.444	.000	***
Teacher used formative assessments	.115	5.249	.000	***

Annex 5 – Comparative Analysis of EGRA Results by Study Group

Letter Recognition

Average scores. In average, the recently joined group could identify 40.5 letters per minute, compared to 36.5 letters per minute among the long-term group. The difference is statistically significant.[59] The proportion of students unable to identify any letters is similar across study groups. Among the long-term group, 5.8% of the students could not identify any letters, compared to 5.4% of those in the recently joined group. The difference is not statistically significant.

The average number of letters identified per minute is significantly higher[60] among female students in both long-term and recently joined schools. There is no statistically significant difference by sex and study group, meaning that the performance patterns by sex are similar across study groups (i.e. girls outperform boys in both groups).

Table 12: Average number of letters identified per minute, by study group and sex

Study group	Male	Female	Total
Long-term	33.9	39.6	36.5
Recently joined	37.9	43.8	40.5
Total	36.1	41.5	38.6

Mentorship. There is no statistically significant difference in the number of letters identified by students whose teachers or school directors/ coordinators self-reported receiving support from an ALMA mentor. However, there is a trend towards improved performance among students attending schools where both teachers and directors/ coordinators self-reported receiving support from an ALMA mentor. The trend is similar in recently joined and long-term schools.

[59] $F_{1,2063}$=15.087, p<0.000
[60] $F_{1,2065}$=30.956, p<0.000

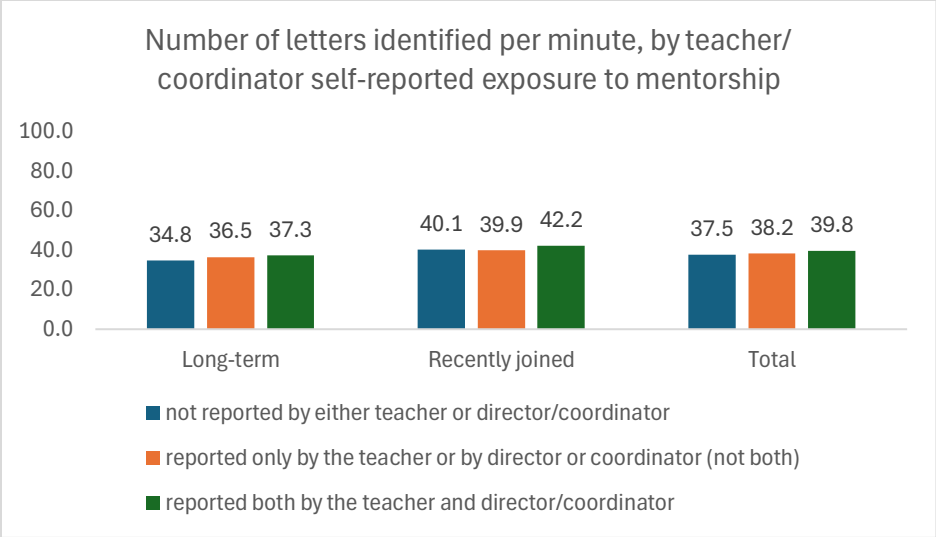


Figure 38: Average number of letters identified per minute by students, disaggregated by level of exposure to mentorship (self-reported by teachers and/or school directors or coordinators)

Invented Word Reading

Average reading fluency. The average number of invented words read per minute by students in recently joined schools is slightly higher, compared to long-term supported schools. The difference is statistically significant.[61] Patterns in the performance of male and female students are similar in both long-term and recently joined schools.

Table 13: Average number of words read per minute by students in long-term and recently joined schools, disaggregated by sex, disaggregated by study group

Study group	Male	Female	Total
Long-term	12.3	16.5	14.3
Recently joined	14.4	18.3	16.1
Total	13.4	17.4	15.2

Non-readers. Over a quarter of the students (27%) were unable to read any invented words. The proportion is similar in long-term supported and recently joined schools.

Table 14: Proportion of students unable to read any invented words, disaggregated by study group

Subgroups	% of zero scores
Long-term	28%
Recently joined	26%
Total	**27%**

[61] $F_{1,2065}=9.789$, $p<0.002$

Mentorship. The exposure to mentorship identifies where ALMA is being implemented as intended – and therefore enables the study to assess if ALMA is having an impact on reading skills development when implemented as intended.

The average invented word reading fluency is **significantly higher[62] among students attending schools whose directors/ coordinators self-reported receiving support from an ALMA mentor.** The effect is similar across study groups (see figure below).

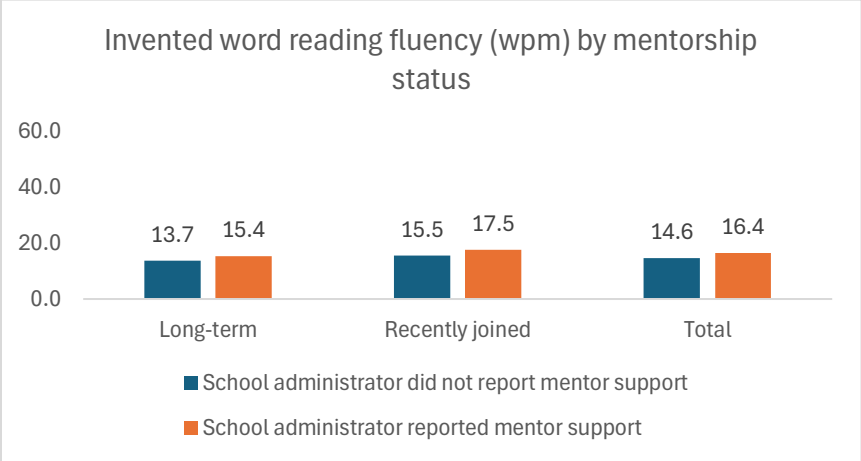


Figure 39: Average invented word reading fluency (words per minute/wpm) in long-term and recently joined schools, disaggregated by self-reported exposure to mentorship by school administrators (directors or coordinators)

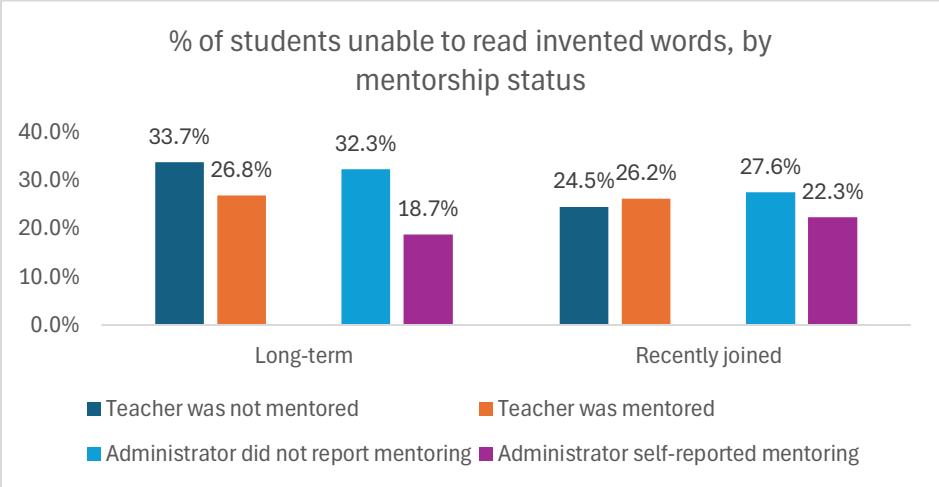


Figure 40: Proportion of students unable to read any invented words, disaggregated by self-reported teacher or administrator exposure to mentorship in each study group

In the long-term group, the proportion of students unable to read any words increases by seven percentage points when teachers did not receive mentorship, and by 14 percentage points in schools where the administrator did not report receiving a mentor's support. The

[62] $F_{1,2065}=7.938$, $p<0.005$

pattern is not observed in the recently joined schools, potentially indicating the need for a 'maturation effect' for mentorship to trigger higher levels of impact on non-readers.

Familiar Word Reading

Average familiar word reading fluency. The average number of familiar words read per minute is significantly higher among students attending recently joined schools.[63] The differences in boys' and girls' familiar word reading fluency are similar across study groups – i.e. girls are reading at a faster pace in both groups.

Table 15: Average number of familiar words read per minute by study group, disaggregated by sex

Study group	Male	Female	Total
Long-term	15.8	21.6	18.5
Recently joined	18.3	24.1	20.9
Total	**17.1**	**22.8**	**19.7**

Non-readers. Overall, 27% of the students could not read any familiar words. The proportion of students who could not read any familiar words is similar in long-term supported and recently joined schools.

Table 16: Proportion of students scoring zero (unable to read any familiar words), disaggregated by study group

Subgroups	% of zero scores
Long-term	28%
Recently joined	26%
Total	**27%**

Mentoring. The average familiar word reading fluency is significantly higher in schools where the administrator identified mentors as a source of support.[64] The effect was similar in long-term supported and recently joined schools.

[63] $F_{1,2065}=10.491$, $p<0.001$
[64] $F_{1,2065}=10.153$, $p<0.001$

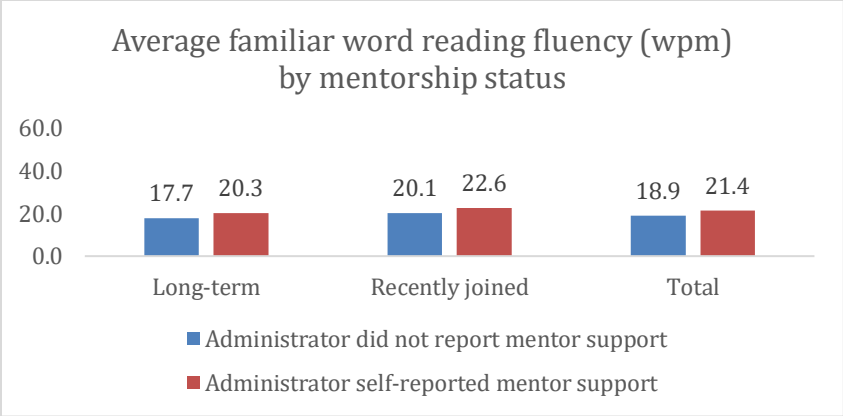


Figure 41: Average familiar word reading fluency in long-term and recently joined schools, disaggregated by administrator self-reported access to mentorship

Mentorship and non-readers. **In long-term supported schools, the access to mentorship is associated with a reduced proportion of non-readers.** In schools where the administrator self-reported access to mentorship, the proportion of non-readers is 14 percentage points lower than in schools where the administrator did not report receiving a mentor's support. In schools where the teacher self-reported being mentored during this school year, the proportion of non-readers is seven percentage points lower than in schools where teachers did not report being mentored. This pattern was not observed in recently joined schools.

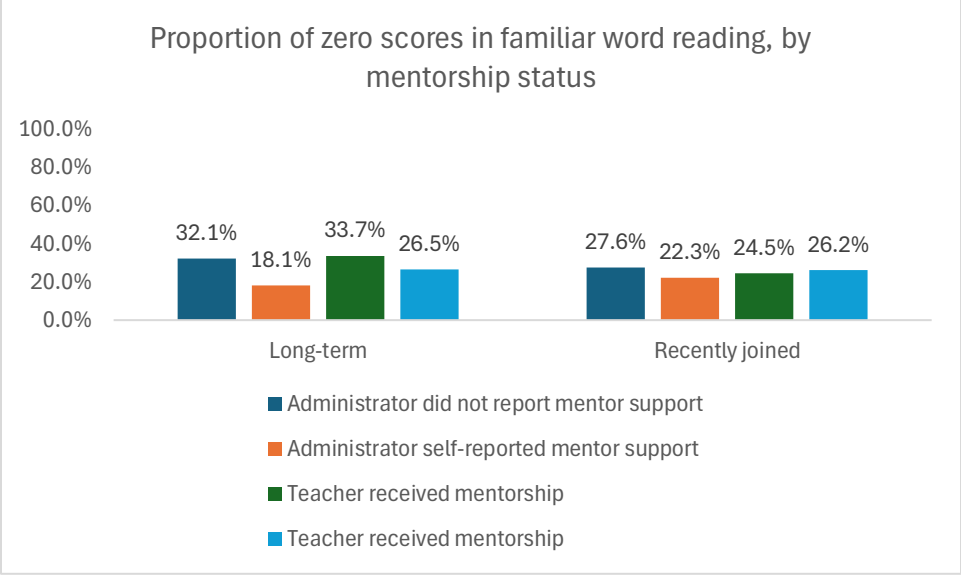


Figure 42: Proportion of students scoring zero in familiar word reading in long-term and recently joined schools where (i) the administrator reported/did not report mentor support and (ii) the teacher received/did not receive mentorship.

Passage Reading Fluency

Average passage reading fluency. Grade 3 students attending recently joined schools have significantly higher average passage reading scores, compared to their peers in long-term supported schools – 25.6wpm compared to 21.8 wpm.[65] The patterns by sex are similar across study groups, with girls showing higher fluency in both long-term and recently joined schools.

Table 17: Average passage reading fluency scores, disaggregated by study group and sex

Study group	Male	Female	Total
Long-term	17.6	26.6	21.8
Recently joined	21.8	30.4	25.6
Total	**19.8**	**28.5**	**23.7**

Non-readers. Overall, 30% of the grade 3 students were unable to read any words in the passage reading task, compared to 27% in the invented word and familiar word reading tasks. The difference between the proportion of non-readers in long-term and recently joined schools is nearly significant[66], with the long-term schools having a high proportion of non-readers.

Table 18: Proportion of students unable to read any words in a passage, disaggregated by study group

Subgroups	% zero scores
Long-term	31.9%
Recently joined	28.8%
Total	**30.3%**

Mentorship. The average passage reading fluency is higher among students attending schools whose administrator (director/ coordinator) reported receiving support from a mentor. The difference is nearly significant.[67] However, there is no statistically significant difference in the average passage reading fluency by study group x mentorship, meaning that the exposure to mentorship has a significant effect in both groups. There is no statistically significant difference of teacher self-reported exposure to mentorship on passage reading fluency.

[65] $F_{1,2064}$=13.780, p<0.000
[66] Significant at the 10% level, χ^2=2.219, p<0.075
[67] Significant at the 10% level; $F_{1,2064}$=3.689, p<0.055

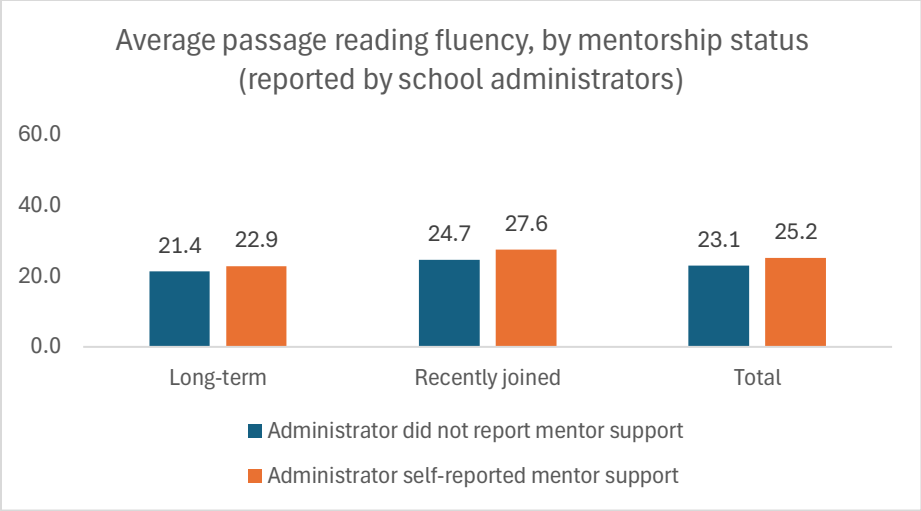


Figure 43: Average passage reading fluency (wpm), disaggregated by study group and school administrator self-reported mentorship status

Mentorship and non-readers. The self-reported provision of mentorship to school administrators (directors/ coordinators) is associated with a significantly lower proportion of non-readers in both long-term and recently joined schools.[68] However, the effect is higher in long-term schools[69] (a reduction of 12 percentage points compared to 9 percentage points in recently joined schools).[70] There is no similar effect when mentorship is reported by teachers.

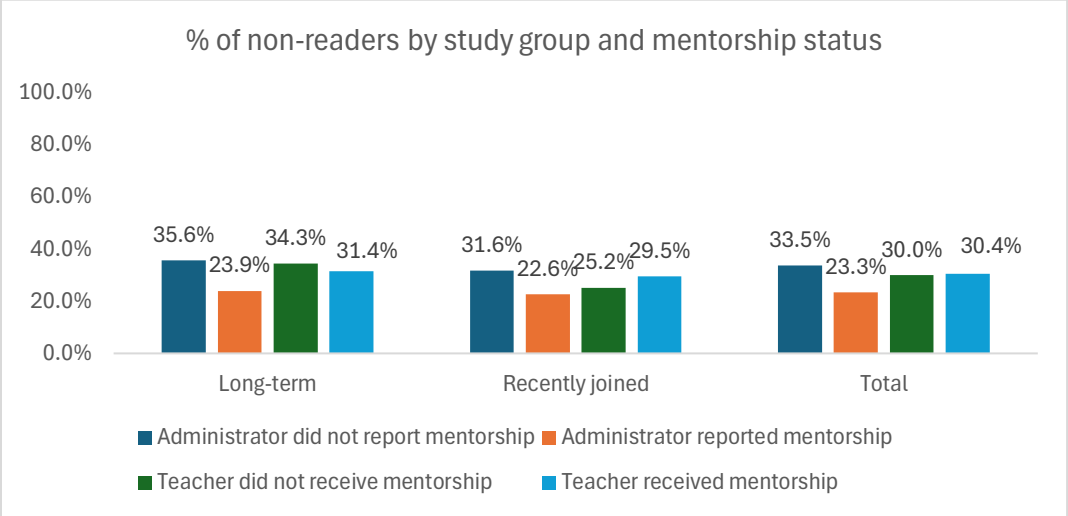


Figure 44: Proportion of non-readers (students unable to read a single word in a passage), disaggregated by study group and mentorship status

[68] χ^2=21.999, p<0.000
[69] χ^2=13.886, p<0.000
[70] χ^2=8.571, p<0.002

Reading Comprehension

Average reading comprehension score. The average reading comprehension score is significantly higher in recently joined schools[71], although the difference is small (41% compared to 37%). There is a much larger and significant difference in the performance by sex, with girls having a higher performance than boys (45% compared to 34%).[72] The pattern is similar in both long-term and recently joined schools, with girls performing significantly better.

Table 19: Average reading comprehension scores (lower level), disaggregated by study group and sex

Study group	Male	Female	Total
Long-term	31.9%	43.4%	37.3%
Recently joined	36.7%	45.8%	40.8%
Total	34.4%	44.6%	39.0%

Zero scores. Overall, 37% of the students were unable to respond reading comprehension tasks. The proportion of zero scores is considerably higher in long-term schools, reaching nearly 40%, compared to 34% in recently joined schools.

Table 20: Proportion of zero scores in reading comprehension, disaggregated by study group

Subgroups	% zero scores
Long-term	39.8%
Recently joined	34.0%
Total	**37%**

Mentorship. Students attending schools where the administrator (director/ coordinator) self-reported receiving mentor support had significantly higher reading comprehension scores.[73] The impact is similar in both long-term and recently joined schools. There is no difference between the performance of students in schools where grade 2-3 teachers reported receiving mentorship and those who did not receive it.

[71] $F_{1,2063}=4.590$, $p<0.032$
[72] $F_{1,2063}=39.824$, $p<0.000$
[73] $F_{1,2063}=19.478$, $p<0.000$

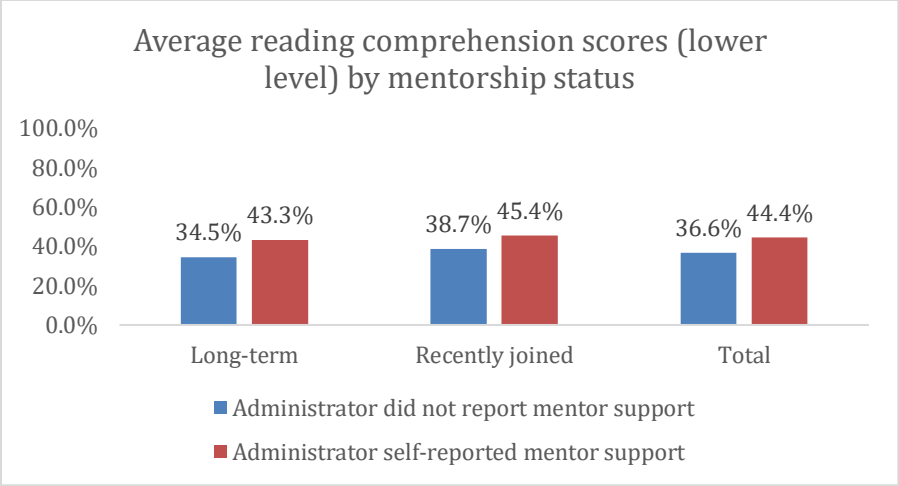


Figure 45: Average reading comprehension scores (lower level) disaggregated by study group and school administrator self-reported mentorship status

Zero scores and mentorship. The proportion of zero scores drops dramatically in long-term-supported schools where the administrator self-reported having mentor support – a reduction of 13 percentage points. A similar pattern is observed in recently joined schools, although less dramatic (a reduction of nine percentage points in the proportion of students scoring zero).

Again, **the result reinforces the trend of improved reading performance among students attending schools where administrators self-reported support from an ALMA mentor**. The result is particularly strong in long-term supported schools.

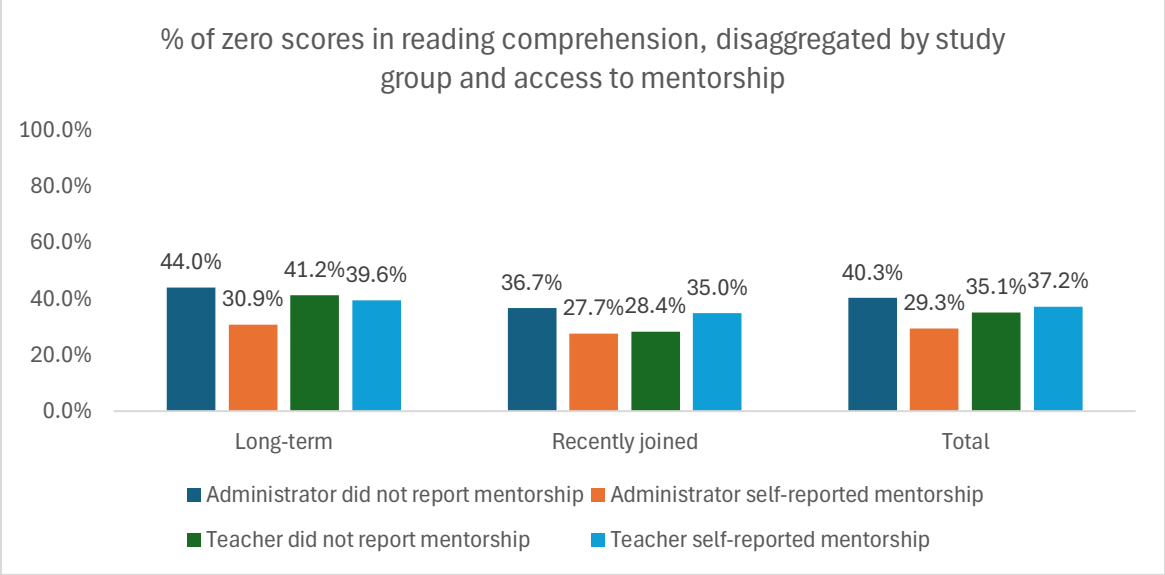


Figure 46: Proportion of zero scores in reading comprehension in schools where the administrator or teacher reported receiving mentorship, disaggregated by study group

Additional Analysis: Adjusted Samples for Comparable Vulnerability Levels

This additional analysis considers the average vulnerability level in the long-term and recently joined schools. The analysis used 10 variables74 to calculate a school-level vulnerability score, which was used to match the two samples. The refined samples included 53 long-term schools (933 students) and 51 recently joined schools (905 students).

The results of this additional analysis show similar trends to those described in Annex 5. The recently joined schools tend to have a slightly higher performance than the long-term supported schools. While the differences are small, they are statistically significant, except for reading comprehension.

Table 21: Average results for each EGRA task, disaggregated by study group, calculated using samples matched by vulnerability level

Study group	Letter recognition[75]	Invented words[76]	Familiar words[77]	Passage reading[78]	Reading comprehension[79]
Long-term	36.4	14.3	18.4	21.7	37.1%
Recently joined	37	15.7	20.4	24.8	39.7%

The proportion of non-readers in the passage reading task is similar in long-term and recently joined schools – 32.5% and 30%, respectively, with 31% of all assessed students unable to read words. The difference is not statistically significant.

Table 22: Proportion of non-readers in the passage reading task, disaggregated by study group, calculated using samples matched by vulnerability level

Study group	% of non-readers
Long-term	32.5%
Recently joined	30%
Total	31.3%

Similarly to the previous analysis, students in schools whose administrator self-reported mentorship have a significantly higher performance both in passage reading fluency[80] as well as reading comprehension.[81] In average, passage reading fluency is 14% higher and reading comprehension is 9 percentage points higher in schools where administrators self-reported mentorship access. There is no statistically significant difference in the pattern by

74 Accessibility (on foot/ via vehicle); size (<4 classes; more than 4 classes), presence of multigrade classes, availability of water, toilets, and female toilets; availability of handwashing stations; active parent-teacher association; availability of reading corner.
75 $F_{1,1838}=10.846$, $p<0.001$
76 $F_{1,1838}=5.424$, $p<0.020$
77 $F_{1,1838}=6.797$, $p<0.009$
78 $F_{1,1837}=8.423$, $p<0.004$
79 Difference is not statistically significant ($p<0.144$)
80 $F_{1,1837}=7.661$, $p<0.006$
81 $F_{1,1836}=25.636$, $p<0.000$

study group. Additionally, there is no difference in the proportion of non-readers by administrator mentorship status.

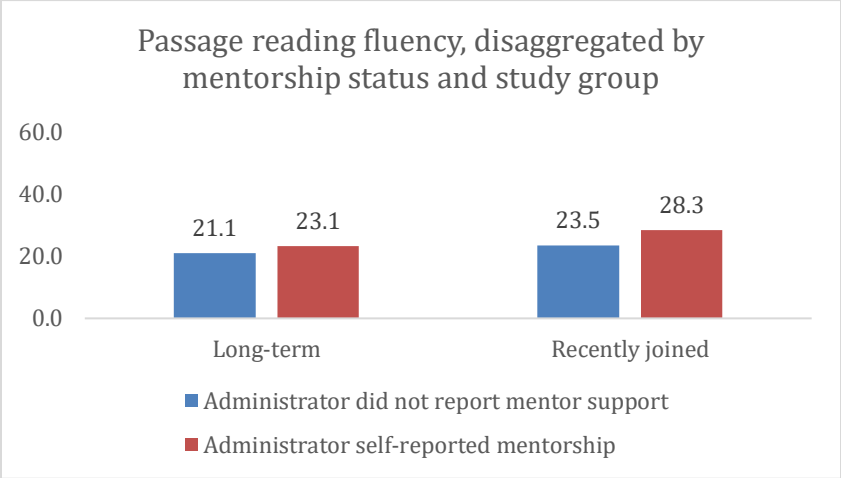


Figure 47: Average passage reading fluency (wpm) disaggregated by study group and mentorship status, calculated using samples matched for vulnerability

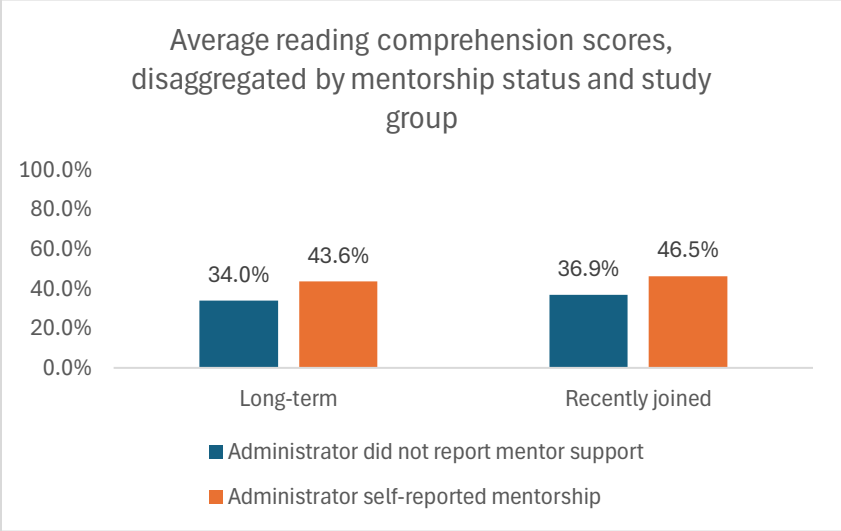


Figure 48: Average reading comprehension scores, disaggregated by mentorship status and study group, calculated using samples matched for vulnerability

Annex 6: Case Study – Longitudinal Patterns in Phase 8 Schools

This case study uses secondary data collected by the HATUTAN II program, which includes only rural and remote schools. All schools involved in this case study were part of HATUTAN II's comparison arm.

The case study provides a longitudinal analysis of the changes in a cluster of six schools (all located in Aileu's Remexio administrative post). These schools were targeted by ALMA in Phase 8 and showed remarkable changes in reading fluency and comprehension. This study compares their longitudinal results with those observed in (i) all Phase 8 schools included in HATUTAN II's sample (total 20 schools), and (ii) in all Phase 9 schools assessed by HATUTAN II (total 38 schools). This cluster of six Phase 8 schools is of particular interest as they did not receive support from any other initiatives during the last two years.[82]

The results show a sharp increase in both passage reading fluency and reading comprehension three years after the implementation of Phase 8. Interestingly, the schools targeted in Phase 8 seem to have faced particularly strong losses after COVID, but the results post-mentorship surpass the performance pre-COVID. The Phase 9 sample does not show similar gains, despite having had lower losses after COVID.

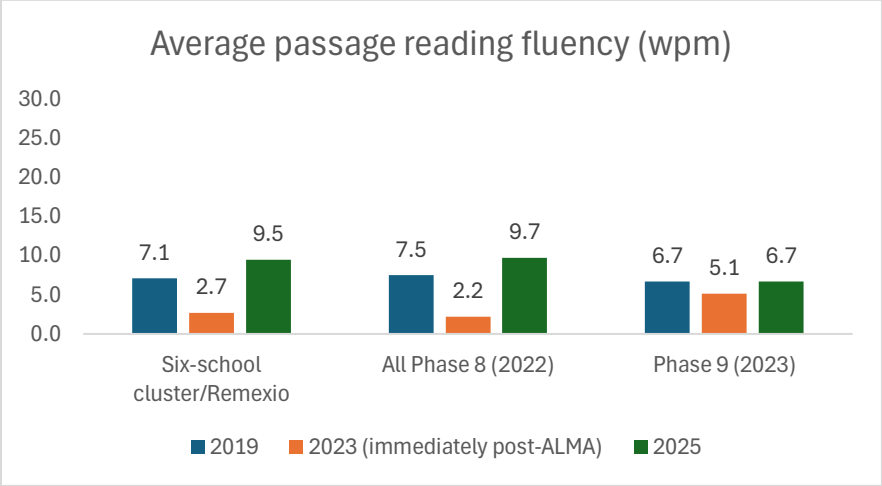


Figure 49: Longitudinal trends in passage reading fluency prior to ALMA's intervention and following implementation in six schools in Remexio, targeted in Phase 8; a larger sample of Phase 8 schools; and a sample of Phase 9 schools.

The gains are particularly remarkable in reading comprehension, with an increase of over 11 percentage points in the six-school cluster in relation to the pre-COVID performance. Overall, the Phase 8 sample had a gain of nearly 7 percentage points over and above the

[82] Once the sharp increase in learning outcomes was identified in the HATUTAN II midterm evaluation, follow-up visits were conducted to validate the data and rule out the possibility of errors, bias in data collection, or contamination by another initiative. These visits validated the results and identified ALMA as the only initiative providing support to these schools.

pre-COVID performance. Comparatively, the Phase 9 sample seems to have only recovered to pre-COVID performance levels.

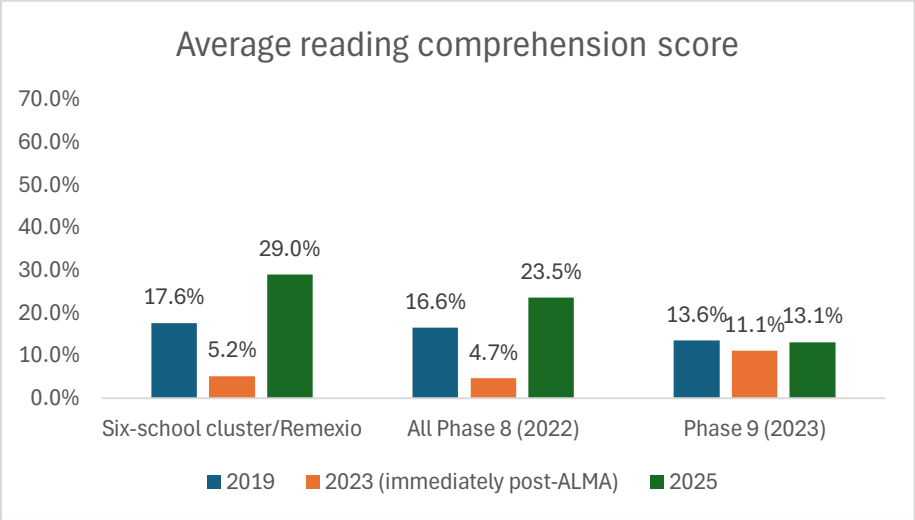


Figure 50: Longitudinal comparison of reading comprehension scores previously and after exposure to ALMA implementation. Data from student samples targeted in Phases 8 and 9, and in a sub-group of six Phase 8 schools in Remexio.

These findings may reflect a "maturation effect" after exposure to mentorship – which has no yet been observed in Phase 9. It may also indicate a particularly strong implementation in Phase 8. While the findings cannot be generalised, they provide an interesting perspective of pre-post patterns associated with ALMA's implementation.

Annexes for ALT TEXT

Annex 1 – **Written version of Report figures**

Written version of Figure 1/37 – Mentorship pathways and literacy outcomes

The figure presents a conceptual diagram illustrating how ALMA supported interventions are associated with improved literacy outcomes and enabling school conditions.
At the centre of the diagram is mentorship of school leaders. This is directly associated with three student literacy outcomes: higher reading fluency, higher reading comprehension scores, and a lower proportion of non readers.
Mentorship of school leaders is also associated with improvements in school level enabling conditions. These include a higher percentage of students with access to water at school, a higher percentage of students consuming protein rich foods in school meals, and a higher percentage of students with access to reading corners.
On the right side of the diagram, two additional pathways are shown. Teacher participation in Teacher Working Groups (GTP) is associated with greater use of improved teaching practices. Separately, mentorship of teachers is associated with greater use of formative assessments.
The diagram shows that the use of improved teaching practices and the use of formative assessments both contribute to improved student literacy outcomes, reinforcing the links between mentorship, teaching practices, school conditions, and learning results.
Overall, the figure illustrates how ALMA supported mentorship and professional learning activities operate through both instructional and school resource pathways to support improvements in early grade reading outcomes.

Written version of Figure 4

Food consumption at home or school, by study group
Title: Food consumption at home or school, by study group
Y-axis:
Percentage (%), ranging from 0.0% to 80.0%
Legend:

- Long-term (blue)
- Recently joined (orange)

Data shown:

1. Did not eat before coming to school
    - Long-term: 14.5%
    - Recently joined: 13.4%
2. Did not have a meal at home or school
    - Long-term: 4.7%
    - Recently joined: 4.2%
3. Did not have a school meal

- Long-term: 22.4%
- Recently joined: 15.4%

Written version of Figure 7

Figure 7. Average number of letters identified per minute by students, disaggregated by level of exposure to mentorship (self-reported by teachers and/or school directors or coordinators)
Chart title:
Number of letters identified per minute, by teacher/coordinator self-reported exposure to mentorship
Y-axis:
Number of letters identified per minute, ranging from 0.0 to 100.0
X-axis:
Total
Legend and data values:

1. Not reported by either teacher or director/coordinator
    - Average letters identified per minute: 37.5
2. Reported only by the teacher or by director or coordinator (not both)
    - Average letters identified per minute: 38.2
3. Reported both by the teacher and director/coordinator
    - Average letters identified per minute: 39.8

Written version of Figure 8

Figure 8. Average number of invented words read per minute, disaggregated by sex
Chart title:
Average number of invented words read per minute, disaggregated by sex
Y-axis:
Average number of invented words read per minute, ranging from 0 to 60
X-axis categories and data values:

- Male: 13.4 invented words read per minute
- Female: 17.4 invented words read per minute
- Total: 15.2 invented words read per minute

Written version of Figure 9

Figure 9. Invented word reading fluency (words per minute) by school administrator self-reported mentorship status
Chart title:
Invented word reading fluency (wpm) by school administrator self-reported mentorship status

Y-axis:
Invented word reading fluency, measured in words per minute, ranging from 0.0 to 60.0
X-axis categories and data values:

- School administrator did not report mentor support:
  Average invented word reading fluency of 14.6 words per minute
- School administrator reported mentor support:
  Average invented word reading fluency of 16.4 words per minute

Written version of Figure 10

Figure 10. Percentage of students unable to read invented words, disaggregated by mentorship status
Chart title:
% of students unable to read invented words, disaggregated by mentorship status
Y-axis:
Percentage of students, ranging from 0.0% to 70.0%
Data shown:

- Administrator did not report mentoring: 29.9%
- Administrator self-reported mentoring: 20.5%
- Teacher was not mentored: 29.4%
- Teacher was mentored: 26.5%

Written version of Figure 11

Figure 11. Average number of familiar words read per minute (wpm), by sex
Chart title:
Average number of familiar words read per minute (wpm), by sex
Y-axis:
Average number of familiar words read per minute, ranging from 0 to 60
Data shown:

- Male: 17.1 words per minute
- Female: 22.8 words per minute
- Total: 19.7 words per minute

Written version of Figure 12

Figure 12. Average familiar word reading fluency (out of a maximum of 60 wpm), disaggregated by administrator self-reported access to mentorship
Chart title:
Average familiar word reading fluency (wpm) by mentorship status
Y-axis:
Average familiar word reading fluency, measured in words per minute, ranging from 0.0 to 60.0
Data shown:

- Administrator did not report mentor support: 18.9 words per minute

- Administrator self-reported mentor support: 21.4 words per minute

Written version of Figure 13

Figure 13. Percentage of students unable to read familiar words, by mentorship status
Chart title:
% of students unable to read familiar words, by mentorship status
Y-axis:
Percentage of students, ranging from 0.0% to 70.0%
Data shown:

- Administrator did not report mentoring: 29.8%
- Administrator self-reported mentoring: 20.2%
- Teacher was not mentored: 29.4%
- Teacher was mentored: 26.4%

Written version of Figure 14

Figure 14. Average passage reading fluency (words per minute), by sex
Chart title:
Average passage reading fluency (wpm) by sex
Y-axis:
Average passage reading fluency, measured in words per minute, ranging approximately from −19 to 61
Data shown:

- Male: 19.8 words per minute
- Female: 28.5 words per minute
- Total: 23.7 words per minute

Written version of figure 15

Figure 15. Average passage reading fluency, disaggregated by mentorship status
Chart title:
Average passage reading fluency, disaggregated by mentorship status
Y-axis:
Average passage reading fluency, measured in words per minute, ranging from 0.0 to 60.0
Data shown:

- Administrator did not report mentor support: 23.1 words per minute
- Administrator self-reported mentor support: 25.2 words per minute

Written version of Figure 16

Figure 16. Percentage of non-readers, disaggregated by mentorship status
Chart title:
% of non-readers, disaggregated by mentorship status
Y-axis:
Percentage of students, ranging from 0.0% to 70.0%

Data shown:

- Administrator did not report mentoring: 33.5%
- Administrator self-reported mentoring: 23.3%
- Teacher was not mentored: 30.0%
- Teacher was mentored: 30.4%

Written version of Figure 17

Figure 17. Average reading comprehension fluency, by sex
Chart title:
Average reading comprehension fluency, by sex
Y-axis:
Percentage score for reading comprehension fluency, ranging from 0.0% to 100.0%
Data shown:

- Male: 34.4%
- Female: 44.6%
- Total: 39.0%

Written version of Figure 18

Figure 18. Average reading comprehension score, by mentorship status
Chart title:
Average reading comprehension score, by mentorship status
Y-axis:
Reading comprehension score expressed as a percentage, ranging from 0.0% to 100.0%
Data shown:

- Administrator did not report mentor support: 36.6%
- Administrator self-reported mentor support: 44.4%

Written version of figure 19

The figure is a bar chart titled ‘’Percentage of zero scores in reading comprehension, by mentorship status.”
The vertical axis shows percentages from 0.0% to 70.0%, increasing in increments of ten percentage points.
The horizontal axis presents four categories related to mentorship status:

- Administrator did not report mentoring: 40.3% of students scored zero in reading comprehension.
- Administrator self-reported mentoring: 29.3% of students scored zero in reading comprehension.
- Teacher was not mentored: 35.1% of students scored zero in reading comprehension.
- Teacher was mentored: 37.2% of students scored zero in reading comprehension.

Overall, the chart shows a lower proportion of zero scores in schools where administrators self-reported receiving mentorship compared to schools where administrator mentorship was not reported, while the difference between teacher mentorship categories is smaller.

Written version of Figure 20

The figure is a clustered bar chart titled “Lesson preparation.”
The vertical axis shows percentages from 0% to 100%.
The horizontal axis compares results for external enumerators and ALMA mentors.
Each group is rated across four categories: Very good, Good, Ok, and Not good.
External enumerator observations:

- Very good: 29%
- Good: 44%
- Ok: 21%
- Not good: 6%

ALMA mentor observations:

- Very good: 53%
- Good: 22%
- Ok: 21%
- Not good: 3%

Overall, ALMA mentors rated lesson preparation as “very good” more frequently than external enumerators, while external enumerators recorded higher proportions of “good” and “not good” ratings.

Written version of Figure 21

The figure is a grouped bar chart titled “Instruction strategies.”
The vertical axis shows percentages from 0% to 100%.
The chart presents three instructional practices, each rated by external enumerators and ALMA mentors, using four categories: Very good, Good, Ok, and Not good.
1. The teacher facilitated the lesson effectively

- External enumerators:
    - Very good: 14%
    - Good: 54%
    - Ok: 27%
    - Not good: 4%
- ALMA mentors:
    - Very good: 47%
    - Good: 19%
    - Ok: 30%
    - Not good: 3%

2. The teacher used appropriate and effective pedagogic approaches

- External enumerators:
    - Very good: 17%
    - Good: 50%
    - Ok: 28%
    - Not good: 4%
- ALMA mentors:

    - Very good: 56%
    - Good: 11%
    - Ok: 27%
    - Not good: 5%

3. The teacher used appropriate assessment approaches

- External enumerators:
    - Very good: 13%
    - Good: 48%
    - Ok: 29%
    - Not good: 6%
- ALMA mentors:
    - Very good: 56%
    - Good: 11%
    - Ok: 27%
    - Not good: 3%

Overall, across all three instruction strategy areas, ALMA mentors rated a higher proportion of observations as "very good" compared to external enumerators, while external enumerators more frequently rated practices as "good" or "ok."

Written version of Figure 22

The figure is a grouped bar chart titled "Implementation of the Lesson Plan."
The vertical axis shows percentages from 0% to 100%.
The chart compares external enumerator observations and ALMA mentor observations across three aspects of lesson implementation.
Each aspect is rated using four categories: Very good, Good, Ok, and Not good.

1. The teacher introduced the lesson appropriately

External enumerators:

- Very good: 12%
- Good: 44%
- Ok: 30%
- Not good: 6%

ALMA mentors:

- Very good: 56%
- Good: 15%
- Ok: 26%
- Not good: 3%

2. The teacher followed the activities defined in the lesson plan

External enumerators:

- Very good: 12%
- Good: 49%
- Ok: 29%
- Not good: 6%

ALMA mentors:

- Very good: 47%
- Good: 13%
- Ok: 36%
- Not good: 4%

3. The teacher ended the lesson appropriately

External enumerators:

- Very good: 10%
- Good: 36%
- Ok: 31%
- Not good: 10%

ALMA mentors:

- Very good: 40%
- Good: 15%
- Ok: 39%
- Not good: 5%

Overall, ALMA mentors rated teachers more frequently as "very good" across all three aspects of lesson plan implementation, while external enumerators more often rated practices as "good" or "ok." Ratings of "not good" were consistently low in both groups but slightly higher in enumerator observations.

Written version of Figure 23

The figure is a grouped bar chart titled "Classroom management."
The vertical axis shows percentages from 0% to 100%.
The chart compares external enumerator observations and ALMA mentor observations across two classroom-management practices.
Each practice is rated using four categories: Very good, Good, Ok, and Not good.

1. The teacher created an inclusive classroom environment

External enumerators:

- Very good: 19%
- Good: 50%
- Ok: 25%
- Not good: 3%

ALMA mentors:

- Very good: 61%
- Good: 20%
- Ok: 18%
- Not good: 2%

2. The teacher promoted positive attitudes

External enumerators:

- Very good: 13%
- Good: 53%
- Ok: 28%
- Not good: 4%

ALMA mentors:

- Very good: 62%
- Good: 15%
- Ok: 21%
- Not good: 2%

Overall, ALMA mentors rated classroom management practices much more frequently as "very good" for both inclusive classroom environment and promotion of positive attitudes, while external enumerators more often rated practices as "good" or "ok." Ratings of "not good" were low across both observer groups.

Written version of Figure 24

The figure is a grouped bar chart titled "Socio-Emotional Skills."
The vertical axis shows percentages from 0% to 100%.
The chart compares external enumerator observations and ALMA mentor observations across two socio-emotional teaching practices.
Each practice is rated using four categories: Very good, Good, Ok, and Not good.
1. The teacher encouraged and supported equitable collaboration and participation

External enumerators:

- Very good: 15%
- Good: 50%
- Ok: 23%
- Not good: 4%

ALMA mentors:

- Very good: 53%
- Good: 22%
- Ok: 21%
- Not good: 1%

2. The teacher encouraged and supported student perseverance and motivation

External enumerators:

- Very good: 10%
- Good: 51%
- Ok: 31%
- Not good: 6%

ALMA mentors:

- Very good: 60%
- Good: 14%
- Ok: 24%
- Not good: 2%

Overall, ALMA mentors rated socio-emotional teaching practices as "very good" substantially more frequently than external enumerators across both dimensions, while external enumerators more often rated practices as "good" or "ok." Ratings of "not good" were low across both observer groups.

Written version of Figure 25

The figure is a bar chart titled "Average teaching practice score, by sex and GTP participation."
The vertical axis shows percentages from 0.0% to 100.0%, representing average teaching practice scores.
The chart presents four bars grouped into two comparisons:
Teaching practice score by sex

- Female teachers: 60.7%
- Male teachers: 51.7%

Teaching practice score by participation in the Teacher Working Group (GTP)

- Teachers who participated in the GTP: 58.2%
- Teachers who did not participate in the GTP: 50.2%

Overall, the figure shows that average teaching practice scores are higher among female teachers than male teachers, and higher among teachers who participated in the GTP compared to those who did not participate.

Written version of Figure 26

The figure is a grouped bar chart titled "Average teaching score by age and GTP participation."
The vertical axis shows percentages from 0.0% to 100.0%, representing average teaching practice scores.
The chart compares teaching scores across three age groups, disaggregated by whether teachers participated in the Teacher Working Group (GTP) or did not participate.
The legend indicates:

- Did not participate in the GTP (dark blue)
- Participated in the GTP (orange)

Average teaching practice score by age group
Under 50 years

- Did not participate in the GTP: 55.8%
- Participated in the GTP: 59.9%

Ages 51–59

- Did not participate in the GTP: 50.2%
- Participated in the GTP: 57.6%

Over 60 years

- Did not participate in the GTP: 42.9%
- Participated in the GTP: 54.6%

Overall, the figure shows that teachers who participated in the GTP achieved higher average teaching practice scores across all age groups. Teaching practice scores decline with age among those who did not participate in the GTP, while participation in the GTP is associated with notably higher scores, particularly among teachers aged over 60 years.

Written version of Figure 27

The figure is a bar chart titled "Average reading comprehension score by availability of reading corner and mentorship."

The vertical axis shows percentages from 0.00% to 100.00%, representing average reading comprehension scores.
The horizontal axis presents four categories describing the availability of reading corners in Grade 2 classrooms and the presence or absence of administrator mentorship.
Average reading comprehension scores by school conditions

- No reading corner in Grade 2: 35.60%
- Reading corner in Grade 2: 41.40%
- Reading corner in Grade 2, but no mentorship to administrator: 37.80%
- Reading corner in Grade 2 and administrator receiving mentorship: 47.30%

Overall, the figure shows higher average reading comprehension scores in schools with a reading corner in Grade 2 compared to those without one. The highest scores are observed in schools where a reading corner in Grade 2 is combined with administrator mentorship, indicating a positive association between learning resources, leadership support, and student reading comprehension outcomes.

Written version of Figure 29

The figure is a bar chart titled "Ability to read words and reading comprehension score, by use of formative assessments."
The vertical axis shows percentages from 0.0% to 100.0%.
The chart compares outcomes for students taught by teachers who used formative assessments and those taught by teachers who did not use formative assessments, across two indicators.
The legend indicates:

- Teacher used formative assessments (dark blue)
- Teacher did not use formative assessments (orange)

Percentage of students able to read words in a passage

- Teacher used formative assessments: 73.5%
- Teacher did not use formative assessments: 65.2%

Average reading comprehension score

- Teacher used formative assessments: 43.0%
- Teacher did not use formative assessments: 34.5%

Overall, the figure shows that students whose teachers used formative assessments were more likely to be able to read words in a passage and achieved higher average reading comprehension scores than students whose teachers did not use formative assessments.

Written version of Figure 31

The figure is a bar chart titled "% students consuming protein-rich foods in school meals."
The vertical axis shows percentages from 0.00% to 100.00%.
The horizontal axis presents two categories based on whether the school administrator self-reported receiving mentor support.
Percentage of students consuming protein-rich foods

- Administrator self-reported mentor support: 89.5%
- Administrator did not report mentor support: 75.0%

Overall, the figure shows a higher proportion of students consuming protein-rich foods in school meals in schools where administrators self-reported receiving mentor support compared to schools where administrator mentorship was not reported.

Written version of Figure 32

The figure is a grouped bar chart titled "Reading comprehension and working memory scores by protein consumption status."
The vertical axis shows percentages from 0% to 80%, representing average scores.
The chart compares two outcomes—reading comprehension and working memory score—disaggregated by whether students consumed protein-rich foods or did not consume protein-rich foods through school meals.
The legend indicates:

- Consumed protein-rich foods (dark blue)
- Did not consume protein-rich foods (orange)

Reading comprehension scores

- Studied consumed protein-rich foods: 41.0%
- Students did not consume protein-rich foods: 31.2%

Working memory scores

- Students consumed protein-rich foods: 42.2%
- Students did not consume protein-rich foods: 37.6%

Overall, the figure shows that students who consumed protein-rich foods through school meals achieved higher average scores in both reading comprehension and working memory compared to students who did not consume protein-rich foods.

Written version of Figure 33

The figure is a bar chart titled "% of students unable to read words, by sex and water availability at school."
The vertical axis shows percentages ranging from –10% to 70%.
The horizontal axis presents two school conditions: "no water" and "has water."
Results are disaggregated by sex, with Male shown in dark blue and Female shown in orange.
Percentage of students unable to read words
Schools with no water

- Male students: 42.7%
- Female students: 36.0%

Schools with water

- Male students: 34.7%
- Female students: 22.7%

Overall, the figure shows that the proportion of students unable to read words is substantially higher in schools without water for both boys and girls. Across both water conditions, male students have a higher proportion of non-readers than female students, with the lowest proportion observed among girls in schools with access to water.

Written version of Figure 34

The figure is a bar chart titled "Proportion of students with access to water at school, disaggregated by administrator self-reported mentor support."
The vertical axis shows percentages from 0% to 100%.
The chart presents two categories:

- Administrator self-reported mentor support: 92%
- Administrator did not self-report mentor support: 88%

Overall, access to water at school is higher in schools where administrators self-reported receiving mentorship.

Written version of Figure 35

The figure is a bar chart titled "Average attendance rates by grade (Cycles 1–2)."
The vertical axis shows attendance rates from 0% to 100%.
Attendance rates by grade are:

- Grade 1: 68%
- Grade 2: 73%
- Grade 3: 79%
- Grade 4: 77%
- Grade 5: 70%
- Grade 6: 74%

Overall, attendance is lowest in Grade 1 and highest in Grade 3, with an overall average attendance rate of 75%.

Written version of Figure 36

The figure is a grouped bar chart titled "Average attendance rates of Cycle 1 students (Grades 1–4), disaggregated by sex and study group."
Attendance rates by grade and sex are:

- Grade 1: Male 68%, Female 66%
- Grade 2: Male 75%, Female 70%
- Grade 3: Male 79%, Female 79%
- Grade 4: Male 76%, Female 77%

Attendance is lower for female students in Grades 1–2, with parity achieved from Grade 3 onward.

Annex 2 Written version of Report Annex figures

Written version of Figure 38

The figure is a grouped bar chart titled *"Number of letters identified per minute, by teacher/coordinator self-reported exposure to mentorship."*
Average number of letters identified per minute is presented by level of reported exposure to mentorship and by duration of exposure.
Long-term exposure

- Not reported by either teacher or coordinator: 34.8
- Reported by teacher or coordinator (not both): 36.5
- Reported by both teacher and coordinator: 37.3

Recently joined

- Not reported by either teacher or coordinator: 40.1
- Reported by teacher or coordinator (not both): 39.9
- Reported by both teacher and coordinator: 42.2

Total

- Not reported by either teacher or coordinator: 37.5
- Reported by teacher or coordinator (not both): 38.2
- Reported by both teacher and coordinator: 39.8

Across all exposure categories, the highest average number of letters identified per minute is observed where mentorship is reported by both teachers and school coordinators.

Written version of Figure 39

The figure is a grouped bar chart titled "Invented word reading fluency (wpm) by mentorship status."

Average invented word reading fluency, measured in words per minute, is shown for long-term schools, recently joined schools, and the total. Results are disaggregated by whether the school administrator reported mentor support.

Long-term schools:

- School administrator did not report mentor support: 13.7 words per minute
- School administrator reported mentor support: 15.4 words per minute

Recently joined schools:

- School administrator did not report mentor support: 15.5 words per minute
- School administrator reported mentor support: 17.5 words per minute

Total:

- School administrator did not report mentor support: 14.6 words per minute

School administrator reported mentor support: 16.4 words per minute

Written version of Figure 40

The figure is a grouped bar chart titled "Percentage of students unable to read invented words, by mentorship status."

The proportion of students unable to read any invented words is shown for long-term and recently joined schools. Results are disaggregated by self-reported mentorship status of teachers and school administrators.

Long-term schools:

- Teacher was not mentored: 33.7 percent
- Teacher was mentored: 26.8 percent
- Administrator did not report mentoring: 32.3 percent

- Administrator self-reported mentoring: 18.7 percent

Recently joined schools:

- Teacher was not mentored: 24.5 percent
- Teacher was mentored: 26.2 percent
- Administrator did not report mentoring: 27.6 percent
- Administrator self-reported mentoring: 22.3 percent

Written version of Figure 41

The figure is a grouped bar chart titled "Average familiar word reading fluency (wpm) by mentorship status."

Average familiar word reading fluency, measured in words per minute, is shown for long-term schools, recently joined schools, and the total. Results are disaggregated by whether the school administrator reported access to mentorship.

Long-term schools:

- Administrator did not report mentor support: 17.7 words per minute
- Administrator self-reported mentor support: 20.3 words per minute

Recently joined schools:

- Administrator did not report mentor support: 20.1 words per minute
- Administrator self-reported mentor support: 22.6 words per minute

Total:

- Administrator did not report mentor support: 18.9 words per minute
- Administrator self-reported mentor support: 21.4 words per minute

Written version of Figure 42

The figure is a grouped bar chart titled "Proportion of zero scores in familiar word reading, by mentorship status."

The proportion of students scoring zero on familiar word reading is shown for long-term schools and recently joined schools. Results are disaggregated by self-reported mentorship status of school administrators and teachers.

Long-term schools:

- Administrator did not report mentor support: 32.1 percent
- Administrator self-reported mentor support: 18.1 percent
- Teacher received mentorship: 33.7 percent
- Teacher did not receive mentorship: 26.5 percent

Recently joined schools:

- Administrator did not report mentor support: 27.6 percent
- Administrator self-reported mentor support: 22.3 percent
- Teacher received mentorship: 24.5 percent
- Teacher did not receive mentorship: 26.2 percent

Written version of Figure 43

The figure is a grouped bar chart titled "Average passage reading fluency, by mentorship status (reported by school administrators)."

Average passage reading fluency, measured in words per minute, is shown for long-term schools, recently joined schools, and the total. Results are disaggregated by whether the school administrator reported mentor support.
Long-term schools:

- Administrator did not report mentor support: 21.4 words per minute
- Administrator self-reported mentor support: 22.9 words per minute

Recently joined schools:

- Administrator did not report mentor support: 24.7 words per minute
- Administrator self-reported mentor support: 27.6 words per minute

Total:

- Administrator did not report mentor support: 23.1 words per minute
- Administrator self-reported mentor support: 25.2 words per minute

Written version of Figure 44

The figure is a grouped bar chart titled "Percentage of non-readers by study group and mentorship status."
The proportion of non-readers, defined as students unable to read a single word in a passage, is shown for long-term schools, recently joined schools, and the total. Results are disaggregated by administrator- and teacher-reported mentorship status.
Long-term schools:

- Administrator did not report mentorship: 35.6 percent
- Administrator reported mentorship: 23.9 percent
- Teacher did not receive mentorship: 34.3 percent
- Teacher received mentorship: 31.4 percent

Recently joined schools:

- Administrator did not report mentorship: 31.6 percent
- Administrator reported mentorship: 22.6 percent
- Teacher did not receive mentorship: 25.2 percent
- Teacher received mentorship: 29.5 percent

Total:

- Administrator did not report mentorship: 33.5 percent

Written version of Figure 45

The figure is a grouped bar chart titled "Average reading comprehension scores (lower level) by mentorship status."
Average lower-level reading comprehension scores, expressed as percentages, are shown for long-term schools, recently joined schools, and the total. Results are disaggregated by whether the school administrator reported mentor support.
Long-term schools:

- Administrator did not report mentor support: 34.5 percent
- Administrator self-reported mentor support: 43.3 percent

Recently joined schools:

- Administrator did not report mentor support: 38.7 percent

- Administrator self-reported mentor support: 45.4 percent

Total:

- Administrator did not report mentor support: 36.6 percent
- Administrator self-reported mentor support: 44.4 percent

Written version of Figure 46

The figure is a grouped bar chart titled "Percentage of zero scores in reading comprehension, disaggregated by study group and access to mentorship."
The proportion of students scoring zero in reading comprehension is shown for long-term schools, recently joined schools, and the total. Results are disaggregated by administrator- and teacher-reported mentorship status.

Long-term schools:

- Administrator did not report mentorship: 44.0 percent
- Administrator reported mentorship: 30.9 percent
- Teacher did not report mentorship: 41.2 percent
- Teacher self-reported mentorship: 39.6 percent

Recently joined schools:

- Administrator did not report mentorship: 36.7 percent
- Administrator reported mentorship: 27.7 percent
- Teacher did not report mentorship: 28.4 percent
- Teacher self-reported mentorship: 35.0 percent

Total:

- Administrator did not report mentorship: 40.3 percent
- Administrator reported mentorship: 29.3 percent
- Teacher did not report mentorship: 35.1 percent
- Teacher self-reported mentorship: 37.2 percent

Written version of Figure 47

The figure is a grouped bar chart titled "Passage reading fluency, disaggregated by mentorship status and study group."
Average passage reading fluency, measured in words per minute, is shown for long-term schools and recently joined schools. Results are disaggregated by whether the school administrator reported mentor support. The results are calculated using samples matched for vulnerability.

Long-term schools:

- Administrator did not report mentor support: 21.1 words per minute
- Administrator self-reported mentorship: 23.1 words per minute

Recently joined schools:

- Administrator did not report mentor support: 23.5 words per minute
- Administrator self-reported mentorship: 28.3 words per minute

Written version of Figure 48

The figure is a grouped bar chart titled "Average reading comprehension scores, disaggregated by mentorship status and study group."
Average reading comprehension scores, expressed as percentages, are shown for long-term schools and recently joined schools. Results are disaggregated by whether the school administrator reported mentor support. The results are calculated using samples matched for vulnerability.
Long-term schools:

- Administrator did not report mentor support: 34.0 percent
- Administrator self-reported mentorship: 43.6 percent

Recently joined schools:

- Administrator did not report mentor support: 36.9 percent
- Administrator self-reported mentorship: 46.5 percent

Written version of Figure 49

The figure is a grouped bar chart titled "Average passage reading fluency (wpm)."
Average passage reading fluency, measured in words per minute, is shown for three groups of schools: a six-school cluster in Remexio, all Phase 8 schools (2022), and Phase 9 schools (2023). Results are presented for three time points: 2019, 2023 (immediately post-ALMA), and 2025.
Six-school cluster (Remexio):

- 2019: 7.1 words per minute
- 2023 (immediately post-ALMA): 2.7 words per minute
- 2025: 9.5 words per minute

All Phase 8 schools (2022):

- 2019: 7.5 words per minute
- 2023 (immediately post-ALMA): 2.2 words per minute
- 2025: 9.7 words per minute

Phase 9 schools (2023):

- 2019: 6.7 words per minute
- 2023 (immediately post-ALMA): 5.1 words per minute
- 2025: 6.7 words per minute

Written version of Figure 50

The figure is a grouped bar chart titled "Average reading comprehension score."
Average reading comprehension scores, expressed as percentages, are shown for a six-school cluster in Remexio, all Phase 8 schools (2022), and Phase 9 schools (2023). Results are presented for three time points: 2019, 2023 (immediately post-ALMA), and 2025.
Six-school cluster (Remexio):

- 2019: 17.6 percent
- 2023 (immediately post-ALMA): 5.2 percent
- 2025: 29.0 percent

All Phase 8 schools (2022):

- 2019: 16.6 percent
- 2023 (immediately post-ALMA): 4.7 percent
- 2025: 23.5 percent

Phase 9 schools (2023):

- 2019: 13.6 percent
- 2023 (immediately post-ALMA): 11.1 percent
- 2025: 13.1 percent