



MIDLINE EVALUATION

McGovern-Dole Food for
Education Program

HATUTAN

IN TIMOR-LESTE



Legal Notice and Disclaimer

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior approval in writing from Consilient/CARE.

In accordance with 2 CFR 200.315 Intangible Property, USDA has the right to

- (1) Obtain, reproduce, publish, or otherwise use the data produced during the evaluation; and*
- (2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes.*

This report is not a legally binding document. It is a collaborative informational and assessment document and does not necessarily reflect the views of any of the contributing partners in all of its contents. Any errors are the sole responsibility of the authors.

TABLE OF CONTENTS

TABLE OF CONTENTS	III
LIST OF TABLES	VI
LIST OF FIGURES	IX
ACRONYMS	X
EXECUTIVE SUMMARY	I
Methodology	i
Literacy and Quality of Education	ii
Health, Nutrition, and Economic Empowerment	iv
Gender and Power	vii
Predictive Analysis	vii
Learning Agenda	viii
Program Implications	xii
Conclusion and Recommendations	xiv
INTRODUCTION	1
HATUTAN Program Overview	1
Timor-Leste Overview	2
METHODOLOGY	12
Research Objectives	13
Data Collection Tools	15
Data Collection	22
Sampling	24
Methodological Analysis	29
Study Limitations	31
LITERACY RESULTS	34
Overall Literacy Scores	34
Literacy Subtask Results	40
Impact of COVID-19 on Learning	50
QUALITY OF INSTRUCTION	52
Teaching Practices	52
Teacher Attendance	60
School Supplies and Materials	63
Literacy Instruction Materials	64
Skills and Knowledge of Teachers	66
Skills and Knowledge of School Administrators	68
STUDENT ATTENTIVENESS	69
Student Attentiveness Indicators	69
Student Hunger Indicators	73
Food Access Indicators	76
STUDENT ATTENDANCE	79
Student Attendance Indicators	79
Health-Related Absences	91
School Infrastructure	92
Other Factors	95

SCHOOL FEEDING PROGRAM	98
Implementation and Management of the SFP.....	98
School Feeding Program Hygiene and Resources.....	104
Access to and Use of Government Funding for SFP.....	111
Purchase of Local Food Items.....	112
Contributions to SFP.....	117
HEALTH AND NUTRITION.....	119
Nutrition Practices	120
Knowledge of Health and Hygiene.....	124
Knowledge of Nutrition.....	128
Access to Clean Water and Sanitation in the Household and in School	130
Access to Preventative Health Interventions.....	134
ECONOMIC EMPOWERMENT	137
Savings	137
VSLAs.....	139
AGRICULTURAL PRACTICES	143
Overview of Agricultural Practices.....	143
Relationship Between Agricultural Practices and Economic and Nutritional Outcomes	146
GENDER AND POWER.....	147
Division of Labor.....	148
Household Decision-making	151
Control of Productive Assets	153
Claiming Rights and Meaningful Participation in Public Decision-making	155
Control Over One's Body	156
Violence and Restorative Justice.....	158
Aspirations and Strategic Interests	160
PREDICTIVE ANALYSIS	163
Predictors of Literacy.....	163
Predictors of Engaging Teaching Practices.....	167
Predictors of Student Attentiveness.....	168
Predictors of Student Attendance	170
Predictors of Health Knowledge	171
LEARNING AGENDA DISCUSSION	174
Education and Literacy	175
School Meal Program Implementation	178
Health, Nutrition, and Agriculture	181
Gender-based Violence.....	183
Sustainability	185
PROGRAM IMPLICATIONS	185
Design/Relevance.....	186
Management and Coordination.....	189
Effectiveness.....	189
Efficiency	191
Sustainability	191
Impact.....	193

Gender	197
COVID-19 Impact	197
CONCLUSION	198
Literacy Development and Quality of Education	198
School Feeding.....	199
Health and Nutrition	200
ANNEX 1: RESEARCH QUESTIONS	202
ANNEX 2: INDICATORS AND OUTCOMES	206
HATUTAN Indicators Assessed at Baseline and Midline.....	206
McGovern-Dole Standard and Custom Outcomes	209
ANNEX 3: DETAILED METHODOLOGICAL ANALYSIS	211
ANNEX 4: DATA COLLECTION TOOLS	221
Early Grade Reading Assessment	221
Classroom Observation Tool	229
School Survey Tool.....	232
Household Survey Tool	239
ANNEX 5: TERMS OF REFERENCE	254
I. Summary.....	254
II. Project Background	255
III. Purpose and Objectives of the Midline Study.....	256
IV. Indicators for Collection and Midline Evaluation Questions	257
V. Midline Study Design and Methodology	260
VI. Midline Study Deliverables and Reporting Outline	263
VII. Contractor Qualifications and Team Composition	265
VIII. Midline Study Management.....	265
IX. Midline Schedule/Timeline	266
X. Budget	266
XI. Evaluation Criteria for Proposals.....	267
XII. Intellectual Property	267
XIII. Submission	267
TOR Annex 1: Learning Agenda Questions.....	267
TOR Annex 2: Results Framework.....	269

LIST OF TABLES

Table 1: EGRA sections and scoring	16
Table 2: Engaging teaching practices	18
Table 3: Dietary diversity food groups	20
Table 4: EGRA sample by cohort	25
Table 5: Sample for school survey, classroom observation, household survey, and farmer's group survey.....	25
Table 6: Demographics of students assessed at midline	27
Table 7: Household demographics at midline	27
Table 8: Household education and livelihoods	28
Table 9: Number of children in households at midline.....	28
Table 10: Student disabilities at midline	29
Table 11: Summary of literacy scores, cross-sectional cohort	35
Table 12: Summary of literacy scores, recontacted students	37
Table 13: Summary of literacy score changes	38
Table 14: Overall literacy scores by gender and cohort	39
Table 15: Accuracy of letter recognition	41
Table 16: Letter name recognition, difference-in-differences results.....	42
Table 17: Invented word fluency, difference-in-differences results.....	44
Table 18: Familiar word fluency, difference-in-differences results.....	45
Table 19: Passage fluency, difference-in-differences results	47
Table 20: Reading comprehension, difference-in-differences results.....	49
Table 21: Change in overall scores by municipality	51
Table 22: Use of engaging teaching practices (% of classrooms using given number of practices) .	52
Table 23: Change in use of specific engaging teaching practices.....	53
Table 24: Change in McGovern-Dole Custom Outcome #5 for teaching practices	54
Table 25: Change in use of traditional teaching practices	55
Table 26: Change in use of negative teaching practices	55
Table 27: Change in caregiver perceptions of negative teaching practices	56
Table 28: Change in teaching practices, male and female teachers	58
Table 29: Change in positive teaching practices towards girls and boys	59
Table 30: Change in negative teaching practices towards girls and boys	59
Table 31: Treatment of male and female students by gender of teacher	60
Table 32: Change in teacher attendance rates	62
Table 33: Change in McGovern-Dole Custom Outcome #6 for teacher attendance	62
Table 34: Change in use of savings for education	64
Table 35: Changes in access to reading corner and reading materials.....	65
Table 36: Changes in teacher training and education	67
Table 37: Change in director provision of coaching	69
Table 38: Student attentiveness scores, cross-sectional cohort.....	70
Table 39: Panel student attentiveness indicators	71
Table 40: Regressions for student attentiveness indicators, panel cohort	72
Table 41: Disaggregated student attentiveness summary statistics	72
Table 42: Disaggregated student attentiveness panel regressions	73
Table 43: Student hunger variables, cross-sectional cohort.....	74
Table 44: Student hunger scores, panel cohort.....	74
Table 45: Student hunger, disaggregated by gender and municipality.....	75
Table 46: Student hunger panel regressions, disaggregated by gender	76
Table 47: Change over time in food access indicators	76
Table 48: Disaggregated food access summary scores.....	78
Table 49: Change in attendance rate by grade	79
Table 50: Student attendance by gender, intervention schools.....	82

Table 51: Reason for missing school	82
Table 52: Student attendance in comparison group, disaggregated by municipality	83
Table 53: Change in dropout rates by grade	84
Table 54: Dropout rate for treatment group, disaggregated by municipality	85
Table 55: Dropout rates for treatment group, disaggregated by gender	89
Table 56: Summary scores for student anxiety and depression	89
Table 57: Student anxiety/depression, disaggregated by gender in intervention areas	90
Table 58: Student anxiety and depression, disaggregated by municipality	90
Table 59: Summary scores for health-related absences	91
Table 60: Difference-in-differences, days missed due to illness	92
Table 61: Change over time in school infrastructure	93
Table 62: School infrastructure, disaggregated by municipality and location	94
Table 63: Water source of schools	95
Table 64: Change over time in schools with improved infrastructure	95
Table 65: Change in access to public spaces and services	96
Table 66: Summary scores for access to school	96
Table 67: Summary scores of school enjoyment	97
Table 68: Mean days of school missed by key indicators	98
Table 69: School feeding program responsibility, by study group	99
Table 70: Meals provided to students, by survey type and study group	99
Table 71: School feeding program menu available, by study group	100
Table 72: Food items on the school feeding menu, by study group	101
Table 73: Food groups served to children in school, by study group	101
Table 74: School menu dietary diversity score, by study group	102
Table 75: Caregivers' reports of meal quality and availability, by study group	102
Table 76: Hygienic preparation of food, by study group	105
Table 77: School facilities at midline	106
Table 78: Storage spaces, by study group	106
Table 79: Storage space materials, by study group	107
Table 80: Storage sanitation, by study group	108
Table 81: Schools buying local produce from farmers for the SFP, by study group	113
Table 82: Local produce schools bought from farmers, by study group	113
Table 83: Schools buying local produce from farmers, by municipality and study group	114
Table 84: PTA involvement in school feeding, by study group	118
Table 85: Frequency of PTA meetings	118
Table 86: Household participation in the PTA, by study group	118
Table 87: Factors contributing to health and nutrition per the results framework	120
Table 88: Food consumed by women of childbearing age (15-49), by study group	121
Table 89: Food consumed by children between the ages of 6 and 23 months, by study group	122
Table 90: Mean dietary diversity score (DDS) of women (15-49 years old) and children (6-23 months old), by study group	123
Table 91: Knowledge of healthy hygiene practices, by study group	125
Table 92: Custom outcome #21: Percentage of participants who can identify at least 17 out of 19 important hygiene/sanitation practices, by study group	126
Table 93: Unhealthy hygiene practices identified as healthy hygiene practices, by study group	126
Table 94: Knowledge of COVID-19 prevention behaviors, by study group	127
Table 95: Caregivers' knowledge of child nutrition practices, by study group	129
Table 96: Custom outcome #23: Percentage of participants in program target groups who can identify at least three important nutrition/dietary recommendations, by study group	130
Table 97: Access to drinking water and toilets, by study group	131
Table 98: Main source of household drinking water, by study group	132
Table 99: Collection of water from improved drinking water sources, by study group	132

Table 100: Collection of water from improved drinking water sources among treatment households, by rural or urban location	Error! Bookmark not defined.
Table 101: Sanitation in school, by study group.....	133
Table 102: Access to healthcare, by study group.....	135
Table 103: Use of savings and loans, by study group.....	135
Table 104: Households with savings, difference-in-differences.....	137
Table 105: Location of savings at midline	138
Table 106: Change over time in use of savings	138
Table 107: Savings indicators disaggregated by municipality	139
Table 108: Savings use by VSLA participation, midline	140
Table 109: Summary scores for VSLA indicators.....	140
Table 110: VSLA loan decision-making by gender.....	141
Table 111: VSLA loan use, midline	141
Table 112: Perceived benefits of VSLA participation	142
Table 113: Impact of VSLA membership on nutrition and literacy outcomes	142
Table 114: Summary scores for main agricultural indicators.....	143
Table 115: Crops grown in keyhole gardens	144
Table 116: Crops grown in permagardens	144
Table 117: Summary scores for sales of homegrown produce	145
Table 118: Main challenges faced by farmers.....	145
Table 119: Access to AES and sources of information	146
Table 120: Relation between agricultural trainings, health, and economic status	147
Table 121: Gendered division of labor within household	148
Table 122: Change in gendered division of labor within household	149
Table 123: Caregivers' perceptions of student time spent on daily tasks	149
Table 124: Caregiver perceptions on tasks and time for education.....	150
Table 125: Time taken for household tasks and impact on schooling, difference-in-differences analysis	150
Table 126: Household decision-making at midline	152
Table 127: Control of productive assets, difference-in-differences analysis.....	153
Table 128: Garden-related decision-making	154
Table 129: Access to AES and garden trainings by gender, midline respondents	156
Table 130: Scenarios in which beating one's wife is justified, difference-in-difference analysis.....	157
Table 131: Teachers' use of positive and negative discipline practices, by gender of student.....	158
Table 132: Change in teachers' positive and negative discipline practices	159
Table 133: Reporting of abuse at school.....	160
Table 134: Caregivers' beliefs in the capacity of male and female students and school equity, by gender of child.....	161
Table 135: Summary of predictors of literacy results	167
Table 136: Predictors of engaging teaching practices.....	168
Table 137: Predictors of self-reported attentiveness	169
Table 138: Predictors of working memory scores	169
Table 139: Predictors of observed student attentiveness.....	169
Table 140: Predictors of student absences	170
Table 141: Predictors of hygiene knowledge.....	172
Table 142: Predictors of nutrition knowledge	172
Table 143: Predictors of dietary diversity	173
Table 144: Predictors of COVID-19 prevention behavior	174
Table 145: PTA involvement at midline	190
Table 146: Differences in native language among cross-sectional sample.....	211
Table 147: Matched intervention and comparison administrative posts	215
Table 148: Re-contact rates by municipality.....	218
Table 149: Re-contact rates by student age.....	219

LIST OF FIGURES

Figure 1: Change in literacy scores, cross-sectional cohort	36
Figure 2: Change in literacy scores, recontacted cohort	38
Figure 3: Change in literacy scores by gender	40
Figure 4: Letter name knowledge, zero scores	Error! Bookmark not defined.
Figure 5: Change in invented word scores	43
Figure 6: Change in familiar word fluency scores.....	45
Figure 7: Change in passage fluency scores	47
Figure 8: Change in passage fluency scores, non-zero scorers.....	48
Figure 9: Change in reading comprehension scores	49
Figure 10: Change in teacher attendance	61
Figure 11: Change in percent of teachers attending literacy training	67
Figure 12: Trend of student attentiveness indicators.....	71
Figure 13: Trend of student hunger, panel cohort	75
Figure 14: Trend of school feeding indicators	78
Figure 15: Change in attendance rate by grade	81
Figure 16: Change in dropout rates by grade.....	85
Figure 17: Mean days missed due to illness	92
Figure 18: Change over time in school infrastructure.....	94
Figure 19: Implementation of school feeding program	100
Figure 20: Use of savings and loans for healthcare, by study group.....	136
Figure 21: Use of savings and loans for healthcare among treatment households, by urban and rural location.....	136
Figure 22: Predictors of literacy, individual level	165
Figure 23: Predictors of literacy, school level.....	166
Figure 24: Predictors of successful re-contact	219

ACRONYMS

AES	Agricultural extension services
BL	Baseline
COVAX	COVID-19 Vaccines Global Access
CBA	Curriculum-based assessment
CDA	Community development agent
DiD	Difference-in-differences
EGRA	Early Grade Reading Assessment
FGD	Focus group discussion
GDP	Gross domestic product
GAT	<i>Gabinete Apoio Técnico (Technical Support Office)</i>
HATUTAN	<i>Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun</i> (Food to Support Nutrition and Education)
HLM	Hierarchical linear modeling
IPC	Integrated Food Security Phase Classification
KII	Key informant interview
MAD	Minimum acceptable diet
MEYS	Timor-Leste Ministry of Education, Youth, and Sports
ML	Midline
NGO	Non-governmental organization
PTA	Parent-teacher association
SFP	School feeding program
SGBV	Sexual and gender-based violence
TVET	Technical and vocational education and training
USDA	United States Department of Agriculture
VSLA	Village savings and loan association
WASH	Water, sanitation, and hygiene
WDDS	Women's dietary diversity score

EXECUTIVE SUMMARY

The HATUTAN program (*Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun* or Food to Support Nutrition and Education) is a five-year initiative to build a partnership between schools and communities in order to improve literacy, learning, healthy, and nutrition for children and adults in Timor-Leste. The program works in partnership with the Government of Timor-Leste and development stakeholders to address two strategic objectives: improved literacy of school-aged children and increased use of health, nutrition, and dietary practices. The HATUTAN program is funded by the U.S. government through the Foreign Agricultural Service of the United States Department of Agriculture under the McGovern-Dole International Food for Education and Child Nutrition Program. The program is implemented by a consortium led by CARE International with Mercy Corps and WaterAid. The lead Timorese government partner is the Ministry of Education, Youth, and Sports, in collaboration with the Ministry of Health, Ministry of State Administration, and Ministry of Agriculture and Fisheries.

To achieve these objectives, the program supports, among a variety of activities, the Government of Timor-Leste's school feeding program (SFP) to fully operate in all basic education and preschools throughout the school year. Key project activities include strengthening and supplementing the government-sponsored SFP and building school capacity through trainings for teachers and administrators and provision of resource materials. Additionally, the HATUTAN program seeks to support farmers to boost the production of local produce to increase yields and help create sustainable sources of nutritious food for local schools. In addition to activities related to literacy and SFPs, HATUTAN seeks to conduct trainings related to nutrition, health, and other topics, and to promote gender equality and the reduction of gender-based violence.

This report presents the midline evaluation of the HATUTAN program, which began in early 2019. It is important to note that restrictions due to the COVID-19 pandemic have had a substantial impact on program activities and target outputs and outcomes. In March 2020, HATUTAN field activities were halted, field offices were temporarily closed, and staff began to work from home due to a State of Emergency issued by the Government of Timor-Leste. This State of Emergency remains in effect to date, with varying levels of restriction on school activities, movement, and group gatherings. As a result, the HATUTAN program is behind schedule in terms of some major deliverables due to COVID-19. Additionally, many program activities pivoted to include a focus on COVID-19 prevention and awareness.

The purpose of the midline evaluation was to understand the preliminary impact of the HATUTAN program on schools, students, families, and other stakeholders in areas including education and literacy, the quality of instruction, student attentiveness and attendance, student feeding programs, nutrition and health practices, economic empowerment, agriculture practices, and gender and power dynamics. Key research objectives include assessment of the progress of the project, the implementation of HATUTAN activities, the efficiency of project organization, and project sustainability; identification of new COVID-related barriers and changes; and provision of recommendations for the future orientation of the project. The findings will inform the continued orientation and emphasis of the program during its remaining time and will provide meaningful information for the participating organizations and groups.

METHODOLOGY

Following the evaluation methodology in the baseline, the evaluation uses a mixed-methods quasi-experimental design, triangulating information from different sources and both quantitative and qualitative methods to enhance the reliability and comprehensiveness of findings. All methods are gender-sensitive and socially inclusive, ensuring that women, men, girls, and boys are able to provide data in a safe, open, and reliable context, and that perspectives from all age and gender groups are adequately represented in data analysis. Additionally, the analysis takes into account disability and

language-related factors. The evaluation compares the progress observed in intervention primary schools with the progress observed in a comparison group of schools selected in neighboring municipalities. Comparing across similar “intervention” schools (those exposed to HATUTAN programming) and “comparison” schools (schools with no HATUTAN programming) allows us to better understand whether improvements in key areas, such as literacy, are due to HATUTAN program activities or are rather due to external factors that may affect all schools in the country, such as the COVID-19 pandemic or the implementation of nationwide government programs.

Five instruments are included in this study: (1) the Early Grade Reading Assessment (EGRA), which assesses the literacy of both students re-contacted from the baseline who are now in grades 3 or 4 and newly-contacted second grade students; (2) a classroom observation to identify teaching practices of second grade teachers; (3) a school survey to assess school resources, teacher and student enrollment and attendance, and the school feeding program; (4) a household survey, which provides data on household composition, savings and loan information, student attendance, gender and power dynamics, and health and nutrition knowledge; and (5) qualitative interviews and focus group discussions with parents, teachers, and school directors/coordinators.

Several limitations affected the study. The most salient of these is non-random assignment: Because treatment schools were selected based on vulnerability criteria, rather than randomly, comparison and treatment groups are not entirely balanced. As a result, when making inferences based on the data, we cannot be absolutely certain that results are due to program interventions rather than due to systematic differences between treatment and control groups. Additionally, students dropping out of school (and thus from the program)—an issue worsened by the COVID-19 pandemic—may reduce the representativeness of reported results of program impact, particularly if drop-out was biased (for example, if the students most likely to drop out were those performing worst in classes).

LITERACY AND QUALITY OF EDUCATION

The first McGovern-Dole strategic objective is to improve the literacy of school-aged children. The HATUTAN results framework highlights three factors that contribute to improved literacy of school-aged children: quality of instruction, student attentiveness, and student attendance. Activities seeking to strengthen the school feeding program are intended to improve outcomes in all of these areas.

LITERACY RESULTS

At midline, due to the impact of COVID-19, overall literacy scores among grade 2 students worsened for both intervention and comparison groups. However, we find a significant and positive effect on literacy scores for the intervention group compared to the comparison group. In other words, while average scores for both groups declined at midline compared to baseline, average scores for treatment students exposed to the program declined significantly less than those for comparison students. The results for the panel cohort of students assessed at both baseline and midline are less suggestive of program impact, with scores improving by similar amounts in both intervention and comparison areas due to exposure to an additional year of education. These results suggest that the HATUTAN program may have had a positive effect on mitigating the impact of COVID-19 on learning, although results are not conclusive.

For grade 2 students’ performance on specific literacy subtasks, analysis found that students in intervention schools performed significantly better than expected given results in comparison areas for the invented word fluency and passage reading subtasks. The program may also have had a positive impact on grade 2 students’ scores for the remaining subtasks—letter recognition, familiar word fluency, and reading comprehension—but results are less conclusive for these subtasks. It is worth noting, however, that scores for grade 2 students on all subtasks remain very low, in part due to the COVID-19 pandemic and in part due to structural issues affecting literacy in Timor-Leste. Furthermore, there remains a large gap between letter recognition skills and word recognition or

reading comprehension skills, suggesting that teachers may still be using ineffective pedagogical approaches to literacy.

At both baseline and midline, female students performed better than male students. HATUTAN program activities do not appear to have had differential impacts by gender, with scores for both male and female students declining overall, but declining less within treatment areas than within comparison areas. Overall, there remains a large gap in scores between male and female students which HATUTAN programming seems not to have affected.

QUALITY OF INSTRUCTION

At midline, there was little change in the overall use of engaging teaching practices in either treatment or control schools. However, teachers in intervention schools were significantly more likely to use games or exercises in class than expected given results in comparison schools, and were substantially—though not significantly—more likely to ask open questions and use the reading corner. There was a substantial and significant decrease in the use of traditional teaching practices in treatment schools as compared to control schools; unfortunately, the HATUTAN program did not appear to have had a likewise positive effect at reducing the use of negative teaching practices. At midline, the observed use of corporal punishment increased in both intervention and comparison schools, and caregiver perceptions of negative teaching practices remained high from baseline to midline. There was no significant change in perceptions of negative teaching practices in treatment groups as compared to control groups. Analyzing the gender-specific prevalence of negative teaching behaviors, at midline, teachers used verbal and physical discipline more frequently with boys than with girls.

Unfortunately, the HATUTAN program also does not appear to have had a positive effect on teacher attendance at midline. Among comparison schools, at midline, teacher attendance taken the day of the survey and recorded the previous day increased substantially; in contrast, among treatment schools, teacher attendance the day of the survey increased by less than in intervention schools, and attendance the day before decreased. This finding can in part be attributed to the fact that data collection took place at the peak of the rainy season which had a greater effect on many intervention schools due to their location in remote, mountainous areas, and which likely decreased attendance rates by making roads and bridges impassable. Furthermore, analysis disaggregated by municipality suggests that several comparison municipalities had large increases in teacher attendance from low starting points, suggesting that interventions in these municipalities may have partially driven results.

Access to school resources and supplies also affects the quality of instruction. At midline, quantitative data suggests that poor families often have to use their savings to pay for student supplies. However, most households did not report that a lack of school supplies is major constraint to school attendance or learning to read. Within schools, access to literacy materials increased in both comparison and intervention areas at midline. Among intervention schools, there was a particularly notable increase in access to reading corners. However, the analysis did not find a significant increase in access to literacy materials in intervention schools as compared to comparison schools, implying that the increase may not be due to HATUTAN program activities or that comparison schools may be receiving similar support from literacy-focused initiatives—indeed, 14 comparison schools received this type of support from organizations such as ALMA, Mary MacKillop, the Alola Foundation, and Plan International.

STUDENT ATTENTIVENESS

Student attentiveness is highlighted in the theory of change as a factor that may influence literacy, and it is expected to be a key causal step along the chain from improved school feeding to improved educational outcomes. Results show a positive but non-significant increase in self-reported and observed student attentiveness in intervention schools relative to comparison schools. However, working memory scores (used as a proxy for attentiveness) showed negative difference-in-difference score for both the panel and cross-sectional analysis—although again, these results were not

statistically significant. Overall, these results suggest an inconclusive effect of program activities on student attentiveness.

The results framework includes two sub-factors that contribute to student attentiveness: student hunger and access to food. The percent of students reporting that they had eaten on the day of the EGRA improved more in intervention schools than in comparison schools at midline, suggesting that program activities targeting student hunger had a positive effect. Similarly, schools in intervention areas were far more likely to have served meals at midline than at baseline—with nearly 90% of schools serving meals at midline compared to only around 2% at baseline—while there was little change in this indicator for comparison schools.

STUDENT ATTENDANCE

Student attendance is highlighted in the results framework as a factor that may affect literacy scores. Improved management of the school feeding program and consequent improved school feeding could act as a pull factor for students attending the school and reduce dropout rates. The norm-change aspects of HATUTAN, especially those targeting gender inequality and other harmful practices, could also result in a better environment for students and therefore higher attendance.

For five out of six grades there were negative difference-in-difference scores for student attendance. This suggests a possible deterioration in student attendance in intervention schools compared to comparison schools. However, the most likely explanation for this is that students in intervention schools were worse-affected by natural disasters.

In contrast to these results, intervention schools generally had relatively fewer students drop out in almost all grades at midline than expected relative to comparison schools. This suggests there was a benefit from the program and students were less likely to drop out, possibly due to parents being less likely to draw children out of school if there is an effective school feeding program or students being less likely to want to drop out. However, there were positive difference-in-difference scores for number of days missed due to illness, meaning the trend over the study period was worse for the intervention group than the comparison group.¹ Of other factors studied, safety on the way to school and at school play a role on attendance. Students within a 30-minute walk and those who felt safe on their walk missed significantly fewer days of school on average, while students who were afraid of school missed significantly more days on average.

HEALTH, NUTRITION, AND ECONOMIC EMPOWERMENT

The second McGovern-Dole strategic objective seeks to increase the use of improved health, nutrition, and dietary practices. The results framework indicates that healthy practices decrease health-related absences in schools, thus improving student attendance and contributing to improved literacy. Improved economic empowerment of households may also lead to improved health and nutrition practices.

SCHOOL FEEDING PROGRAM

At the midline, most comparison schools (71%) said that there was no school feeding for the day, while only 12% of treatment schools said the same. At the baseline, only 1% of the treatment schools reported having a school feeding program, compared to 30% of the comparison schools; therefore, only the intervention schools had a sharp increase in the provision of meals. Most schools consistently

¹ We note that at baseline, households were asked about the number of days missed due to illness over the past month, while at midline, they were asked about the number of days missed over the past *week*. However, this measurement change applied equally to the treatment and comparison groups. As such, in the absence of any changes, we would expect absences at midline to fall to around 25% of those reported at baseline for both treatment and comparison students. Any relative variation from this expected decline in treatment areas compared to comparison areas comprises the difference-in-differences.

served carbohydrates (91%), legumes (including beans) and nuts (64%), and dark green vegetables such as spinach, lettuce, and mustard greens (36%) at the midline, which is consistent with the baseline findings and with the commodities provided by HATUTAN (rice and pinto beans). Increased consumption of vitamin A-rich vegetables was also observed. However, food served to children in school are lacking in fruits. Intervention schools served a higher-quality menu than comparison schools, scoring significantly higher on dietary diversity.

Most parents felt that food for their children in school is prepared in a hygienic manner, which is an important consideration to ensure that children do not get sick and miss classes. Most schools were observed to have access to clean water for preparing meals (86% overall; 87% treatment, 83% comparison) and storage spaces that are at least somewhat clean (93% overall; 93% treatment, 93% comparison), and that the kitchen is cleaned using detergent (98% overall; 99% treatment, 96% comparison) at midline. These results generally indicate that more schools are practicing hygienic food preparation over time. However, the difference-in-differences analysis suggests that some changes in hygienic food preparation have occurred at similar rates in both intervention and comparison schools. For example, there were no significant differences in whether the school had a kitchen or in improved usage of clean water or detergent at midline in intervention schools compared to comparison schools.

To boost local production and maintain a sustainable source of food for the SFP, schools are encouraged to purchase their produce locally. Most schools that provided meals indicated during the baseline that they purchase food locally sometimes (57%) or all the time (39%). More than half of the schools at the midline continued to buy goods locally sometimes (45%) or all the time (13%).

Most of the schools reported having a PTA that provides oversight of school feeding (75%; 91% treatment, 57% comparison). PTA involvement increased at midline, a potentially positive sign for program effectiveness. Furthermore, at midline, PTAs were significantly more involved in almost every factor affecting schools in intervention areas relative to comparison areas. 63% of the treatment school PTAs had met during the current school year, while only 30% of the comparison school PTAs did the same. The level of participation in the PTA remains low, indicating limited participation and potential influence in activities involving school feeding. Most households (64%; 63% treatment, 66% comparison) do not have a member who participates in the PTA.

NUTRITION AND HEALTH PRACTICES

Caregivers interviewed during the household survey reported the types of food they consumed the prior day. Nearly all of the women (99%) of childbearing age at the midline said they consumed grains such as rice, maize, or bread, followed by vitamin A-rich dark leafy greens (73%) and other vitamin A-rich vegetables and fruits (58%). This is consistent with the food consumption trend observed at the baseline. A dietary diversity score was computed for women of childbearing age. On average, women in the treatment group consumed 3.3 food groups out of nine on the previous day at baseline, and 2.7 food groups at midline. Caregivers also reported the foods consumed by their children between 6 and 23 months old the day prior. Children's dietary diversity scores were calculated to determine if they meet the minimum acceptable diet (MAD) for children ages 6 to 23 months, which requires that they consume four out of the seven food groups. Among treatment households, only 6% of children met the MAD requirement at baseline and 4% at midline, while for the comparison group 7% of children at baseline and 4% at midline met the MAD requirement. On average, children in the treatment group consumed 1.9 food groups at the midline.

The results framework identified six factors that contribute to increased use of health, nutrition, and dietary practices: knowledge of health and hygiene practices, knowledge of safe food prep and storage practices, knowledge of nutrition, access to clean water and sanitation services, access to preventative health interventions, and access to requisite food prep and storage tools and equipment. Three of these sub-factors relate to knowledge. Caregivers were able demonstrate their knowledge of healthy hygiene practices and knowledge of childcare nutrition recommendations. At the midline, most caregivers (94% comparison, 93% treatment) were able to identify at least 17 out of 19 healthy hygiene practices. However, knowledge of nutrition practices is lower; around seven of ten caregivers (69%

comparison, 65% treatment) were able to identify at least three important nutrition/dietary recommendations. Additionally, handwashing with soap was not observed in 75% of intervention households at midline and 24% of intervention schools did not have handwashing stations with soap; this may be linked to low access to water infrastructure.

The remaining three sub-factors relate to access issues. Less than half of the respondents (44%; 47% treatment, 42% comparison) said that they had water available for the whole year, but most households (83%; 84% treatment, 83% comparison) had access to a toilet of some kind. However, it is worth noting that while the proportion of schools and households with toilets increased relative to baseline, 16% of intervention households and 20% of intervention schools still did not have toilets at midline. In terms of access to preventative health interventions, the percentage of caregivers that said they could not afford to take someone to the doctor or a clinic decreased between baseline (35%) and midline (16%). Most of the respondents in the midline said that they do not need to pay (54%), which was not an option in the baseline survey. Savings and loans were not commonly used for medical expenses but were used primarily for food and education expenses. Finally, on access to requisite food preparation and storage tools, most of the treatment (97%) and comparison (81%) schools had a kitchen, but access to equipment (such as utensils or pots) decreased slightly in treatment schools while increasing in comparison schools. A higher number compared to baseline. More than half of the schools (69%; 78% treatment, 56% comparison) also said they had some or enough storage space available. Most of these schools (89%) were observed to have proper ventilation in the storage spaces and no leaks on the roof (84%).

ECONOMIC EMPOWERMENT

There are several ways in which the HATUTAN program aims to be beneficial for economic empowerment, including spillover effects from other aspects of the program, improved gender equality leading to more equalized intra-household decision-making, and through establishing or improving village savings and loan associations (VSLAs). While key indicators for savings had changed from baseline to midline—for example, there was a large decrease in the proportion of households with savings who used those savings for business investment—there was no significant difference in differences. This suggests that the program had limited impact on savings, including whether a household had any savings and whether they used those savings for food, education, debt, or investing in assets.

Disaggregating savings use by VSLA participation showed little difference between the behavior of VSLA and non-VSLA participants. Small differences in VSLA loan use by treatment and comparison were found: The intervention group spent statistically significantly more on agriculture, business investment, and debt. However, given we only have midline results for VSLA participants and cannot assess trends over the program lifetime, we cannot draw strong conclusions as to program impact. The midline also examined whether the program improved how VSLAs function and whether this had an effect on education and nutritional outcomes. There were no statistically significant results, again suggesting limited impact. However, it is important to point out that a richer analysis will be possible at later rounds of program evaluation.

AGRICULTURAL PRACTICES

The main focus of the HATUTAN program regarding agriculture is to support farmers in adopting improved techniques—particularly keyhole gardening and permagardens—to boost yields and create sustainable sources of foods, including through farmer trainings. Most of the data for this section was collected at midline and the sample was comprised of households that were farmers. This means we are unable to assess changes over time and estimate difference in differences; furthermore, analyses are limited by small sample size.

Farmers in the treatment group who received training were significantly less likely to have had someone in their household go without eating in the past 30 days. Farmers in the treatment group who

had a keyhole garden were more likely to spend savings on investment. At this stage we must be cautious in attributing results for these indicators to program impact, however, due to limited sample size.

GENDER AND POWER

Overall, there were few significant changes to most gender and power dynamics among intervention areas at midline. There were no significant changes to the gendered division of labor for children within households; few differences in caregivers' perceptions of how much time male and female students spent on daily tasks or whether tasks made students late for school; few differences in the control of productive assets; no differences in attitudes towards gender-based violence; no differences in caregivers' opinions of the skills and capacities of girls and boys; and few differences in the use of violence in schools. In contrast, there did appear to be a substantial positive change in children's safe access to schools at midline among treatment groups. This improvement appears to have been slightly more salient for boys than for girls, although both male and female students appear to have benefited. Furthermore, in intervention areas, the rate at which teachers questioned and encouraged girls in the classroom increased substantially more at midline than for boys. Overall, however, impact on gender and power dynamics, particularly within households, was likely low because, as they are rooted in social norms, these dynamics tend to change very slowly over time. Additionally, planned gender trainings were not implemented as per schedule due to COVID-19 restrictions, resulting in limited impact in this area.

As described at baseline, girls generally have higher engagement and achievement in school than boys, and many parents even consider girls to have more capacity for reading, writing, and math than boys. However, adult women have limited decision-making power in households, are often limited to working as caregivers, and are often less involved in community organizations such as farmer's group or receive fewer benefits from local services such as agricultural extension services. Clearly, despite the great potential evidenced by young female students, gender norms reduce the options available to girls as they grow older.

Overall, these findings suggest a need for interventions that sustain girls' successes at young ages through adulthood. Activities that provide an opportunity for dialogue with communities to challenge traditional gender roles may help achieve this goal. However, as noted above and in the "Gender and Power" section, gender norms tend to change slowly over time; at midline, there is thus little evidence to date of the impact of program interventions that seek to improve the status and opportunities of girls and women.

PREDICTIVE ANALYSIS

LITERACY AND QUALITY OF EDUCATION

The analysis finds that age and gender are both significant predictors of literacy scores with substantial effect sizes: Older students, female students, and students who speak Tetum (the language of instruction) tend to have better overall literacy scores. Working memory scores and caregiver education were also significant predictors of overall literacy scores, though with smaller effect sizes: Better working memory and a more highly-educated caregiver are correlated with higher overall literacy scores. Most school-level predictors, including availability of reading materials, student attentiveness, engaging teacher practices, and student-teacher ratio were not found to be significant predictors of literacy scores. However, whether a school was supported by a literacy project and whether books were lent to students were found to be significant predictors of improved literacy scores for grade 3 and 4 students, suggesting that these materials may become useful once students have acquired more foundational literacy skills with additional years of schooling.

Class size was found to be a significant predictor of engaging teaching practices, with larger class sizes associated with less use of engaging teaching practices. Other predictor variables, including teacher gender, teacher education (secondary school or greater than secondary school), teacher experience (in years), availability of reading materials, whether the school had electricity, whether the school had a PTA, and whether the director provides coaching to teachers, were not found to be significant at midline.

Linear regression models found five statistically significant predictors of student attentiveness. Whether a student had eaten on that day predicted a higher probability that the student reported they were paying attention. The school purchasing local produce always or sometimes was associated with higher working memory scores than schools who never bought produce. Teaching practices that were associated with an increase in the observed number of students paying attention were reading alone, participating in group work, and using educational games.

Four significant predictors of number of student absences in the past week (as reported by the parent/caregiver at the household survey) were found. Spending at least half a day on chores and having difficult with self-care such as washing were associated with more days missed from school. Experiencing anxiety was associated with fewer days missed from school—a possible reverse causal effect, as students with more responsibilities may be more likely to face anxiety. Improved sanitation in the household (defined as at least a covered pit latrine) was associated with fewer days absent.

HEALTH, NUTRITION, AND ECONOMIC EMPOWERMENT

Caregivers' level of education is a statistically significant predictor of hygiene knowledge, nutrition knowledge, and dietary diversity at the baseline, but only for dietary diversity at the midline. Generally, the more education a household has, the more knowledge they have about hygiene and nutrition knowledge. A link between nutrition knowledge and dietary diversity was also observed, indicating translation of knowledge into practice through dietary diversity. A statistically significant predictor for nutrition knowledge includes household size at the baseline and midline—although the direction of the effect changed from baseline to midline, with a smaller household size significantly correlated with nutrition knowledge at baseline but a larger size at midline—and dietary diversity at baseline only. Other predictors for dietary diversity include household size and nutrition knowledge (i.e., number of nutrition recommendations identified), but they are only statistically significant for the midline.

A regression model for predictors of COVID-19 prevention behavior was also developed, and the number of COVID-19 prevention practices known and availability of a handwashing station at home statistically significantly predict people's COVID-19 prevention practices.

LEARNING AGENDA

The McGovern-Dole Learning Agenda aims to answer questions related to school meal program implementation and education. The HATUTAN program-specific learning agenda also focuses on questions related to literacy, health, nutrition, agriculture, gender-based violence, and sustainability.

LITERACY

The implications of midline findings for the learning agenda are two-fold. First, HATUTAN program activities and the school feeding program may be relatively more effective at reducing the number of students with no letter recognition ability, rather than improving the letter recognition skills of students who already have some ability in this area. In this case, it seems that the HATUTAN program had more of an impact on more disadvantaged students—those who were more likely to be “left behind” entirely—than on students who faced fewer challenges to learning. One possible mechanism for this effect is that school feeding decreased dropout rates with little effect on attendance; as a result, students with low base levels of literacy may have been able to gain some letter recognition skills—the most basic literacy task—but may have been unable to progress beyond this level of understanding

given limited contact hours. However, given that school feeding was only provided from January through March, these results are also likely to be due to other changes, such as changes in teaching practices or access to learning materials.

In contrast, the significant improvement in word recognition and passage fluency among students with some literacy abilities but lack of relative improvement in the percent of students with no abilities to perform these tasks may have been driven by the feeding program's effects on attentiveness and memory. Among students with some word recognition ability, higher levels of attentiveness due to improved nutrition, greater exposure to improved teaching practices, or increased access to learning materials may have allowed for relatively improved learning outcomes in intervention schools for more difficult literacy skills. However, while students without any prior word recognition ability may still have benefitted from increased attentiveness, their low base skill level may not have allowed them to substantially improve their word recognition abilities, particularly while schools were closed.

The midline results point to several specific challenges to improved literacy outside of the broader impact of the COVID-19 pandemic and issues facing the education system of Timor-Leste. In general, students seem to recognize letters relatively well, but struggle with fluency. Students also struggle with the relationship between letter sounds and words, and may benefit from future activities that seek to improve the ability to decode words.

QUALITY OF EDUCATION

The predictive analysis suggests that whether a student has eaten is, indeed, a significant predictor of student attentiveness. However, in general, we find few significant increases in student attentiveness or decreases in student hunger at midline in intervention areas compared to comparison areas, suggesting that the program has not yet had a substantial impact on attentiveness. It is worth noting, however, that the COVID-19 pandemic has substantially impacted household food security and both the pandemic and other factors, such as the delayed approval of the national budget, have affected the provision of school meals across the country; results may thus be affected by these dynamics.

There appears to be little evidence of positive program impact on quality of instruction or teacher attendance, outside of a reduction in the use of ineffective traditional teaching practices. Overall, these results do not show a clear link between school feeding and teacher capacity. However, particularly among longer-tenured teachers with more ingrained teaching habits, teaching practices may be expected to change rather slowly, and only as a result of continued training, feedback from school directors or other relevant professionals, and positive classroom results (as evidenced by higher levels of engagement from students). Endline findings may thus be more suggestive of any links between school feeding and the use of engaging teaching practices.

SCHOOL MEAL PROGRAM IMPLEMENTATION

The McGovern-Dole learning agenda aims to understand the community-level systems of governance and management that are required to successfully implement school meal programs. This relates to the HATUTAN program-specific learning agenda questions about the kinds of partnerships and exit strategies that are most effective at ensuring program sustainability. Most treatment schools said PTAs oversee the feeding program, either exclusively or in collaboration with school staff. However, there are limitations to PTA activities related to school feeding program due to limited participation of parents and irregularity of meeting frequency.

Both the McGovern-Dole and HATUTAN program-specific learning agendas ask about the sustainability of meal program components, such as food production, local procurement, and food preparation. The number of schools with kitchen space and access to clean water for food preparation increased, and there were indications that schools are improving their facilities to ensure hygienic preparation of meals for students. However, in relation to local procurement of food, more than half (54%) of the respondents from treatment schools at the midline reported that their school does not

buy local food from farmers for school feeding. The primary reason given by most of the 47 treatment schools that have a school feeding program but do not purchase produce from local farmers was not having the budget to buy local produce (75%), followed by insufficient supply of produce for preparing daily student meals at school (6%), which leads to reliance on other sources.

The learning agenda aims to understand the impact of school feeding programs on student and community health. The study assesses community health using data on knowledge and awareness of healthy practices, hygienic practices (e.g., availability of handwashing stations at home and at the school kitchen), and food consumption (e.g., types of food consumed the previous day and breastfeeding practices). In relation to food consumption, most of the caregivers in the four intervention municipalities reported poor dietary diversity. Regarding health-related absences, 22% of students were reported by caregivers in the treatment group to have missed at least one day of school in the last week due to sickness.

Another area of focus of the McGovern-Dole learning agenda is on effective methods for food safety. Similarly, the HATUTAN program looks at water, sanitation, and hygiene (WASH) interventions. Effective school-level WASH interventions may contribute to ensuring the hygienic preparation and safety of food served to children. The school and household survey data imply that there had been improvements in level of hygiene maintained in school during food preparation. More schools reported having hygienic food storage spaces and kitchens at the midline compared to the baseline, an indication of increasing attention given to ensuring hygienic preparation of school meals. Moreover, most of the caregivers of children in treatment schools said that food served to children in school is prepared in a hygienic manner, which increased from the baseline. However, it is not clear to what extent parents are aware of or certain about the hygienic preparation of food in schools, especially given the low participation rate in PTAs and limited number of PTA meetings in the early part of the year.

HEALTH, NUTRITION, AND AGRICULTURE

In addition to the McGovern-Dole learning agenda questions, the midline study also aims to answer learning questions on how WASH projects impact learning and literacy outcomes. At the household level, the general trend observed is that there is an improvement in the percentage of caregivers that were able to identify healthy hygiene practices between baseline and midline in both the comparison and treatment groups. However, this is not a strong indicator of change in behavior but rather change in knowledge of healthy hygiene practices.

Regarding nutrition, despite COVID-related disruptions in program implementation and data collection, there is strong evidence that the school feeding program was effective in increasing school meals served by schools, and that this had a positive effect on nutritional outcomes for students. Despite further complicating factors in analyzing health-related absences, we also find a link between school feeding and student nutrition as explanatory variables with fewer absences due to illness as the outcome variable, although the statistical significance was slightly weaker. We can therefore cautiously conclude that school meals seem to be effective in reducing absences due to illness, likely due to improved nutritional status.

School meals have a relatively low level of dietary diversity, although intervention schools served a higher-quality menu than comparison schools. In part, this is because more than half of intervention schools did not purchase produce from local farmers, most of them due to the lack of funds to do so. The program should work to increase linkages between local farmers and schools and access to SFP funding to improve SFP sustainability. Currently, many program activities that seek to improve farmer-school linkages operate on the supply side by providing farmers with training to increase their production of healthy foods. However, an increase in supply will be ineffective if it is not accompanied by the corresponding means to purchase produce. Additional work with authorities, schools and PTAs may be necessary to encourage increased purchase of local produce.

Caregivers are highly knowledgeable about hygiene and nutrition practices, but these are often not reflected in actual behaviors due to multiple barriers.² Dietary diversity among women of childbearing age declined between baseline and midline in intervention municipalities, potentially linked to the economic crisis triggered by COVID-19, and caregivers reported consuming a predominantly carbohydrate-based diet with limited protein intake. Most households had access to a toilet at home but less than half of treatment households reported having access to drinking water all year at midline and handwashing with soap was not observed in 75% of the treatment households. These findings suggest that the program should strengthen activities that address gendered and economic barriers to health and nutrition. Knowledge of good health and nutrition practices is high, but household incomes appear to remain a substantial barrier to the implementation of these practices. Additionally, women report low levels of decision-making power over major household decisions, including large household purchases and the sell or consumption of livestock. This dynamic may influence nutrition and health outcomes for children and other family members, as women have been found to spend a greater portion of household incomes on children.

GENDER-BASED VIOLENCE

The HATUTAN program-specific learning agenda aims to understand how the synchronization of school meal programming with activities addressing sexual and gender-based violence (SGBV) and gender norms can affect learning outcomes and health practices. While the majority of the program's activities intended to address gender-based violence at community and school level had not yet been implemented at the time data collection had taken place, this study provides insight on current practices and how they may be affecting education outcomes. Analyzing violence within schools, at both baseline and midline, caregivers of male grade 2 children were significantly more likely to say that teachers use corporal punishment than caregivers of female children. Accordingly, at both baseline and midline and among both treatment and comparison groups, caregivers of both male and female children were more likely to say that corporal punishment was justified against boys than against girls. The use of corporal punishment and other negative discipline practices may affect student learning outcomes and decrease student attendance and motivation; future program activities should continue to address this issue, and should particularly focus on violence towards boys in schools, which appears to be more normalized than violence towards girls.

Notably, at midline, a higher percent of caregivers (15%) reported that they would not be able to report abuse of their child at school than at baseline (9%). However, only a very small percent of directors stated that abuse could not be reported. This suggests that there may remain barriers to accessing restorative justice for children abused at school, and that while directors may believe that there are effective avenues for reporting abuse and obtaining justice, caregivers do not necessarily agree or are not necessarily aware of these avenues for reporting; additionally, power dynamics between the school and community may prevent the use of reporting mechanisms.

Analyzing gender norms affecting children, we find that there were no significant changes in the gendered division of labor in treatment households at midline when compared to the changes seen in comparison households for either male or female students. There is still a strong gendered division of labor for children, with boys more likely to participate in agricultural activities and girls more likely to participate in household and caregiving activities. We do not find any significant correlations between the time spent on household tasks and literacy scores, attendance rates, or dropout rates, but note some limitations to this analysis, as time spent on tasks is estimated by caregivers who may face social desirability bias to underreport the time their children spend on chores.

Additionally, we find that there is a gendered gap in decision-making power over productive assets which has not changed significantly since baseline: Men often make the primary decisions that have

² Respondents were asked basic questions about hygiene and nutrition practices, but basic knowledge may not be an indicator of high-level knowledge or familiarity with hygiene and nutrition.

major implications for household finances or food security, and women are often limited to making decisions that have smaller financial implications. This can have substantial implications for children's learning outcomes and nutritional status, as studies have shown that, when given decision-making power over household spending, women tend to make purchases that benefit children more often than men.

PROGRAM IMPLICATIONS

The midline study was designed to inform the continued development and implementation of the HATUTAN workplan. Program implications fall under seven key areas: design/relevance, management and coordination, effectiveness, efficiency, sustainability, impact, and gender and power relations. Program impact and gender and power considerations are described above; here, we focus on the other five areas.

DESIGN/RELEVANCE

The HATUTAN workplan includes a variety of activities seeking to improve literacy and health. Activities that target education and schools include training school administrators, establishing and strengthening PTAs, training teachers, producing books and supplemental reading materials, providing school meals or take-home rations, and supporting extracurricular activities.

Midline data suggests that school directors are often responsible for school feeding programs—88% of respondents to the school survey stated that school directors were responsible for oversight of SFPs among all midline schools. Furthermore, in cases where a student is harassed at the toilets or a male teacher pays unwanted attention or harasses a girl, most respondents to the school survey reported that the school director was the first person to whom the student should report the abuse. As such, training of school administrators is particularly relevant to improve SFPs and to reduce the prevalence of SGBV and increase avenues for reporting abuse at schools.

PTAs are involved in many activities related to schools, especially in intervention areas. Fifty-five percent of the treatment school administrators and 25% of the comparison schools have also stated that the PTA is responsible for oversight of the school feeding program. More than half of PTAs reported being involved in improving school infrastructure, overseeing the SFP, monitoring safety and security, and monitoring student and teacher attendance. However, data also suggests that participation in PTAs remains relatively low in many areas, even during implementation of the SFP, which should correspond to a peak of parent participation and oversight. Activities that seek to strengthen PTAs may thus have particular relevance for improving school infrastructure, SFPs, and student and teacher attendance, and, potentially, for reducing the incidence of abuse or violence against students if PTA members have negative views of corporal punishment or have participated in trainings or other programs that seek to raise awareness of the negative effects of violence and abuse on students.

Teachers seem to face challenges to using child-centered teaching strategies, with only a limited increase in the use of engaging teaching practices among intervention schools. Data on literacy scores and subtask scores also suggests that current teaching practices may not be effectively imparting knowledge on phonemes and the relations between letters and words to young students, and that overall reading fluency remains low. Given these results, teacher training remains of high relevance to the program. It may, however, be useful to analyze the design of teacher training activities to make sure that trainings target best practices for improving student fluency and helping young students not only to recognize letters, but to read words.

Provision of reading materials remains a relevant program activity. At the midline, 78% of the treatment schools and 49% of the comparison schools reported that the school lent story books for students to take home and 40% of the treatment schools had students borrowing books the previous week (compared to 18% of the comparison schools). Furthermore, 65% of households at midline were

observed to have any children's books or magazines in the house. Among the treatment households which had reading materials for children at home, 96% were observed to have copies of Lafaek magazines at home, but only 17% had books; among comparison households with reading materials at home, 98% had Lafaek magazines, and 11% had books. The provision of reading materials may improve teaching practices, as teachers are more able to incorporate activities that use the reading corner or have students read alone or in groups; this may then further improve learning scores.

Results suggest that the provision of school meals may have helped mitigate the negative impact of COVID-19 on learning by increasing student attentiveness and decreasing dropout rates. The provision of school meals continues to be highly relevant for students in Timor-Leste, helping to improve nutrition and literacy—both of which remain at relatively low levels throughout the country.

The HATUTAN program intended to incorporate reading and health-focused extracurricular activities and games to increase contact time with students in a fun and cooperative environment and to develop students' self-confidence and voice, particularly for girls. Unfortunately, due primarily to restrictions due to the COVID-19 pandemic, HATUTAN had limited opportunities to work with teachers to initiate extracurricular activities. Given the low number of contact hours in most schools over the past year, extracurricular activities may be particularly useful to help students catch up after successive school closures.

Other activities that supplement these school-related program activities include partnering with farmers' groups; forming VSLAs; training for good health, hygiene, and nutrition practices; and undertaking capacity building and advocacy.

Schools need to have a reliable budget for local purchases and build better linkages with farmers. Ensuring consistent and sufficient supply of goods for the school feeding program will help in serving well-balanced, nutritious meals to students.

Because VSLAs are the foundation for other trainings, their continued establishment and support is of strong relevance to program activities. However, it is worth noting that households that are not involved in VSLAs may also not be able to receive trainings if this is the primary mechanism by which training participants are recruited. As such, it may be worth exploring other mechanisms to recruit households for trainings outside of VSLA participation or expanding access to VSLAs among parents of schoolchildren.

The training on health, hygiene, and nutrition is relevant to understanding the importance of related practices to ensure that schools meals are prepared in a hygienic manner to prevent absences due to sickness, that storage spaces are clean and secure from pests, and that well-balanced and nutritious meals are served to children. These practices are also relevant to the school feeding program, which contributes to improving student attendance, attentiveness, and performance. Moreover, hygienic practices are promoted to children in school with the availability of handwashing stations and toilets. However, given that knowledge of health, hygiene, and nutrition appears fairly high, it may be worth pivoting program activities to focus on behavior change, rather than knowledge.

With the onset of the COVID-19 pandemic, the HATUTAN program increased its focus on capacity building and advocacy at the national level for COVID-19 prevention and awareness. The program also included activities focused on advocating for policy changes related to the national SFP and education policies, including strengthening the government of Timor-Leste's capacity to deliver the SFP. After delays due to the COVID-19 pandemic, national budget issues further delayed provision of school feeding funds to schools. This remains a highly relevant issue for school feeding across Timor-Leste; program activities should continue to advocate for flexible funding for the SFP and build capacity for SFP implementation.

MANAGEMENT, COORDINATION, EFFECTIVENESS, AND EFFICIENCY

Overall, program effectiveness has been somewhat undercut by the COVID-19 pandemic, which has had substantial negative effects on education outcomes, nutrition, the provision of school feeding, and

health, among other areas. However, other factors are also crucial to ensure program effectiveness, including school infrastructure, PTAs, and school administrators.

While the HATUTAN program aimed to address infrastructure issues by building the capacity of school administrators to apply for infrastructure-related grants or funding, infrastructure remains a constraint to the establishment of effective and hygienic SFPs in all intervention schools. PTA involvement improved in some respects at midline, but still needs further improvement to bolster program effectiveness. Similar to the role of the PTAs, school administrators were also identified as crucial for program effectiveness due to many activities' reliance on the ability and motivation of administrators for implementation. The effectiveness of program activities through the end of the program will rely on continued commitment from school administrators; as such, HATUTAN program staff should continue to pay close attention to administrators' involvement.

Midline data also revealed several other constraints to program effectiveness. First, the wide range of students' literacy abilities within classrooms presents a challenging teaching environment; literacy results on each subtask suggest that currently, teaching practices are not effective at improving the literacy skills on both of these groups of students at the same time. To improve program effectiveness, it may therefore be useful to provide teacher training on effective ways to engage all students in classes and teach to a wide range of skill levels and literacy abilities.

A second constraint relates to activities seeking to increase access to reading materials at school and at home. While the number of grade 2 classrooms with reading materials increased at midline, it is not clear that this always translates into increased levels of access to reading materials for students. Many schools did not report lending students books at midline, often because students were viewed as careless or because it was believed that the books would be damaged. This suggests that mistrust of students or beliefs that books are "too nice" for student use might reduce the effectiveness of interventions seeking to increase access to reading materials.

SUSTAINABILITY

In order to ensure that the benefits of HATUTAN program activities continues after conclusion of the program, sustainability must be at the forefront of all program-related decisions. The HATUTAN work plan focuses on sustainability through advocacy and capacity-building activities at the local, regional, and national levels. Additionally, the program intends to improve sustainability by training school administrators to better manage schools and apply for funding, training school cooks to provide more nutritious school meals, and training teachers to encourage the use of better teaching practices. The establishment of VSLAs and use of VSLAs as a platform to provide trainings to community members may also help improve sustainability.

The HATUTAN program incorporates training within most of its activities: Ten of twelve activities rely on training as a critical element of the work. The program has also partnered with the government of Timor-Leste, international development partners, farmers' groups, community development agencies, and PTAs in implementation of various activities. At the local level, the HATUTAN program seeks to enhance community participation in monitoring school feeding and education outcomes. The involvement of community members, teachers, and parents is critical for the sustainability of program activities. Top-down policy implementation that only involves, for example, central school directors or government officials, is less likely to succeed as there may be less buy-in and more resistance from those responsible for carrying out the policy at the local level.

CONCLUSION AND RECOMMENDATIONS

The COVID-19 pandemic had a substantially negative impact on literacy scores, but the HATUTAN program seems to have mitigated some of this effect. However, there remains a major gap in literacy ability between letter recognition and word recognition, and between word recognition and reading comprehension. Furthermore, teaching quality appears to have changed little at midline as a result of

program activities. As a result of these dynamics, it is recommended that the program increases activities seeking to improve the quality of instruction. The above results suggest that teachers still face challenges to effectively teaching literacy skills, particularly to young students. As at baseline, teachers appear to teach literacy by focusing first on letter recognition before moving on to words. Effective literacy development, in contrast, occurs from more well-rounded instruction that includes concurrent focus on sounds, vocabulary development, and comprehension. It also necessitates engaging teaching practices that increase student interest in the content. Program activities that seek to strengthen both the use of engaging teaching practices and pedagogical strategies may help to improve student literacy.

School attendance did not improve in intervention areas relative to comparison areas. However, dropout rates in intervention municipalities decreased significantly compared to comparison municipalities. Additionally, there is an enormous difference between attendance rates in intervention schools that provided meals (72%) and those that did not (54%). This effect on dropout rates and attendance rates among schools that provided meals suggests two issues with comparing aggregate attendance rates across all intervention and comparison areas: First, some schools started school feeding late, reducing its impact on attendance, and second, other barriers to reach school (such as strong rains washing away bridges and affecting students' ability to attend, which have a disproportional impact on the most remote areas, where students often have to cross rivers to reach the school) may have reduced student attendance particularly in intervention areas. Furthermore, students in intervention municipalities also appeared to have somewhat higher attentiveness than expected given changes in comparison municipalities, suggesting further impact of school feeding. It may be useful to further examine ways to improve student health and attendance by increasing the regularity or quality of school meals, among other interventions.

School meals have a relatively low level of dietary diversity, although intervention schools served a higher-quality menu than treatment schools. In part, this is because more than half of intervention schools did not purchase produce from local farmers, mostly due to the lack of budget to do so. The program should work to increase linkages between local farmers and schools to improve SFP sustainability. Currently, many program activities that seek to improve farmer-school linkages operate on the supply side by providing farmers with training to increase their production of healthy foods. However, an increase in supply will be ineffective if it is not accompanied by a corresponding increase in demand. Additional work with schools and PTAs may be necessary to encourage increased purchase of local produce, in parallel with advocacy for timely release of SFP funds to schools to enable local purchases.

Caregivers are highly knowledgeable about hygiene and nutrition practices. However, dietary diversity among women of childbearing age declined between baseline and midline in intervention municipalities, and caregivers reported consuming a predominantly carbohydrate-based diet with limited protein intake. Most households had access to a toilet at home but less than half of treatment households reported having access to drinking water all year at midline. These findings suggest that the program should strengthen activities that address gendered and economic barriers to health and nutrition. Knowledge of good health and nutrition practices is high, but household incomes appear to remain a substantial barrier to the implementation of these practices. Additionally, women report low levels of decision-making power over major household decisions, including large household purchases and the sell or consumption of livestock. This dynamic may influence nutrition and health outcomes for children and other family members, as women have been found to spend a greater portion of household incomes on children.

INTRODUCTION

HATUTAN PROGRAM OVERVIEW

The HATUTAN program (*Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun* - Food for the Improvement of Nutrition and Education), funded through the Foreign Agricultural Service of the United States Department of Agriculture (USDA) under the McGovern-Dole International Food for Education and Child Nutrition Program and implemented by a consortium including CARE, Mercy Corps, and WaterAid, was officially launched to improve education, nutrition, health, hygiene/sanitation, economic empowerment, and gender equality in 449 schools and communities in the municipalities of Ermera, Liquiçá, Ainaro and Manatuto. The program aims to build a partnership between schools and their communities to improve literacy, learning, health, and nutrition for children and adults. The program works in partnership with the Government of Timor-Leste, including the Ministry of Education, Youth, and Sports (MEYS) and the Ministries of Health, State Administration, and Agriculture and Fisheries, and development stakeholders to achieve two key strategic objectives: (1) improved literacy of school-aged children, and (2) increased use of health, nutrition, and dietary practices.

To achieve these objectives, the program supports, among a variety of activities, the Government of Timor-Leste's school feeding program (SFP) to fully operate in all basic education and preschools throughout the school year. Key project activities include strengthening and supplementing the government-sponsored SFP and building school capacity through trainings for teachers and administrators and provision of resource materials. Additionally, the HATUTAN program seeks to support farmers to boost the production of local produce to increase yields and help create sustainable sources of nutritious food for local schools. In addition to activities related to literacy and SFPs, HATUTAN seeks to conduct trainings related to nutrition, health, and other topics, and to promote gender equality and the reduction of gender-based violence.

In light of these issues, the program focuses its interventions in four key areas:

1. Increasing the capacity of government agencies, school administrations, and community-based organizations (such as PTAs, village savings and lending associations (VSLAs), etc.) to better manage, fund, and monitor a comprehensive school feeding program and support nutrition, health, and hygiene improvements in homes and schools.
2. Improving tools, techniques, and learning environments to increase literacy skills.
3. Overcoming social norms to increase gender equality, reduce sexual and gender-based violence, ensure equal learning opportunities for girls, and improve nutrition and WASH practices through targeted social behavior change communications.
4. Increasing food production and income-generating activities through farmer trainings, establishing VSLAs, and enabling community development agents to profitably provide agriculture inputs and technical services.

The program's theory of change argues that by providing schools meals, teacher training, and related support, school enrollment and academic performance will improve. This effect will be amplified and sustained by improving children's health and learning capacity before they enter school by offering nutrition support programs for pregnant and nursing women, infants, and preschoolers and by addressing issues of gender dynamics and gender-based violence.

The program's four target municipalities—Ainaro, Ermera, Liquiçá, and Manatuto—were selected due to having the worst education and health indicators in the country. The program aims to reach an estimated 462,806 target beneficiaries, including 368,548 school-aged children, 1,351 teachers, 502 school administrators, around 2,200 PTA members at 220 PTAs, 280 VSLAs, 48 community development agents, and 4200 farmers. In total, HATUTAN operates in 449 schools, which include

every primary and preschool in the four target municipalities, with the exception of a small number that either opted out of participation or have closed. In addition to these localized activities, HATUTAN also has a national-level advocacy component to address barriers to SFP implementation and improved education outcomes.

HATUTAN provides two packages of support in target areas: “partial support” and “full support.” Partial support includes provision of commodities for school feeding (oil, rice, and beans) between January and March and copies of supplementary literacy materials (including magazines and exercise books for early grade readers) and encompasses all the pre-school and basic education schools in the four municipalities. Full support, implemented in 219 schools and their surrounding communities, representing about half of the preschools and primary schools in the target municipalities. The support includes provision of commodities for school feeding (oil, rice, and beans) between January and March; provision of literacy materials (storybooks and educational magazines); coaching of headmasters and teachers; mobilization and training of PTAs; implementation of the school dialogue and improvement plan (Community Scorecard); support for extracurricular activities; and training of parents on VSLAs, agriculture, health, WASH, and gender. The 219 communities and schools were selected for full support based on location in rural and remote areas, and include 173 primary schools in vulnerable conditions and 46 preschools. Importantly, we note that the evaluation assesses only areas in which the full support package was provided.

TIMOR-LESTE OVERVIEW

After more than three centuries of colonial rule by Portugal followed by decades of Indonesian occupation, Timor-Leste gained independence in 2002, making it the world’s second-youngest sovereign state. The transition to independence was marked by widespread violence committed by militias supported by the Indonesian military, which killed around 1,300 Timorese, displaced around 500,000,³ and destroyed approximately 70% of the country’s infrastructure.⁴ Timor-Leste has made considerable progress since independence, building stability and democratic institutions and rebuilding infrastructure. However, the country continues to face many challenges: Around 42% of Timor-Leste’s population of 1.3 million live below the national poverty line,⁵ and the country’s score on the Human Capital Index, which measures key indicators of health and education, is below the average for both the East Asia and Pacific region and for other lower-middle income countries.⁶ Demographically, Timor-Leste has a young population: Over 50% of its population is younger than 24, and 20% of the population is between the ages of 15 and 24.⁷ This poses a substantial challenge—and opportunity—for the country to ensure that youth have sufficient access to education and job opportunities.

Timor-Leste has established itself as a stable democracy since gaining independence, holding free and fair elections with high voter participation in 2012, 2017, and 2018. The country has also made substantial progress towards upholding the rule of law, enacting and overseeing a nationwide legal framework, and protecting human rights. However, governing institutions remain weak, and more progress is needed to strengthen the judiciary, improve access to justice and the efficacy of local governance, support gender equality, establish a strong and independent media, and continue to

³ “Timor-Leste: Background and U.S. Relations,” *Congressional Research Service*, June 27, 2019, <https://fas.org/spp/crs/row/IF10320.pdf>.

⁴ Nicole Stout, “Infrastructure in Timor-Leste Growing According to Strategic Plan,” *The Borgen Project*, February 23, 2018, <https://borgenproject.org/infrastructure-in-timor-leste>.

⁵ As of 2014; the poverty rate declined from 50% in 2007. “The World Bank in Timor-Leste,” *World Bank*, April 28, 2020, <https://www.worldbank.org/en/country/timor-leste/overview>.

⁶ “Timor-Leste,” *World Bank Human Capital Index 2020*, October 2020, <https://www.worldbank.org/en/publication/human-capital>.

⁷ Jessica Gardner, *Timor-Leste Population and Housing Census 2015: Thematic Report Volume 14: Analytical Report on Youth* (Dili, Timor-Leste: Timor Leste General Directorate of Statistics and United Nation Population Fund, 2018).

uphold human rights.⁸ Governance and state-building have, in general, been highly centralized since independence,⁹ but the constitution of Timor-Leste includes explicit provisions for administrative decentralization in order to “avoid bureaucratization, bring services closer to the population and ensure the participation of those interested in their effective management.” As a result, Timor-Leste has adopted a variety of policies in recent years to decentralize governance to the country’s 13 municipalities, which are further subdivided into administrative posts. However, institutional capacity at the subnational (and national) level remains low, and the main source of improved government services still comes from national-level investment.¹⁰

The COVID-19 pandemic has severely affected Timor-Leste’s economy, with GDP expected to contract by 6.8% in 2020, the largest fall since independence. This large decline in economic activity would represent the country’s third recession in four years. The economic impacts of COVID-19 in Timor-Leste come primarily from indirect costs due to public health measures and voluntary changes in behavior that have lowered both demand for and supply of goods and services, as well as from external conditions such as declines in the prices of petroleum, rather than from direct costs due to mortality and illness-driven absences from work, which have been relatively low to date.¹¹ The dependence of the Timorese economy on oil and gas exports makes the country particularly vulnerable to fluctuations in gas prices, such as those that occurred during the pandemic: The oil and gas sector contributes 36% of the country’s total GDP, more than 90% of government revenue, and 98% of exports. Revenue from oil and gas is deposited into the country’s Petroleum Fund, from which a limited amount of money can be withdrawn annually to fund government projects, such as investment in infrastructure and human capital.¹²

COVID-19 IMPACT AND RESPONSE

The first case of COVID-19 in Timor-Leste was reported on March 21, 2020, and the government declared a state of emergency on March 28, enacting public health measures—including restrictions on international travel; school closures; restrictions on gatherings, the use of public transport, and businesses; and hygiene measures—to reduce the spread of the virus.¹³ These measures, and the speed with which they were enacted, helped reduce transmission of COVID-19, giving Timor-Leste one of the lowest COVID-19 incidence rates in the world, at around 77 cases per million inhabitants (compared to a world average of around 14,000 cases per million population), until March 2021, when a surge in cases resulted in new lockdowns.¹⁴

While the country was initially successful at preventing the spread of COVID, data from small-scale surveys suggests that public health measures and voluntary changes in behavior that led to reductions in economic activity had a substantial impact on personal incomes, employment, food security, education, and use of health services. As a result, the World Bank estimates that the pandemic increased the poverty rate in Timor-Leste by 5 to 7 percentage points. Poverty is expected to increase more in rural areas and areas that rely on the tourism or petroleum sectors for employment.¹⁵

To address the impact of COVID-19 on poverty, the government of Timor-Leste has spent more than \$120 million to finance preventative health expenditures and economic relief measures through a COVID-19 fund created in April 2020. Most of this money was spent on cash transfers to households,

⁸ “Timor-Leste: Democratic Governance,” *United Nations Development Program*, accessed January 27, 2020.

⁹ M. Anne Brown, “State Formation and Political Community in Timor-Leste – The Centrality of the Local,” *RCCS Annual Review* 7, no. 7 (2015).

¹⁰ Terry Russell, “Decentralization and rural development in Timor-Leste,” *East Asia Forum*, April 3, 2015, <https://www.eastasiaforum.org/2015/04/03/decentralisation-and-rural-development-in-timor-leste>.

¹¹ World Bank, *October 2020 Timor-Leste Economic Report: Towards a Sustained Recovery* (Washington, D.C.: World Bank, 2020).

¹² “Timor-Leste,” *Extractive Industries Transparency Initiative*, February 10, 2021, <https://eiti.org/timorleste>.

¹³ World Bank, *October 2020 Timor-Leste Economic Report*.

¹⁴ “COVID-19 Coronavirus Pandemic,” *Worldometer*, February 23, 2021, <https://www.worldometers.info/coronavirus>.

¹⁵ World Bank, *October 2020 Timor-Leste Economic Report*.

such as a two-month \$100 cash transfer to households in which every member earned less than \$500 per month. The government also provided wage subsidies for firms and electricity and water credits for public utility customers, and purchased a three-month emergency supply of rice to shore up food stocks. Surveys suggest that the economic measures had a positive impact, and that the cash transfer program was generally successful in bolstering household incomes; however, these measures also faced implementation constraints which reduced their impact and efficacy, such as a lack of comprehensive administrative records.¹⁶

A complete economic recovery will depend on national and worldwide management or elimination of the virus through vaccination. On February 15, 2021, the government of Timor-Leste approved a national vaccination plan against COVID-19. As part of this vaccination plan, Timor-Leste has joined the COVAX facility, which promises free access to the COVID-19 vaccine for 20% of the Timorese population. The process for acquisition of vaccines for the remaining 80% of the population is still under discussion. Vaccination is planned to occur in three phases: The first will include essential workers, people residing near the land border with Indonesia, and those with preexisting conditions; the second will cover the elderly and critical but non-essential workers, such as teachers and market workers; and the third will cover the remainder of the population.¹⁷ The vaccination plan has been accelerated with the recent arrival of 20,000 doses from Australia and 100,000 SINOVAR doses from China; Australia also plans to send as many as 6 million vaccine doses to the Pacific region and Timor-Leste in 2021. However, full rollout of this plan still relies on timely and sufficient distribution of vaccines through COVAX, but the facility has, to date, struggled to purchase and distribute vaccinations as high-income countries have focused on securing their own vaccine supplies rather than contributing to COVAX as planned. As of May 3, 2021, 28,575 doses had been administered in country.¹⁸

EDUCATION

The education system in Timor-Leste consists of four levels: pre-school, primary education, secondary education, and higher education (university and polytechnic). Primary and secondary schooling comprise “basic education,” which is universal, free, and compulsory according to the National Education Strategic Plan. Basic education is divided into three cycles: grades 1-4, grades 5-6, and grades 7-9. The school system includes basic schools, which provide education for all three cycles, and filial schools, which offer first and second cycle grades (and, in limited cases, only early grades). Filial schools are generally located in remote and rural areas, and are directly associated with a nearby central basic school. Each cluster of central and filial schools is managed by a school director; a school council, consisting of representatives of schools, parents, and local authorities; and an academic council, consisting of teaching staff representatives of all schools. The school council is responsible for the achievement of educational targets and development of strategic education plans, and for encouraging the establishment of parent-teacher associations in all schools. The academic council is responsible for strengthening curriculums, providing pedagogical support and training, and improving teacher performance and professionalization.¹⁹

The government has made a strong commitment to education, pushing for universal enrollment in basic education and committing around 10% of the annual national budget to expenditure related to education.²⁰ Government expenditure on education has included investments in infrastructure, which increased the number of preschools, primary schools, and secondary schools from 943 in 2002 to

¹⁶ Ibid.

¹⁷ “Government approves national vaccination plan against COVID-19,” *Government of Timor-Leste*, February 15, 2021, <http://timor-leste.gov.tl/?p=26919&lang=en&n=1>.

¹⁸ “COVID-19: Timor-Leste,” *World Health Organization*, accessed May 11, 2021, <https://covid19.who.int/region/searo/country/tl>.

¹⁹ Timor-Leste Ministry of Education, *National Education Strategic Plan 2011-2030* (Dili, Timor-Leste: Ministry of Education, 2011).

²⁰ World Bank, *Timor-Leste Basic Education Strengthening and Transformation* (Washington, D.C.: World Bank, 2020).

1,715 in 2017;²¹ teacher training; curriculum design; and operational decentralization, in order to improve support for remote and rural areas. In accordance with this investment, participation in education has increased in recent years, with the number of out-of-school adolescents declining from more than 20,000 in 2010 to around 9,500 in 2019.²²

PRIORITIES FOR EDUCATION SYSTEM

The 2002 Constitution of Timor-Leste established that the state “will do everything within its means to help education, health, and vocational training for youth” and states that “the state recognizes and guarantees the right to education for all citizens.”²³ In line with this commitment, Timor-Leste’s National Education Strategic Plan 2011-2030 established three key priorities for education: achieving universal completion of basic education by 2030, eliminating illiteracy (particularly among youth ages 15-24), and achieving gender parity by 2015 (including by increasing the number of female teachers and administrators). Pursuit of these priorities is guided by seven general goals: quality, equity, access, social and economic relevance, co-participation (in which families participate in education management and decision-making), social partnership, and flexibility.²⁴

EDUCATION OUTCOMES

Despite improvements to enrollment and infrastructure, education outcomes remain relatively poor for most of the country. Student learning, as measured through standardized tests including the Early Grade Reading Assessment (EGRA) and a curriculum-based assessment, is low. Results from the EGRA administered in 2017 showed that 15.5% of grade 1 students were not able to identify a single letter and 54.5% were not able to read a single word. Among students who had completed grade 1 who were able to read, the mean reading fluency score was only 6 words per minute, and the mean reading comprehension score was only 14.4%.²⁵ Notably, this is an improvement from the results of an EGRA administered in 2011, in which 27% of grade 1 students were not able to identify a single letter and 64% were not able to read a single word. The 2011 EGRA showed improvement at higher grade levels—as would be expected—but with still low overall results: 12% of grade 2 students and 8% of grade 3 students could not identify a single letter, 28% of grade 2 students and 7% of grade 3 students could not read a single word, and the average reading comprehension scores for students who scored above zero were 41% in grade 2 and 67% in grade 3.²⁶ Similarly, the curriculum-based assessment (CBA) showed that less than 50% of students in grades 1 and 2 achieved the competencies outlined in the curriculum, including competencies in both literacy and math.²⁷

In accordance with these poor educational outcomes, repetition rates are high for students in primary school, at an average of 12.5% for all grades. Repetition rates are highest for grade 1 students, at 24%, and lowest for grade 6 students, at 5%.²⁸ Additionally, in 2018, on average, around 5% of students dropped out of each of grades 1 through 5, grade 7, and grade 8, suggesting further challenges to student retention. Dropout rates were highest in Liquiçá, Bobonaro, Covalima, and Aileu municipalities, and lowest in Dili and Lautem municipalities.²⁹

²¹ “Número de escolas e de alunos em Timor-Leste quase duplicou nos últimos 15 anos – PM,” *Diário de Notícias*, May 15, 2017, <https://www.dn.pt/lusa/numero-de-escolas-e-de-alunos-em-timor-leste-quase-duplicou-nos-ultimos-15-anos---pm-8476453.html>.

²² “Timor-Leste: Education and Literacy,” *UNESCO Institute for Statistics*, accessed February 23, 2020, <http://uis.unesco.org/en/country/tl>.

²³ World Bank, *Timor-Leste Basic Education*.

²⁴ Timor-Leste Ministry of Education, *National Education Strategic Plan*.

²⁵ Tazeen Fasih, Stephen L. Walter, Karla J. Smith, Pedro Ximenes, and Adelaide Camões, *Using EGRA for an Early Evaluation of Two Innovations in Basic Education in Timor-Leste* (Washington, D.C.: World Bank, 2019).

²⁶ Steph de Silva and Luc Gacougnolle, *The Timor Leste 2011 EGRA: Tetum Pilot Results* (Washington, D.C.: World Bank, 2011).

²⁷ World Bank, *Timor-Leste Basic Education*.

²⁸ Ibid.

²⁹ Timor-Leste Ministry of Education, “Statistical Data: Drop-Out Rate by Grade,” accessed February 23, 2020, <http://www.moe.gov.tl/pt/emis/dados-estatistico>.

Teacher training and the quality of instruction remain central issues for education outcomes in Timor-Leste. There are between 11,000 and 12,000 teachers working in the country; most of these teachers have university degrees or an equivalent qualification, but some have only secondary education. Teacher certifications range from full teacher training qualifications to bacharelato (government-provided in-service course towards the minimum teaching qualification).³⁰ As a result, some Timorese teachers have weak pedagogical skills and require further professional development, and teachers are often insufficiently prepared to teach in challenging contexts, or to support students with diverse skill levels. Further challenges come from the linguistic diversity of Timor-Leste: 32 languages were identified within the country in the 2015 census, and students whose mother tongue is not Tetum—the language of instruction in the first four years of basic education, before instruction transitions to Portuguese in upper grades—are at a disadvantage. In general, teachers are often unprepared to facilitate the transition of non-Tetum speaking students to a classroom where Tetum is the language of instruction. Particularly in rural areas with low population densities and small school sizes, teachers may also be required to teach multigrade classes, presenting a further challenge.³¹

In addition to instruction challenges, many students face low levels of access to learning resources and insufficient infrastructure. While initiatives by the Ministry of Education, Youth, and Sports (MEYS) have substantially improved access to teaching and learning materials, including textbooks and workbooks in line with the current curriculum, there is evidence that available resources are often not used by students or teachers, either because teachers prefer not to teach using the current curriculum or because it is believed that students will damage the materials. Furthermore, while many schools have libraries or book corners, these often lack age-appropriate reading materials.³² In basic education schools, average class sizes can also be high; average class sizes in 2019 ranged from 20 students per class in Covalima to 36 students per class in Dili. Average class sizes increase dramatically in secondary schools in all municipalities, to as many as 87 students per class in Dili.³³

There are major disparities in education outcomes across rural and urban areas and different municipalities. Additionally, girls consistently outscore boys on standardized tests including the EGRA and CBAs for both math and language learning, and have lower dropout and repetition rates than boys.³⁴ Other key factors affecting reading fluency and comprehension at the student level include the availability of printed materials at home, whether a student reads with family members, whether Tetum is spoken at home, and the frequency of student absences. At the school level, school feeding programs and in-service training for teachers were also found to have a positive effect on learning outcomes.³⁵

SCHOOL FEEDING PROGRAM

The government of Timor-Leste established a national school feeding program (SFP) for all basic education students to improve school attendance, address nutritional needs, improve student attention and performance, and boost the local economy through linkages between schools and farmers. The school feeding program was managed by World Food Program in six municipalities and by the government in seven municipalities until 2009, when the two programs were merged.³⁶ The government has fully managed the program since 2011. In addition to government support, SFPs rely on PTAs for regular monitoring and selection of cooks.

³⁰ Fasih et al., *Using EGRA for an Early Evaluation*.

³¹ World Bank, *Timor-Leste Basic Education*.

³² Ibid.

³³ Timor-Leste Ministry of Education, “Statistical Data: Average Class Size,” accessed February 23, 2020, <http://www.moe.gov.tl/pt/emis/dados-estatistico>.

³⁴ Fasih et al., *Using EGRA for an Early Evaluation*.

³⁵ de Silva and Gacougnolle, *The Timor Leste 2011 EGRA*.

³⁶ Stephen Lister, Jane Keylock, and Trish Silkin, *Timor Leste: An evaluation of WFP's portfolio (2008-2012)* (Rome: World Food Program, 2013).

The government-led SFP provided unfortified rice and equipment as well as 25 cents per child per day in budgetary support for schools to buy local produce to supplement meals, although budget transfers have been consistently late over the past years. However, observations suggest that many schools do not buy produce from local farmers daily due to both budgetary constraints and procurement choices of cooks and school administrators, specifically including delayed transfers of funding, limited and seasonal local production, a lack of linkages with farmers, and insufficient budget to purchase more expensive local products. The most common reason that schools do not buy local produce is lack of budget; the resulting irregularity of purchases means that farmers do not see SFPs as a reliable market with which to build a business relationship, thus leading to lack of linkages. As a result, the food composition of school meals is frequently poor, with only 35% of basic education school meeting a recommended composition of two vitamins, one protein, and one carbohydrate as of 2019.³⁷

The SFP suffers from systematic challenges that limit the availability of funds and reduce the number of actual school feeding days. In 2017, school meals were only delivered on about one-third of school days, and in 2018, on substantially less than one-third of school days due to delayed budget approval; in 2019, delivery improved, but was only funded for about 43% of total school days. In general, school feeding often does not occur during the first school trimester due to regular delays in approval of the annual national budget; this is particularly problematic as this trimester occurs during the most food-insecure time of the year. School feeding is also often interrupted during the school year because of delays in the reporting system. Despite these challenges, the program receives widespread support from parents and students, and has had positive effects on absenteeism and student attention.³⁸

IMPACT OF COVID-19 ON EDUCATION SYSTEM

After confirming its first positive case of COVID-19 on March 21, 2020, the government of Timor-Leste closed schools on March 23 and quickly implemented a remote learning program called “School Goes Home” (*Eskola Ba Uma*). The program was established in a challenging environment for remote learning, with limited internet and television access and few traditions of schooling at home or parents helping children to learn. On May 29, the MEYS established a consultative commission to coordinate the education response to COVID-19, and the government issued guidelines for reopening schools on June 6, but most schools did not complete the requirements for reopening until July. As a result, most students were out of school from March until July, participating only in home-based learning programs.³⁹ Qualitative data also suggests that teachers and students both felt unmotivated to return to the classroom after lockdowns due to a variety of factors, including the psychological impact of new habits (for teachers, not working; and for students, playing rather than studying), fear of the pandemic, and issues adapting classrooms and teaching practices to physical distancing requirements. This combination of school closures and loss of motivation led to enormous learning losses, as observed in the present study.

Upon reopening, the MEYS instructed schools to divide any classes with more than 25 students into shifts, with students either attending school for 2.5 hours per day (instead of five) or attending school every other day for the full five hours. A survey conducted by the CARE HATUTAN team prior to the midline evaluation found that 30% of grade 1 classes and 27% of grade 2 classes were operating in shifts. The survey also found that most schools were only providing students in shifts with two hours of instruction per day. Correspondingly, the survey found that among 170 basic education schools receiving a full package of HATUTAN interventions, in 65% of schools, students had only received half or less of normal class hours. In addition, teacher training courses for contract teachers also resumed after the end of lockdown, taking these teachers out of school on Thursdays, Fridays, and Saturdays and further reducing contact hours. In response to these issues, the MEYS instructed

³⁷ CARE and Julie Imron, *School Feeding Program Study Report: Timor-Leste* (Dili: CARE, 2019).

³⁸ Ibid.

³⁹ CARE, *Timor-Leste HATUTAN Project, Semi-Annual Report FY2020 (April – September 2020)* (Atlanta: CARE, 2020).

schools to hold remedial classes on Saturdays; the CARE survey found that 23% of schools were conducting remedial classes as of October 2020.⁴⁰

In the face of school closings, no nationwide measures were taken to compensate for the absence of the school feeding program. HATUTAN provided take-home rations to students in 416 schools in May 2020, using an existing commodity balance due to the early interruption of in-person classes, although the amount provided was limited, particularly considering the prolonged period of school closure. A food security assessment conducted in May found that household food security had been impacted by COVID-19, with 81% of households reporting that COVID restrictions had affected their food and income sources and more than 40% of households reporting engaging in coping strategies, such as limiting the amount of food that they eat.⁴¹ Once schools reopened, there were also substantial challenges to resuming the SFP. Due to COVID-19 and unrelated political tensions, the national budget was not approved until October, delaying the release of SFP funds to schools. As a result, the SFP was not operational in most municipalities for most of the 2020 school year. Survey data from the HATUTAN team suggests that as of October 2020, most schools had not yet received funds to purchase food for student meals since reopening; furthermore, only 16% of the schools provided meals during the day of the survey visit. The delays in school feeding represent a major missed opportunity to attract students back to school, motivate students, and reduce the food security burden on households. The lack of school meals, the COVID-induced lockdown, and financial hardships faced by households have contributed to low student attendance rates. However, schools where the school feeding program continued had significantly higher attendance rates than those not providing meals.⁴²

HEALTH, NUTRITION, AND SANITATION

Timor-Leste has made substantial progress towards improving health outcomes and building its healthcare system since independence, when over 75% of health facilities were damaged and many health professionals left the country. Life expectancy has increased by around 10 years, to 70.⁴³ Infant mortality has declined from 60 deaths per 1,000 live births in 2003 to 30 deaths per 1,000 live births in 2016; under-five mortality similarly declined by 2016 to about half the 2003 rate, and maternal mortality declined by more than half over a similar time period, to 218 deaths per 100,000 live births. Nearly half of children aged 12-23 months have received all basic vaccinations,⁴⁴ and in 2018, Timor-Leste was declared free of measles. The country is also on track to eliminate malaria after aggressive use of indoor residual spraying and insecticide-treated mosquito nets for more than a decade.⁴⁵ These improvements have been underpinned by a steadily increasing number of doctors and other health care professionals in the country and by increasing government health expenditure as a percent of GDP.

However, coverage of essential health services is uneven, and health service utilization is low—one reason why few households report spending savings on medical expenses. Rural and poor households receive, on average, poorer quality healthcare than urban or wealthier households.⁴⁶ Additionally, Timor-Leste has one of the highest tuberculosis incidence rates in the world, and the incidence of non-communicable diseases has risen; these diseases now account for 62% of all deaths in the country.⁴⁷

⁴⁰ Ibid.

⁴¹ Food and Agriculture Organization of the United Nations (FAO), *National agrifood systems and COVID-19 in Timor-Leste Effects, policy responses, and long-term implications* (Rome: FAO, 2020).

⁴² CARE, *Timor-Leste HATUTAN Program*.

⁴³ Sophie Cousins, “Health in Timor-Leste: 20 years of change,” *The Lancet World Report* 394 (2019): 2217-8.

⁴⁴ General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016* (Dili, Timor-Leste: General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, 2016).

⁴⁵ Cousins, “Health in Timor-Leste”

⁴⁶ World Bank, *Timor-Leste COVID-19 Emergency Support Project: Project Information Document* (Washington, D.C.: World Bank, 2020).

⁴⁷ Cousins, “Health in Timor-Leste”

Malnutrition also remains a severe problem in the country. The 2016 Timor-Leste Demographic and Health Survey found that 46% of children under 5 were stunted, or too short for their age, an indication of chronic undernutrition; 24% of children under age 5 were wasted, or too thin for their height, an indication of acute malnutrition; and 40% of children under age 5 were underweight. While rates of stunting declined between 2009 and 2016, rates of wasting actually increased slightly in this time period, pointing to malnutrition as a persistent problem. Malnutrition rates are also high among adults; adult malnutrition is particularly problematic when occurring among women, as children of malnourished women are more likely to also be malnourished. In 2016, 27% of women were underweight, and 23% of women age 15-49 were anemic.⁴⁸ In general, rates of malnutrition and undernutrition are higher in rural areas than urban areas.⁴⁹

Several factors contribute to high rates of malnutrition in Timor-Leste. Only half of children age 0-6 months are exclusively breastfed and only 35% are exclusively breastfed at age 4-5 months.⁵⁰ Lack of dietary diversity and food insecurity mean that only 13% of children age 6-23 months eat a minimum acceptable diet (which includes at least four food groups and between two to four meals a day, depending on age and whether the child is breastfed).⁵¹ High levels of food insecurity exacerbate this situation: 36% of the population of Timor-Leste is chronically food insecure and an additional 39% are mildly food insecure in part due to low levels of agricultural productivity and high rates of poverty that limit households' abilities to purchase high-quality food.⁵²

Low levels of access to improved sanitation and poor hygiene practices exacerbate health and nutrition challenges. As of 2016, 79% of households in Timor-Leste had access to an improved source of drinking water (such as piped water, public taps, or boreholes); urban households were substantially more likely to have access to an improved source than rural households. Only 50% of households had access to an improved sanitation facility, again with a substantial urban-rural gap in access rates. Handwashing practices are also generally weak: Among the 90% of households observed to have a place for washing hands during the 2016 Demographic and Health Survey, only 28% of these households had both soap and water at the handwashing area.⁵³ Access to safe drinking water and improved sanitation facilities, in addition to good hygiene practices, prevents diarrheal disease, a major cause of child mortality and malnutrition, as well as other diseases borne through contaminated water.

Preliminary assessments suggest that the COVID-19 pandemic has had a major impact on food security in Timor-Leste. Eighty one percent of households reported that COVID restrictions had affected their food and income sources, and 70% of households reported having reduced meal sizes or skipped a meal in the past 30 days because they did not have enough money for food.⁵⁴ The pandemic's effects on food insecurity are likely to have long-term, wide-reaching effects on health outcomes, particularly due to a potential increase in levels of anemia and malnutrition in mothers and children.

GENDER AND POWER

Timor-Leste has, in general, high levels of gender inequality, with strong patriarchal cultural norms that enforce gender inequality. Cultural practices that perpetuate gender inequality include polygamy,

⁴⁸ General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016*.

⁴⁹ USAID, "Timor-Leste: Nutrition Profile," USAID, March 2018, <https://www.usaid.gov/sites/default/files/documents/1864/Timor-Leste-Nutrition-Profile-Mar2018-508.pdf>.

⁵⁰ Ibid.

⁵¹ General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016*.

⁵² Integrated Food Security Phase Classification (IPC), *Timor-Leste: Chronic Food Insecurity Situation 2018-2023* (Rome: IPC, 2018).

⁵³ General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016*.

⁵⁴ FAO, *National agrifood systems*.

the payment of bride prices, and customary rules regarding property rights, inheritance, and succession to traditional offices. Although its prevalence has declined over time, early marriage is also a persistent gender issue, as women who marry early tend to have less education and bear more children; a relatively high proportion of women are married by age 20, while the average age of marriage for men is much higher.⁵⁵

Gender norms mean that men are more likely to work outside of the home, and generally have higher incomes, more employment opportunities, and fewer barriers to paid work than women. Correspondingly, social norms dictate that women and girls are responsible for unpaid work in the house, for bearing and raising children, and for caring for the elderly, while men are responsible for providing financial support for the household through agricultural or paid work. As a result of these and other gender dynamics, on average, men have higher levels of literacy, education, and employment than women.⁵⁶

Timor-Leste has successfully increased girls' enrollment in primary and secondary schools, with girls' enrollment rates now exceeding boys' at lower primary school levels. However, girls may face gender-related barriers to education, such as sexual harassment, violence in schools, early pregnancies, and lack of adequate sanitation facilities.⁵⁷ Women are also less likely to attend and complete tertiary studies and technical and vocational education and training (TVET) than men. Additionally, there are relatively few women working as teachers in primary and secondary schools or working in the Ministry of Education, particularly in decision-making positions, which poses a significant challenge to improving challenging gender dynamics within the education system.⁵⁸

Overall, women's participation in national government is relatively high: 38% of parliamentary seats are held by women, the highest rate in the Asia-Pacific region. However, local governance remains male-dominated, and only 5% of suco (village) chiefs are women.⁵⁹ Women are also rarely involved in community decision-making, in part due to social norms in which women are expected to be subordinate to men and not express their opinions.⁶⁰

The COVID-19 pandemic is likely to exacerbate many of these negative gender and power dynamics. Because women are generally the primary caregivers for family members and, furthermore, are often frontline responders in the healthcare system, they are at increased risk of infection from COVID. Women may have to spend more time on domestic duties and child rearing due to school closures, and may have to reduce their food consumption due to heightened levels of food insecurity and gender dynamics in which women generally eat after men. Furthermore, women's maternal, sexual, and reproductive health needs may get sidelined as the healthcare system pivots to focus on the COVID-19 crisis.⁶¹ Before the COVID-19 crisis, 60% of women already reported experiencing at least one challenge in accessing healthcare.⁶²

GENDER-BASED VIOLENCE AND VIOLENCE AGAINST CHILDREN

Timor-Leste has one of the highest rates of gender-based violence in the world. The 2016 Timor-Leste Demographic and Health Survey found that nearly three-quarters of women and over 50% of men

⁵⁵ Asian Development Bank (ADB), Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment* (Mandaluyong City, Philippines: ADB, 2014).

⁵⁶ Athena Nguyen, Alison Darcy, and Louise Kelly, "CARE Rapid Gender Analysis: COVID-19 Timor-Leste," CARE, April 27, 2020.

⁵⁷ Ibid.

⁵⁸ ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*.

⁵⁹ Timor-Leste electoral law requires a minimum of one-third of the party lists for members of parliament to be women, while suco chief elections have no specific gender requirements for candidates, explaining this gap in national-level and local-level participation.

⁶⁰ Nguyen, Darcy, and Kelly, "CARE Rapid Gender Analysis."

⁶¹ Ibid.

⁶² General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016*.

believe that a husband is justified in beating his wife in at least some cases. The survey also found that 33% of women age 15-49 had experienced physical violence since the age of 15, 29% of women had experienced physical violence in the last year, and 5% of women had ever experienced sexual violence. The most common perpetrator of physical violence among women who were or had been married was their current husband; 40% of women who had ever been married had experienced spousal violence (physical, sexual, or emotional). Only 20% of women who had experienced physical or sexual violence sought help to stop the violence, with an additional 6% telling someone, but not seeking help. Women most commonly went to family members for help to stop the violence.⁶³ Women generally report reluctance going to the police for help due to fear of repercussions, low levels of trust in the police, pressure from family members, lack of confidence, and self-blame.⁶⁴

Children also face violence (physical and otherwise) both at home and at school. While little data exists on violence against children, a 2019 study found that 87% of children have experienced physical or emotional violence at home, and an estimated 75% of boys and 67% of girls had experienced physical punishment by a teacher.⁶⁵ A study on causes of school dropouts found that 35% of girls in grades 4-6 feel unsafe traveling to and from school, and 26% do not feel safe at school. Some girls also reported that boys harass girls in schools. In 2011, the Ministry of Education implemented a zero-tolerance policy towards sexual violence, corporal punishment, and other forms of violence in schools. However, more effort is needed to successfully implement this policy across Timor-Leste.⁶⁶

Timor-Leste has several laws and policies enacted to penalize gender-based violence and violence against children and encourage reporting by survivors, including a law against domestic violence,⁶⁷ a child and family welfare system to protect children, and a National Commission on the Rights of the Child.⁶⁸ However, in many cases, community leaders and elders are responsible for dispensing justice rather than police or the judicial system. This system is problematic in cases when customary justice does not provide sufficient safeguards for women's and children's rights.⁶⁹ More work remains to be done to harmonize the customary and formal justice systems to ensure that women's and children's rights are upheld, and to implement laws and policies currently in place.

Worldwide, the COVID-19 pandemic has been associated with increased rates of gender-based violence and violence against children. Studies find that economic stressors, low levels of social support, unemployment, substance abuse, and poor mental health were associated with increased rates of spousal violence, and that parenting stress, job losses, and lack of support were associated with increased rates of violence against children.⁷⁰ Correspondingly, it is likely that the COVID-19 pandemic may be associated with worsening rates of gender-based violence and violence against children in Timor-Leste.

IMPACT OF COVID-19 ON PROGRAM ACTIVITIES

Restrictions due to the COVID-19 pandemic have had a substantial impact on program activities and target outputs and outcomes. In March 2020, HATUTAN field activities were halted, field offices were temporarily closed, and staff began to work from home due to a State of Emergency issued by the Government of Timor-Leste. This State of Emergency was ended in late May but was re-imposed in July and has remained in effect to date. As a result, the HATUTAN program fell behind schedule in

⁶³ Ibid.

⁶⁴ ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*.

⁶⁵ "Unseen, Unsafe; The Underinvestment in Ending Violence Against Children in the Pacific and Timor-Leste," *World Vision*, August 15, 2019, <https://www.wvi.org/newsroom/timor-leste/unseen-unsafe-underinvestment-ending-violence-against-children-pacific-and-timor-leste>.

⁶⁶ ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*.

⁶⁷ Ibid.

⁶⁸ Nguyen, Darcy, and Kelly, "CARE Rapid Gender Analysis."

⁶⁹ ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*.

⁷⁰ Amber Peterman and Megan O'Donnell, "COVID-19 and Violence against Women and Children: A Third Research Round Up for the 16 Days of Activism," *Center for Global Development*, December 7, 2020, <https://www.cgdev.org/publication/covid-19-and-violence-against-women-and-children-third-research-round-16-days-activism>.

terms of some major deliverables due to COVID-19 and the workplan was revised in September 2020. Additionally, many program activities have pivoted to include a focus on COVID-19 prevention and awareness.

At the national and municipal levels, the HATUTAN program has formed a close partnership with the Ministries of Health and Education to help inform the COVID-19 response and public information campaign. HATUTAN also coordinated with MEYS to provide data to support its application to the Global Partnership for Education for accelerated funding to support the COVID-19 response. Additionally, HATUTAN supported municipal-level COVID-19 task forces led by health department officials to improve public WASH facilities and public information campaigns, reproduced Ministry of Health messages for rapid dissemination, and produced posters and flipbooks to fill information gaps.

At the local level, funds were redeployed from stalled project activities to urgently address small-scale WASH repairs at schools, markets, and health clinics. In general, school and community WASH activities were redesigned to respond to the COVID-19 pandemic. As a result of the effort to improve handwashing practices and prepare schools for safe reopening, access to handwashing stations at schools has greatly increased. Additionally, during school closures due to COVID-19, HATUTAN provided students with take-home meal rations from the balance of commodities available for the SFP.

Specific program activities that have changed due to the COVID-19 pandemic include training of school administrators (changed to focus on guidance for COVID-19); training of PTAs on hygiene and WASH (changed slightly to focus on issues relevant to COVID-19); coaching of teachers (delayed but later implemented); production of books and supplemental reading materials for schools, including Lafaek student and teacher magazines (changed to support remote learning); provision of school meal commodities (HATUTAN permitted schools to divide the remaining food stock at the schools and distribute those to students as take-home rations); development of partnerships with farmers' groups to supply food to schools, including by providing trainings and support to produce nutritious foods (delayed due to the COVID-19 State of Emergency, but later implemented); development and implementation of a social behavior change strategy for health and nutrition (changed to focus on COVID-19 issues); development and implementation of a multi-sectoral nutrition training curriculum (delayed due to COVID-19, but later implemented); provision of trainings on optimal behavior in health, nutrition, WASH, and gender equality (changed to include COVID-19 issues); cross-visits of water user groups (changed to support WASH infrastructure improvements and repairs) and trainings of school cooks on the safe preparation of nutritious meals (delayed due to COVID-19, but later implemented). Activities that were scheduled to occur but have not yet occurred due to the COVID-19 pandemic include training of school administrators on nutrition, gender awareness, and SGBV prevention; training of PTAs to raise awareness of the importance of education and SGBV; development of video resources on positive teaching practices; and formation and training of VSLA groups.

METHODOLOGY

This section provides an overview of the research design of the HATUTAN midline evaluation. Following the evaluation methodology in the baseline, the evaluation uses a mixed-methods quasi-experimental design, triangulating information from different sources and both quantitative and qualitative methods to enhance the reliability and comprehensiveness of findings. All methods are gender-sensitive and socially inclusive, ensuring that women, men, girls, and boys are able to provide data in a safe, open, and reliable context, and that perspectives from all age and gender groups are adequately represented in data analysis. This includes conducting gender-specific focus group discussions (FGDs) with mothers and fathers, using appropriate approaches for the engagement of child respondents, and using an analysis framework that allows for the assessment of differential impacts based on gender as well as the extent to which the HATUTAN program addresses gender-, disability-, and other subgroup-specific barriers and cultural constraints to its project objectives.

The evaluation compares the progress observed in “full support” primary schools with the progress observed in a comparison group of schools selected in neighboring municipalities. Comparing across similar “treatment” or “intervention” schools (those exposed to HATUTAN programming) and “comparison” schools (schools with no HATUTAN programming) allows us to better understand whether improvements in key areas, such as literacy, are due to HATUTAN program activities or are rather due to external factors that may affect all schools in the country, such as the COVID-19 pandemic or the implementation of nationwide government programs. Comparison schools and communities were purposefully selected to match the socio-economic characteristics of the primary schools in which the “full support” program activities were implemented, particularly considering linguistic backgrounds, livelihoods, and geographies. Comparison schools and communities were also individually checked for the existence of similar project interventions in order to avoid bias, although similar interventions (including interventions providing reading materials) exist in some comparison areas. The selection of highly comparable schools and communities with, in many cases, no similar project interventions allows for more confident attribution of any findings to the impact of the HATUTAN program, rather than to any external factors.

While tools were generally designed in order to replicate the baseline survey and ensure comparability of results, the midline evaluation incorporates additional tools to analyze the impact of COVID-19 on learning patterns and food security, among other considerations. Additionally, minor modifications were made to the EGRA tasks in order to prevent pre-exposure bias⁷¹ and ensure that learning assessment results are both valid and comparable to the baseline results.

Overall, the evaluation seeks to provide valid and reliable data in order to assess the HATUTAN program. Additionally, the evaluation provides data on gendered dimensions of program impact and implementation as well as the impact of COVID-19 on both the program and on students, teachers, and households. Data obtained during the evaluation provides key recommendations for CARE, Mercy Corps, WaterAid and Timorese government officials to learn from and adapt the HATUTAN program.

RESEARCH OBJECTIVES

The midline study was designed to understand the preliminary impact of the HATUTAN program on schools, students, families, farmers, and other stakeholders. The findings will inform the continued orientation and emphasis of the program during its remaining time and will provide meaningful information for the participating organizations and groups. Correspondingly, the midline study pursues the following research objectives:

1. Assess and highlight factors affecting effective, quality, and efficient implementation of HATUTAN activities/interventions.
2. Assess the progress (strengths and weaknesses) of the project (per each component) against stated outputs and outcomes to date; this will include an assessment of the relevance of the outputs and outcomes through a gender lens.
3. Identify new barriers and trends associated to the ongoing COVID-19 crisis, providing recommendations on how to refine the intervention to mitigate its impact on project outcomes.
4. Assess early evidence of changes in behavior and practices (both intended and unintended) and compare these with the changes that were expected to be promoted by project activities. Identify factors in the implementation or context that hold back or promote observed and intended changes.

⁷¹ For example, the passage reading section was revised and letters were presented in a different order for the letter recognition subtask.

5. Assess the efficiency and effectiveness of project organization, management and coordination mechanisms, including quality and usefulness of partnership.
6. Assess sustainability efforts to date and potential factors that may impede schools' ability to graduate and sustain activities post-project.
7. Recommend the future orientation and emphasis of the project during its remaining time, including course corrections and adjustments to the Results Framework, project design, resource allocation, or implementation process as necessary.

Overall, the midline study attempts to determine impact, address key research questions listed in Annex 1, and provide evidence towards answering a subset of questions in the McGovern-Dole Learning Agenda⁷² for program activities that aim to affect literacy, the quality of instruction, student attentiveness and attendance, student feeding programs, nutrition practices, economic empowerment and VSLAs, agriculture practices, and gender and power dynamics. The study further determines program implications for design relevance, management and coordination, effectiveness, efficiency, sustainability, impact, and gender sensitivity.

Specific Learning Agenda questions and other learning questions addressed within the report as requested in the Terms of Reference (see Annex 5) include:

1. School meal program implementation
 - a. What community-level systems of governance and management are required for the successful implementation and sustainability of school meal programs?
 - b. Which components of school meal programs, including food production, procurement, and preparation of meals, are the most sustainable in terms of operational efficiency and why? Does the cost-effectiveness of HATUTAN change over time?
 - c. What are the most effective methods for ensuring food safety within school meal programs taking into consideration the different systems of national, regional, local and community governance?
 - d. In what ways do school meal programs impact health equity in terms of poverty, gender, or geography?
 - e. *How do health and educational outcomes linked to school meal programs differ in rural versus urban school settings?*⁷³
 - f. *What are the most successful policies affecting the success of school meal programs, and what are the necessary conditions for these policies to be implemented and to be effective?*
2. Education/literacy
 - a. In what ways do the combination of school meal interventions and educational interventions improve education and literacy levels? How can these combinations improve cost-effectiveness?
 - b. In what ways do school meal interventions impact the resources of teachers, such as classroom time, teacher incentives, and teacher capacity?
 - c. **What are the effects of school meal interventions on the acquisition of reading skills in multilingual settings?**⁷⁴
3. Health
 - a. *Which combination of school meal programs and disease treatment (i.e. malaria) and/or prevention interventions (i.e. diarrheal illnesses) reduce and/or prevent health related absences?*

⁷² The McGovern-Dole Learning Agenda is available online here: https://www.fas.usda.gov/sites/default/files/2020-03/mgd_learning_agenda_final.pdf.

⁷³ Italicized learning questions were not included in the Terms of Reference but are addressed, at least in part, within the report.

⁷⁴ Bold learning questions are not McGovern-Dole Learning Agenda questions, but are additional learning questions included within the Terms of Reference.

- b. **How do WASH programs impact learning and literacy outcomes?**
 - c. **What are the best practices in sustaining WASH interventions delivered through a school meal project?**
- 4. Nutrition
 - a. *What is the effect of school based micronutrient supplementation and/or fortification on long-term educational outcomes such cognition and knowledge acquisition?*
 - b. **How effective are school based meals at preventing and/or decreasing health related absences?**
 - c. **To what extent has Social Analysis and Action (SAA) been effective in addressing barriers to improved nutrition? Are there specific groups that benefitted more than others? Who are they and why did this happen?**
- 5. Agriculture
 - a. *How do the impacts of local procurement models and other community and nationally sourced models compare with those that rely on international food sources?*
 - b. **How can a combination of local procurement during harvest time be supplemented with international food aid to promote locally and/or nationally sustainable school meals programs?**
 - c. **What kinds of partnerships with the private sector and/or host country governments are the most effective at ensuring project sustainability? Among successful partnerships, who are the key players and what are their roles? In what contexts do private sector and/or government partnerships work best and which contexts may be more challenging?**
- 6. Gender-based violence
 - a. **To what extent does the synchronization of school meal programming with activities to address sexual and gender-based violence (SGBV) and school-based affect literacy outcomes and health/nutrition practices at household level?**
- 7. Sustainability
 - a. **The following combination of four strategies will be assessed for their combined impact on sustainability: (i) strengthened capacity of school staff and authorities to manage school feeding projects; (ii) enhanced community participation in monitoring school feeding and education outcomes; (iii); private sector engagement; and (iv) building advocacy skills of stakeholder groups.**

DATA COLLECTION TOOLS

As described above, the midline study generally followed the methodology of the baseline study in order to ensure comparability of results, with some revisions to tools in order to prevent pre-exposure bias and to incorporate additional areas of interest, such as the impact of COVID-19. Overall, quantitative tools included a learning assessment administered across 2,695 students in treatment municipalities and 1,965 students in comparison locations,⁷⁵ classroom observations conducted in 98 treatment and 87 comparison schools, school surveys conducted in 98 treatment and 88 comparison schools, and household surveys conducted in 982 households in treatment municipalities and 625 households in comparison locations. A subsample of 248 respondents additionally responded to a farmer's group survey. Qualitative tools included focus group discussions conducted with mothers, fathers, and teachers and key informant interviews conducted with school directors.

⁷⁵ The total number of learning assessment administered was slightly higher than the number reported here; however, due to data issues in which student IDs were not correctly recorded, a small number of learning assessments were dropped from the analysis.

EARLY GRADE LEARNING ASSESSMENT

As in the baseline study, the Early Grade Learning Assessment (EGRA) used in this evaluation was administered in Tetum-Prasa, the language of instruction in grade two. At baseline, consultation of native speakers of the main local languages in target areas took place to identify letters and sounds which are uncommon or nonexistent in their native languages, but which are found in Tetum. Additionally, Tetum speakers were consulted to identify more than 80 common words that would be universally relevant to Timorese children regardless of their location of residence. The choice of these generally familiar words for use in the assessment was validated through consultation with speakers of other major languages in target areas (Tetum-Terik, Mambae, Tokodede, Kemak, Galolen, and Bunak). The list of words was then refined to exclude words that included uncommon sounds in one or more other major language and words that could have an ambiguous meaning in another language. In addition to the consultation with native language speakers in target areas, the Ministry of Education shared the EGRA tools with advisors who had worked on previous reading assessments in Timor-Leste, whose combined feedback was incorporated into the tool in order to make the assessment comparable to previous EGRA tests conducted in Timor-Leste. The EGRA tool and its adaptations follow the structure and procedures recommended in the 2016 EGRA Toolkit (second edition).

The EGRA consists of five sections: letter name knowledge, invented word reading, familiar word reading, passage reading, and reading comprehension (including two levels of increasing difficulty). For the letter name knowledge, invented/familiar word reading, and passage reading sections, students were given one minute to read as many letters/words as possible; they were then given a score based on the number of letters/words they were able to correctly read. For the two reading comprehension tests, students were provided with short passages and then asked to answer five comprehension questions for each passage; there was no time limit, and students were given a score based on the number of correct answers to the reading comprehension questions. Students received instructions for each task in their mother tongue,⁷⁶ although the tasks themselves were conducted in Tetum, in order to preclude the possibility of poor results due to misunderstanding of instructions, rather than due to poor reading skills.

Tasks on letter recognition, invented word reading, and familiar word reading have a progressive increase in the level of difficulty of the letter/word. Subsequent section groupings (letter knowledge, invented and familiar word reading, passage reading, and reading comprehension) also generally have a progressive increase in the level of difficulty; as such, students who were unable to read any letters were not asked to attempt to answer any subsequent sections, and students who were unable to read any words were not asked to attempt the passage reading or reading comprehension sections.

Raw scores were calculated for each section based on the number of correct responses. The raw score was divided by the total possible score to produce a percent correct score for each section. Each section's percent correct score was then weighted equally to calculate an overall literacy score.

Table 1: EGRA sections and scoring

Section	Items	Total Possible
Letter name knowledge	100 letters	100
Reading invented words	60 words	60
Reading familiar words	60 words	60

⁷⁶ Every team includes local language speakers who could provide the instructions in the child's mother tongue. There might have been potential and rare exceptions where students spoke other minority languages (not commonly spoken in the area) and translation was not possible, although those were not recorded as a challenge by team leaders.

Passage reading	60 words (61 at midline) ⁷⁷	60 (61)
Reading comprehension	10 questions (two groupings of 5 questions, related to two different passages with increasing levels of complexity)	10

A reliability analysis was conducted using Cronbach's alpha in order to determine the extent to which the five sections of the test measured the target outcome (literacy) consistently, and thus whether they can be used to create an overall literacy score. The EGRA was found to have a high reliability score of 0.85 for all tests conducted at baseline and midline, which indicated that the sections are consistent measures of literacy and justifies the construction of an overall literacy score.

Given the multilingual context of Timor-Leste and the range of teaching practices in the country, enumerators were instructed to accept any correct response to letter identification regardless of the language in which the letter was identified (Tetum, Portuguese, or other local language). Similarly, enumerators were instructed to accept different accents in word and passage reading sections. Enumerators were trained in the recognition of different letter names across languages and local accents in order to ensure that these instructions were applied in the field.

In addition to the five EGRA sections described above, students were also administered a pictorial working memory tests in the assessment. The working memory test was included as a proxy for attentiveness. After completing the five EGRA sections, students were presented with a set of 19 images representing common objects and animals. The enumerator showed the child each image individually, mentioned the name of the object/animal in the image, and instructed the child to remember the image for later. The child was then asked to recall as many images as possible without looking at the images. This test was based on standard working memory tests used in clinical psychology and was adapted for administration in the field.

CLASSROOM OBSERVATION

Team leaders conducted classroom observations in second grade Tetum language classes in 98 treatment schools and 87 comparison schools. The classroom observation tool was developed based on existing tools used by CARE, and includes items on teacher background, reading practices in class, child-centered teaching practices, student participation, student access to materials, gendered practices, use of physical and verbal violence against students, and use of formative assessments. During the classroom observation, data collectors observed whether a set of teaching practices occurred during class, including copying from the board, reading to students, engaging students in classroom activities, and using games. In addition, data collectors observed teacher behavior towards girls and boys, including whether they encouraged, asked questions to, used angry voices with, or used corporal punishment towards girls and boys.

Data collected in the classroom observation was used to measure teachers' use of engaging teaching practices, traditional teaching practices, and negative teaching practices. Thirteen engaging practices (two of which observed the same behavior but were disaggregated by gender in the data), two traditional practices, and two negative practices were observed. Given the large number of engaging practices, we analyze whether an index measuring the use of engaging teaching practices is reliable. We find that the Cronbach's alpha of all thirteen items is 0.76, an indication of an acceptable level of internal reliability. We also conduct a reliability analysis and calculate the corrected item-total correlation for each of the 13 items, another indication of whether an item can reliably be included in the overall index; a value of at least 0.40-0.50 is recommended for indexes that measure a narrow

⁷⁷ 61 words were included in the passage at midline to create a logical and complete passage.

range of characteristics.⁷⁸ We find that many engaging teaching practices do not have a corrected item-total correlation of at least 0.4; some, such as whether the teacher reads to students or uses the reading corner, are as low as 0.15. As such, while we do compare changes in the total number of engaging teaching practices used, we also analyze the prevalence with which specific teaching practices are used across groups, rather than focusing on an index score.

Table 2: Engaging teaching practices

Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Removed
Students participate in reading activities with others	0.307	0.758
Students read by themselves	0.297	0.758
Teacher uses games or exercises	0.452	0.742
Teacher calls on inactive students to engage them in activities	0.518	0.735
Teacher asks the students' opinion	0.415	0.746
Teacher asks open questions	0.256	0.762
Teacher reads to the students	0.149	0.772
Students work together in groups	0.295	0.759
Teacher uses the reading corner for literacy activities	0.151	0.770
Teacher encourages male students	0.520	0.735
Teacher encourages female students	0.517	0.735
Teacher asks questions to male students	0.581	0.728
Teacher asks questions to female students	0.567	0.729

SCHOOL SURVEY

The school survey was administered with school directors or coordinators in 98 treatment schools and 88 comparison schools to collect data on the number of teachers employed, teacher attendance, teacher training and qualifications, student enrollment and attendance, PTA activity, COVID-19 related restrictions, and school infrastructure, including water, electricity, toilets, kitchens, and storage space. The school survey also recorded the sources of food for the SFP (with particular emphasis on whether schools obtained food from local farmers) as well as the types of food served that day, if any.

Student enrollment and dropout data was copied from school records. Student attendance and teacher attendance was collected through head counts; student attendance was additionally recorded for cross-checks by copying data from school records.

HOUSEHOLD SURVEY

The household survey was conducted with families of second grade students included in the EGRA. The questions, asked to the head of household and caregiver, covered a wide variety of topics. The head of household was asked to answer questions about the number of people living in the household and its composition. Caregivers were asked about student attendance at school, the SFP, gender-based violence, nutrition knowledge and practices, and hygiene knowledge and practices, among other topics.

The household survey included a module on savings and VSLA use and participation for households who reported having savings, and a module on breastfeeding and child nutrition for households with

⁷⁸ See L. A. Clark and D. Watson, "Constructing validity: Basic issues in objective scale development," *Psychological Assessment* 7 (1995): 309-19.

a child under the age of 2 and babies under six months of age. The module on savings and VSLAs included questions on the use of savings and VSLA loans, the frequency and benefits of VSLA participation, and decision-making about the use of VSLA loans. The module on breastfeeding and child nutrition included questions about the frequency of breastfeeding, whether the child was breastfed exclusively, other foods or drinks given to the child, and reasons for giving the child foods or drinks other than breastmilk.

At baseline, questions on hygiene knowledge and practices were found to have validity issues, and an additional set of questions was developed partway through the baseline study which used a pictorial approach to improve question clarity. In this new approach, respondents were asked to identify if they should wash their hands before and/or after doing the activity depicted in a picture, and to identify which pictures demonstrated good hygiene behaviors. As this new set of questions was developed and validated in the midst of baseline data collection, the sample size of respondents who answered these questions is smaller at baseline than for other questions in the household survey. The midline sample size was not affected and used the pictorial approach from the beginning of data collection.

The household survey also included questions on dietary diversity, for which overall dietary diversity scores were calculated for women caregivers of childbearing age (15-49) and for children between the ages of 6 months and 23 months. Caregivers reported the foods that they and their child consumed during the prior day, which were coded according to seven food group categories (Table 3). Scores were calculated based on the number of food groups each caregiver/child ate the previous day, from a minimum of 0 to a maximum of 7 for children and a maximum of 9 for caregivers.

Table 3: Dietary diversity food groups

Food Group	Food Item	Respondent
Grains, roots, and tuber	Maize, rice, bread, cereals/porridge, noodles, rice, mash/residue, or other foods made from grains such as maize or wheat	Caregiver and child
	White potatoes, white yams, white sweet potato, cassava, or any other foods made from roots	Caregiver and child
	Thin porridge	Child
Beans, legumes and nuts	Any foods made from beans, peas, lentils, peanuts	Caregiver and child
	Any foods made from nuts and seeds such as pumpkin, sunflower seeds	Caregiver and child
Dairy products	Milk or food prepared with milk (not including sweetened condensed milk)	Caregiver and child
	Infant formula	Child
	Milk such as tinned, powdered, or fresh animal	Child
Eggs	Eggs	Caregiver and child
Organ meat	Any liver, kidney, heart, blood, or other organ meats from domesticated animals such as cow, pig, goat, chicken, or duck	Caregiver
	Any organs from wild animals, such as game meat, bush rats, birds, wild pigeons, guinea fowl, deer, wild boar	Caregiver and child <i>Note: Counted as "flesh food" for children</i>
Flesh foods	Any meat such as beef, pork, lamb, goat, chicken, or duck	Caregiver and child
	Any flesh from wild animals, such as game meat, bush rats, wild birds, deer, wild boar, wild goat	Caregiver and child
	Fresh or dried fish, shellfish, or seafood	Caregiver and child
	Grubs, snails, or insects	Caregiver
Vitamin A-rich dark leafy greens	Dark green leafy vegetables such as spinach, kangkung, lettuce, mustard greens, pumpkin leaves, cassava leaves, or potato leaves	Caregiver and child <i>Note: Counted as "Vitamin A-rich fruits and vegetables" for child</i>
Other vitamin A-rich vegetables and fruits	Pumpkin, carrots, squash, orange fleshed sweet potatoes or any other dark yellow or orange fleshed roots, tubers, and vegetables	Caregiver and child <i>Note: Counted as "Vitamin A-rich fruits and vegetables" for child</i>
	Ripe mangoes, ripe papaya, melon, passionfruit, or other fruits that are dark yellow or orange inside	Caregiver and child <i>Note: Counted as "Vitamin A-rich fruits and vegetables" for child</i>
	Foods made with red palm oil	Caregiver
Other fruits and vegetables	Any other vegetables, like cucumbers, tomatoes, cabbage, eggplant, etc.	Caregiver and child
	Any other fruits like watermelon, tamarind, jackfruit, etc.	Caregiver and child
	Any indigenous/wild fruits	Caregiver and child

FARMERS' GROUP SURVEY

A subset of household survey respondents who were provided with training on VSLAs, keyhole gardens, or permagardens were asked to answer an additional module about farming practices and farmers' groups. This module included questions on the types of crops grown in the garden, the use

of the crops (for sale or own consumption), challenges faced with the garden, and support received from agricultural extension services.

The proportion of VSLA participants, and consequently farmers receiving HATUTAN agriculture trainings, was found to be low among parents of students who took the EGRA at midline. As a result, midway through the midline data collection process, a shorter version of the household survey was created that included only questions for the head of household, the farmer's group survey, nutrition questions, and hygiene questions. A separate sample of VSLA participants who had received training on improved agricultural techniques was then created, and these VSLA members were administered the short version of the household survey in order to obtain sufficient information on farmers' groups and the effects of training. In total, 45 farmers were surveyed using the full household survey, and an additional 248 VSLA members were surveyed using the shortened version of the household survey focusing on VSLA participants and farmers' groups.

FGDS AND KIIS

Focus group discussions were conducted with parents (mothers and fathers separately) and teachers. Key informant interviews were conducted with school directors and coordinators. KIIs were also conducted with teachers in cases where there was only one teacher available at a school.

FGDs with parents were designed to provide in-depth information on student absenteeism or tardiness, including causes, patterns, and potential solutions; perceptions of student learning and the factors affecting learning; school feeding practices; perceptions of school management; PTA engagement in school governance and participatory school management; work-sharing practices at the household and the impact of any gendered divisions of work on health, hygiene, and nutrition; traditional practices on nutrition and childcare; breastfeeding; healthcare seeking behaviors; savings practices and decision-making related to the use of savings; perceptions of farmers' groups and ability to increase farming production; occurrence of and attitudes towards violence and GBV; and perceptions of the use of physical and verbal violence against students.

FGDs (or KIIs) with teachers included questions on attendance, learning, and classroom management.

KIIs with school directors and coordinators included questions on perceived management responsibilities; previous training and perceptions of training; PTA engagement and perceptions of its value; student and teacher attendance; effective teaching strategies for student literacy and major challenges to teaching literacy; classroom management and student participation; and school feeding practices, management, and challenges. Additionally, female directors and coordinators were asked to answer questions about potential gender-specific challenges faced when undertaking their duties.

TOOL MAPPING

The below table maps the data collection methods used to the key research areas addressed with each tool. For the quantitative tools, the table designates main and secondary data sources used for the analysis. The secondary sources, in this case, refer to data sources used mostly to validate, triangulate, or add nuance to findings; for example, for the case of literacy, most results are drawn from the EGRA, but data from the household survey and school survey provides further information on potential reasons for results, including students' access to literacy materials and homework help at home.

Table 4: Tool mapping

Research area	Quantitative tools		Qualitative tools
	Main data sources used	Secondary data sources used	
Literacy	EGRA	Household survey, school survey	KIIs, FGDs
Quality of instruction	Classroom observation, school survey	Household survey	KIIs, FGDs

Student attentiveness	Classroom observation	Household survey, school survey, EGRA	KIIs, FGDs
Student attendance	School survey, household survey	Classroom observation	KIIs, FGDs
School feeding program	Household survey, school survey		KIIs, FGDs
Health	Household survey		FGDs
Nutrition	Household survey		FGDs
Economic empowerment	Household survey, farmer's group survey		FGDs
Agricultural practices	Farmer's group survey		FGDs
Gender and power	Household survey	School survey, classroom observation	KIIs, FGDs

DATA COLLECTION

ENUMERATOR SELECTION AND TRAINING

Upon receiving applications for enumerator positions (including team leaders and data collectors), CARE preselected applicants based on previous data collection experience and skills in local languages. The preselected group took a practical test and those meeting the cut-off point on the test were then interviewed. After the interview process, final enumerators selections were made.

Team leaders were trained for 11 days, including five days of joint training with data collectors. Training topics included an introduction to the program, child protection, prevention of sexual harassment and abuse, research ethics and informed consent for adults and children, confidentiality and data security, using electronic data collection forms, working with children, a review of the quantitative tools, and data quality control practices. Training also included a mock practice session, field practice, and a final test. Team leaders received additional modules on team management and reporting, the work plan, data quality control, and qualitative data collection, as well as a more extensive field practice. Data collectors received six days of training, including an introduction to the program, child protection, prevention of sexual harassment and abuse, research ethics and informed consent for adults and children, confidentiality and data security, using electronic data collection forms, working with children, a review of the quantitative tools, data quality control practices, mock practices, and field practices. Two assessments and a final field practice observation were conducted to finalize the selection of enumerators and team leaders; only those who had reached minimum cut-off scores in assessments and demonstrated proficiency in the administration of the reading assessments according to the protocol were contracted for data collection.

Before the training, all data collectors received a detailed explanation of CARE's policies on child protection, sexual harassment, and abuse. Data collectors were provided with copies of these policies and were required to sign their agreement with both policies.

In order to comply with the Government of Timor-Leste's COVID-19 prevention protocols and in line with CARE's own guidance and do-no-harm principles, the following measures were observed during training: (1) guidance to enumerators and staff on COVID-19 prevention; (2) mandatory mask wearing; (3) social distancing; and (4) handwashing stations with soap positioned in multiple locations in the training facility.

FIELDWORK OVERVIEW

Data collection began on February 8, 2021 and ended on March 29, 2021. Teams were distributed to various locations based on their linguistic skills. Teams spent an average of one and a half days at

each school collecting data with students, teachers, school coordinators/directors, and families. They used electronic data collection tools to allow for real-time data verification and cleaning.

During data collection, enumerators were required to wear masks at all times when interacting with respondents in schools or households. When conducting surveys or qualitative data collection, enumerators were also required to practice social distancing.⁷⁹ Enumerators were asked to conduct surveys, learning assessments and qualitative interviews outdoors or in semi-open spaces (verandas) to minimize risk for the respondents and themselves. Throughout data collection, CARE monitored emerging COVID cases in targeted areas, conducted daily checks with all field teams and tracked potential risks for field teams and respondents.

A week into the fieldwork, it was found that the proportion of VSLA members covered by agriculture trainings was low among the grade 2 students' parents contacted in the household survey. As a result, fieldwork was adjusted to include an additional sample of VSLA participants who received agriculture training in target locations, who were asked a shortened version of the household survey.

Two weeks into the fieldwork, data collection was suspended in Bobonaro after a number of illegal border crossing incidents resulted in COVID-19 cases being identified in the municipality. The government restricted movement in and out of two municipalities (including Bobonaro and Covalima) and two teams were instructed to stay in place in Maliana, Bobonaro's municipal capital, to reduce risk for themselves and schools. All schools were closed in Covalima and some schools were closed in Bobonaro while the Ministry of Health conducted mass screening across all three municipalities. After imported COVID cases were identified in Covalima, the entire municipality was placed under lockdown. Data collection in schools in Covalima and Bobonaro was affected by these restrictions; however, a sufficient sample of schools and households was still obtained in these areas to allow for statistically robust comparisons.

Provision of qualitative data was delayed due to COVID-19 lockdowns, which restricted data collectors' abilities to download files into computers in municipal offices and send them to Dili for transcription. Additionally, Timor-Leste was hit by a major cyclone in early April, which resulted in most of the capital being underwater, widespread destruction of infrastructure across the country, and loss of power. This tragic event resulted in a major delay in translation of qualitative data.

DATA QUALITY CONTROL

Tools were translated into Tetum by Tetum-speaking CARE staff. The EGRA was originally developed in Tetum and was backtranslated into English for quality control purposes. All translations were checked by an independent translator.

All tools were reviewed by a working group formed by representatives of the MEYS, MOH, MAF and development partners. Comments and requests for additional items/removal of items were incorporated into the tools.

For quantitative data, several quality checks were scripted into the survey tools to reduce the data-entry related errors and ensure only eligible respondents would be interviewed, such as choice filters, age restrictions, constraints for the numeric values and calculations for the learning assessment scores.

During the fieldwork, teams were provided with several tracking tools, such as individual tracking sheets and tracking sheets for each community/school, containing the identifier and demographic information for the target respondents. Research processes were monitored in the field by Consilient and the CARE team to ensure that protocols were being followed, address any data quality issues in

⁷⁹ Due to the requirements of EGRA application with young children, social distancing was not possible, but enumerators were required to wear masks at all times and to conduct data collection with children outdoors or, when not possible, in spaces with high ventilation (next to an open window).

a timely manner, and enable team leaders to rapidly clarify any procedural questions. A quality control tracking tool was specifically developed in Stata and used on the daily basis to track the number of submitted surveys, results by school/community and enumerator, and any changes/information related to the quantitative data collection. Quality control checks of the submitted data were conducted on a daily basis, and checked for issues such as implausible EGRA results, possible cases of EGRA misadministration by enumerators, contradictory attendance and enrollment records, and logical and coherent text-based responses, among other possible issues. All inconsistencies and mistakes were discussed with the teams in the field, and if necessary, corrected in the data.

DATA MANAGEMENT AND CLEANING

For the quantitative data, to ensure secure data management, the evaluation team used an online data management platform (ONA) and all teams were required to submit the surveys to the ONA servers once they were completed. The submitted data were downloaded on a daily basis for regular quality control and data cleaning.

Daily data cleaning focused on general inconsistencies, duplicate observations, variables in which numeric answers were hand-entered (rather than selected from a list), school attendance and enrollment variables, and learning assessment scores. While household survey and EGRA data were reviewed daily, the review and cleaning of the data from other surveys were done bi-weekly. On a weekly/bi-weekly basis, depending on the specific survey data, a more in-depth data cleaning was conducted by our team. All the variables were separately examined and cross-tabulated to identify any possible inconsistencies in the data.

SAMPLING

In this section we describe the sample used for the midline evaluation, including the sample of schools and municipalities, student cohorts, and the demographics of the achieved samples. In January 2019, in preparation for the baseline assessment, CARE selected target schools in four municipalities (Ainaro, Ermera, Liquica, and Manatuto) considering vulnerability criteria such as absence of similar interventions other than the national SFP, education outcomes, location (rural/urban), and distance from the main road. The list of schools formed the sampling framework from which, at baseline, 104 treatment schools were selected. Sampled schools were selected using stratified random sampling, considering the distribution of the student population per municipality in intervention schools.

Following the selection of treatment schools, comparable sub-districts were matched with treatment sub-districts, and comparison schools were selected to have a similar average “remoteness” as intervention schools. The table below describes the geographic breakdown of samples for the EGRA, comparing baseline to midline and intervention group to comparison group. As per the evaluation plan, the sample size for students was calculated considering a 0.2 standard deviation effect size, 5% significance level, 80% power, and a design effect of 2. The sample size at baseline was calculated assuming 30% attrition and was powered to allow for gender-disaggregated data. The effect size was purposefully set at a relatively low level to avoid underpowering the sample, and also takes into consideration the potential for losses in learning due to prolonged absenteeism during the midline data collection, which took place at the peak of the rainy season.

Table 5: EGRA sample by cohort

	Midline cross-sectional cohort		Midline panel cohort		Baseline cohort	
Municipality (treatment)	n	%	n	%	n	%
Ainaro	362	14.0%	284	13.7%	360	14.6%
Ermera	641	24.8%	536	25.9%	631	25.6%
Liquica	204	7.9%	158	7.6%	190	7.7%
Manatuto	267	10.3%	243	11.7%	266	10.8%
Total	1,474	57.0%	1,221	58.9%	1,447	58.7%
Municipality (comparison)						
Aileu	238	9.2%	205	9.9%	234	9.5%
Baucau	48	1.9%	44	2.1%	48	2.0%
Bobonaro	409	15.8%	270	13.1%	356	14.5%
Covalima	168	6.5%	118	5.7%	146	5.9%
Manufahi	245	9.5%	211	10.2%	230	9.4%
Total	1,108	42.9%	848	41.0%	1,014	41.3%

The table below describes the achieved sample for the school survey, classroom observation, household survey, and farmer's group booster survey. At midline, the school survey and classroom observation were intended to occur in every school in which students were administered the EGRA. At midline, one school was not re-contacted in the municipalities Ainaro, Bobonaro, and Manufahi; additionally, a classroom observation did not occur in one school in Manufahi which was re-contacted. The sample size for households was calculated considering a 5% significance level, 80% power, and a design effect of 2, and was calibrated based on a 15 percentage point change in knowledge of improved nutrition practices.

Table 6: Sample for school survey, classroom observation, household survey, and farmer's group survey

	School survey		Classroom observation		Household survey		Farmer's group survey	
Midline municipality (treatment)	n	%	n	%	n	%	n	%
Ainaro	27	14.5%	27	14.6%	189	13.9%	86	34.7%
Ermera	41	22.0%	41	22.2%	321	23.7%	62	25.0%
Liquica	11	5.9%	11	5.9%	86	6.3%	55	22.2%
Manatuto	19	10.2%	19	10.3%	137	10.1%	45	18.2%
Total	98	52.6%	98	53.0%	733	54.1%	248	100.0%
Midline municipality (comparison)								
Aileu	23	12.4%	23	12.4%	150	11.1%	-	-
Baucau	3	1.6%	3	1.6%	24	1.8%	-	-
Bobonaro	31	16.7%	31	16.8%	220	16.2%	-	-
Covalima	10	5.4%	10	5.4%	80	5.9%	-	-
Manufahi	21	11.3%	20	10.8%	148	10.9%	-	-
Total	88	47.4%	87	47.0%	622	45.9%	-	-
Baseline municipality (treatment)								
Ainaro	28	14.8%	27	18.9%	128	14.9%	-	-
Ermera	41	21.7%	41	28.7%	208	24.2%	-	-
Liquica	11	5.8%	11	7.7%	55	6.4%	-	-

Manatuto	19	10.1%	19	13.3%	91	10.6%	-	-
Total	99	52.4%	98	68.5%	482	56.0%	-	-
Baseline municipality (comparison)								
Aileu	23	12.2%	12	8.4%	98	11.4%	-	-
Baucau	3	1.6%	2	1.4%	14	1.6%	-	-
Bobonaro	32	16.9%	13	9.1%	136	15.8%	-	-
Covalima	10	5.3%	8	5.6%	39	4.5%	-	-
Manufahi	22	11.6%	10	7.0%	91	10.6%	-	-
Total	90	47.6%	45	31.5%	378	44.0%	-	-

Grade 2 students in selected schools were randomly sampled from attendance lists for the EGRA. At baseline, the original sample was set at 20 students per school, but the actual average sample was 13 students per school due to absenteeism, dropout, and small class sizes in remote schools. At midline, among the sample of grade 2 students, an average of 14 students per school were sampled. For the sample of students re-contacted at midline, an average of 11 students per school were sampled. In small classes where the total number of students was equal to or less than the desired sample size of 20, a “take all” approach was used.

The household sample was randomly selected from the list of assessed students, with an overall sample of five households per location. At midline, midway through data collection, it became evident that the household sample did not include enough VSLA participants trained on improved agricultural techniques to make robust conclusions about the impact of farm-related HATUTAN program activities. As such, an additional sample of 248 VSLA participants trained on improved agricultural techniques was added within treatment municipalities. The VSLA participants were selected randomly as a sample of those who received training in the treatment locations.

DEMOGRAPHICS OF ACHIEVED SAMPLE

In this section, we describe the demographic composition of the midline samples. We further analyze demographic differences between midline and baseline samples, as well as between intervention and comparison areas, in the “Methodological Analysis” section below.

Students Assessed with EGRA

At midline, 4,651 students were successfully assessed across two cohorts: the cross-sectional cohort of newly selected grade 2 students, and the panel cohort of students re-contacted from the baseline. Table 7 shows the demographic characteristics of these students, disaggregated by cohort and by intervention/comparison group. The midline sample was relatively balanced by gender, with about 51% male respondents and 49% female respondents. The average age for the cross-sectional cohort of grade 2 students was, as expected, substantially lower than that of the panel cohort. Students spoke Tetum as their native language at relatively similar rates across all groups, with around two-thirds of midline students speaking Tetum natively.

Table 7: Demographics of students assessed at midline

	Cross-sectional cohort		Panel cohort	
	Intervention	Comparison	Intervention	Comparison
n	1,474	1,108	1,221	848
Male	50.8%	51.8%	50.5%	51.5%
Female	49.2%	48.2%	49.6%	48.5%
Average age (years)	7.7	7.5	9.7	9.4
Native Tetum speaker	66.8%	65.0%	68.3%	69.2%

Households

From the cross-sectional cohort of students selected for the EGRA, about seven students' households from each school were selected for the household survey, for a total of 1,355 households surveyed at midline. An additional four households were accidentally assessed with students from the panel cohort, and 10 households were unable to be matched with student data from the EGRA. Table 8 shows that caregivers in sampled households were almost entirely female. The average age of caregivers was around 39 years old.⁸⁰ Tetum was spoken in over two-thirds of households; Mambae was also commonly spoken in both intervention and comparison households and Kemak was commonly spoken in comparison households. In most households, more than one language was spoken.

Table 8: Household demographics at midline

	Intervention	Comparison
n	734	625
Caregiver gender		
Male	2.3%	2.4%
Female	97.7%	97.6%
Caregiver age		
Average age (years)	39.7	39.2
Language spoken in household		
Tetum-Prasa	71.4%	68.0%
Mambae	55.7%	31.5%
Tokodede	11.6%	0.2%
Kemak	10.8%	33.0%
Idate	8.2%	0.2%
Bunak	4.6%	15.2%
Galolen	4.2%	0.8%
Tetum-Terik	4.0%	13.8%
Midiki	0.7%	0.0%
Makasae	0.0%	0.0%
Other	4.2%	6.9%

⁸⁰ Unfortunately, at midline, data on head of household gender and age was not collected, so this demographic data cannot be provided.

Table 9 further breaks down the education levels and occupations of heads of household and caregivers. The majority of heads of household and caregivers at midline had low levels of education—either no education or incomplete primary school. Caregivers were less likely to have an education than heads of household. The vast majority of heads of household worked as farmers, either for own consumption or for sale; caregivers also frequently worked as farmers, but were more likely to be unemployed than heads of household.

Table 9: Household education and livelihoods

	Head of household		Caregiver	
	Intervention	Comparison	Intervention	Comparison
n	734	625	734	625
Education				
No education	29.8%	28.8%	39.8%	35.7%
Incomplete primary	23.2%	21.1%	18.4%	18.4%
Complete primary	7.9%	6.1%	6.4%	6.4%
Incomplete pre-secondary	5.7%	7.7%	7.6%	9.1%
Complete pre-secondary	8.5%	5.3%	7.6%	8.0%
Incomplete secondary or technical school	4.2%	5.3%	4.4%	5.4%
Complete secondary or technical school	14.7%	19.4%	12.4%	15.4%
University	5.5%	6.1%	2.3%	1.4%
Occupation				
Farmer (own consumption)	41.4%	39.7%	34.1%	33.0%
Farmer (sale and own consumption)	24.9%	29.4%	19.4%	18.4%
Unemployed	4.5%	5.4%	18.5%	23.8%
Other	29.2%	25.5%	28.0%	24.8%

Among all midline households, the average household had 7.4 members. Households in intervention areas were, on average, slightly larger than those in comparison areas, with 7.6 and 7.2 average members respectively. In intervention areas, 10% of households had 4 or fewer members, 59% had 5-8 members, and 31% had 9 or more members. In comparison areas, 13% of households had 4 or fewer members, 62% had 5-8 members, and 25% had 9 or more members.

Table 10 shows the number of children in intervention and comparison households at midline. The majority of households had either zero or one child under 3. On average, midline households had three children age 5-15, half of whom were girls. Around 2.7 children age 5-15 were attending school on average, of whom, again, half were girls.

Table 10: Number of children in households at midline

	Children under 3		Children age 5-15		Girls age 5-15		Children in school		Girls in school	
	Int.	Comp.	Int.	Comp.	Int.	Comp.	Int.	Comp.	Int.	Comp.
n	734	625	734	625	734	625	734	625	734	625
0	47.7%	50.4%	0.0%	0.0%	16.2%	22.1%	0.0%	0.2%	19.4%	24.8%
1	41.3%	39.7%	12.0%	15.5%	37.7%	38.4%	16.8%	20.3%	39.1%	39.2%
2	9.4%	8.5%	28.3%	30.7%	28.3%	26.1%	30.8%	34.2%	28.1%	25.0%
3	1.1%	1.1%	23.6%	26.2%	12.7%	9.0%	24.9%	24.6%	10.0%	8.3%
4+	0.5%	0.4%	36.2%	27.7%	5.0%	4.4%	27.5%	20.7%	3.4%	2.7%

Households were also asked whether there were any pregnant or lactating mothers among household members. Within intervention households, 9.5% reported that the household members included a pregnant mother, and 7.5% of comparison households reported that there was a pregnant mother within the household. Around 32% of intervention households and 31% of comparison households reported that there was a lactating mother in the household.

In the household survey, caregivers were also asked if the student participating in the EGRA had some form of disability (physical or mental/cognitive). Table 11 shows that the reported prevalence of physical disabilities was relatively low; the most common physical disabilities were related to hearing. Cognitive disabilities and mental health issues were reported far more frequently. More than one-third of caregivers reported that students have difficulty remembering or concentrating and with self-care in both intervention and comparison areas; the prevalence of these disabilities may be related to nutritional issues. More than one-quarter of caregivers also stated that the child has trouble communicating. A relatively high percent of caregivers also stated that the assessed child has anxiety or depression on a daily, weekly, or monthly basis.

Table 11: Student disabilities at midline

Disability	Intervention	Comparison
n	734	625
Vision	0.9%	2.1%
Hearing	4.2%	4.5%
Mobility	2.6%	1.8%
Memory	38.7%	36.2%
Self-care	35.3%	35.0%
Communication	25.1%	27.0%
Anxiety ⁸¹	13.1%	19.7%
Depression	6.0%	10.6%

Farmers

The farmer's survey booster sample took place with an additional 248 VSLA participants trained on improved agriculture techniques outside of those already identified in the household survey in intervention municipalities at midline. Male respondents comprised a much larger percentage of the farmer's survey than the household survey; 36% of respondents to the farmer's survey were male. The language profile of respondents was similar to that of the household survey in intervention areas, with most respondents speaking one or more of Tetum-Prasa, Mambae, or Tokodede. The education profile of respondents was also similar to that of the household survey, with 35% of respondents stating they had no education and 19% that they had an incomplete primary education. Sixty-eight percent of respondents listed their occupation as farmer (either for subsistence or for both sale and own consumption); an additional 13% of respondents stated that they were unemployed.

METHODOLOGICAL ANALYSIS

In this section, we consider critical methodological issues related to the midline evaluation and investigate their potential impact on the results presented throughout this report. We do not analyze every potential methodological pitfall of the evaluation; rather, we focus on those that may be particularly problematic for drawing causal inferences regarding the program's impact.

⁸¹ Anxiety and depression are calculated as the percent of students reported to feel very anxious or worried/very sad or depressed daily, weekly, or monthly.

The discussion below is not intended to imply that the methodology of the evaluation is systematically flawed or invalid. Instead, it intends to systematically consider potential threats to inference that can be discounted through supplementary analysis; to briefly discuss common issues that are not actually true threats to inference, given the overall design of the evaluation; and to highlight genuine threats to inference so that, in discussing our substantive results below, we can make clear the extent to which methodological challenges are actually problematic. In short, the goal of this section is to make clear the extent to which and under what circumstances methodological issues are of true concern, so that readers can interpret our findings with the appropriate degree of caution.

The HATUTAN midline evaluation relies on two samples for analysis: a cross-sectional sample and a panel sample. The cross-sectional sample includes grade 2 students assessed at baseline and a new cohort of grade 2 students who were assessed at midline. The panel sample, in contrast, includes grade 2 students assessed at baseline who were also re-contacted for assessment at midline (and who, at midline, were primarily in grades 3 or 4).

Both the cross-sectional and panel samples have methodological advantages and disadvantages. Because the cross-sectional sample assesses grade 2 students at both baseline and midline, we would expect these students' assessment results and schooling to be broadly similar, since they are of similar ages and have had similar amounts of exposure to education (or at least would have had in the absence of COVID-19). However, because the cross-sectional sample is comprised of an entirely different group of students at midline, it is vulnerable to bias that may occur due to observed or unobserved differences between groups of students. For example, if the group of students observed at midline in intervention schools is, coincidentally, more motivated or has greater aptitude for reading than those students observed at baseline, the midline students will perform better than expected on the assessment. In a situation such as this, we may thus mistakenly attribute the improvement in scores to the impact of the HATUTAN program, rather than to characteristics innate to this new cohort of students.

In other words, in a repeated cross-sectional design, a particularly unusual sample of intervention students at midline could result in positive or negative estimates of impact that are driven by the unusual nature of the sample, rather than actual impact. On average, a repeated cross-sectional design is still unbiased, but there is no guarantee that individual iterations of the design will produce unbiased results, due to the potential for sampling variation.

Analysis of the cross-sectional sample may also be affected if the HATUTAN program has an impact on the types of students who are enrolled in or regularly attend schools. For example, benefits provided by the program, such as school meals, may increase school enrollment or attendance in intervention schools among the most-disadvantaged students who are likely to have lower literacy abilities, but have little impact on enrollment or attendance of more-advantaged students for whom school meals are less of a draw. If this is the case, increased enrollment or attendance of disadvantaged students in areas affected by the program, but not in comparison areas, would result in lower literacy scores in intervention schools despite overall positive program impact.

A panel design, on the other hand, eliminates this specific threat to inference by following up with the same students over baseline and midline. As a result, we can be more assured that any change in the intervention group that is statistically different from the comparison group can be attributed to program impact, rather than to innate differences between student groups. We note that at midline, students in the panel sample are around two years older and have been exposed to an additional 1.5 years of education (considering school closures). As such, there is a natural improvement in these students' assessment results. This is not a methodological limitation, as comparisons of changes across intervention and control groups allow for understanding of relative improvement due to the program, rather than general improvement due to additional exposure to schooling. However, if we are more interested in the impact of the HATUTAN program on very young (grade 2) students, this may be a drawback of the panel sample.

Of more methodological concern among the panel sample is panel attrition—students who were contacted at baseline but who were not able to be contacted at midline. A panel design is only inferentially valid if the rate of panel attrition is minimized. If many students fall out of the sample and are not replaced, the design is weakened in terms of statistical power. Minimizing attrition by maximizing re-contact rates is of the utmost importance, as any bias in the type of students who are re-contacted compared to those who are no longer in school, and who thus “fall out” of the sample, can influence our findings. If, for example, students who are performing worse in school are more likely to drop out, and thus less likely to be found by enumerators at midline, then scores at midline would improve—but only because the students who were successfully re-contacted were those who were already performing better in school. In this scenario, we could potentially misattribute improvement in scores to program impact, rather than to the result of bias due to attrition from the sample.

In Annex 2, we analyze the cross-sectional and panel samples for any observable demographic differences across groups to be controlled for in our later regression analysis. We also assess the intervention and comparison samples to see if observable differences in students, schools, or households across these groups may introduce bias to our results. Within the cross-sectional sample, overall, we find some differences in student, school, and household characteristics across treatment and comparison groups. The net impact of these differences between treatment and comparison areas is difficult to unpack. Some differences—for example, changes in students’ native languages—may lead us to underestimate the impact of the HATUTAN program on learning outcomes. Others, such as a relative improvement in student-to-teacher ratios in treatment schools compared to comparison schools, may lead us to overestimate program impact. We bring up these issues not to imply that the methodology or results used in this report are invalid, but rather to systematically analyze potential pitfalls to inference, justify the use of control variables (such as those for student language or school fixed effects) where needed, and caution against overinterpretation of results.

Among the panel data, in contrast, we find that there is little bias in the type of students who drop out of the sample across treatment and comparison schools. Overall, while these findings suggest that attrition is not entirely as-if random, it appears to operate similarly across both intervention and comparison schools. As a result, while we rely first and foremost on the cross-sectional sample for analysis, the panel sample provides strong data through which to check the robustness of our results.

As a result of our methodological analysis, we use a variety of regression specifications within the report of increasing methodological rigor, particularly focusing on EGRA results and the school survey. For the EGRA, we report results using a difference-in-differences model without controls. We then check for robustness using a difference-in-differences model which controls for student gender, age, and native language, and an additional model which controls for the aforementioned variables as well as school fixed effects. For outcomes related to schools, such as the quality of instruction, our robustness models include school fixed effects which control for the potential impact of variables such as director experience and teacher-to-student ratio, as well as, when relevant, classroom-level controls including teacher gender, education, and experience. We also control for the type of school—central or filial—as outcomes may vary across these school types given different levels of remoteness and access to resources.

STUDY LIMITATIONS

The above analysis is associated with a number of limitations, some of which are described here.

Non-random assignment: While sampled schools were selected using stratified random sampling, the list of all eligible target schools in the four program municipalities were not selected at random. Target schools were selected based on vulnerability criteria such as the absence of similar interventions, overall education outcomes, location, and distance from a main road. The sample design has paired treatment and comparison schools such that they are as balanced as possible in terms of several potentially relevant characteristics: Comparison schools were selected in order to have a similar level of “remoteness” and to lack interventions focused on reading and school feeding.

However, treatment and comparison schools are not perfectly balanced—for example, in the baseline, heads of household in the comparison group were more likely to be educated than those in the treatment group. It is almost certain that treatment and comparison schools are also imbalanced in terms of other potentially important, but unobserved, factors that may bias analysis.

The main implication of this limitation is that, when making inferences based on these data, we cannot be absolutely certain that observed results are a product of program interventions and not at least partly a product of unobserved, systematic differences between the treatment and comparison groups. We attempt to mitigate this problem in our analysis using statistical controls in regressions to adjust findings for the influence of observable factors that are significantly different between treatment and comparison groups. However, we can never be certain that we have accounted for all potential confounders, and thus we can never claim that our estimates are completely unbiased.

Attrition and dropout: As with most longitudinal surveys, attrition poses a significant threat to drawing valid inferences. The COVID-19 pandemic has led to substantial decreases in school attendance and increases in dropout rates. Minimizing attrition remained a goal of data collection at the midline, but if the actual attrition rate exceeds the anticipated attrition rate embedded in the sample size calculation, the project's ability to assess impact will be compromised. Furthermore, if attrition is biased—if, for example, the students most likely to drop out are those performing worst in classes—results may not be representative of program impact. To attempt to address these issues, the evaluation team prescribed a set of formal procedures to attempt to recontact students.

Heterogeneous effects of COVID-19: The COVID-19 pandemic has had a substantial impact on households, schools, and the government in Timor-Leste. The national budget, for example, was not approved until December, resulting in most municipalities not implementing school feeding. Of more concern for research validity, however, are heterogeneous effects of the pandemic. For example, dates of school closing and reopening have varied across municipalities and strategies to enforce social distancing have varied by school, with some schools adopting class shifts on alternate days or weeks. These varying strategies to mitigate the effects of COVID-19 may significantly affect key program outcomes such as literacy and school attendance, making it difficult to determine results driven by program activities as opposed to those driven by responses to COVID-19. Collecting data on the COVID-19 response at the school level may help disentangle these effects.

Estimating attendance – inaccuracy of school record-keeping: Collecting attendance data for both students and teachers from school records can be challenging, as school records of attendance are often of poor quality and consist of either partially or entirely incomplete records. Furthermore, due to COVID-19, many schools have split classes into shifts, which has affected attendance and enrolment and may result in confusing or inaccurate record-keeping. The midline evaluation triangulates attendance across multiple sources to provide an overall picture of attendance rates, rather than relying on school records for a precise count of attendance over the previous year.

Accessing schools and respondents: Families in Timor-Leste frequently travel for extended periods of time in order to attend traditional ceremonies, which may make it difficult to contact some respondents for household surveys or to administer the learning assessment. Additionally, field work was conducted during the rainy season; as a result, some schools in remote areas were only accessible by foot or during specific times of day. This increased the time needed for data collection, and resulted in some remote schools only being visited at the end of data collection.

Social desirability bias: Some respondents' answers, especially to questions that are potentially sensitive, may not be wholly accurate or truthful. In cases where respondents are asked to self-report on behaviors and practices, there is often a strong desire to respond in a socially desirable manner. For example, parents may recognize that it is socially desirable for children to spend only a limited amount of time on household tasks; as such, rates of child participation in household labor may be underreported. While the design of the instruments and the interview process attempted to account for this by using clear language and creating a comfortable environment for respondents, response

bias is unavoidable. In the report, we note instances where this may have occurred, and triangulate responses for validation wherever possible.

Errors or limitations in data: Wherever inconsistent patterns were observed or data was not properly recorded, the data was removed from the analysis. An example of this occurred with the student age variable, where some second grade students' ages were unusually high or recorded as 99 ("don't know"). This reduces the sample size for some variables.

Additionally, some variables were recorded correctly at baseline but not at midline, or vice-versa. Data on the gender of the head of household, for example, is missing for some households at baseline and all households at midline. This limitation reduces our ability to compare results over time for some variables, or to disaggregate results. Specific limitations are noted in the relevant analysis sections.

Generalizability of results: Sample sizes are too small within each municipality and language group to generalize the results at those levels, as powering the sample adequately to disaggregate results by municipality and native language would have required an extremely large number of interviews. As a result, this report disaggregates by study group and gender (where possible and relevant), but not by municipality or language.

Floor and ceiling effects: Within the EGRA and each of its subtasks, there is a minimum and maximum possible score. If subtasks are too easy for students, most scores will tend to be clustered around the maximum possible score with little variation; similarly, if subtasks are too hard, most scores will be clustered around 0% with little variance. Floor and ceiling effects can dampen our ability to differentiate between intervention and comparison schools, thus reducing our ability to draw conclusions about the potential impact of the program in intervention schools.

LITERACY RESULTS

OVERALL LITERACY SCORES

Students who participated in the Early Grade Reading Assessment (EGRA) were assessed in five areas: letter name knowledge, reading invented words, reading familiar words, passage reading, and reading comprehension. Each task was scored as a percent of items correctly answered out of the total number of items (100 letters, 60 invented/familiar words, 60 or 61 passage words,⁸² and 10 reading comprehension questions in two groups of five) within a time limit of one minute, except for the reading comprehension task, which was not timed. An overall literacy score was also calculated as the simple average of the five task scores. This means that all subtasks are equally weighted in the calculation of the overall score; however, individual test items have different weights, since each subtask includes a different number of total questions.

Table 12 presents a summary of overall scores for the new cohort of midline students. Among the new cohort of midline students, the average literacy score was 7%. In comparison, the average literacy score at baseline was 12%. Overall literacy scores worsened for both treatment and comparison groups. This pattern of decreasing scores from baseline to midline for both comparison and intervention groups is followed across all subtasks: On average, students performed worse at midline on every subtask compared to baseline. The standard deviation for both the overall scores and all sub-tasks is generally high—often greater than the mean scores themselves—indicating that scores tend to vary substantially, with many students receiving scores far from the mean, and relatively few students scoring close to the mean.

Students' performance declined in terms of both mean scores and the percent of students scoring zero points on each subtask;⁸³ in other words, more students were unable to, for example, identify a single letter at midline than at baseline. However, the mean score of students who scored greater than zero points on a subtask improved for all intervention schools, and although the mean score of these students still declined in comparison schools, the scores declined by fewer percentage points when compared to the scores of all students. This pattern suggests a dichotomous effect on learning: At midline, more students are getting left behind entirely and are unable to read a single letter or word, but, in contrast, scores are stable or improving among students who have some literacy ability. We examine this finding further in the sections for each subtask below and the section "Impact of COVID-19 on Learning."

⁸² At baseline, the passage was 60 words long. At midline, the passage was 61 words long, an increase made in order to ensure the passage was logical and complete.

⁸³ Notably, students who were unable to identify a single letter (i.e., received a score of zero on letter name knowledge) were not asked to identify any invented or familiar words. Similarly, students who were unable to read a single invented or familiar word were not asked to read the passage. As a result, a student who received a zero score on these subtasks also received a zero score on subsequent subtasks.

Table 12: Summary of literacy scores, cross-sectional cohort

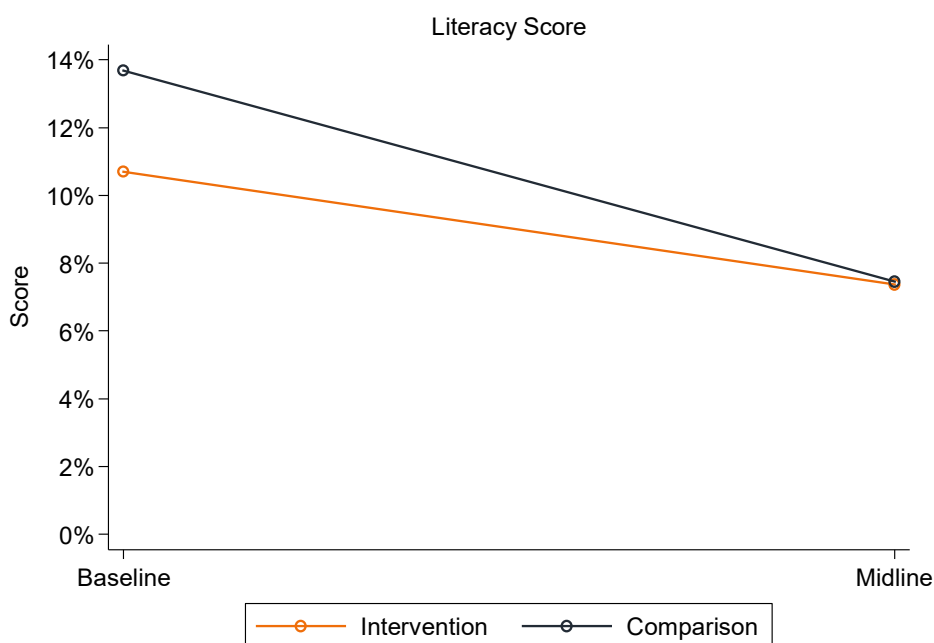
Comparison Schools				Intervention Schools		
Overall score	BL	ML	Difference	BL	ML	Difference
Number of respondents	1014	1108		1447	1474	
Mean score	13.7	7.5	-6.2	10.7	7.4	-3.3
Standard deviation	17.3	13.0	-	14.8	12.7	-
Letter name knowledge	BL	ML	Difference	BL	ML	ML
Mean score	22.3	16.0	-6.3	19.5	15.9	-3.6
Standard deviation	18.0	17.6	-	17.2	16.4	-
Percent zero scores	9.0	31.1	22.1	8.7	25.9	17.2
Mean score without zeros	24.5	23.2	-1.3	21.3	21.5	0.2
Invented word reading	BL	ML	Difference	BL	ML	Difference
Mean score	8.4	4.1	-4.3	6.1	4.4	-1.7
Standard deviation	13.6	9.6	-	11.3	11.0	-
Percent zero scores	60.3	77.8	17.5	64.0	78.7	14.7
Mean score without zeros	21.2	18.3	-2.9	17.0	20.5	3.5
Familiar word reading	BL	ML	Difference	BL	ML	Difference
Mean score	12.6	5.5	-7.1	9.6	5.3	-4.3
Standard deviation	18.3	12.6	-	15.5	13.2	-
Percent zero scores	54.1	77.3	23.2	58.2	79.5	21.3
Mean score without zeros	27.5	24.1	-3.4	22.9	25.8	2.9
Passage reading	BL	ML	Difference	BL	ML	Difference
Mean score	11.2	5.2	-6.0	8.2	5.3	-2.9
Standard deviation	20.1	14.1	-	17.0	14.8	-
Percent zero scores	62.7	81.4	18.7	69.5	82.8	13.3
Mean score without zeros	30.1	27.9	-2.2	26.8	30.9	4.1
Reading comprehension	BL	ML	Difference	BL	ML	Difference
Mean score	13.9	6.6	-7.3	10.2	6.0	-4.2
Standard deviation	21.4	18.5	-	21.4	17.7	-
Percent zero scores	70.3	85.3	15.0	76.5	87.1	10.6
Mean score without zeros	46.7	44.6	-2.1	43.2	46.3	3.1

While these results generally suggest a decline in learning outcomes for many students, a large portion of this decline in scores is likely to be attributable to external factors—namely, the COVID-19 pandemic and related school closures—rather than to program impact. To better understand the impact of the HATUTAN program on literacy scores we use a difference-in-differences model comparing the change in scores among treatment groups at baseline and midline to the change in scores among comparison groups at baseline and midline. This allows us to disentangle the impact of the program from the more general, negative impact of COVID-19 on all students.

To understand these effects, we run a difference-in-differences regression model that analyzes how the change in literacy scores varies by round (baseline or midline) and treatment group (treatment or comparison). We find a significant and positive effect for the treatment group compared to the comparison group. In other words, while average scores for both groups declined at midline compared to baseline, average scores for treatment students exposed to the program declined *less than* those for comparison students. Average scores for the comparison group declined by around 6 points (from 14% to 8%), while average scores for the treatment group declined by only around 3 points (from less

than 11% to 7%) (see Figure 1). This suggests that, in the absence of program intervention, the treatment group would have had, on average, even worse scores on overall literacy at midline; it seems likely that the HATUTAN program had some positive effect in mitigating the negative impacts of COVID-19 on learning.

Figure 1: Change in literacy scores, cross-sectional cohort



To test the robustness of this finding, we add variables to the regression to control for differences in student age, gender, and whether a student's mother tongue is Tetum. We include these variables as even small differences in, for example, Tetum abilities across subgroups might bias our results because Tetum speakers are more likely to have greater literacy ability, especially in early grade, due to Tetum's status as the language of instruction and examination (see "Methodological Analysis" for further explanation). In general, older students, female students, and students who speak Tetum perform better on the EGRA; any differences in these demographic variables across baseline and midline or across treatment and comparison groups may therefore bias our results. After adding these control variables, we find that there remains a statistically significant, positive, and substantive effect for the treatment group compared to the control group. As above, on average, at midline, the treatment group performed around 3 percentage points better than would be expected compared to the comparison group, even when controlling for demographic differences.

Finally, we add additional variables to the regression to control for any school-specific differences that may have varied across groups and biased results, as well as for student-specific differences. We do so by adding school fixed effects, which control for any observed or unobserved differences across schools that may affect student learning, such as teacher quality or availability of learning resources. Findings are robust to the inclusion of these variables; in fact, our estimate of the program's impact actually increases in our most conservative model. Using this model, we find that the treatment group performed around 4 percentage points better than would be expected compared to the comparison group, a significant difference.

To further examine this finding, we utilize the panel dataset of all students who were assessed at baseline and then re-contacted and assessed again at midline. Table 13 presents a summary of scores for this cohort of students who were re-contacted from the baseline. In contrast to the declining scores for the new cohort of midline students, among this group, scores generally improved at midline—as

expected, given that these students have been exposed to an additional 1.5 years of learning. There was a substantial increase in overall literacy scores and scores on all subtasks, and a decrease in the number of students scoring zero on each subtask.

Table 13: Summary of literacy scores, recontacted students

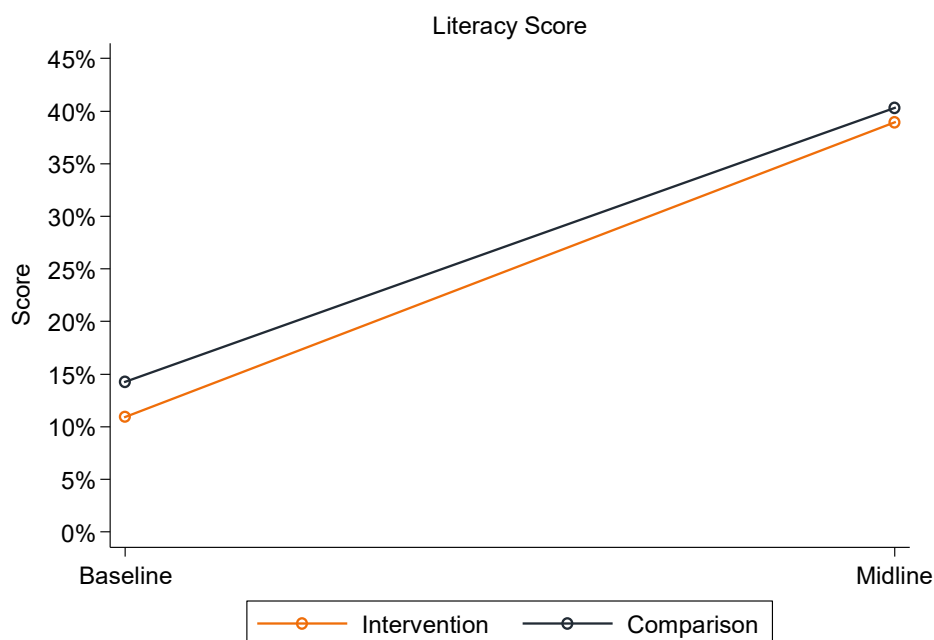
	Comparison Schools			Intervention Schools		
Overall score	BL	ML	Difference	BL	ML	Difference
Number of respondents	848	848	-	1221	1221	-
Mean score	14.3	40.3	26.0	10.9	38.9	28.0
Standard deviation	17.6	26.7	-	14.9	25.3	-
Letter name knowledge	BL	ML	Difference	BL	ML	Difference
Mean score	23.0	41.7	18.7	19.9	39.6	19.7
Standard deviation	18.3	22.9	-	17.1	20.7	-
Percent zero scores	8.1	3.3	-4.8	7.6	4.2	-3.4
Mean score without zeros	25.1	43.2	18.1	21.5	41.3	19.8
Invented word reading	BL	ML	Difference	BL	ML	Difference
Mean score	8.7	28.5	19.8	6.2	27.7	21.5
Standard deviation	13.8	23.8	-	11.3	22.9	-
Percent zero scores	58.8	22.9	-35.9	63.7	21.4	-42.3
Mean score without zeros	21.2	37.0	15.8	17.0	35.3	18.3
Familiar word reading	BL	ML	Difference	BL	ML	Difference
Mean score	13.1	37.8	24.7	9.7	37.3	27.6
Standard deviation	18.6	29.6	-	15.6	28.6	-
Percent zero scores	52.8	22.1	-30.7	57.7	21.1	-36.6
Mean score without zeros	27.8	48.5	20.7	23.0	47.3	24.3
Passage reading	BL	ML	Difference	BL	ML	Difference
Mean score	11.7	47.4	35.7	8.4	46.4	38.0
Standard deviation	20.6	37.5	-	17.2	36.3	-
Percent zero scores	61.7	24.1	-37.6	69.0	22.1	-46.9
Mean score without zeros	30.6	62.4	31.8	27.0	59.5	32.5
Reading comprehension	BL	ML	Difference	BL	ML	Difference
Mean score	14.7	46.1	31.4	10.4	43.6	33.2
Standard deviation	25.4	32.0	-	21.7	31.4	-
Percent zero scores	69.1	25.5	-43.6	76.0	25.6	-50.4
Mean score without zeros	47.6	61.8	14.2	43.3	58.5	15.2

The overall substantial improvement in scores in both groups can primarily be attributed to their exposure to an additional 1.5 years of schooling since baseline; however, as above, differences in the amount of improvement between the treatment group, who were exposed to HATUTAN programming, and the comparison group can help us better understand program impact. Furthermore, because this dataset includes the same students assessed at two different times, we can be more confident that any findings are due to program impact rather than to unobserved and uncontrolled-for differences between baseline and midline student groups.

Using a panel regression that compares the EGRA results of treatment and comparison groups with recontacted students across baseline and midline, we find a positive but smaller overall effect: On average, the treatment group performed around 2 percentage points better than would be expected

given the results of the comparison group (see Figure 2). However, this result is not significant. The result remains insignificant when controlling for differences in age, gender, and native language across treatment and control groups, and when controlling for potential school-specific differences.

Figure 2: Change in literacy scores, recontacted cohort



Overall, the findings from the cross-sectional and panel data suggest that while HATUTAN programming may have had some effect on mitigating the negative impacts of COVID-19 on learning outcomes, these findings are not entirely conclusive (see Table 14 for a summary of results). We find that intervention students performed on average 2 to 4 percentage points better at midline than expected given the results of comparison students. This corresponds to a small but meaningful improvement in literacy; for example, for a student who could only read letters but scored 0 on every other subtask, a 3 percentage point increase in overall score would correspond to 15 more recognized letters. However, there are significant differences in cross-sectional and panel students that may have implications for results: For example, students in the cross-sectional cohort (those randomly selected at midline) are significantly younger than students in the panel cohort (those re-contacted at midline). It is possible that HATUTAN programming is more effective for younger students or students in earlier grades, thus resulting in significant effects on learning outcomes for these students, but not for the students in the panel data. This could occur if the program's impact is greatest for students at a relatively low level of reading ability (i.e., those who may be able to identify some letters, but lack substantial other skills), but does not help students advance to high levels of literacy (such as reading comprehension).

Table 14: Summary of literacy score changes

	Cross-sectional cohort		Panel cohort	
	Difference in differences	p	Difference in differences	p
No controls	2.9	0.04*	2.0	0.21
Student-level controls	3.1	0.04*	2.6	0.12
Student- and school-level controls	3.7	0.01*	2.2	0.18

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In addition to analyzing the changes in literacy scores over time for all students, we further disaggregate by gender to uncover any potentially heterogeneous program impacts on male and female students. At baseline, female students were found to have small but significantly better literacy scores than male students. At midline, among the new cohort of students, Table 15 shows that female students still perform better than male students, but, as above, scores for both male and female students decreased on average. Furthermore, the gap between male and female students seems to have decreased slightly for this cohort.

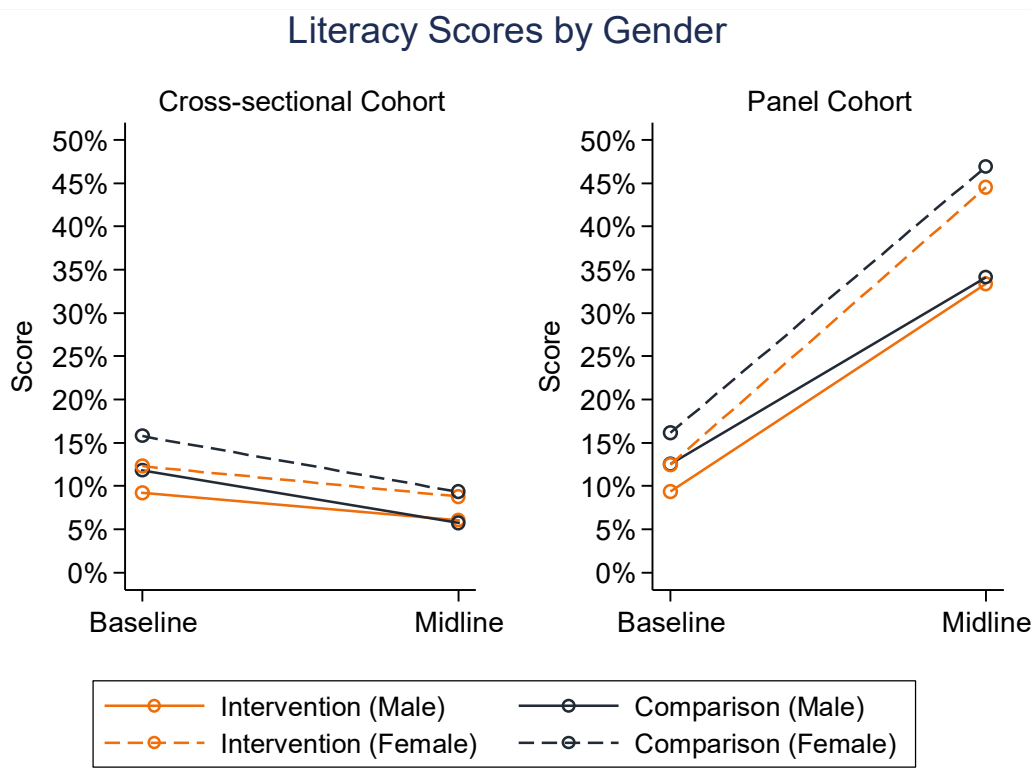
Table 15: Overall literacy scores by gender and cohort

	Comparison Schools			Intervention Schools			Difference in Differences	
Male	BL	ML	Difference	BL	ML	Difference	DiD	p
Number of respondents	533	574		748	749			
Mean score	11.8	5.7	-6.1	9.2	6.0	-3.2	2.9	0.07
Female	BL	ML	Difference	BL	ML	ML	DiD	p
Number of respondents	481	534		699	725			
Mean score	15.8	9.3	-6.5	12.3	8.8	-3.5	2.9	0.11

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

An analysis of both the cross-sectional and panel data suggests that HATUTAN programming may have had some effect at mitigating the negative impacts on COVID-19 on learning outcomes for both genders—as evidenced by somewhat better learning outcomes for treatment students than would be expected given comparison students—but that, as above, findings are not fully conclusive (see Figure 3). As with the aggregate scores discussed above, evidence of impact is stronger among the cross-sectional group than the panel group: Among the cross-sectional cohort, male and female students in treatment groups both scored on average 3 percentage points higher than would be expected given the results of the comparison groups. These results are not significant when differences in groups are not controlled for or when only student-specific differences are controlled for, but become significant for male and female students, with a higher effect size of around 4 percentage points, when controlling for both student- and school-specific differences. In contrast, among the panel cohort, the effect size was smaller for female students—who had only around a 1 percentage point improvement in scores for the treatment group compared to the comparison group—than for male students, who scored around 3 percentage points better than would be expected given the comparison group's results. The results for male students in this cohort were significant when controls were added for student-specific differences in groups, but were not significant for the other regression specifications.

Figure 3: Change in literacy scores by gender



Overall, these results do not conclusively suggest that the program had differential impacts by gender. While female students in the cross-sectional cohort appear to have fared slightly better than male students, in the panel data, scores for male students appear to have improved more than scores for female students, although these results are not significant. Overall, there remains a large gap in scores between male and female students which HATUTAN programming seems not to have affected.

LITERACY SUBTASK RESULTS

To better understand the effects of HATUTAN programming on literacy, we now analyze changes in scores on specific sub-tasks.

LETTER NAME KNOWLEDGE

For the letter name knowledge subtask, students were given a list of 100 letters and asked to read as many as possible within a time limit of one minute. They were then scored based on the number of letters they were able to read accurately within the time limit. Overall, 28% of all students tested in the new midline cohort could not read a single letter. The mean letter name knowledge score among this cohort was 16%. In contrast, at baseline, only 9% of students were unable to read a single letter, and the mean letter score was 21%.

At baseline, results suggested that students knew the names of letters relatively well—in general, when presented with a letter, far more students tended to name that letter correctly than incorrectly—but struggled with fluency, as they were not able to name very many letters within one minute. Midline

results confirm this finding: An analysis of the first 20 letters of the task⁸⁴ show that most students did not identify letters incorrectly. Rather, most low overall scores are due to low reading speed, suggesting low levels of fluency.

Table 16: Accuracy of letter recognition

Letter	Comparison Schools		Intervention Schools	
	n	% accurate	n	% accurate
m	804	77.2	1,163	75.7
i	808	85.5	1,156	84.3
a	731	63.9	1,065	59.3
L	751	80.0	1,110	76.9
T	749	83.3	1,106	80.4
s	766	88.6	1,116	85.0
u	740	83.8	1,074	82.1
N	721	81.3	1,057	76.7
e	683	79.1	1,012	73.8
R	690	85.4	1,006	82.7
B	675	93.6	1,001	89.4
o	672	92.9	982	90.1
k	645	91.0	945	87.9
t	584	71.8	872	65.0
d	560	65.4	831	57.4
v	533	74.1	786	67.3
E	528	91.5	773	86.9
F	510	90.0	731	83.7
U	499	92.8	710	90.6
N	494	90.5	689	87.7

As with overall scores, in general, students in the cross-sectional cohort performed worse on letter name knowledge at midline than at baseline due, most likely, to the impact of COVID-19 on learning. However, the difference-in-differences regression analysis of the change in scores by treatment and control group suggests that HATUTAN programming had an inconclusive, but possibly positive, effect on students' knowledge of letter names (see Table 17). Among the cross-sectional cohort, on average, students in treatment schools scored around 3 percentage points higher than would have been expected given the results of students in comparison schools. In treatment schools, the average score at baseline was 19.5, while the average score at midline was 16, a decrease of around 3.5 points; in contrast, in comparison schools, the average score at baseline was 22, while the average score at midline was 16, a decrease of around 6 points. These results are not significant for the regression models without controls or with only student-level controls; however, the results become significant, and the effect size increases, when both student- and school-level controls are added to the model. Results for the panel cohort, as with the results for overall scores, suggest a generally much smaller and insignificant effect size of around 1 percentage point, although total letter scores for this cohort increased as expected due to exposure to an additional year of education.

⁸⁴ Only the first 20 letters are analyzed because most students did not continue to read after 20 letters. Students who did continue to read after the first 20 letters are, in general, stronger readers, and are thus more likely to recognize letters correctly, potentially biasing our findings.

Table 17: Letter name recognition, difference-in-differences results

	Cross-sectional cohort		Panel cohort	
	Difference in differences	p-value	Difference in differences	p-value
No controls	2.8	0.10	1.0	0.51
Student-level controls	2.8	0.09	1.5	0.36
Student- and school-level controls	3.4	0.04*	1.1	0.48

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In addition to this analysis of scores for all students within each cohort, we further analyze potential differences in program impact between all students in the cross-sectional cohort and the group of students who scored greater than zero on this subtask—i.e., those students with some literacy ability. As discussed above, at midline, more students are getting “left behind” entirely and are unable to read a single letter. However, within treatment schools, while there is still a larger percent of students unable to read letters at midline than at baseline, there are fewer of these students than in comparison schools. Figure 4 shows that among treatment and comparison schools, a similar percent of students (around 9%) were unable to recognize any letters at baseline. In contrast, at midline, far more students in the comparison group—31%—were unable to recognize any letters compared to the intervention group (26% of students).

Figure 4: Letter name knowledge, zero scores



In contrast, a difference-in-differences analysis of just students who scored greater than zero on this subtask suggests that the HATUTAN program may not have had a substantial effect on these students' letter recognition ability. The regression analysis shows that the students in the treatment group who scored above zero scored only around 1.4 percentage points higher at midline than would be expected given the results for comparison group students who scored above zero—substantially less than the 3 percentage point difference found above for the entire cross-sectional cohort. This score difference

is not significant and remains insignificant when controls are added for student- and school-level differences.

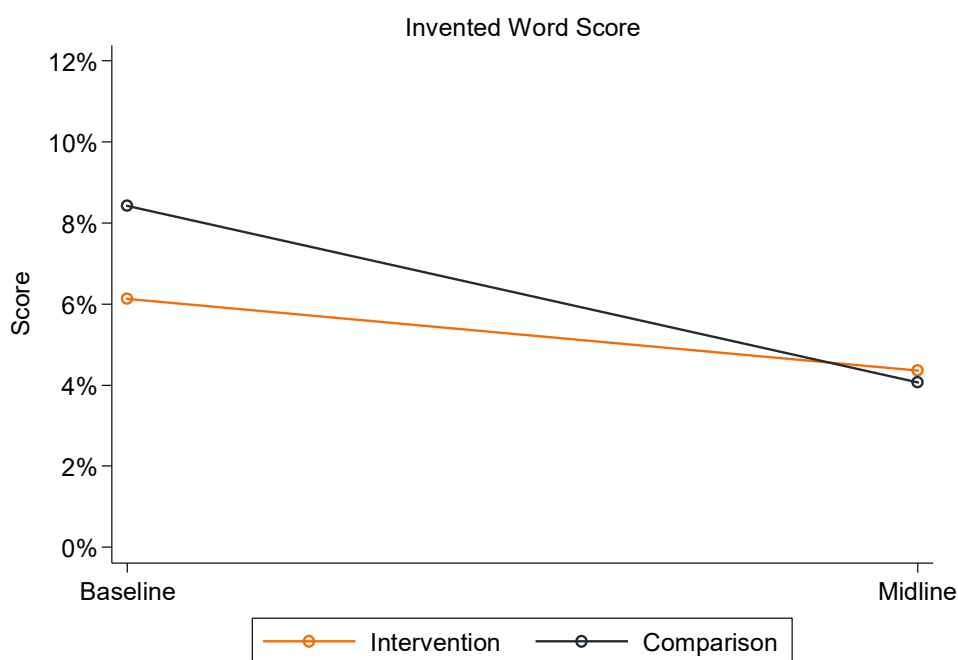
These results suggest that the HATUTAN program may be relatively more effective at reducing the number of students with no letter recognition ability, rather than improving the letter recognition skills of students who already have some ability in this area. The results also reinforce that the HATUTAN program likely mitigated some of the negative impact of the COVID-19 pandemic, but that this effect was heterogenous. In this case, it seems that the HATUTAN program had more of an impact on more disadvantaged students—those who were more likely to be “left behind” entirely—than on students who faced fewer challenges to learning.

INVENTED WORD FLUENCY

For the invented word fluency subtask, students were given a list of 60 invented words and asked to read as many as they could within a time limit of one minute. They were then scored based on the number of invented words they were able to read correctly within the time period. On average, students performed worst on this subtask. At midline, among the cross-sectional cohort, 78% of students were unable to read a single invented word. The mean score for this cohort at midline was only 4%.

Figure 5 shows, however, that the HATUTAN program seems to have been relatively successful at reducing the decline in invented word fluency among students in the intervention group compared to the comparison group. Among the comparison group, scores declined from around 8% at baseline to around 4% at midline, a 4 percentage point decrease. In contrast, among the treatment group, scores declined from around 6% at baseline to around 4% at midline, a decrease of only around 2 percentage points. The difference-in-differences regression analysis produces significant results for the model without control variables and for the models with student-specific and school-specific controls, with a slightly higher effect size of 3 percentage points for the model with both student- and school-specific control variables (see Table 18). In other words, controlling for differences in student and school characteristics across groups, at midline, the treatment group performed around 3 percentage points better than expected given the results of the comparison group.

Figure 5: Change in invented word scores



As above, however, among the panel cohort, results are less conclusive. For this cohort's comparison group, at baseline, the average invented word fluency score was 9%, while at midline, the average score was 29%, an increase of around 20 percentage points. Similarly, for the intervention group, at baseline, the average invented word fluency score was 6%, while at midline, the average score was 28%, an increase of slightly more than 21 percentage points. While this suggests that the HATUTAN program may have had a slight effect on invented word scores among the treatment group, these results are not significant (Table 18).

Table 18: Invented word fluency, difference-in-differences results

	Cross-sectional cohort		Panel cohort	
	Difference in differences	p-value	Difference in differences	p-value
No controls	2.6	0.02*	1.8	0.18
Student-level controls	2.8	0.02*	2.2	0.11
Student- and school-level controls	3.2	0.01*	1.0	0.17

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

It is worth noting that despite these results suggesting that the HATUTAN program may have had some positive, though inconclusive, impact on invented word fluency, average scores remain extremely low, and the majority of grade 2 students remain unable to read any invented words. Invented words require a strong understanding of phonemes in order to apply them to words that are unfamiliar and lack any meaning; this suggests that students may need more work in recognizing the sounds that specific letter groups make.

In contrast to the results for letter name knowledge, HATUTAN programming also had a less conclusive impact on the percent of zero scores for invented word fluency among the cross-sectional cohort. While around 3 percentage points more treatment students were able to recognize any invented words than expected given the results from the comparison group, this result was not significant. However, a difference-in-differences analysis of students who scored greater than zero on this subtask suggests that the program had a significant and positive effect on scores among these students. On average, among this cohort of non-zero scorers, students in the treatment group scored around 6 percentage points higher than would be expected given the results in the comparison group. This result is significant for all the regression specifications. These results suggest that, unlike the letter name recognition results, the HATUTAN program was more effective at improving invented word recognition skills among students with some ability in this area than at improving skills for those students with the lowest levels of knowledge.

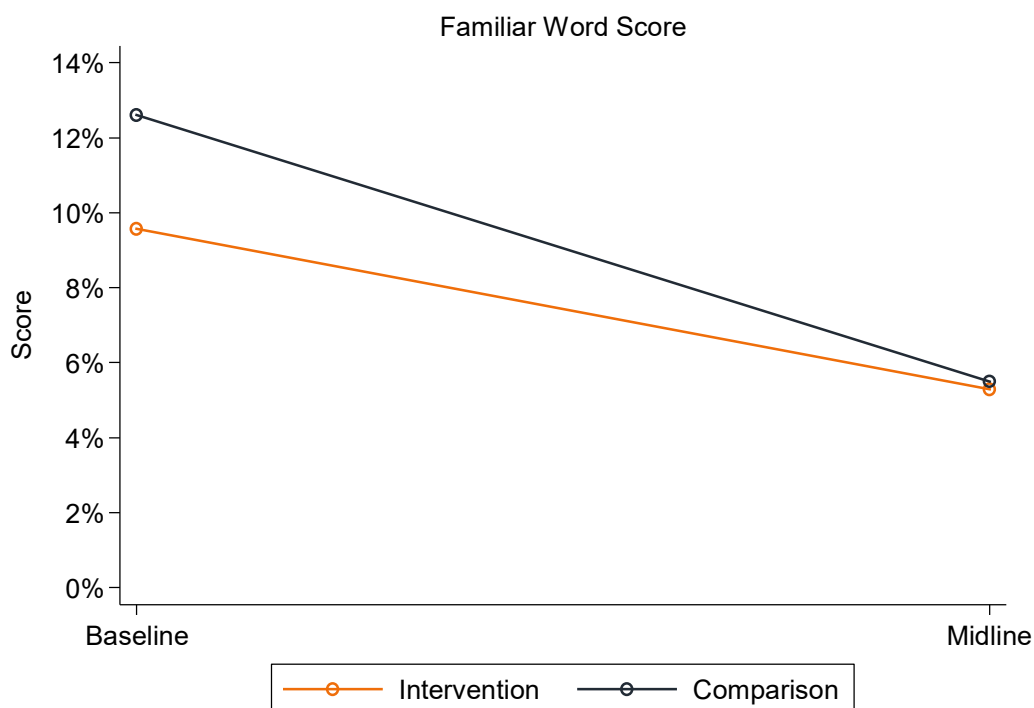
FAMILIAR WORD FLUENCY

As with the invented word fluency subtask, for the familiar word fluency subtask, students were given a list of 60 familiar words and asked to read as many as they could within a time limit of one minute. They were then scored based on the number of invented words they were able to read correctly within the time period. Among the cross-sectional cohort, 79% of students were unable to read a single familiar word at midline, similar to the results for invented word fluency. The mean familiar word fluency score for this cohort at midline was 5%.

These results stand in contrast to those at baseline, in which students generally performed substantially better on the familiar word fluency task than on invented word fluency. The convergence of the familiar and invented word fluency scores at midline suggests that there may be a floor effect to the scoring: Students cannot score less than 0%, but because students have very low levels of reading ability on average, most students score 0%, and there is relatively little variance in scores around a low level. This floor effect can dampen our ability to differentiate between intervention and comparison schools and to understand the potential impact of the program in intervention schools, and applies not only to familiar word fluency, but to subsequent subtasks as well as overall scores.

Figure 6 shows that, as with most results so far, the HATUTAN program appears to have had some mitigating effect on the worsening of learning outcomes over the past year. For the comparison group, average scores decreased by over 7 percentage points from baseline to midline, while for the treatment group, average scores decreased by only around 4 percentage points. This result is significant when controls are added for both student- and school-specific differences across groups (see Table 19).

Figure 6: Change in familiar word fluency scores



As above, however, the results from the panel data are inconclusive. For this cohort of students, midline average scores were relatively similar for both treatment and comparison groups, at 37% and 38% respectively. However, the treatment group scored relatively worse on average at baseline (10%) than the comparison group (13%), and therefore improved more than the comparison group over time. These results, however, are not significant (Table 19).

Table 19: Familiar word fluency, difference-in-differences results

	Cross-sectional cohort		Panel cohort	
	Difference in differences	p-value	Difference in differences	p-value
No controls	2.8	0.05	2.9	0.10
Student-level controls	3.1	0.05	3.5	0.06
Student- and school-level controls	3.7	0.02*	3.1	0.09

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

As discussed in the invented word fluency section, although the HATUTAN program appears to have had some effect on familiar word fluency, average scores are still extremely low. This indicates that second grade students still lack adequate development of sound recognition, as this is a fundamental skill needed for the recognition of both familiar and invented words.

Furthermore, as with invented words, HATUTAN programming did not have a significant effect on reducing the number of zero scores among students in the cross-sectional treatment group as opposed to the comparison group. However, the program had a significant and positive effect on improving scores among those students who did not score 0%. For these students, the difference-in-differences regression analysis suggests that average scores in the treatment group were around 6 percentage points higher than would be expected given the results of the comparison group. Indeed, at midline, scores for non-zero scorers improved in the treatment group, but worsened in the control group. When considered in conjunction with the similar results for invented word fluency, these results suggest that the HATUTAN programming may not be effectively improving teaching of phonemes and the relation between letter sounds and words to the weakest students, but may be improving these skills among students who already have some understanding of the concept.

Within the focus group discussions, a female teacher stated that many children have difficulties learning when words are long; however, if words are broken down into syllables, students can understand.⁸⁵ Similarly, a father stated that children often struggle with long words, and have to sound out the words letter-by-letter.⁸⁶ Additionally, a mother described how children may be able to read some letters in pairs—i.e., understand the sounds made by letter groupings, an important skill to be able to read familiar or invented words—but struggle with some combinations of letters.⁸⁷ Overall, these findings validate the idea that students lack adequate sound recognition of letters and letter groupings.

PASSAGE FLUENCY

For this subtask, students were given a 61-word passage⁸⁸ and asked to read as many words in the passage as they could in one minute. Students were scored based on the number of words they were able to read correctly in this time period. As expected from the low scores in both the invented and familiar word fluency subtasks, passage fluency scores were generally very low. Among all students in the cross-sectional cohort at midline, 82% of students could not read a single word. The average score at midline for this cohort was only 5%.

Notably, even among students who were able to read some invented and familiar words, 16% were unable to read a single word of the passage at both baseline and midline. For students who scored zero on passage fluency but greater than zero on both invented and familiar word fluency, scores on the invented and familiar word fluency subtasks were, on average, low (around 12% for invented words and 14% for familiar words), but not so low as to suggest that these students simply guessed words correctly but have no actual word recognition ability. Rather, these students seem to have some ability to read words in isolation, but struggle to read words in the context of a passage; future program activities that focus on training teachers should attempt to address this issue. This pattern also reinforces the finding that students generally have low levels of fluency: Although they may be able to identify individual letters or words, they struggle to apply those basic skills to more difficult reading tasks.

As with previous sub-tasks, HATUTAN programming seems to have had some success in mitigating negative trends in learning outcomes among the cross-sectional treatment group (Figure 7). For the comparison group, scores declined by around 6 percentage points from baseline to midline—from 11% to 5%. In contrast, among the treatment group, scores declined by only around 3 percentage points, from 8% at baseline to around 5% at midline. The difference-in-differences regressions show a significant difference in these outcomes across groups, with the treatment group performing around 3 percentage points better than would be expected given the results of the comparison group.

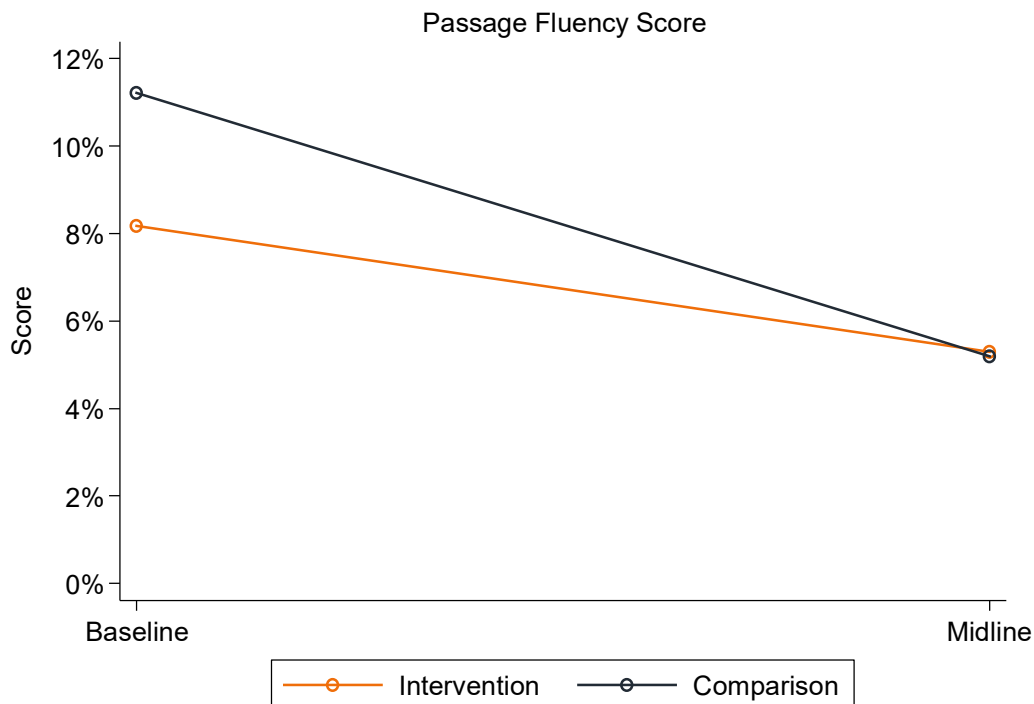
⁸⁵ FGD with teachers, female, Ermera municipality, Int. 104

⁸⁶ FGD with fathers, male, Ermera municipality, Int. 105

⁸⁷ FGD with mothers, female, Manatuto municipality, Int. 134

⁸⁸ At baseline, the passage was 60 words.

Figure 7: Change in passage fluency scores



For the panel cohort of students, scores were generally similar at midline across treatment and comparison groups, at 46% and 47% respectively. At baseline, the treatment group performed somewhat worse than the comparison group, at 8% and 12% respectively; the treatment group thus improved somewhat more than the comparison group by midline. However, this result is not significant (Table 20). Interestingly, at midline for this cohort, scores exhibited both a ceiling and a floor effect. Sixteen percent of students scored 100% on the subtask and 23% scored 0%, but the distribution of scores greater than 0% and less than 100% was relatively uniform. This distribution of scores suggests that, at later grades, some students continue to get “left behind” entirely and have very little literacy ability. However, current teaching practices may be relatively successful at improving the reading skills of students who already have some literacy ability—thus resulting in a relatively high percentage of students who are able to read the entire passage.

Table 20: Passage fluency, difference-in-differences results

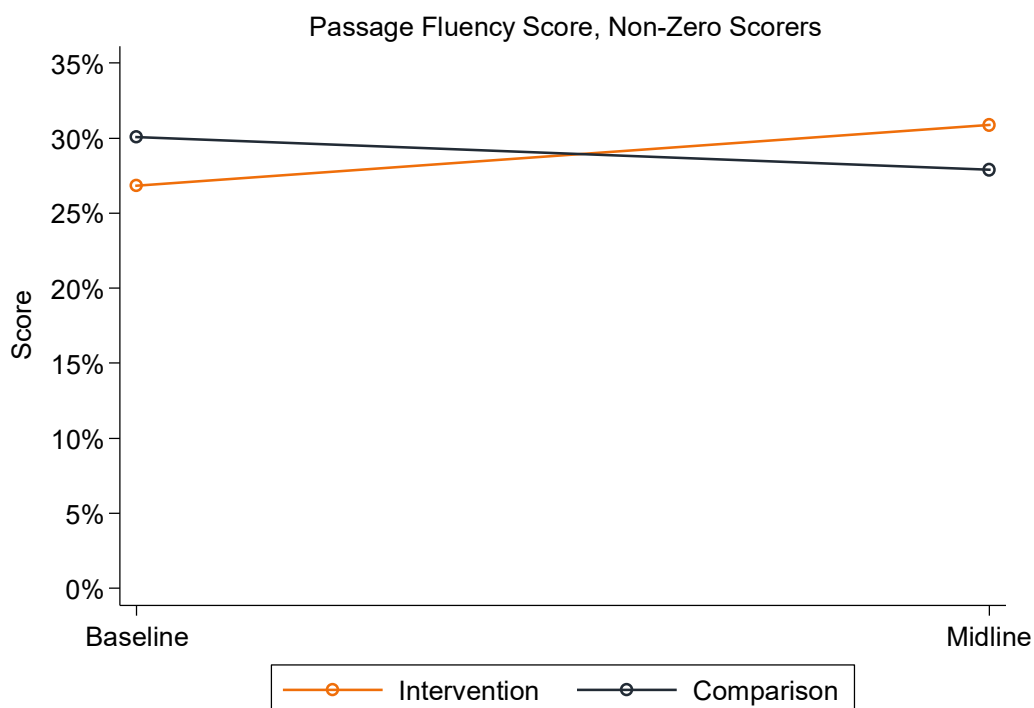
	Cross-sectional cohort		Panel cohort	
	Difference in differences	p-value	Difference in differences	p-value
No controls	3.2	0.04*	2.4	0.31
Student-level controls	3.4	0.04*	3.0	0.22
Student- and school-level controls	4.2	0.02*	2.6	0.29

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

There was no significant difference in the change in the percent of zero scores between treatment and comparison groups from baseline to midline. HATUTAN programming seems, however, to have had a significant and positive effect on passage fluency among students who scored greater than 0% (Figure 8). For these non-zero scorers, the difference-in-differences regression suggests that students in the treatment group scored around 6 percentage points higher at midline than would be expected given the results of non-zero scorers in the comparison group, and this effect size increases to 10

percentage points when controls for student- and school-specific differences are added to the regression. This difference in scores among non-zero scorers is significant.

Figure 8: Change in passage fluency scores, non-zero scorers



READING COMPREHENSION

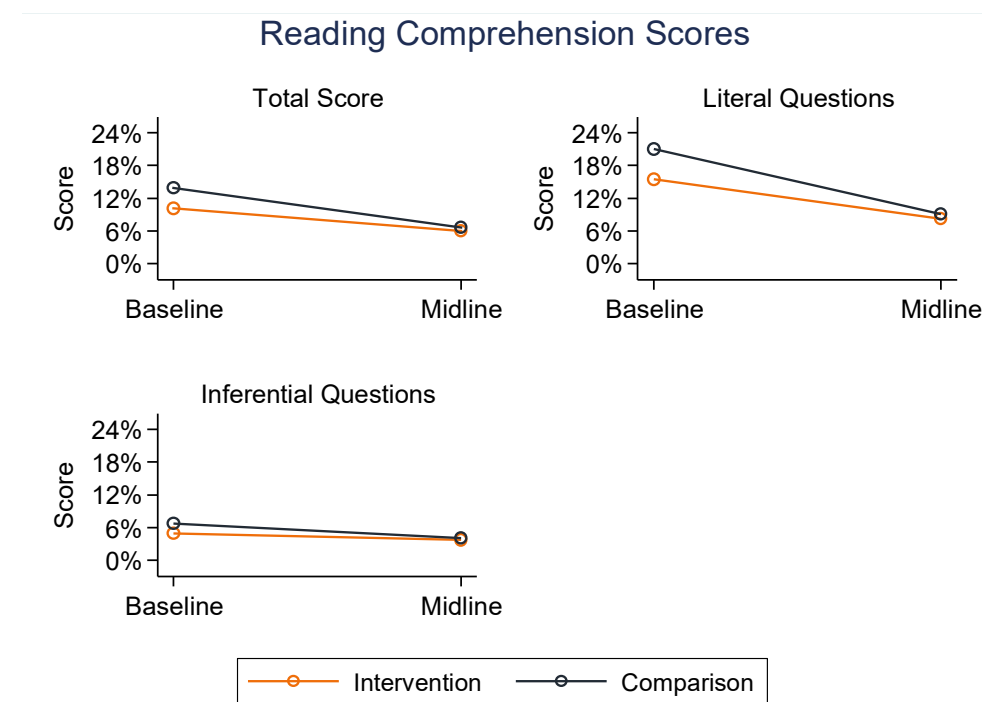
For the reading comprehension subtask, students were asked a total of 10 comprehension questions split into two groups of five. The first five questions were asked after reading the passage used for the passage fluency sub-task, and the second five after reading a second, unscored passage. Among the first passage, four out of five comprehension questions were literal (i.e., answers could be found within the text) and one was inferential (i.e., students were required to draw conclusions using clues in the text), while the second five included two literal and three inferential questions. We again note that, because there are only 10 total comprehension questions, scores on this section are relatively higher than might be expected given results on other subtasks, as a score of 20% only requires correctly answering two questions, rather than identifying, for example, 20 letters or 12 familiar words.

At midline, 86% of students in the cross-sectional cohort did not answer a single reading comprehension question correctly. The average score for the midline cross-sectional cohort was 6%. As expected, students performed relatively worse on the second, more difficult reading comprehension task than on the first: At midline, over 90% of students were unable to answer a single question on the second reading comprehension task compared to 87% on the first task, and the average score on the second task was 4% compared to 9% on the first task.

HATUTAN programming had a less conclusive impact on reading comprehension results than on most previous subtasks. While students in the treatment group performed, on average, about 3 percentage points higher at midline than would be expected given the results of the comparison group (Figure 9), these results were not significant for the difference-in-differences regression models without controls and with only student-specific controls. It is worth noting that, due to the limited number of questions

asked for reading comprehension, a 3 percentage point difference is relatively less substantial for this subtask than for the other subtasks.

Figure 9: Change in reading comprehension scores



Among the panel cohort, the difference-in-differences regression analysis showed no significant difference between the treatment and control groups' results from baseline to midline (Table 21). For this cohort, the treatment group's scores improved by around 33 percentage points at midline, while the comparison group's scores improved by around 31 percentage points—an average of around 3 more correctly answered questions at midline than at baseline. As expected, this cohort scored significantly better on the first five questions, with an average score of 61% for the comparison group and 59% for the treatment group at midline, than the second five questions, with an average score of 31% for the comparison group and 28% for the treatment group.

Table 21: Reading comprehension, difference-in-differences results

	Cross-sectional cohort		Panel cohort	
	Difference in differences	p-value	Difference in differences	p-value
No controls	3.1	0.09	1.8	0.42
Student-level controls	3.5	0.08	2.7	0.24
Student- and school-level controls	4.2	0.04*	2.4	0.31

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The McGovern-Dole Standard Outcome #1, related to improved literacy of school-aged children, is measured through an indicator for the percent of students who, by the end of grade two, demonstrate that they can read and understand the meaning of a grade-level passage. For this indicator, the general standard as measured through the EGRA is that students should respond to at least 80% of reading comprehension questions correctly; however, because a large portion of Timorese children are learning in a second language, students are considered to meet this standard if they answer at least one reading comprehension question correctly. Among the newly selected cohort of midline

students, slightly less than 14% of students were able to meet this criterion for all reading comprehension questions. Only around 2% of this cohort were able to answer at least 80% of all reading comprehension questions correctly, and only 7% were able to meet this criterion for the first reading comprehension test.

Unlike with other subtasks, HATUTAN programming had no significant effect on either the percent of students scoring zero on this subtask or on the reading comprehension of students scoring above 0%. It may be the case that HATUTAN programming does not sufficiently address reading comprehension skills. Alternatively, students may have too little general reading ability (in terms of letter and, especially, word recognition) for programming that targets reading comprehension to have a large impact.

IMPACT OF COVID-19 ON LEARNING

Overall, these results point first and foremost to a large, negative impact of COVID-19 on learning. Among the cross-sectional cohort, both treatment and comparison groups saw substantive declines in learning outcomes at midline compared to baseline. While results suggest that the HATUTAN program did help to mitigate this effect, the impact of COVID-19 and school closures is undeniable.

School closures were implemented uniformly on the national level by the Timorese government; the timing and extent of closures is thus unlikely to have had heterogeneous effects on treatment and control groups. However, municipal-level differences in characteristics such as access to water, access to electricity, and remoteness may have allowed some municipalities and schools to better cope with the negative impacts of COVID-19 than others. In municipalities where most students have a radio at home and a means to power that radio, for example, the *Eskola ba Uma* program may have more successfully allowed for at-home learning. Qualitative data suggests that this dynamic did occur: A school director mentioned that while the establishment of an at-home learning program had been positive, in many areas there is no electricity or children do not have access to televisions, and so students cannot access the program.⁸⁹

While it is difficult to fully unpack these complicated dynamics, we analyze the change in overall scores over time for the cross-sectional cohort in each municipality to better understand if learning outcomes in some municipalities may have been more affected by COVID-19 than others. Because treatment and comparison groups were divided by municipality—i.e., each of the nine municipalities in which data collection occurred was either assigned to the treatment group *or* to the comparison group, but not to both—a difference-in-differences regression cannot be run. Rather, we run a simple regression comparing scores across baseline and midline for the cross-sectional cohort, controlling for student- and school-specific differences.

Table 22 shows the changes in mean overall scores by municipality and suggests that students in some municipalities may have been more affected by COVID-19 than others. Most notably, in Baucau, the average overall score declined by nearly 28 percentage points from baseline to midline among the cross-sectional cohort (although we note that the sample size for this municipality was low, at only 48 students). In contrast, in Covalima and Manatuto, scores declined by essentially 0 percentage points.

⁸⁹ FGD with school directors, female, Ermera municipality, Int. 103

Table 22: Change in overall scores by municipality

Municipality	Group	n	Coefficient	p-value
Aileu	Comparison	238	-10.1	<0.001***
Ainaro	Treatment	362	-6.1	0.004**
Baucau	Comparison	48	-27.7	0.01*
Bobonaro	Comparison	409	-6.7	0.002**
Covalima	Comparison	168	-0.2	0.94
Ermera	Treatment	641	-3.0	0.009**
Liquiçá	Treatment	204	-7.4	0.002**
Manatuto	Treatment	267	0.2	0.90
Manufahi	Comparison	245	-6.6	0.002**

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In addition to this municipality-level analysis, we also analyze whether COVID-19-imposed school shifts had an impact on learning. Schools were required to hold classes in shifts in any cases where classrooms had more than 25 students in an effort to socially distance and reduce the potential spread of COVID-19. As a result, at midline, around 84 percent of all schools reported that grade 2 classes were operating in shifts.

Among the cross-sectional cohort, students in grade 2 classes operating in shifts scored slightly worse on average for overall literacy, at around 8%, than students in schools without shifts, who scored around 7%. These results, however, were not significant, and remained insignificant when controlling for differences in student gender, age, native language, and school characteristics across groups. These results suggest that, while class shifts may have had an impact on learning, the majority of the decline in learning outcomes may be due to school closures, rather than changes to the learning environment once students returned to school. It is also possible that schools that are larger, and thus more likely to have to operate in shifts, also have more resources due to their large size—for example, most central schools have better resources and more teachers due to their central locations, but may have been more likely to operate in shifts than filial schools. The opposite direction of these effects—the tendency for more advantaged schools to need to operate in shifts—makes it difficult to ascertain the full effect of shifts on learning outcomes.

Qualitative data, however, suggests that many schools struggled to effectively teach students due to COVID-19 restrictions and the need for shifts. For example, a teacher stated:

Previously we followed a teaching plan where each plan is allocated 50 minutes of learning time, but we no longer follow this. As there are so many children in this school one class accommodates many of them. If everyone must observe the social distancing then we have to distribute them. When we distribute them into first shift and second shift we cannot follow the plan anymore and this is a challenge.⁹⁰

A father similarly stated that with the state of emergency, schools operate for just two hours for students in some grades, and as a result, many lessons are missed.⁹¹ One school coordinator described compensating for missed lessons by organizing recovery lessons to teach children content that was missed due to COVID;⁹² however, no other coordinators or teachers mentioned a similar program in the qualitative interviews.

⁹⁰ FGD with teachers, male, Ermera municipality, Int. 110

⁹¹ FGD with fathers, male, Ermera municipality, Int. 132

⁹² FGD with school coordinators, male, Manatuto municipality, Int. 118

QUALITY OF INSTRUCTION

Quality of instruction is a key factor contributing to the improved literacy of students, and includes consistent and frequent teacher attendance, access to school supplies and materials, high-quality and accessible literacy instruction materials, and skilled and knowledgeable teachers and school administrators. For both the baseline and midline surveys, quality of instruction was measured through classroom observations conducted in grade 2 Tetum language classes. Data collectors observed whether teachers used engaging or ineffective teaching practices and whether there was gender bias in teaching practices, as well as collecting data on teacher attendance, school supplies, and the backgrounds of teachers and school administrators.

TEACHING PRACTICES

Engaging teaching practices include asking open questions, reading to students, calling on inactive students to engage them, using games or exercises, asking students' opinions, having students participate in reading activities with others, having students read by themselves, having students work together in groups, using a reading corner for literacy activities, encouraging students, and asking questions to students. Ineffective teaching practices, which have traditionally been overused in Timorese schools, include having students spend most of their time copying from the board and having students spend most of their time repeating the teacher. Negative teaching practices include using an angry voice or harsh tone with students and using corporal punishment against students.

On average, at midline, there was little change in the use of engaging teaching practices in either treatment or control schools. At baseline, teachers in both control and treatment schools used an average of 4.8 engaging teaching practices. At midline, teachers in control schools used an average of 4.4 engaging teaching practices, while those in treatment schools used 4.5 on average. Table 23 shows that, while the average number of engaging teaching practices did not change substantially at midline, the distribution of the number of teaching practices used changed somewhat.⁹³ Across both treatment and control groups, at midline, teachers were somewhat more likely to use no engaging teaching practices and four, five, or six engaging teaching practices, and somewhat less likely to use seven or eight engaging teaching practices.

Table 23: Use of engaging teaching practices (% of classrooms using given number of practices)

	Comparison Schools		Intervention Schools	
	BL	ML	BL	ML
n	45	87	98	98
Average # practices	4.8	4.4	4.8	4.5
0 practices	0.0%	1.2%	0.0%	4.1%
1 practice	8.9%	3.5%	6.1%	10.2%
2 practices	8.9%	13.8%	10.2%	7.1%
3 practices	11.1%	14.9%	19.4%	9.2%
4 practices	15.6%	19.5%	12.2%	12.2%
5 practices	15.6%	13.8%	10.2%	24.5%
6 practices	13.3%	23.0%	17.4%	14.3%
7 practices	15.6%	4.6%	8.2%	9.2%
8 practices	8.9%	4.6%	15.3%	7.1%
9 practices	2.2%	1.2%	1.0%	2.0%

⁹³ The nine practices referred to in this table exclude "encouraging students" and "questioning students." Data for these two teaching practices was collected separately for boys and girls, and is discussed further below.

The difference-in-differences regression analysis confirms that there was no significant change in the number of engaging teaching practices used at midline in treatment schools as compared to comparison schools. For this analysis, two specifications are used: a regression that does not control for any potential differences across treatment and control or baseline and midline groups, and a regression that controls for the education and total experience of the teacher, the gender of the teacher, the number of students in the classroom, and the type of school (central or filial), all of which may affect the use of engaging teaching practices. For both of these specifications, we find that teachers in treatment schools only used around 0.1 more engaging teaching practices on average than would be expected given the use of these practices in control schools; this finding was not significant in either regression.

In addition to this regression analysis of all schools in which a classroom observation was conducted at either baseline or midline, we also restrict the sample to just schools in which a classroom observation was conducted at *both* baseline and midline. This allows us to control for any potential unobserved differences between schools which may affect the use of engaging teaching practices and thus bias our results. This restricted sample includes all 98 treatment schools and 43 comparison schools that were assessed at both baseline and endline. This analysis confirms the findings above: There were no significant changes in the number of engaging teaching practices used in treatment schools compared to comparison schools at midline.

Although there were no significant differences in the overall use of engaging teaching practices, there do appear to have been some changes in the types of engaging teaching practices used. At baseline, teachers were most likely to ask open questions (observed in 87% of all classes), read to the classroom (67%), and engage inactive students (62%), and least likely to use the reading corner (25%) or have students participate in group work (38%). In contrast, at midline, teachers were relatively less likely to ask open questions (observed in 58% of all classes) but used games or exercises more frequently (61%).

The difference-in-differences analysis, however, suggests that there was little significant change in the use of engaging teaching practices because of HATUTAN programming. Table 24 shows that among all engaging teaching practices, there was only a significant difference in teachers' use of games or exercises within treatment schools as compared to control schools at midline. At midline, this practice was observed 16 percentage points more frequently in treatment schools, while in comparison schools, the practice was observed 8 percentage points *less* frequently. In other words, data collectors observed this practice on average 24 percentage points more often than would be expected given its use in comparison schools. While the use of some other teaching practices changed somewhat substantially in treatment schools as compared to control schools—such as asking open questions and use of the reading corner, which were observed 9 to 10 percentage points more frequently in treatment schools than expected given comparison schools—these results were not significant.

Table 24: Change in use of specific engaging teaching practices

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Teacher asks open questions	86.7%	52.9%	-33.8	87.8%	63.3%	-24.5	9.3	0.36
Teacher reads to students	68.9%	73.6%	4.7	66.3%	64.3%	-2.0	-6.7	0.53
Teacher calls on inactive students	66.7%	69.0%	2.3	60.2%	67.4%	7.2	4.8	0.63
Teacher uses games or exercises	64.4%	56.3%	-8.1	49.0%	65.3%	16.3	24.4	0.02*
Teacher asks students' opinions	48.9%	37.9%	-11.0	51.0%	42.9%	-8.1	2.8	0.78

Students read with others	42.2%	58.6%	16.4	51.0%	54.1%	3.1	-13.3	0.25
Students read by themselves	42.2%	34.5%	-7.7	46.9%	28.6%	-18.3	-10.6	0.31
Students work in groups	35.6%	42.5%	6.9	39.8%	37.8%	-2.0	-9.0	0.41
Teacher uses reading corner	24.4%	14.9%	-9.5	25.5%	25.5%	0.0	9.5	0.31

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In order to check for the robustness of these results, we then run the regression controlling for teacher, classroom, and school characteristics that may impact the use of engaging teaching practices. The increase in use of games or exercises in treatment schools remains significant when these controls are added. Changes in the use of all other specific engaging teaching practices remain insignificant for this additional regression specification. As a further robustness check, we limit the comparison sample to only schools assessed at both baseline and midline and find no significant change in the use of any engaging teaching practice in treatment schools compared to control schools. Overall, these results suggest that while the HATUTAN programming may have had a minor effect on the types of engaging teaching practices used in treatment schools, the impact was inconclusive.

McGovern-Dole Custom Outcome #5 relates to improved teaching practices and is measured by the percent of teachers adhering to improved learning practices in schools. Teachers must demonstrate a minimum of four engaging teaching practices during the classroom observation to meet this standard. Table 25 shows that at midline, achievement of this indicator improved substantially more in treatment schools, where the use of four or more engaging practices was observed in around 5 percentage points more classrooms than at baseline, than in comparison schools, where this indicator was achieved less frequently at midline than at baseline by about 4 percentage points. However, this result is not significant.

Table 25: Change in McGovern-Dole Custom Outcome #5 for teaching practices

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Less than 4 practices	28.9%	33.3%	4.4%	35.7%	30.6%	-5.1%	-0.10	0.33
4 or more practices	71.1%	66.7%	-4.4%	64.3%	69.4%	5.1%	0.10	0.33

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Teachers were also asked about their use of formative assessments, another positive teaching practice that helps teachers conduct in-process evaluations of student comprehension, learning needs, and academic progress. There was a substantial decrease in reported use of formative assessments from baseline to midline, particularly in intervention schools. At baseline, 35% of teachers in intervention schools reported using formative assessments; at midline, this had decreased to 14%, a decline of 21 percentage points. In contrast, in comparison schools, 20% of teachers reported using formative assessments at baseline and 11% at midline, a decline of around 9 percentage points. While this relatively larger decrease in intervention schools was not found to be significant in the difference-in-differences regression, the overall decrease suggests that at the national level, teachers may not currently be effectively encouraged or trained to use formative assessments.

In addition to observing the use of engaging teaching practices, data collectors also observed the use of traditional and negative teaching practices in classrooms. In contrast to the results for engaging teaching practices, at midline, there was a substantial and significant decrease in the use of traditional teaching practices in treatment schools as compared to control schools. Table 26 shows that the average number of traditional teaching practices used (a value of 0, 1, or 2) decreased by almost 0.4 at midline in intervention schools, compared to only 0.1 in comparison schools. Looking at individual

practices, there was a particularly substantive decrease in the practice of students copying from the board in treatment schools: In these schools, at midline, students were observed copying from the board around 28 percentage points less frequently than at baseline, while in comparison schools, students were observed copying from the board slightly more frequently at midline than at baseline. In contrast, while a similar pattern was observed with students repeating the teacher, a practice which declined in prevalence in intervention schools but increased in prevalence in comparison schools at midline, this change was not significant. These results are robust to the inclusion of control variables for teacher, classroom, and school characteristics, and the finding for students copying from the board is additionally robust for the sample limited to only schools assessed at both baseline and midline. Overall, it appears that while the HATUTAN program had limited success at encouraging the use of engaging teaching practices, the program did have significant success in reducing the use of ineffective traditional teaching practices.

Table 26: Change in use of traditional teaching practices

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Average # practices	1.2	1.3	-0.1	1.6	1.2	-0.4	-0.4	0.02*
Students copy from the board	60.0%	63.2%	3.2%	81.6%	54.1%	-27.5%	-30.8	0.005**
Students repeat the teacher	64.4%	67.8%	3.4%	74.5%	68.4%	-6.1%	-9.5	0.38

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

However, program impact on the use of negative teaching practices appears to be less positive. Table 27 shows that the use of negative teaching practices seems to have slightly increased in treatment schools at midline compared to a slight decrease in comparison schools, although the results are not significant. Looking at specific negative teaching practices, the use of corporal punishment increased by around 3 percentage points in both intervention and comparison schools. Overall, the use of corporal punishment was higher in intervention schools than in comparison schools at both baseline and midline. In contrast, teachers' use of an angry voice or harsh tone towards students decreased among comparison schools at midline, but increased in treatment schools—although this result, again, was not significant. While no decisive conclusions can be drawn due to the lack of significance, these results suggest that, at best, the HATUTAN program was not particularly effective in discouraging the use of negative teaching practices, and that more program activities are needed to reduce the prevalence of these practices among teachers.

Table 27: Change in use of negative teaching practices

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Average # practices	0.4	0.3	-0.04	0.5	0.6	0.09	0.14	0.33
Teacher uses angry voice	36.4%	27.6%	-8.8	34.7%	40.8%	6.1	14.9	0.19
Teacher uses corporal punishment	2.2%	5.7%	3.5	11.2%	14.3%	3.1	-0.5	0.94

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Furthermore, as discussed in the baseline report, the prevalence of verbal and physical discipline is likely considerably higher than observed in classroom observations due to social desirability bias. Indeed, in the household survey at midline, 42% of caregivers reported that teachers discipline children by shouting at them, which was observed in only 35% of classrooms; 31% reported that teachers use corporal punishment, which was observed in only 10% of classrooms; and 15% reported that teachers assign chores to students as a form of discipline. Additionally, at midline, 28% of caregivers reported that their children sometimes feel afraid to go to school; however, it is important to note that this number may be inflated due to COVID-19 and related student perceptions that schools are unsafe due to the pandemic. These dynamics are discussed further in the section “Gender and Power.”

Table 28 shows that caregiver perceptions of negative teaching practices remained high from baseline to midline, and there was no significant change in perceptions of negative teaching practices in treatment groups as compared to control groups. While perceptions of teacher use of harsh tones and corporal punishment decreased slightly (but insignificantly) in intervention municipalities, perceptions of the use of chores as a form of discipline increased slightly—though again, insignificantly. These results reiterate that more work is needed to reduce the use of negative teaching and discipline practices in schools, and that HATUTAN programming has, to date, not significantly reduced the prevalence of these practices.

Table 28: Change in caregiver perceptions of negative teaching practices

	Comparison Municipalities			Intervention Municipalities			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	625		482 ⁹⁴	734			
Teacher uses angry voice	39.4%	44.3%	4.9	40.2%	39.5%	-0.7	-5.6	0.31
Teacher uses corporal punishment	30.4%	30.1%	-0.3	36.1%	31.7%	-4.4	-4.0	0.50
Teacher assigns chores	28.8%	13.9%	-14.9	26.3%	15.1%	-11.2	3.7	0.39
Student afraid to attend school	10.1%	29.4%	19.3	16.5%	26.2%	9.7	-9.7	0.62

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Qualitative interviews revealed a wide range of views on the effectiveness of teaching practices and varying prevalence of the use of engaging, traditional, and negative practices. Multiple respondents, including a mother,⁹⁵ female and male teachers,⁹⁶ and school directors and coordinators⁹⁷ mentioned that patience and encouragement were more effective to improve children’s performance in school than violence or punishment, as these negative teaching practices scared the children and reduced their participation in classes. One male teacher in Ermera municipality stated that if children did not understand his teaching, he looked for other ways to teach the subject,⁹⁸ while a male teacher in Manatuto municipality mentioned playing games in class⁹⁹ and many teachers mentioned using group work. Regarding traditional teaching practices, a male school coordinator in Manatuto municipality also stated that teachers had moved away from the traditional practice in which they wrote on the board and students read the lesson; however, he stated that students still copied lessons from the

⁹⁴ n = 479 for the “student afraid to attend school” indicator.

⁹⁵ FGD with mothers, female, Ainaro municipality, Int. 113

⁹⁶ FGD with teachers, male, Ermera municipality, Int. 110; FGD with teachers, female, Ermera municipality, Int. 104

⁹⁷ FGD with school directors, female, Ainaro municipality, Int. 109; FGD with school coordinators, male, Manatuto municipality, Int. 118

⁹⁸ FGD with teachers, male, Ermera municipality, Int. 104

⁹⁹ FGD with teachers, male, Manatuto municipality, Int. 119

board.¹⁰⁰ Similarly, several teachers and school coordinators in multiple locations,¹⁰¹ both female and male, mentioned relying on repetition—whereby students repeat what the teacher says—as a teaching strategy.

However, multiple respondents also mentioned use of or desire to use corporal punishment and other negative teaching practices. A school coordinator mentioned, for example, that it was now forbidden to beat students, but suggested that without the use of corporal punishment, it was more difficult to teach students.¹⁰² Many respondents also seemed to distinguish between “light” corporal punishment—such as twisting a student’s ear—and “serious” corporal punishment, such as beating or punching students. Overall, these findings suggest that in many schools, corporal punishment is still fairly normalized—especially since, given social desirability bias, we expect the use of corporal punishment to be underreported in the data.

In addition to these engaging, traditional, and negative teaching practices, educators also discussed the methods they used to teach literacy. Most educators mentioned that they first teach students the alphabet followed by, in some cases, syllables formed by letter combinations, and only move on to words once students have completed learning the alphabet.¹⁰³ These findings help explain the gap between students’ abilities to recognize letters and their abilities to read words.

Teachers, parents, and school coordinators also mentioned that large classroom sizes could pose a major challenge to the use of engaging teaching practices and to classroom management more broadly. For example, one school coordinator mentioned that: “As teachers, it is difficult to control students if in one classroom there are more than 70 students. We can control if the total students in a classroom below 30, like 25 or 20.”¹⁰⁴

Many teachers also mentioned that large class sizes also present an issue as there are a wide range of skill levels among students in the classroom, which makes it difficult to ensure that all students are engaged in lessons and learning effectively.

We now analyze gender-specific differences in teaching practices in order to understand whether teaching practices differ by the gender of the teachers and, furthermore, whether teachers tend to treat male and female students differently. At midline, on average, female teachers used slightly more engaging teaching practices than male teachers, slightly fewer traditional practices, and approximately the same number of negative practices. Examining the difference in use of teaching practices across baseline and midline for treatment and comparison groups, in general, HATUTAN programming does not appear to have had a significant effect on female or male teachers’ practices compared to teachers unexposed to the program. The only exception to this finding is that there was a significant increase in the use of negative teaching practices among male teachers at intervention schools at midline than among comparison schools at midline; in fact, the average number of negative teaching practices used increased among male teachers in intervention schools, while it decreased in comparison schools.¹⁰⁵ This finding may, in part, be due to male teachers feeling less stigma or shame to be observed using negative teaching practices; however, it also suggests that HATUTAN programs may need to more carefully assess the differences between male and female teachers when addressing the use of negative teaching practices, in order to ensure that programming is effective for teachers of both genders.

¹⁰⁰ FGD with school coordinators, male, Manatuto municipality, Int. 114

¹⁰¹ FGD with school coordinators, female, Ermera municipality, Int. 129; FGD with teachers, male, Ermera municipality, Int. 104

¹⁰² FGD with school coordinators, male, Manatuto municipality, Int. 126

¹⁰³ FGD with school coordinators, male, Ermera municipality, Int. 107; FGD with teachers, female, Ermera municipality, Int. 104; FGD with teachers, female, Ermera municipality, Int. 110

¹⁰⁴ FGD with school directors, male, Ermera municipality, Int. 102

¹⁰⁵ We note that the sample size for male teachers at baseline in comparison schools is small; these results should thus not be taken as entirely definitive.

Table 29: Change in teaching practices, male and female teachers

		Comparison Schools			Intervention Schools			Difference in Differences	
		BL	ML	Difference	BL	ML	Difference	DiD	p
Average # engaging practices	Male	4.2 (n = 17)	4.1 (n = 45)	-0.07	4.7 (n = 51)	4.4 (n = 47)	-0.28	-0.21	0.76
	Female	5.2 (n = 28)	4.7 (n = 42)	-0.47	4.9 (n = 47)	4.5 (n = 51)	-0.30	0.16	0.78
Average # traditional practices	Male	1.4 (n = 17)	1.4 (n = 45)	0.03	1.5 (n = 51)	1.2 (n = 47)	-0.38	-0.41	0.11
	Female	1.1 (n = 28)	1.2 (n = 42)	0.02	1.6 (n = 47)	1.3 (n = 51)	-0.30	-0.32	0.18
Average # negative practices	Male	0.5 (n = 17)	0.4 (n = 45)	-0.17	0.3 (n = 51)	0.5 (n = 47)	0.24	0.41	0.04*
	Female	0.3 (n = 28)	0.3 (n = 42)	0.02	0.6 (n = 47)	0.6 (n = 51)	-0.07	-0.09	0.64

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Data collectors were also asked to observe potentially gender-biased positive and negative teacher behaviors, including whether teachers encourage or ask questions to male and female students at different rates and whether teachers use an angry voice or corporal punishment with male and female students. On average, among all midline schools, teachers encouraged girls somewhat more than boys (observed in 64% of classrooms and 59% of classrooms respectively) and asked questions to girls somewhat more than boys (observed in 66% of classrooms and 64% of classrooms respectively). Furthermore, in 24% of classrooms, teachers were observed asking questions primarily to either boys or girls, rather than to students of both genders.¹⁰⁶

HATUTAN programming, however, appears to have had little significant effect on teachers' use of questions and encouragement with boys and girls. Table 30 shows that there were no significant changes in gender-specific positive teaching practices among the treatment group at midline compared to the comparison group at midline. However, notably, encouragement of girls did increase rather substantially, although not significantly, among intervention schools; encouragement of girls was observed in 11 percentage points more classrooms at midline than at baseline among treatment schools, while it was observed in around 5 percentage points fewer classrooms among comparison schools.

¹⁰⁶ Enumerators were not asked to specify whether it was girls or boys to whom questions were primarily directed, only whether questions were primarily directed to students of one gender, rather than both. Analyzing this indicator in conjunction with observations of whether teachers question boys or girls, we can conclude that in 10 classrooms, teachers primarily asked questions to female students, and in seven classrooms, teachers primarily asked questions to male students. However, there remain 27 classrooms for which we cannot conclude whether teachers primarily asked questions to male or female students.

Table 30: Change in positive teaching practices towards girls and boys

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Teacher encourages boys	57.8%	56.3%	-1.5	66.3%	62.2%	-4.1	-2.6	0.82
Teacher encourages girls	62.2%	57.5%	-4.7	58.2%	69.4%	11.2	16.0	0.16
Teacher questions boys	60.0%	60.9%	0.9	57.1%	66.3%	9.2	8.3	0.48
Teacher questions girls	60.0%	64.4%	4.4	52.0%	68.4%	16.4	12.0	0.32

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Analyzing the gender-specific prevalence of negative teaching behaviors, at midline, teachers used angry voices or harsh tones more often with boys (observed in 26% of classrooms) than with girls (observed in 16% of classrooms). Teachers were also slightly more likely to use corporal punishment with boys (observed in 6.5% of classrooms) than with girls (observed in 6% of classrooms). As with positive teaching behaviors above, the HATUTAN program appears to have had little impact on the use of negative teaching behaviors towards girls or boys (Table 31). In intervention schools, there was a slight reduction in the use of harsh tones towards girls and corporal punishment towards boys at midline, as well as a reduction in the use of corporal punishment on girls relative to the change in comparison schools, but these changes were not significant.

Table 31: Change in negative teaching practices towards girls and boys

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	45	87		98	98			
Teacher uses angry voice with boys	24.4%	21.8%	-2.6	26.5%	29.6%	3.1	5.7	0.55
Teacher uses angry voice with girls	22.2%	16.1%	-6.1	23.5%	16.3%	-7.2	-1.0	0.91
Teacher uses corporal punishment on boys	2.2%	5.7%	3.5	8.2%	7.1%	-1.1	-4.5	0.37
Teacher uses corporal punishment on girls	0.0%	3.4%	3.4	6.1%	8.2%	2.1	-1.4	0.75

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Finally, we analyze any difference between whether male and female teachers tend to treat their male and female students differently. At baseline, it was found that female teachers behaved similarly towards their male and female students, while male teachers appeared to treat boys more positively. In contrast, at midline, we find that while female teachers still generally used most positive and negative behaviors at similar rates on all students—with the exception of a harsh tone, which they used more frequently towards male students—male teachers tended to treat their female students somewhat more positively than male students. Table 32 shows that this pattern of preferential treatment of female students by male teachers was particularly common in intervention schools. This suggests that programming may have been effective at encouraging male teachers to improve their behaviors towards female students, but, as an unintended consequence, made teaching behaviors somewhat less equitable towards male students.

Table 32: Treatment of male and female students by gender of teacher

	Comparison Schools		Intervention Schools	
	Female teachers	Male teachers	Female teachers	Male teachers
n	42	45	51	47
Encourages female students	61.9%	55.6%	74.5%	63.8%
Encourages male students	59.5%	51.1%	72.6%	51.1%
Questions female students	69.1%	53.3%	72.6%	63.8%
Questions male students	76.2%	53.3%	74.5%	57.5%
Uses angry voice with female students	14.3%	17.8%	19.6%	12.8%
Uses angry voice with male students	19.1%	24.4%	31.4%	27.7%
Uses corporal punishment against female students	7.1%	4.4%	7.8%	6.4%
Uses corporal punishment against male students	2.4%	4.4%	11.8%	4.3%

TEACHER ATTENDANCE

Consistent and frequent teacher attendance contributes to the quality of education by increasing the number of hours of instruction received by children and because teachers who regularly attend classes may have a better understanding of the needs and abilities of their students, thus allowing them to adjust lessons as necessary to improve learning outcomes. Teacher attendance was collected at the school level as part of the school survey. Data collectors recorded the number of permanent, contract, and volunteer teachers at each school, as well as the number of teachers in attendance on the day of the visit and the previous day.¹⁰⁷

At midline, on average, school records showed that 76% of teachers had attended on the day prior to data collection. On the day of data collection, in contrast, 88% of teachers were in attendance on average. These results differ somewhat from those found at baseline, in which school records showed that 74% of teachers were in attendance the day prior to data collection but only 61% were in attendance on the day of collection. We note, however, that there are some limitations to this midline analysis. While the number of teachers who should have been in attendance was recorded on the day of data collection, 19 schools reported that zero teachers were assigned to teach that shift, and 36 schools had more teachers in attendance than were assigned to teach. These values, representing day-of attendance at around 25% of midline schools, were removed from the analysis.

It was noted in the baseline report that school records may overreport teacher attendance rates compared to headcounts, thus explaining the gap at baseline between previous-day and day-of attendance and making the day-of attendance recorded through headcounts a more reliable figure. That the opposite pattern occurs at midline—with a higher percentage of teachers in attendance the day of data collection as opposed to the day before—may indicate data reliability issues, or may have occurred if school directors knew that a teacher headcount would occur during data collection and encouraged teachers to attend. As such, results should not be taken as conclusive, but only indicative of a possible overall improvement in teacher attendance rates.

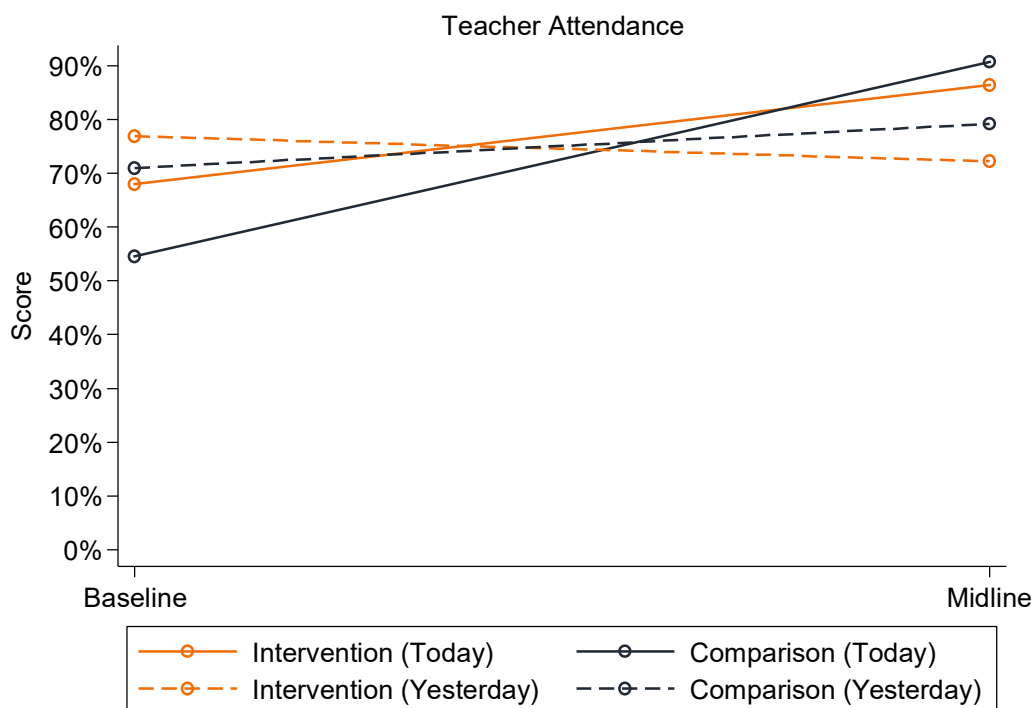
Data collectors were also asked to record headcounts for grade 2 teachers; at midline, on average, 93% of grade 2 teachers were in attendance on the day of data collection. This is similar to findings at

¹⁰⁷ At baseline, due to a data collection error, teacher attendance data was unreliable and thus removed for 46 schools.

baseline and is potentially explained by teachers' knowledge that data collectors would be present and observing grade 2 classes on the day of data collection.

Disaggregating by treatment and comparison schools shows significant differences in changes to teacher attendance rates at midline. Among comparison schools, at midline, teacher attendance taken the day of the survey increased by 36 percentage points and previous-day attendance increased by around 8 percentage points compared to baseline. In contrast, among treatment schools, teacher attendance the day of the survey increased by around 19 percentage points and previous-day attendance decreased by around 5 percentage points compared to baseline (Figure 10).

Figure 10: Change in teacher attendance



The difference-in-differences regression analysis finds that there was a significantly higher increase in attendance for comparison schools as compared to intervention schools (Table 33). These results remain significant when controlling for differences in school type (which may be a proxy for remoteness, and thus affect teacher attendance) and for whether the school has a PTA (which may be involved in enforcing teacher attendance). These results may be due to the location of many treatment schools in remote areas with poor infrastructure, which, particularly during the rainy season (in which data collection took place and during which roads may be washed out), may reduce teacher attendance as it becomes difficult for teachers to access schools.

Table 33: Change in teacher attendance rates

Day of data collection	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	69	61		67	70			
Teacher attendance day of survey	54.5%	90.6%	36.1	68.0%	86.5%	18.5	-17.7	0.004**
Day before data collection								
n	68	88		67	98			
Teacher attendance previous day	71.0%	79.2%	8.2	76.9%	72.2%	-4.7	-12.9	0.04*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Further analysis of the data suggests that this result for teacher attendance is not explained by teachers attending trainings; at midline, only slightly more treatment schools (18 schools) reported that grade 1 or grade 2 teachers were attending trainings compared to comparison schools (13 schools with teachers at trainings). Additionally, within the household survey, respondents were asked how many times their child's teacher had been absent that week; however, the majority (60%) of respondents stated that they did not know, reducing the usefulness of this data for triangulation.

Examining results by municipality suggests that some municipalities were more affected than others. In the comparison municipalities Baucau and Covalima, for example, day-of attendance increased by 48 percentage points and 54 percentage points respectively, from low baseline levels of 42% and 32%. In contrast, in the treatment municipality Liquica, day-of attendance was already high at baseline (75%) and increased by only 5 percentage points at midline. This suggests that the teacher attendance results found above may have been driven by municipality-level factors that increased attendance in some comparison municipalities, rather than by HATUTAN program impact. Alternatively, teacher attendance in intervention areas may have been particularly affected by weather and natural disasters, as intervention schools were generally more remote and in areas with worse infrastructure; roads in intervention areas are thus often more likely to be washed out during the rainy season, reducing attendance.

McGovern-Dole Custom Outcome #6 for teacher attendance measures the percent of schools with at least 80 percent of teachers in attendance on both the day of data collection and the prior day.¹⁰⁸ At midline, around 50% of schools achieved this outcome, an improvement from the 32% of schools who achieved the outcome at baseline.¹⁰⁹ Table 34 further shows, however, that treatment schools were around 12 percentage points less likely to achieve this outcome at midline than would be expected given the results of the comparison schools, although this result was not significant.

Table 34: Change in McGovern-Dole Custom Outcome #6 for teacher attendance

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	67	61		66	70			
Achieved	31.3%	55.7%	24.4	31.8%	44.3%	12.5	-11.9	0.30

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

¹⁰⁸ This outcome originally intended to measure the percent of schools in which at least 80 percent of teachers were in attendance on 90 percent of school days; however, data collection limitations did not allow for accurate teacher attendance records spanning 90 days, so the indicator was adjusted to cover only the day before and day of data collection.

¹⁰⁹ We do not include schools that did not have teacher attendance records in the analysis.

Qualitative data suggests that teacher attendance is highly affected by trainings. A father in Ermera municipality, for example, stated that:

Parents observe that it is one of the problems because teachers sometimes only teach for 2 days and the other 3 days are used to attend trainings. This situation really implicates our children because they don't often get all the lessons they are supposed to get for the week. In [the local school], there are few teachers (1 -2 teachers) that are in this situation. We parents are not satisfied with this situation.¹¹⁰

This respondent further emphasized the need for teachers and schools to better plan for trainings, so that another teacher covers the lessons for the teacher who is absent due to training and students are not left without teachers. In Manatuto municipality, a father described an example of this happening: He stated that if teachers were attending a training, they talked to the school coordinator to ensure that their lessons were covered by another teacher.¹¹¹ However, most respondents who mentioned training stated that it had a negative impact on teacher attendance, which affected students' learning. Despite its negative impact on teacher training, one school coordinator stated that teachers had to attend trainings because otherwise they would not be able to effectively teach students.¹¹²

In addition to training, several respondents mentioned that bad weather or road conditions affected teacher attendance, and that understaffing of schools exacerbated these issues. One school coordinator in Manatuto municipality described needing to merge classes when teachers were absent due to understaffing:

The impact is that when a teacher asks permission for sick leave, then there are only two teachers left. And if the other also asks permission to attend a training, then only one teacher left. So, our policy is to accumulate six classes into two classes. Grade 1, 2, 3 in a single class, and grade 4, 5, 6 together in one class.¹¹³

SCHOOL SUPPLIES AND MATERIALS

Access to school supplies and materials, such as notebooks, pens, and school uniforms, contributes to the quality of the learning environment, as students without these materials may not be able to effectively take notes in class, do homework, or perform other tasks necessary for learning. Additionally, schools without sufficient furniture (chairs and desks) for classrooms may not be able to provide students with an effective learning environment.

At baseline, parents described in interviews how poor families often struggle to pay for student supplies and school materials, and often use their savings to do so. Quantitative data suggests that this issue persists, and may in fact have worsened, at midline. Among all households surveyed at midline, 91% reported using savings for education expenses and 72% reported using a loan from a VSLA to pay for education expenses. In contrast, at baseline, 86% of households reported using savings for education and 69% using a VSLA loan for education. The difference-in-differences analysis suggests that the HATUTAN program had no significant effect on whether households use savings to pay for education expenses (Table 35); while the use of VSLA loans for education expenses did increase substantially more in intervention municipalities at midline than in comparison municipalities, this increase was not

¹¹⁰ FGD with fathers, male, Ermera municipality, Int. 136

¹¹¹ FGD with fathers, male, Manatuto municipality, Int. 135

¹¹² FGD with school coordinators, male, Manatuto municipality, Int. 126

¹¹³ FGD with school coordinators, male, Manatuto municipality, Int. 116

significant, and may reflect greater participation in VSLAs, rather than any change in affordability of school supplies.

Table 35: Change in use of savings for education

	Comparison Municipalities			Intervention Municipalities			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Used savings for education	86.1% (n = 202)	91.2% (n = 377)	5.1	85.4% (n = 233)	90.5% (n = 401)	5.1	0.0	1.0
Used VSLA loan for education	75.6% (n = 78)	70.0% (n = 50)	-5.6	65.7% (n = 134)	75.0% (n = 40)	9.3	15.0	0.21

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

However, data from the household survey also suggests that, while families may find school supplies expensive, lack of school supplies is not a major constraint to school attendance or learning to read. At midline, only one household reported that their child did not attend school because they did not have school supplies, and only two households reported that lacking stationery was a challenge faced by their child to learn to read.

Qualitative data at baseline also suggested that PTA members contribute furniture to schools in some cases, but that PTA support was primarily for infrastructure repairs. Quantitative data from the school survey suggests that at midline, PTAs were still involved in infrastructure improvements at approximately similar rates as at baseline: 76% of PTAs reported involvement in infrastructure repairs at midline compared to 74% at baseline. PTA involvement in infrastructure improvement varied across treatment and control groups, however: At midline, 87% of PTAs in treatment areas reported involvement in infrastructure, compared to 64% of PTAs in comparison areas. The difference-in-differences regression finds a significant increase in PTA involvement in treatment areas compared to comparison areas; this increase could be due to HATUTAN program activities or municipality-level differences that influence PTA priorities.

Similarly, qualitative data from the midline suggests that PTAs remain involved in infrastructure improvement, although at lesser rates than their involvement in enforcing teacher and student attendance or managing the school feeding program. A variety of respondents, for example, mentioned building fences for the school,¹¹⁴ repairing walls and classrooms,¹¹⁵ and repairing kitchens.¹¹⁶ A smaller number of respondents mentioned contributions to school furniture, including chairs and desks.¹¹⁷

LITERACY INSTRUCTION MATERIALS

Literacy instruction materials contribute to the quality of instruction by improving teacher's access to materials that can strengthen lesson plans and classroom activities and by improving students' access to supplementary materials that contribute to learning. Within the school survey, data collectors were asked to observe whether grade 2 classrooms had a reading corner and reading materials (books and magazines). At midline, 47% of grade 2 classrooms were observed to have a reading corner, and 63% had reading materials. This is a substantial increase from baseline, when only 34% of classrooms had a reading corner and 52% had reading materials.

At midline, both comparison and intervention schools saw substantial increases in access to literacy instruction materials. For intervention schools, the increase was particularly notable for reading

¹¹⁴ FGD with school directors, female, Ermera municipality, Int. 103; FGD with mothers, female, Ainaro municipality, Int. 108

¹¹⁵ FGD with teachers, female, Ermera municipality, Int. 104

¹¹⁶ FGD with school coordinators, male, Ermera municipality, Int. 107

¹¹⁷ FGD with teachers, male, Ermera municipality, Int. 110

corners; almost 19 percentage points more schools had a reading corner at midline than at baseline. For comparison schools, the increase was more substantial for reading materials, which over 12 percentage points more schools had at midline than at baseline. The difference-in-differences analysis suggests that in treatment schools, access to a reading corner or both a reading corner and reading materials increased substantially more at midline than would be expected given the results in comparison schools. However, while these differences were relatively large in absolute terms—12 and 11 percentage points more than would be expected respectively—they were not significant (Table 36). Despite the lack of significance, the substantial size of the results suggests that the HATUTAN program may have had a meaningful effect on, in particular, access to reading corners in grade 2 classrooms.

Table 36: Changes in access to reading corner and reading materials

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	90	88		99	98			
Reading corner	31.1%	37.5%	6.4%	37.4%	56.1%	18.7%	12.4	0.25
Reading materials	47.8%	60.2%	12.4%	56.6%	66.3%	9.7%	-2.7	0.82
Reading corner and materials	30.0%	35.2%	5.2%	34.3%	51.0%	16.7%	11.4	0.23

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

McGovern-Dole Custom Outcome #7 measures the percent of schools that have age-appropriate reading materials in classrooms. As the survey did not assess the age-appropriateness of reading materials at baseline or midline,¹¹⁸ the outcome is measured by just the percent of schools with reading materials in classrooms. Table 36 shows that at midline, 60% of comparison schools and 66% of intervention schools achieved this outcome. Both intervention and comparison schools achieved this outcome at significantly higher rates at midline than at baseline; however, access to reading materials improved slightly more in comparison schools than in intervention schools.

Within the household survey, caregivers were also asked to report their perceptions of their children's access to reading materials at school. In contrast to the positive results from the school survey, at midline, caregivers reported slight decreases in access to literacy materials. At baseline, 95% of caregivers agreed strongly or somewhat that their child had enough books at school; at midline, only 91% of caregivers agreed with this statement, a significant decrease. In contrast, however, only 1% of caregivers reported that "no reading materials at home" was a challenge impacting their child's ability to learn to read. Overall, it is likely that many caregivers are not fully familiar with the number of literacy materials available at their children's schools; these findings should thus not be taken as a definitive sign that access to reading materials declined (or did not improve) at midline, especially when considered in contrast to the positive and more reliable findings from the school survey.

Qualitative data reveals several challenges to the use of literacy materials in and out of schools, however. Many respondents mentioned that they did not lend books or magazines to students because the students damaged the books or forgot to return them:

We have not let them to borrow books because, I will give you one example, the Lafaek magazine, when the Lafaek magazines arrive, they [students] receive them, and then

¹¹⁸ As described in the baseline report, enumerators do not necessarily have education qualifications, and would not have been able to reliably assess the age-appropriateness of reading materials available in the classroom.

*they will take the magazines home but they wouldn't take it back to school to read [or to do more reading]. They took it home and then they tore it or just chuck it there.*¹¹⁹

This dynamic may explain the contrast between the increase in access to literacy materials reported in the school survey and the slight decrease in reported access among caregivers. While schools may have more literacy materials, an unwillingness to lend them to students may mean that the materials are not effectively utilized.

Several additional challenges to the use of literacy materials were mentioned. One school coordinator stated that his school did not have enough books for students.¹²⁰ Another school director stated that while his school had sufficient books, there was not enough space to organize the books, and so they were left in boxes.¹²¹ Lastly, a school coordinator stated that there was no benefit of literacy materials for children who could not yet read, as they would just look at the pictures and drawings in the books or magazines.¹²² All of these findings suggest the need for improved access to age-appropriate literacy materials and related infrastructure and a need to change attitudes around literacy materials, so that schools do not refuse to lend books or magazines to students.

SKILLS AND KNOWLEDGE OF TEACHERS

The education and skill levels of teachers may strongly affect quality of instruction, particularly in challenging learning environments like those found in much of Timor-Leste. At baseline, findings suggested that many teachers were at a transitional moment, and were learning new, non-violent ways of managing classes, learning how to work with young students, incorporating participatory pedagogical methods in classroom practice, gaining awareness of the need to engage students who were not participating in class, and trying new approaches to engage students. Engagement in literacy trainings and education courses can help expedite this transition and improve the skills and knowledge of teachers, and thus the learning outcomes of students.

At midline, a higher average percent of teachers reported having attended training on literacy education (45%) in the school survey than at baseline (33%).¹²³ Treatment schools saw a relatively higher increase in the percent of teachers attending literacy training than comparison schools: At treatment schools, there was a 14 percentage point increase in the percent of teachers attending literacy training, from 27% to 42%, while at comparison schools, there was only around an 8 percentage point increase, from 41% to 48% (Figure 11). This result, however, was not significant (see Table 37).

Looking at teacher education, at midline, a slightly lower number of male teachers and a slightly higher amount of female teachers reported having concluded a bacharelato or teacher training college program. The difference-in-differences regression analysis suggests that teachers' completion of a bacharelato or teacher training college program improved somewhat less than would be expected in treatment schools as compared to comparison schools (Table 37). However, as above, these results are not significant. We further note that due to inconsistencies in data collection and indicator design, as well as varying numbers of schools with zero male or female teachers across baseline and midline,

¹¹⁹ FGD with school directors, male, Ermera municipality, Int. 102

¹²⁰ FGD with school coordinators, male, Manatuto municipality, Int. 133

¹²¹ FGD with school directors, male, Ermera municipality, Int. 102

¹²² FGD with school coordinators, male, Ermera municipality, Int. 107

¹²³ The school survey recorded the number of teachers in grades 1-3 who had attended literacy training. Unfortunately, the total number of teachers was only recorded for grades 1, 2, and for the entire school. Therefore, the sample is limited to 135 baseline schools and 140 midline schools for which the highest school grade is 1, 2, or 3, for which the relative number of teachers who have attended literacy training can be accurately calculated. The baseline sample is further limited by the data collection errors that necessitated the removal of some teacher enrollment data; only 100 baseline schools are therefore analyzed for this indicator.

the sample size of schools used for this analysis varies substantially. These results should thus not be taken as indicative of any definitive, larger patterns.

Figure 11: Change in percent of teachers attending literacy training

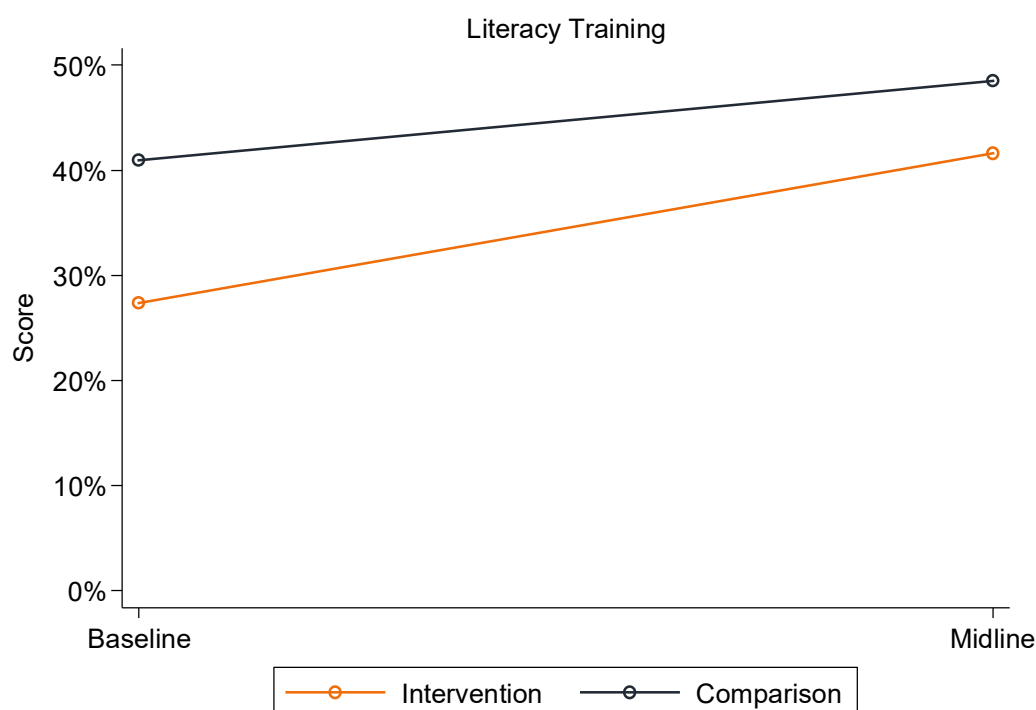


Table 37: Changes in teacher training and education

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Attended literacy training	40.9% (n = 45)	48.5% (n = 62)	7.6	27.4% (n = 55)	41.6% (n = 78)	14.2	6.7	0.44
Completed bacharelato or teacher training college (male)	56.5% (n = 67)	57.0% (n = 87)	0.5	49.2% (n = 67)	44.9% (n = 98)	-4.3	-4.8	0.65
Completed bacharelato or teacher training college (female)	41.4% (n = 58)	46.3% (n = 76)	4.9	30.6% (n = 57)	33.4% (n = 88)	2.8	-2.0	0.83

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In the qualitative data, while trainings were most frequently mentioned as a detractor from teacher attendance, they were also discussed as an important way for teachers to gain knowledge and experience, particularly for teachers who do not have degrees in education.¹²⁴ Teacher's working group meetings were also mentioned as an effective way to share teaching and learning materials

¹²⁴ FGD with school directors, female, Ermera municipality, Int. 103

among teachers and improve the skills of those teachers with less experience or education.¹²⁵ One school coordinator additionally emphasized that the education of teachers was not the most important thing contributing to their success in the classroom; rather, mastery of teaching methodology was important:

*What is most important for teaching 1st grade is especially the mastery of methodology and the way of teaching must be of a high standard, so that the children can grasp it easily. The most important thing is that children must read and write, they must be able to do math. So our interest is that children must be adequately taught. One should not look at the baccalaureate degree or whatever, the important thing is to use a good methodology for children to adapt quickly and learn.*¹²⁶

SKILLS AND KNOWLEDGE OF SCHOOL ADMINISTRATORS

In addition to the skills and knowledge of teachers, experienced and knowledgeable school administrators can help improve the quality of instruction by providing training to teachers and ensuring that practices used in classrooms are effective. At midline, directors had on average around one year more experience than at baseline—an expected finding, given that midline data collection took place around two years after the baseline. There were small differences in the education level of directors: At midline, directors were slightly more likely to have a secondary school degree only, and slightly less likely to have a teacher training institute or Faculty of Education degree.

In addition to these descriptive statistics of directors' experience and education, the school survey collected data on whether school directors provide coaching to teachers. At midline, only around 5% of directors said that they had never provided coaching. Overall, compared to baseline, more directors stated that they provided monthly coaching and fewer directors stated that they provided weekly coaching. Within comparison and intervention schools, the difference-in-differences analysis suggests that at midline, directors reported never providing coaching at significantly lower rates in treatment schools than would be expected given rates in comparison schools. Within treatment schools, directors were 8 percentage points less likely to report never providing coaching, while within comparison schools, directors were 2 percentage points *more* likely to have never provided coaching (Table 38). While there were no other significant difference-in-differences results, there was a substantial increase in provision of coaching every trimester in intervention schools as compared to control schools. These results suggest that the HATUTAN program may have been successful at encouraging school directors to provide training to teachers, but that the frequency of training remains low. The COVID-19 pandemic may also have influenced the frequency of trainings, and thus may have increased the likelihood that directors provide training on only a monthly or once-a-trimester basis.

¹²⁵ FGD with teachers, male, Manatuto municipality, Int. 119

¹²⁶ FGD with school coordinators, male, Ermera municipality, Int. 107

Table 38: Change in director provision of coaching

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	89	88		99	98			
Weekly coaching	42.7%	36.4%	-6.3	42.4%	28.6%	-13.8	-7.5	0.54
Monthly coaching	19.1%	25.0%	5.9	22.2%	31.6%	9.4	3.5	0.65
Coaching every trimester	33.7%	31.8%	-1.9	23.2%	34.7%	11.5	13.4	0.22
Never provided coaching	4.5%	6.8%	2.3	12.1%	4.1%	-8.0	-10.4	.04*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Note: Percentages may not sum to 100 due to a “don’t know” option excluded from the table.

Few school directors or coordinators explicitly mentioned providing training to teachers in the qualitative data; instead, more coordinators described assigning teachers to classes based on their skills and experience with teaching various topics and ages. However, one school director mentioned the following strategy for coaching a teacher who did not have good classroom management skills:

*We sit down to talk, [I ask] if in his subject he made a complete plan. When we come into the class [I discuss] how should we greet the students, and ask if they remember the subject of the previous class and what it covered. After asking about the previous day's subject we go into the day's subject, after the class is over we give an oral or written test. Oral test we ask questions on our plans and goals.*¹²⁷

In addition, several school coordinators/directors mentioned meeting with teachers who were observed using corporal punishment or yelling at students in order to encourage them not to use such methods.¹²⁸

STUDENT ATTENTIVENESS

This section analyzes student attentiveness, which is highlighted in the log frame as a factor that may influence literacy scores, and factors that are turn expected to influence student attentiveness. Initially we analyze three headline indicators of attentiveness, presenting midline summary scores and employing a panel regression to capture changes since baseline by cohort. This is followed by looking at both student hunger and school-level indicators of food access, both of which are expected to have an effect on student attentiveness and are ways in which the activities of the program could contribute to increased literacy scores. Again, we present summary figures at midline and changes since baseline, as well as an analysis on how these factors are associated with attentiveness in this study. The scores are also disaggregated as suitable to gain a fuller picture.

STUDENT ATTENTIVENESS INDICATORS

Three indicators are used to measure student attentiveness. First, in a self-reported measure of student attentiveness, students were asked if they felt they were able to pay attention in class or not. As this measure is likely prone to inaccuracy and desirability bias, this is supplemented by two

¹²⁷ FGD with school directors, female, Ermera municipality, Int. 103

¹²⁸ FGD with school coordinators, male, Manatuto municipality, Int. 116; FGD with school coordinators, male, Manatuto municipality, Int. 126

additional measures – (1) observed student attentiveness, where an enumerator observed 10 students in class and reported how many were paying attention, and (2) working memory, which was assessed in a memory test and is reported on a scale from 1-100. Based on discussions with CARE, we use working memory score as a proxy measure as it is thought to depend on student attentiveness, and it is an objective measure in contrast to observed or self-reported attentiveness. The test is designed to mirror classic working memory tests which typically include non-sequential digit, word, and sentence recall. It is adapted to use images instead of words or digits to avoid potential misinterpretation and only includes tasks related to phonological span, which is most important in measuring the extent to which working memory affects reading skills.

We first report the scores for the cross-sectional panel, where the baseline scores for treatment and comparison schools are compared against a new sample collected at midline. In this midline sample the students were in the second grade, the same as the baseline group, meaning we can compare students of the same age who have benefitted from the program against students who have not. The summary scores can be seen in Table 39 below. The observed attentiveness score was school-based, not individual, and thus we include this indicator in the panel section rather than the cross-sectional.

Table 39: Student attentiveness scores, cross-sectional cohort

	Comparison Schools			Intervention Schools			Difference in Differences	
Student attentiveness (self-reported)	BL	ML	Difference	BL	ML	Difference	DiD	p
n	1,004	1,101		1,409	1,457			
Score (% paying attention)	96.5	94.6	-2.1	95.5	95.8	0.3	2.3	0.10
Working memory score	BL	ML	Difference	BL	ML	Difference		
n	998	1,108		1,412	1,474			
Mean score	40.9	34.2	-6.7	39.9	31.7	-8.2	-1.5	.41

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

As expected, self-reported attentiveness is very high and relatively uniform across rounds and groups. There is a small decrease in the percentage paying attention for the comparison group and a very small increase for the intervention group. Working memory, which is a more objective measure, shows interesting patterns: There was a substantial decrease in working memory both the comparison and intervention groups.

We test for the statistical significance of these results using a difference-in-differences regression model. Self-reported student attentiveness has a positive difference in difference measure of 2.3, but this result is not significant. The difference-in-differences for working memory is negative but the finding is not significant. It is unclear whether student attentiveness captures meaningful information about students' academic performance, despite potential desirability bias in reported data. If this is the case it could provide evidence of positive impact on students in intervention areas, with a possible causal mechanism of effective school feeding leading to higher concentration.

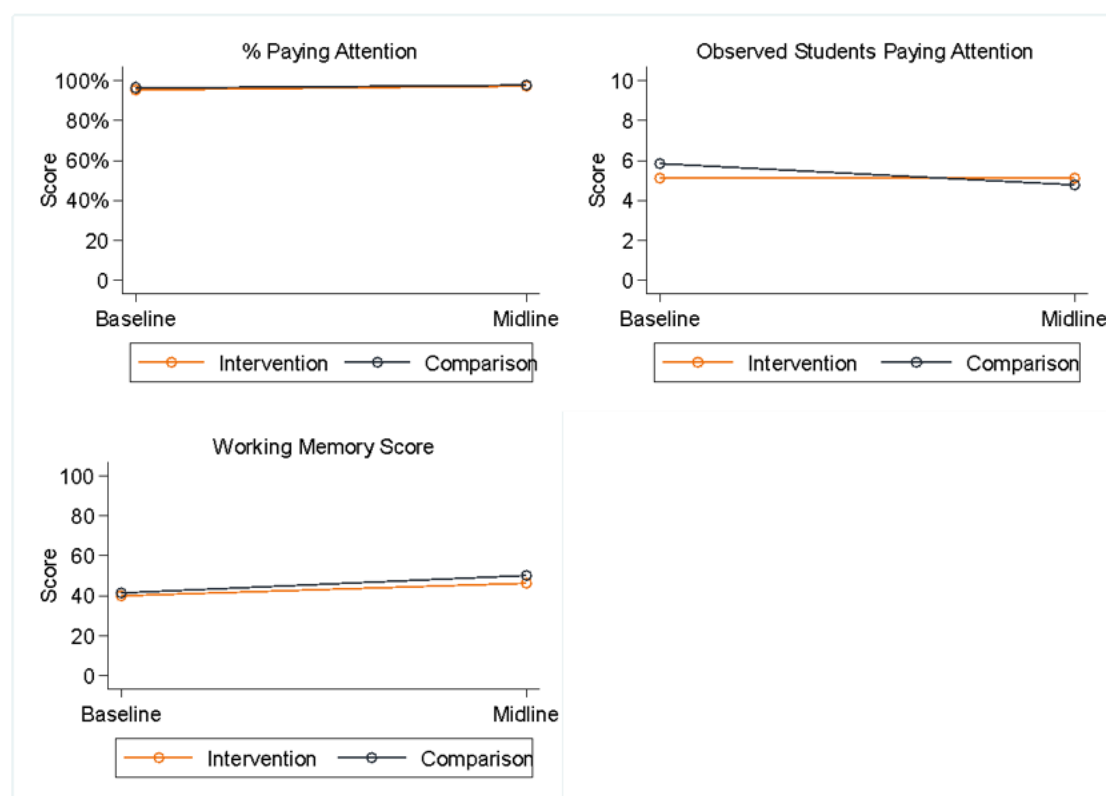
Next we report the same results from the panel analysis. In this section, the same students were tracked from baseline to midline, which allows for more rigorous analysis. The summary scores can be seen in Table 40 below. Self-reported student attentiveness was again very high, with a score of over 95% in intervention schools at both baseline and midline. There was a slight increase in the scores at midline for both treatment and comparison areas. Observed student attentiveness, likely the more accurate measure, shows an average of between 4-6 students paying attention over the 10-minute observation period. Working memory scores range from approximately 40 to 50 out of 100. Both comparison and treatment groups saw an increase across the study period.

Table 40: Panel student attentiveness indicators

	Comparison Schools		Intervention Schools	
Student attentiveness (self-reported)	BL	ML	BL	ML
n	838 ¹²⁹	845	1,190	1,215
Score (% paying attention)	96.3%	97.5%	95.4%	97.0%
Student attentiveness (observed)	BL	ML	BL	ML
n	44	43	95	98
Mean Score (out of 10 students)	5.8	4.8	5.1	5.1
Working memory score	BL	ML	BL	ML
n	834	848	1,191	1,221
Mean Score	41.2%	49.9%	39.8%	46.4%

Changes across baseline and midline are analyzed in more detail with the graphs below in Figure 12. For self-reported attentiveness, both cohorts improved slightly and any difference in rates of improvement was negligible. For observed attentiveness, the comparison group mean dropped by 1 while the treatment group remained mostly stable. Conversely, while both groups improved in working memory score, the gap widened and the comparison group grew by a greater amount. This mirrors the results from the cross-sectional analysis, which showed an improvement in attentiveness and a deterioration in working memory in the treatment group with respect to changes in the comparison.

Figure 12: Trend of student attentiveness indicators



¹²⁹ Number of respondents varies due to exclusion of students who did not answer the question from the table.

The panel regression testing for the statistical significance of these changes is reported in Table 41 for each variable of interest. The difference in difference for self-reported student attentiveness is quite small at 0.4, an even weaker effect than that found in the cross-sectional section for the same indicator, and is not statistically significant. This could be an indication that the variable is not capturing any meaningful information and that the minor changes between rounds are mostly random. As well as the desirability bias noted above, this indicator is also unable to capture whether a student's attentiveness had increased, only whether a student was attentive or not. Given the high values reported, this means there is little room for possible improvement that the variable would be able to measure. The difference in difference for observed student attentiveness of 1, on the other hand, is more substantial given the overall range of possible scores from 1-10 students. Our proxy measure for attentiveness, working memory score, has a negative difference in difference, meaning that the comparison group improved at a faster rate than the treatment group from baseline to midline. However, the result was not statistically significant. Furthermore, this result is to be expected given that more households experienced food insecurity in the treatment group than the comparison group, thus resulting in a negative impact of hunger on working memory scores. The result suggests a need for further reinforcement of household-level interventions to reduce food insecurity.

Table 41: Regressions for student attentiveness indicators, panel cohort

	Intervention			Comparison			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Student attentiveness (self)	95.4%	97.0%	1.6	96.3%	97.5%	1.2	0.4	0.72
Student attentiveness (observed)	5.1	5.1	0.0	5.8	4.8	-1.0	1.0	0.09
Working memory score	39.8%	46.4%	6.6	41.2%	49.9%	8.7	-2.0	0.25

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The main total midline scores are disaggregated below in Table 42 by gender and municipality in order to gain a fuller understanding of the variance in scores. It was not possible to disaggregate observed student attentiveness by gender, as the data is at school level. Female students at midline scored slightly higher than male students on both self-reported student attentiveness and working memory. Across municipalities there was little variability in self-reported student attentiveness and working memory score. There were, however, large differences across observed student attentiveness, with some municipalities showing average scores close to 6, two municipalities (Bobonaro and Manatuto) showing scores just below 4, and one (Baucau) having an average of just 1.5 students paying attention. This is a very large drop from the baseline figure of 8.5, although both of these extreme figures are likely caused by the very small sample size for this municipality – there were only two observations from Baucau.

Table 42: Disaggregated student attentiveness summary statistics

	Student attentiveness (self)	Student attentiveness (observed)	Working memory score
Overall	97.2%	5.0	47.8%
Gender			
Male	96.9%	-	46.4%
Female	97.5%	-	49.3%
Municipality			
Aileu	97.1%	5.3	49.8%
Ainaro	98.6%	5.7	46.1%
Baucau	95.5%	1.5	54.9%
Bobonaro	97.4%	3.8	49.0%
Covalima	99.2%	5.1	49.7%

Ermera	97.6%	5.1	44.9%
Liquica	95.6%	5.8	46.8%
Manatuto	95.0%	3.8	49.8%
Manafahi	97.6%	5.8	50.1%

Below we disaggregate the panel analysis by running a separate regression after dividing the students by gender. Difference in difference for self-reported student attentiveness is slightly higher for boys, but not significant for either boys or girls. Similarly, the working memory difference in difference is lower for female students than for male students, although neither result was statistically significant.

Table 43: Disaggregated student attentiveness panel regressions

	Student attentiveness (self)		Working memory	
	DiD	p	DiD	p
Male	0.33	0.85	-1.9	0.35
Female	0.55	0.74	-2.2	0.27

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Working memory was highlighted as a key indicator to track against targets laid out at the baseline in order to assess program performance. The target at baseline for year 3 (October 2020 – September 2021) was that 35% of students would have a working memory score of more than 50%, up from 29% at baseline. This figure for the midline treatment group of the panel cohort was 44%, which exceeds this target. However, this analysis does not account for improvement in working memory simply due to students being older and more educated. For the cohort of newly-contacted grade 2 students, at midline, only 17% of students achieved this target. Achievement of the target thus cannot be fully attributed to program impact, especially as the difference in difference was negative.

STUDENT HUNGER INDICATORS

In this section and the following, we study factors that are expected to have an effect on the headline indicators for student attentiveness and how they have been impacted by the intervention. Student hunger is identified in the log frame as a factor that is expected to influence attentiveness and, in turn, literacy scores. Before the EGRA was carried out each student was asked to state whether they had eaten any food that day (McGovern-Dole Custom Outcome #13). As above, we initially report scores for the cross-sectional analysis and then compare these with the results from the panel section.

Summary scores are shown in Table 44. Most students had eaten on the day of the EGRA test, with scores between 85% and 90%. In comparison schools there was a small drop in the percentage of students having eaten, while in the intervention this percentage was stable from baseline to midline. There was also an increase in the percentage of households that had gone at least one day in the past 30 days without eating in both comparison and intervention areas. This deterioration could possibly be due to COVID-related disruptions to food security.

Table 44: Student hunger variables, cross-sectional cohort

	Comparison			Intervention			Difference in Differences	
Student eaten	BL	ML	Difference	BL	ML	Difference	DiD	p
n	1,012	1,107		1,442	1,470			
Score (% eaten on day)	88.3%	89.5%	1.2	86.6%	89.5%	2.9	1.7	.43
Household not eaten	BL	ML		BL	ML			
n	366	619		466	960			
% not eaten at least once	4.9%	11.3%	6.4	8.4%	14.0%	5.6	-0.8	.75

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

These findings are tested for statistical significance in the last column of Table 44. The difference in difference for students having eaten on the day of the test is positive, while the difference in difference score for whether a household had not eaten for a day in the past 30 days was very small and negative. Neither score was statistically significant.

The summary statistics for the analysis of the panel cohort are presented in Table 45. Households in the treatment group were not recontacted at midline in the data collection so the only indicator in this section is whether a student had eaten on the day of the survey. At both baseline and midline in both comparison and intervention areas, most students had reported eating that day, with between 86% and 93% answering positively. There was a very slight increase over the study period of 0.6 percentage points for the comparison group, and a larger increase of around 6 percentage points for the treatment group, suggesting a positive impact from the program from baseline to midline.

This finding is tested for significance using a panel regression (Table 45). The total difference in differences is 5.1 percentage points, which is a substantial increase given the small number of students reporting that they had not eaten—no more than 15% percent for any of the study groups. Most of the difference in difference is attributable to a positive increase in the score in the intervention group, and not a deterioration in the comparison group. This is a much stronger effect than the improvement in the cross-sectional analysis; since we expect the panel data to provide a more accurate picture of changes due to the program, this suggests that program activities may have had a positive impact on increasing the percent of students who had eaten in intervention schools.

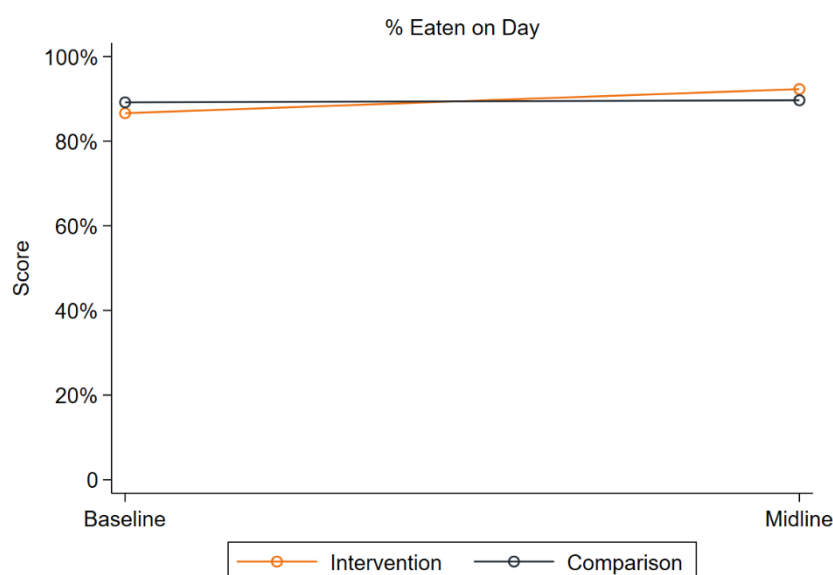
Table 45: Student hunger scores, panel cohort

	Comparison Schools			Intervention Schools			Difference in Differences	
Student eaten	BL	ML	Difference	BL	ML	Difference	DiD	P
n	846	847		1,217	1,218			
Score (% eaten on day)	88.9	89.5	0.6	86.5	92.2	5.7	5.1	0.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

This effect is further shown in Figure 13 below, which compares the scores for this question for each group across the study period. The trends described above can be seen quite clearly, with the orange line for intervention starting with a slightly lower score than the comparison line and finishing in the midline with a higher score.

Figure 13: Trend of student hunger, panel cohort



We also test whether there was a link between whether a student had eaten that day and whether the school had provided meals. Students in schools with food served were approximately 3% more likely to have eaten and the effect was statistically significant. This suggests a clear relationship between program activities and outcomes: The program improves school feeding and school feeding increases the probability that the student has eaten.

As most of the difference in difference is driven by changes in the treatment group, the baseline and midline scores for the treatment are disaggregated below by gender and municipality to see what is driving the change. There is little difference between genders within rounds, although female students scored slightly higher than male students. In terms of the change from baseline to midline, male students increased slightly more than female students. There was more variability in change across round by municipality: Municipalities with lower proportions of students having eaten at baseline showed greater improvements, with Ainaro increasing 10 percentage points from 86% and Manatuto increasing 7 percentage points. This suggests the program is effective at improving school feeding habits in areas where the need is greater.

Table 46: Student hunger, disaggregated by gender and municipality

	Student eaten	
	Baseline	Midline
Overall	86.5%	92.2%
Gender		
Male	85.6%	91.8%
Female	87.4%	92.6%
Municipality		
Ainaro	86.2%	96.1%
Ermera	87.7%	91.8%
Liquica	93.0%	94.3%
Manatuto	80.1%	87.2%

The panel analysis is disaggregated by gender to analyze separate difference in difference scores for significance. The total difference in difference for both males and females is around 5 percentage

points; these results are not significant. This result suggests the overall positive difference in difference is driven by improvements across both genders.

Table 47: Student hunger panel regressions, disaggregated by gender

	DiD	p-value
Student eaten (male)	4.8	0.09
Student eaten (female)	5.3	0.07

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

As mentioned above, student hunger is a McGovern-Dole custom outcome indicator and has been highlighted at baseline as a particularly important indicator to track the impact of the program. The target for year 3, October 2020 – September 2021, was that only 9% of students would report that they had not eaten. The reported figure for the treatment group of the panel cohort at midline was 7.8% reporting that they had not eaten, meaning the target was met. The positive results from the difference in difference analysis further suggests that this is at least in part attributable to the impact of the intervention. For the cross-sectional cohort, however, this target was not met, as around 10.5% of students reported that they had not eaten.

FOOD ACCESS INDICATORS

The results framework also highlights food access as a factor that may influence student attentiveness. We use three main indicators—whether the school had a menu for school feeding that day (indicating a level of preparedness and organization), whether and how often the school purchased produce from local farmers for school feeding, and whether the school provided meals to the students that day. We further analyze these and other indicators in the section “School Feeding Program,” but present a brief summary of findings here. Table 48 below summarizes these indicators.

Table 48: Change over time in food access indicators

	Comparison Schools			Intervention Schools			Difference in Differences	
School menu	BL	ML	Difference	BL	ML	Difference	DiD	p
n	89	87		99	98			
Score (% of schools with menu)	94.4%	81.6%	-12.8	92.9%	67.3%	-25.6	-12.9	0.08
Purchase produce	BL	ML		BL	ML			
n	27	26		1	86			
No	3.7%	3.9%	0.2	0.0%	54.7%	-	-	-
Sometimes	59.3%	76.9%	17.7	0.0%	34.9%	-	-	-
All the time	37.0%	19.2%	-17.8	100.0%	10.5%	-	-	-
School feeding today	BL	ML		BL	ML			
n	90	88		99	98			
Score (% of schools with meals)	30.0%	29.5%	-0.5	1.0%	87.8%	86.8	87.9	<0.001***

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At baseline most schools had a menu for school feeding, although the score was lower for intervention schools. Both cohorts, however, saw quite a large fall in this score at midline. At baseline most schools

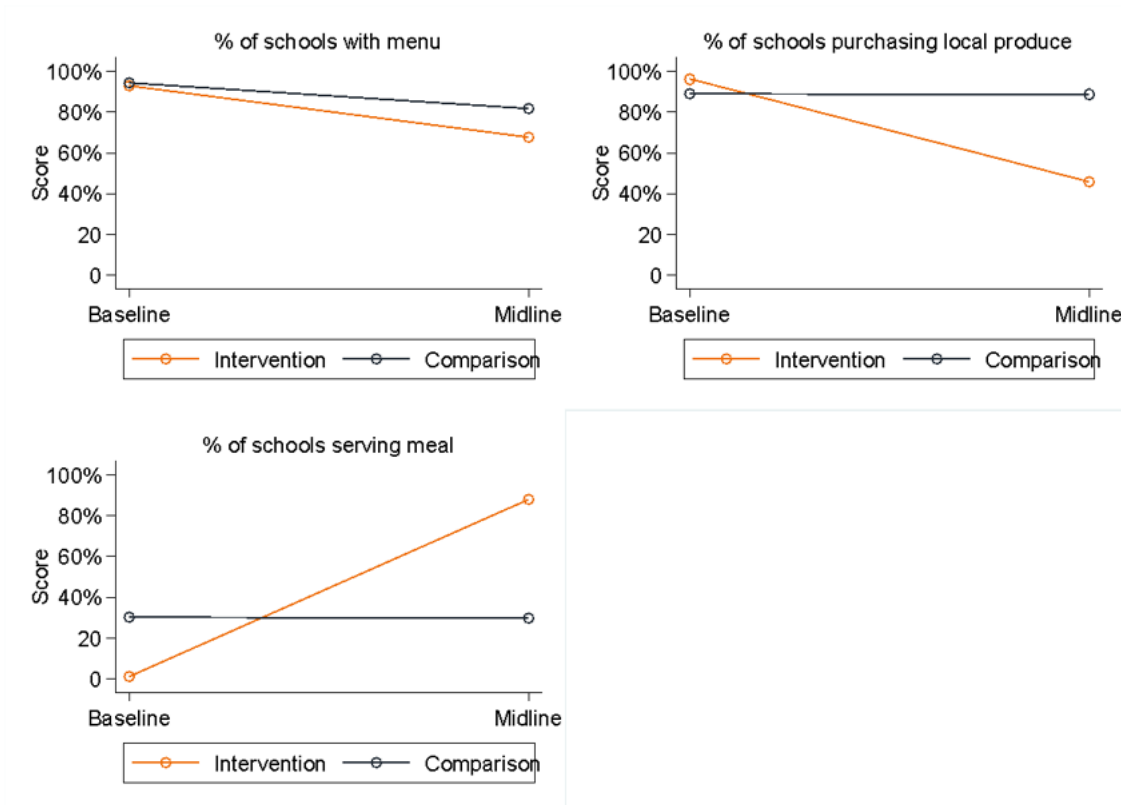
claimed to purchase produce from local farmers, with only 11% and 4% of schools buying no local produce for the comparison and treatment groups respectively. In the treatment group, however, there were quite significant changes in this variable, with answers “sometimes yes” and “all the time” falling and “no” answers increasing from 4% to 54%. We note, however, that these numbers are not fully comparable; during baseline data collection, most schools were not providing school meals, and their response to this question was based on past practices and may have been influenced by social desirability bias. In contrast, at midline, responses were based on current practices of schools, and were influenced by the fact the commodities were provided to schools but there was no state budget for food purchases.

The percentage of schools that had provided meals on the day of the survey was mostly unchanged in the comparison group, but there was an extremely large increase in the treatment group from 1% to 88%. This rise is partly a result of the very low baseline score, which is explained by the fact that baseline data collection took place 1-2 months earlier than comparison and just before a budget transfer for school feeding had been received by many schools. This means this indicator is slightly biased.

These overall differences are visualized in Figure 14 below. The extremely large rise in the percentage of schools serving a meal for the intervention cohort is apparent, with the score rising from close to 0% up to nearly 90%, as the comparison group falls slightly. Despite the fact that this indicator is slightly biased, as explained above, the achieved level is still higher than would be expected even given this bias, and clearly reflects the impact and effective implementation of the school feeding program. The proportion of treatment schools serving meals and purchasing from local producers is far greater (40%) than the proportion of comparison schools serving meals and purchasing from local producers (28%), indicating that the provision of commodities is not affecting local purchases.

The difference-in-differences regression analysis is also presented in Table 48; the change in the percent of schools with a menu is not significant, while the changes in purchasing school feeding are significant. Overall, these results suggest that the program contributes to far higher levels of schools serving meals to students.

Figure 14: Trend of school feeding indicators



In Table 49 below the summary figures are disaggregated by municipality. The percentage of schools with a menu fell across all municipalities, and the decrease was particularly pronounced in Ainaro and Ermera, where it fell from approximately 90% to 70% and 54% respectively. HATUTAN's programming and training focused on creating flexible menus. Differences in schools purchasing local produce vary even more by municipality; however, due to the very low number of schools providing meals in treatment municipalities at baseline and lack of SFP budget, purchase of local produce cannot be compared over time. Regardless, this high level of variation suggests that this indicator is influenced by region-specific factors. School feeding on the day of the survey rose in all municipalities relatively uniformly aside from Ainaro, which had a rate of 78% at midline, meaning that the school feeding program may not have been implemented as effectively in this municipality.

Table 49: Disaggregated food access summary scores

	School menu		Purchase produce ¹³⁰		School feeding today	
	BL	ML	BL	ML	BL	ML
Overall	92.9%	67.4%	96.0%	45.9%	1.0%	87.8%
Municipality						
Ainaro	92.9%	70.4%	-	71.4%	3.4%	77.8%
Ermera	90.2%	53.7%	-	16.2%	0.0%	90.2%
Liquica	100.0%	81.8%	-	80.0%	0.0%	90.9%
Manatuto	94.7%	84.2%	-	55.6%	0.0%	94.7%

¹³⁰ Restricted to only schools providing meals.

As well as whether a meal had been served, the quality of the menu is likely to influence student attentiveness measures. We analyze quality of school meals in detail in the section “School Feeding Program.” Here, we note that treatment schools were much more likely to provide dark green vegetables and beans, peas, soybeans or peanuts, and somewhat more likely to serve pumpkin, carrot or sweet potato, rice, maize, or bread, and “other” foods. Treatment schools did not provide sweetened condensed milk as part of the school feeding, but 17 comparison schools (65%) did.

On this measure, it is difficult to give an overall assessment of food quality across the treatment and comparison as both menus have strengths relative to the other. However, as discussed further in “School Feeding Program,” analysis of dietary diversity scores for schools suggests that treatment schools had significantly higher mean scores than comparison schools. As such, we tentatively conclude that the program contributed to a higher quality menu.

STUDENT ATTENDANCE

In this section we analyze student attendance. Headline indicators of student attendance, including attendance rates, dropout rates, and reasons for missing school, will be analyzed initially. This will be followed by a focus on health-related absences, as this is a McGovern Dole custom indicator. Finally, factors that may affect attendance are analyzed, including school infrastructure and access of students to the school.

Student attendance is highlighted in the results frame as a factor that may affect literacy scores. The program is expected to influence student attendance through a number of ways. Improved management of the school feeding program and consequent improved school feeding could act as a pull factor for students attending the school and also lower dropout rates. Improved learning environments could increase the perceived utility among parents or caregivers of sending the students to school. The norm-change aspects of HATUTAN, especially those targeting gender inequality and other harmful practices, could also result in a better environment for students and therefore higher attendance.

STUDENT ATTENDANCE INDICATORS

The overall attendance rate for each grade is reported in Table 50. Attendance rate is calculated as the total number of students observed in a class divided by the total number of students recorded as being enrolled. This analysis excluded classes that recorded a higher number of students attending class than were enrolled in those classes. Also, since the data is school-level, this analysis tracks the same units over time and therefore takes a panel-based approach.

Table 50: Change in attendance rate by grade

	Comparison Schools			Intervention Schools			Difference in Differences	
Grade 1	BL	ML	Difference	BL	ML	Difference	DiD	p
n	44	78		70	93			
Attendance rate	55.6%	63.3%	7.7	60.1%	62.8%	2.7	-5.8	0.32
Grade 2	BL	ML		BL	ML			
n	58	83		69	91			
Attendance rate	64.8%	75.1%	10.3	67.4%	71.2%	3.8	-6.4	0.22
Grade 3	BL	ML		BL	ML			
n	41	79		63	94			
Attendance rate	68%	73.5%	5.5	69.8%	70.3%	0.5	-5	0.39

Grade 4	BL	ML		BL	ML		
n	41	76		71	94		
Attendance rate	73.1%	74.5%	1.4	71.8%	73.9%	2.1	0.7 0.9
Grade 5	BL	ML		BL	ML		
n	28	61		69	83		
Attendance rate	67.2%	78.0%	10.8	75.7%	74.6%	-1.1	-11.8 0.06
Grade 6	BL	ML		BL	ML		
n	28	58		58	83		
Attendance rate	68.5%	83.8%	15.3	77.0%	76.2%	-0.8	-16.1 0.01*

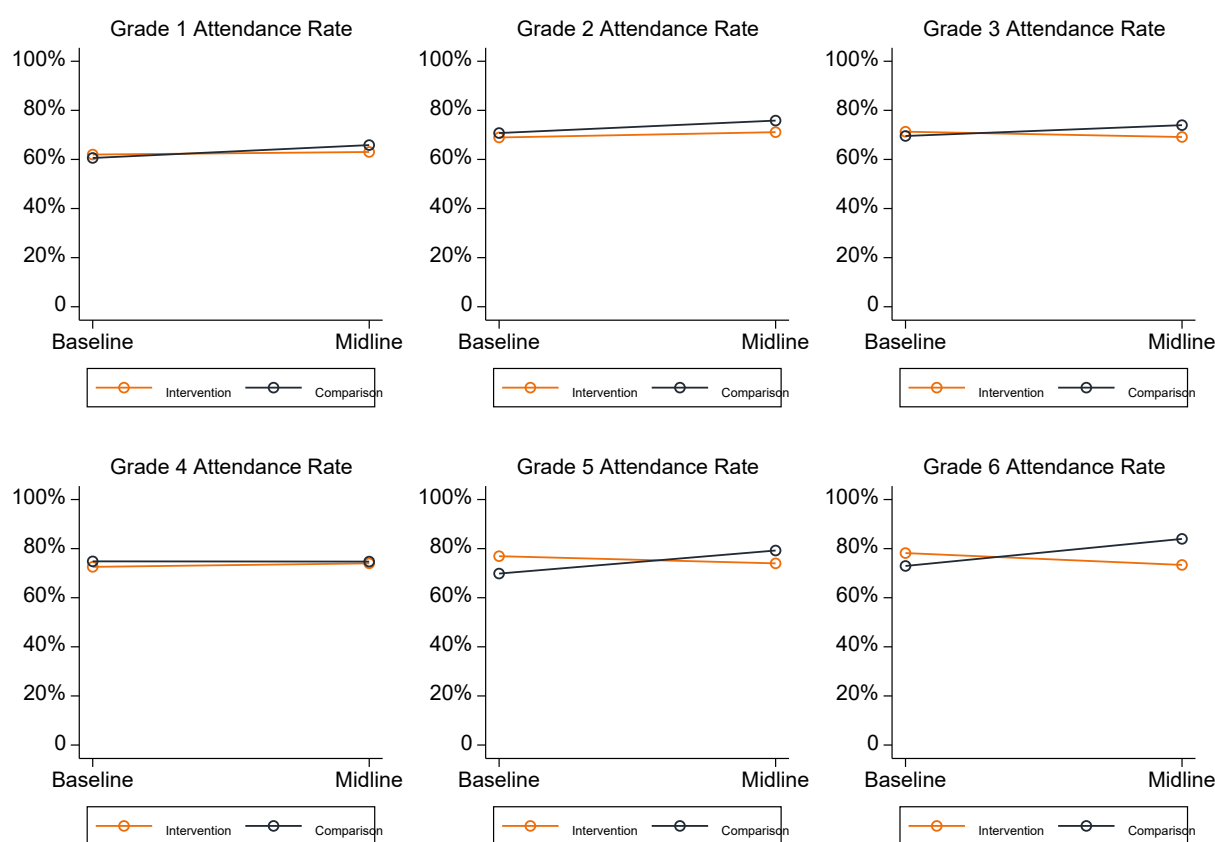
* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The first key thing to note is that the sample size, particularly at baseline and in the comparison group, is small. This matters from the point of view of achieving statistical significance, but more importantly from the point of view of introducing bias into the sample as those that responded may not be representative of the whole sample. We must therefore be cautious in interpreting changes over time and difference-in-difference statistics, as an observed increase in the comparison group could simply be attributable to the effect of the new schools at midline, rather than a real increase.

Statistical significance for these findings is tested in the table above. All difference in difference scores are negative, but most are not statistically significant, except for grade 6 students, for whom there was a 16 percentage point difference in attendance rates between intervention and comparison groups from baseline to midline. However, it is important to note that the bias discussed above is unrelated to whether the finding is statistically significant and we must be cautious in attributing the results to program impact.

Across most grades and for both treatment and comparison groups, the attendance rate increased over the study period, although there were declines in the treatment group for grades 5 and 6. Rates ranged quite widely, from 56% to 84%, and attendance rate tended to be higher in higher grades compared to lower. The rate of increase with respect to treatment can be seen more closely in Figure 15 below. For Grades 5 and 6, the decrease in attendance in treatment was matched with an increase in attendance for comparison, leading to a negative difference-in-difference score. However, in all other grades where there was an increase in attendance for the treatment group (aside from grade 4), the increase was greater for the comparison group, again leading to negative difference-in-difference scores. The data, therefore, does not seem to provide evidence for positive program impact on student attendance. However, as noted above, it is difficult to draw conclusions given the small sample size at comparison baseline; without sufficient data for this group, we do not have a robust counterfactual to estimate what would have happened without treatment.

Figure 15: Change in attendance rate by grade



We additionally analyze attendance rates by gender among intervention schools. We find that at midline, in all grades, girls had consistently higher average attendance rates than boys (Table 51). In grades 2 and 6, girls' attendance rates were significantly higher than boys', at around 6 percentage points greater in both grades. Similarly, at baseline, girls' attendance rates were higher than those of boys in all grades except grade 2, although girls' attendance was only significantly higher among grade 1 students. These results suggest that the gap between girls' and boys' attendance may have widened slightly at midline, but that overall, there was little significant change in the difference between girls' and boys' attendance.

Table 51: Student attendance by gender, intervention schools

	Female	Male	Difference	p
Midline				
Grade 1	64.0% (n = 94)	62.0% (n = 95)	2.0	0.45
Grade 2	74.4% (n = 93)	68.7% (n = 92)	5.7	0.04*
Grade 3	72.8% (n = 98)	68.9% (n = 94)	3.9	0.05
Grade 4	76.0% (n = 93)	72.1% (n = 95)	3.9	0.11
Grade 5	76.1% (n = 86)	72.3% (n = 89)	3.8	0.07
Grade 6	78.3% (n = 83)	72.7% (n = 89)	5.6	0.02*
Baseline				
Grade 1	62.7% (n = 73)	58.4% (n = 72)	4.3	0.03*
Grade 2	66.3% (n = 73)	68.2% (n = 72)	-1.9	0.44
Grade 3	68.7% (n = 67)	68.1% (n = 69)	0.6	0.72
Grade 4	72.6% (n = 72)	70.6% (n = 74)	2.0	0.33
Grade 5	76.0% (n = 73)	72.7% (n = 72)	3.3	0.10
Grade 6	77.4% (n = 62)	74.1% (n = 61)	3.3	0.24

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The main reasons given for school absence can be seen below. This data is from the household survey and the analysis is cross-sectional. If a student missed a school day last week, parents were asked to give the main reason for the absence. By far the main reason given overall was sickness, with natural disaster, not wanting to go to school, and “other” (where the reason was not specified) also scoring highly. In both comparison and treatment, the proportion missing school due to natural disaster rose greatly, as Timor-Leste experienced much heavier rains and winds, especially in mountain areas, during the rainy season at midline. The increase in absence did not affect the groups evenly – it rose by around 13 percentage points in the comparison group and 23 percentage points in the treatment, as expected, given that many treatment schools are in more remote areas where students must cross rivers or walk across washed-out roads to reach schools. This is likely a contributing factor to the negative difference-in-difference scores seen above.

Table 52: Reason for missing school

	Comparison		Intervention	
	BL	ML	BL	ML
n	121	205	227	324
Sick	57.0%	35.1%	49.8%	33.0%
Natural disaster	4.1%	16.6%	11.5%	34.9%
Other unspecified	13.2%	13.7%	16.7%	11.7%
Did not want to go to school	18.2%	16.1%	7.9%	10.2%
Funeral, marriage, traditional ritual	2.5%	4.9%	4.0%	3.4%

Household chores or caring	0.0%	2.4%	4.0%	3.1%
Teacher did not attend	0.8%	2.9%	3.5%	1.9%
Other (farm work, help with business, etc.)	4.2%	8.3%	2.6%	1.8%

Below, we disaggregate attendance rates by municipality. The negative difference in difference scores above were attributable to increases among the comparison group more so than decreases in the treatment group, so we check the comparison group for patterns. Most grades saw increases in attendance across all municipalities aside from Baucau, suggesting the rise was evenly distributed and not driven by particular circumstances in a single municipality. Among intervention municipalities, we note that attendance fell in Liquica at midline for all grades, while it rose or remained fairly steady in most grades in Ainaro, Ermera, and Manatuto. This suggests that attendance in Liquica may have been particularly affected by municipality-specific factors, such as severe local weather.

Table 53: Student attendance in comparison group, disaggregated by municipality

	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
	BL	ML	BL	ML	BL	ML	BL	ML	BL	ML	BL	ML
Comparison Municipality												
Aileu	53.6%	59.6%	63.2%	73.1%	68.7%	71.0%	69.9%	78.0%	71.7%	69.0%	73.6%	82.9%
Baucau	83.3%	85.4%	94.4%	88.8%	90.0%	95.0%	100.0%	59.2%	100.0%	98.3%	100.0%	96.7%
Bobonaro	48.9%	61.9%	59.1%	75.9%	68.2%	69.3%	73.6%	69.0%	60.8%	75.4%	65.6%	81.5%
Covalima	61.4%	62.8%	86.7%	82.5%	71.8%	78.2%	75.7%	79.0%	60.8%	82.7%	68.1%	87.5%
Manufahi	60.5%	66.0%	56.7%	70.6%	57.7%	77.3%	69.7%	78.5%	71.4%	82.2%	63.8%	83.3%
Intervention Municipality												
Ainaro	48.2%	62.6%	59.8%	70.3%	59.6%	70.7%	61.4%	74.8%	66.6%	74.2%	68.5%	75.5%
Ermera	62.6%	59.9%	64.4%	69.2%	69.9%	65.6%	71.2%	70.7%	75.6%	75.7%	77.4%	77.4%
Liquica	63.6%	61.6%	76.2%	69.1%	80.9%	66.9%	83.1%	69.8%	79.7%	59.1%	83.6%	63.8%
Manatuto	66.9%	70.2%	84.3%	78.2%	75.5%	81.3%	79.1%	82.4%	82.3%	82.0%	81.4%	80.9%

The McGovern-Dole standard outcome #2 assesses the percentage of schools that had an average attendance rate of at least 80%. At baseline the overall attendance rate in schools was 36%, with treatment schools having a higher attendance rate at 39% compared to only 24% for comparison (although the sample size was quite small). At midline the overall percentage of schools meeting this target was again 36%, with comparison schools rising to 30% and treatment schools falling to 33%. The target for year 3 set out in the baseline was that 34% of schools would have attendance rates over 80%. This target has not been met, but this must be viewed in the context of increased absences due to natural disasters as discussed above.

Next, dropout rates are analyzed. This is calculated as the number of dropouts (as recorded by the school) divided by the number of students enrolled in each class. Again, some sample sizes are small, especially for later grades in the comparison group, which will lead to the same difficulties related to statistical significance and bias mentioned above. Given this caveat, quite a clear pattern emerges from the table below – the average dropout rate increased for all grades in the comparison group, while the average dropout rate decreased across all grades for the treatment group.

In Table 54 we also test for the statistical significance of the difference-in-differences results. The difference in difference is negative for every grade, and is not statistically significant for grade 3 but is statistically significant for all other grades. This may suggest a positive program impact, possibly due to parents or caregivers being less likely to draw their children out of school if there is a strong school feeding program or children being less likely to want to drop out. However, due to the high non-

response rate at baseline, care must be taken in attributing impact to this factor or other possible factors from the program, as increases in the dropout rate in the comparison could have feasibly been due to the new schools at midline having higher dropout rates rather than dropout rates actually increasing across the study period.

This finding is visualized in Figure 16 below. A negative difference in difference (which for this indicator is regarded as a positive outcome) is visible across all grades, with the orange intervention group line starting above the blue comparison group line and ending below the comparison group line for every grade. The effect seems stronger for grades 1, 2, and 6.

Table 54: Change in dropout rates by grade

	Comparison Schools			Intervention Schools			Difference in Differences	
Grade 1	BL	ML	Difference	BL	ML	Difference	DiD	p
n	54	88		71	98			
Dropout rate	2.8%	4.2%	1.4	4.5%	0.8%	-3.8	-5.1	0.01*
Grade 2	BL	ML		BL	ML			
n	63	88		73	98			
Dropout rate	2.5%	4.0%	1.5	3.8%	0.8%	-3.1	-4.6	0.03*
Grade 3	BL	ML		BL	ML			
n	56	87		70	98			
Dropout rate	1.7%	2.2%	0.5	2.8%	1.2%	-1.6	-2.1	0.20
Grade 4	BL	ML		BL	ML			
n	50	80		74	97			
Dropout rate	1.4%	2.8%	1.4	3.0%	0.8%	-2.2	-3.6	0.04*
Grade 5	BL	ML		BL	ML			
n	38	67		73	94			
Dropout rate	0.6%	1.8%	1.2	3.0%	0.3%	-2.7	-3.8	0.03*
Grade 6	BL	ML		BL	ML			
n	36	62		61	92			
Dropout rate	0.7%	4.3%	3.6	3.5%	0.4%	-3.1	-6.7	0.03*
Overall rate								
n	50	83		55	95			
Dropout rate	1.7%	3.0%	1.3	3.6%	0.7%	-2.9	-4.1	0.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Figure 16: Change in dropout rates by grade



The negative difference in difference scores was driven to a greater extent by decreases in the dropout rate among intervention schools, so the scores for intervention schools are disaggregated by municipality in Table 55 below. Most grades across all municipalities saw a fall in dropout rates over the study period, meaning it was well dispersed across the intervention group and not due to a peculiarity in a single municipality. It is worth noting, however, that decreases in the dropout rate are particularly noticeable in Ermera and are weakest in Liquica.

Table 55: Dropout rate for treatment group, disaggregated by municipality

	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
	BL	ML	BL	ML	BL	ML	BL	ML	BL	ML	BL	ML
Municipality												
Ainaro	2.2%	1.2%	2.4%	1.0%	1.8%	0.0%	3.1%	1.1%	1.6%	1.0%	2.5%	0.6%
Ermera	6.9%	0.4%	6.4%	0.8%	3.0%	1.4%	4.1%	0.9%	4.6%	0.1%	5.0%	0.4%
Liquica	3.3%	1.8%	0.0%	0.0%	0.0%	0.8%	1.1%	1.4%	0.7%	0.0%	0.2%	0.9%
Manatuto	1.3%	0.3%	0.8%	0.8%	5.3%	2.4%	0.6%	0.09%	1.0%	0.0%	2.3%	0.2%

Based on qualitative data collected, school coordinators/directors highlighted the value of the school feeding program in improving student attendance, as well as other household factors that cause children to skip school such as helping with chores or family business, death in the family, and sickness. Some of the school coordinators/directors noted that sometimes parents do not inform the school when children are missing school due to sickness. A number of teachers also said that PTA members also look for students who have been missing from school for a while to determine the reason for skipping school. However, there were also parents who pointed out that children going to school is also influenced by the parents' willingness to send their children to school.

When there is no school feeding, the presence of the children decreases remarkably. But when there is school lunch, then the children - because at home they only eat corn, cassava, bananas, our local products - like to come to school. There are fewer absences. But there is another problem, the parents are not conscious on the days when there is a bazaar, they take their children to the bazaar. Usually parents go to the market, other times they send their children to the market. Sometimes they tell their children "take care of your little brother, I'm going to the market." These are things that have an impact on children.¹³¹

Children often missed school because when parents go to traditional ceremonies, they also take their children with them, and so their children cannot come to school. In the months of August and September, many traditional ceremonies took place. Parents, instead of bringing their children to school, took them with them to these ceremonies. So, their children didn't come to school. On the other hand, parents also have no control over their children, and they are constantly missing classes. Parents don't pay attention to their children, so they miss school. And from our side, from the teachers' side, when they don't come to school, we impose some sanctions on the children to call their parents to school.¹³²

The number of attendances has increased. When there is school meal provided and when we prepare meal, we check the attendance list and compared them. The previous numbers were smaller, sometimes there were only around 100 students. Now and maybe in the coming days, when the students assemble in the front yard and we counted them, there were more than 200 students. This shows that the result of the collaboration between the teachers and the PTAs is working.¹³³

Some of the parents highlighted the importance of parental support and teacher attendance in student attendance and performance. Parental support would include provision of school uniforms and supplies. However, there is also recognition that financial constraints may prevent parents from providing these.

Sometimes parents are not supportive, and so sometimes children are worried that they are going to school without parental support. Therefore, when they [children] think like this [that parents are not supportive], they just spend time playing at home.¹³⁴

The other reason is the school uniform. Some parents could not afford to buy the uniform. In the school, their friends [wore] uniforms, but they did not. They felt bad and said "I am [don't look] like a student. My school friends wear complete uniform but not me. It is better to stay home and support my parents [with chores]."¹³⁵

¹³¹ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

¹³² FGD with mothers, female, Ermera municipality, Int. 103

¹³³ FGD with teachers, male, Ermera municipality, Int. 137

¹³⁴ FGD with mothers, female, Ermera municipality, Int. 101

¹³⁵ FGD with mothers, female, Ermera municipality, Int. 101

*It is good [easy] for parents who have a job, but [for] those who don't have a job it will be difficult for them to help their children. Everything that people sell in the market, no one buys. If the parents are not selling, how can they make money to sustain their children to go to school? Things like notebooks, pens and shoes require money. Having so many children in a family, some unfortunately cannot go to school. It would be fine to have many children if the father has a job. Many fathers do not have a job. We have to carry our vegetables and chilis to market to sell. Only then we can make money to buy their needs.*¹³⁶

*In this case [where schools reopened and students did not show up], I would say that as parents we would probably blame the teacher. Students have interest to come to school, but if there was no one to look after them [when] they would do some activities at school. Instead, [students] go back home because there were no teachers there [in school].*¹³⁷

The closure of schools due to the COVID-19 pandemic prolonged students' stay at home. Based on qualitative data, when the schools reopened, school coordinators/directors observed a huge drop in student attendance. Parents said that they had to force their children to go back to school as it was perceived safe to do so. It is also evident in the FGDs that parents want their children to go back to school instead of staying home to play.

*Our difficulties at the moment are this state of emergency is very prolonged for the children, so the children stay [at home and] not going to school and end up not coming back to school anymore. They stay [at home] too long without coming to school, and if we have taught them something – like earlier when I showed them how to share – but when they come back and if we ask the same question again, they don't remember anymore.*¹³⁸

*The more they are playing at home, the more they get used to it. If we want them to go to school, we need to force them by threatening [them] or [using] the broomstick to scare them. Then they will go to school. If they wake up late in the morning that would be a day off for them. If they go to school everyday, they said we wake up earlier because we should go to school. If they just stay/play at home, they get used to it and they do not have desire [or lost the desire] to go to school. We have to force them again to go back. If we don't then they won't go back. This is what impacts [or affects] their school attendance.*¹³⁹

It is because they did not send us the notice. There should be any notice from school. Because if the teachers notice that the students do not show up, they should send out notice to the parents so they could have a discussion on how to send their children back to school. Find out why they did not go back, so that parents could have some

¹³⁶ FGD with mothers, female, Ermera municipality, Int. 101

¹³⁷ FGD with fathers, male, Ermera municipality, Int. 105

¹³⁸ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

¹³⁹ FGD with mothers, female, Ermera municipality, Int. 101

explanations why their children did not return to school, and teachers would be aware of it. From the school, they would always say school is open. ¹⁴⁰

The role of the PTA was also highlighted, where the PTA provides support to the teachers in informing parents regarding their children's absences so that related challenges could be addressed. However, this was not the case for all schools.

We have established the PTA, and their job is to remind or inform the parents if the children are absent very often. They will prepare and send notification letters to their parents, directly to their homes. But there has not been any case where a child is absent too often and thus require the PTA to apply such measure for their parents to respond. Since there is no case of frequent absence, no parents have been asked to respond. ¹⁴¹

If a child is not coming to school or dropped out of school, it becomes PTA's priority. They will go out there and collect information from parents and then they will try to convince the parents [to send the child back to school]. Together, they will encourage their child to continue his/her study. This is their job. They act as CCTV, so once a child is frequently absent means that something causes his/her absence. The PTA will directly visit the child's home and they will do everything possible to bring the child back to school. ¹⁴²

The PTA didn't track down the children. Here we are the ones who selected PTA members and there were no PTA members or teachers who went to our house and asked for our children to return to school. If they have our telephone numbers with them, then they will contact us directly. ¹⁴³

Below, we further analyze dropout rates by gender for intervention schools. At midline, girls had lower average dropout rates in every grade, although the difference between girls' and boys' dropout rates was not significant. Likewise, at baseline, there were no significant differences between girls' and boys' dropout rates in any grade, although girls' dropout rates were marginally higher than boys in grades 3, 5, and 6. It is worth noting as above that, especially at midline, dropout rates were very low for both girls and boys, and that dropout rates appear to have decreased at fairly similar rates for both girls and boys from baseline to midline.

¹⁴⁰ FGD with fathers, male, Ermera municipality, Int. 105

¹⁴¹ FGD with fathers, male, Manatuto municipality, Int. 115

¹⁴² FGD with fathers, male, Manatuto municipality, Int. 115

¹⁴³ FGD with mothers, female, Ermera municipality, Int. 131

Table 56: Dropout rates for treatment group, disaggregated by gender

	Female	Male	Difference	p
Midline				
Grade 1	0.6% (n = 98)	0.9% (n = 97)	-0.3	0.39
Grade 2	0.3% (n = 98)	1.2% (n = 96)	-0.9	0.05
Grade 3	1.0% (n = 98)	1.4% (n = 98)	-0.4	0.51
Grade 4	0.8% (n = 95)	1.2% (n = 98)	-0.4	0.78
Grade 5	0.2% (n = 94)	0.5% (n = 94)	-0.3	0.21
Grade 6	0.3% (n = 90)	0.6% (n = 92)	-0.3	0.49
Total	0.6% (n = 95)	0.9% (n = 96)	-0.3	0.13
Baseline				
Grade 1	4.3% (n = 72)	4.6% (n = 71)	-0.3	0.72
Grade 2	3.5% (n = 73)	4.4% (n = 73)	-0.9	0.18
Grade 3	2.8% (n = 71)	2.3% (n = 72)	0.5	0.51
Grade 4	2.8% (n = 74)	3.1% (n = 75)	-0.3	0.67
Grade 5	3.2% (n = 75)	2.9% (n = 75)	0.3	0.56
Grade 6	3.5% (n = 62)	3.3% (n = 63)	0.2	0.89
Total	3.5% (n = 55)	3.6% (n = 56)	-0.1	0.96

Student anxiety and depression, both factors that may be linked to dropout and school absence, are analyzed in Table 57 below. Parents/caregivers were asked how often students displayed signs of either. Student anxiety declined slightly in comparison and intervention groups, while changes in student depression were more mixed.

Table 57: Summary scores for student anxiety and depression

	Comparison Areas		Intervention Areas	
	BL	ML	BL	ML
Student anxiety				
n	378	622	482	733
Daily	13.5%	6.9%	8.7%	6.3%
Weekly	8.2%	6.3%	10.0%	4.1%
Monthly	3.7%	6.6%	5.0%	2.7%
A few times a year	33.6%	37.9%	40.9%	43.6%
Never	41.0%	42.3%	35.5%	43.3%
Student depression				
n	378	622	482	733
Daily	6.1%	2.6%	1.7%	1.8%

Weekly	3.4%	4.2%	4.8%	1.9%
Monthly	3.7%	3.9%	3.9%	2.3%
A few times a year	31.2%	31.4%	33.4%	32.7%
Never	55.6%	58.0%	56.2%	61.3%

We further disaggregate reported student anxiety and depression in intervention areas by the gender of the student and the municipality in which they live to better understand how anxiety/depression may vary by demographic characteristics. Table 58 below shows that there are no significant differences in rates of anxiety/depression by gender at midline. At baseline, however, caregivers of female students were significantly less likely to report that their student experienced depression monthly, although we note that rates for both girls and boys were relatively low, with the vast majority reported to experience depression a few times a year or never.

Table 58: Student anxiety/depression, disaggregated by gender in intervention areas

	Anxiety				Depression			
	Female	Male	Difference	p	Female	Male	Difference	p
Midline								
n	372	361			372	361		
Daily	6.5%	6.1%	0.4	0.84	2.2%	1.4%	0.8	0.43
Weekly	3.2%	5.0%	-1.8	0.23	2.2%	1.7%	0.5	0.63
Monthly	3.0%	2.5%	0.5	0.70	2.7%	1.9%	0.8	0.50
A few times a year	41.1%	46.3%	-5.2	0.16	32.5%	33.0%	-0.5	0.90
Never	46.2%	40.2%	6.0	0.10	60.5%	62.1%	-1.6	0.67
Baseline								
n	234	248			234	248		
Daily	8.1%	9.3%	-1.2	0.65	2.6%	0.8%	1.8	0.13
Weekly	9.4%	10.5%	-1.1	0.69	4.3%	5.2%	-0.9	0.62
Monthly	4.3%	5.7%	-1.4	0.49	1.7%	6.1%	-4.4	.01*
A few times a year	41.5%	40.3%	1.2	0.80	35.0%	31.9%	3.1	0.46
Never	36.8%	34.3%	2.5	0.57	56.4%	56.1%	0.3	0.94

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Disaggregating by municipality for treatment areas, we find that caregivers in Ermera and Manatuto were significantly less likely to report that their student experiences anxiety daily, weekly, or monthly at midline than at baseline. There were no significant differences for Ainaro and Liquica. This suggests that municipality-level factors may be driving changes in student anxiety and depression, rather than broader program effects.

Table 59: Student anxiety and depression, disaggregated by municipality

	Experiences anxiety				Experiences depression			
	BL	ML	Difference	p	BL	ML	Difference	p
Ainaro	10.9%	16.9%	6.0	0.14	3.1%	5.8%	2.7	0.27
Ermera	36.1%	13.0%	-23.1	<0.001***	17.8%	7.5%	-10.3	<0.001***
Liquica	9.1%	7.0%	-2.1	0.65	3.6%	8.1%	4.5	0.29
Manatuto	22.0%	11.7%	-10.3	.04*	7.7%	1.5%	-6.2	.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

It is not possible to directly assess the link between anxiety/depression (which was surveyed at individual-level) and dropout and attendance rates (which was surveyed at school-level). However, observing the link between these two and between students not wanting to go to school reveals

associations. Across all students, 11% of those whose parents said they never seemed anxious or only seemed anxious a few times a year were reported to have missed school because they did not want to go, compared to 17% among those who seemed anxious at least monthly. With that said, changes across time for these indicators were modest, so its relevance regarding whether the program had impact or not is limited.

HEALTH-RELATED ABSENCES

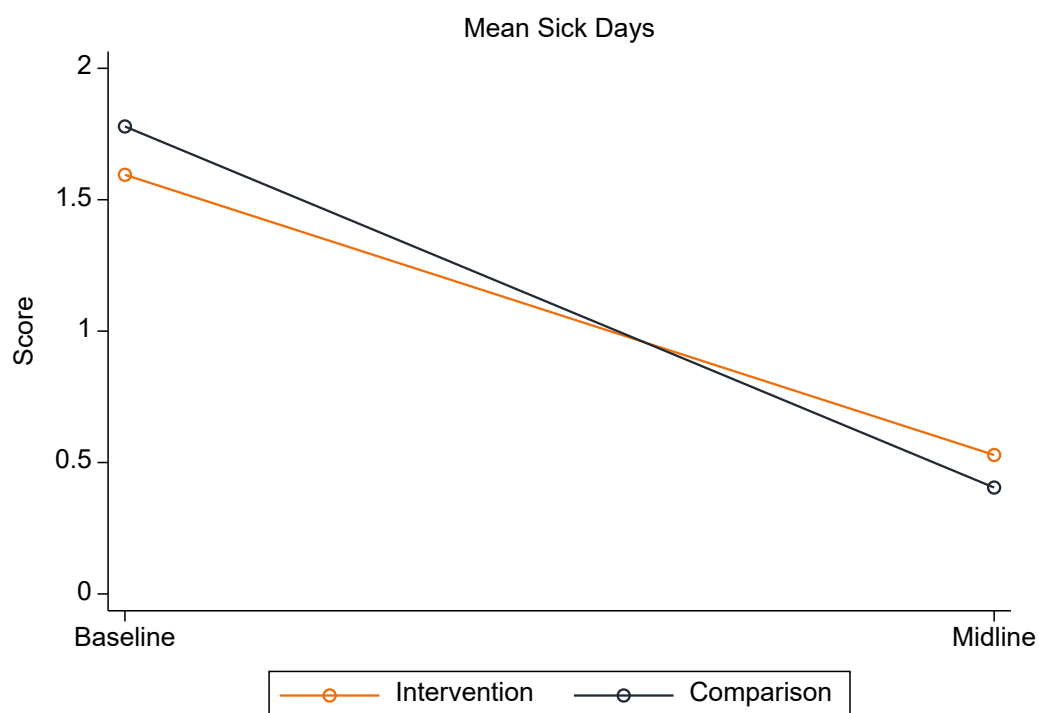
Of particular interest are health-related absences, which are a key sub-factor in student attendance overall. Parents and caregivers were asked in the household survey to report how many days the student had missed due to illness. It is important to note that the period asked in the question changed from baseline to midline – in baseline the question referred to absences in the past month, while at midline the question referred to absences in the past week. While we expect a fall in absences reported at midline (to approximately 25% of the baseline score), we do not expect this to be different across treatment and comparison. We therefore expect to still extract a meaningful result from the difference in difference score.

Table 60: Summary scores for health-related absences

	Comparison Schools		Intervention Schools	
Days missed due to illness	BL	ML	BL	ML
Number of respondents	378	620	482	732
None	47.6%	82.6%	47.1%	78.1%
1 to 2 days	25.9%	12.1%	30.7%	14.9%
3 to 5 days	18.8%	4.5%	14.5%	5.7%
6 or more days	7.7%	0.8%	7.7%	1.2%
Means days missed	BL	ML	BL	ML
Number of respondents	375	620	479	732
Means days missed total	1.8	0.4	1.6	0.5
Mean days missed female	1.8	0.4	1.8	0.5
Mean days missed male	1.7	0.4	1.4	0.5

As expected, the percentage missing school due to sickness falls from baseline to midline. Observing the mean days missed, the fall seems greater for the comparison group than the intervention group, with the comparison group scores falling to less than the expected 25% and the intervention group falling to more than the expected 25%. This pattern can be seen in Figure 17 below – the comparison group starts with a higher score for mean days absent due to illness and ends at midline with a lower value for mean days absent due to illness. In part, the overall fall in days missed due to absence may be due to COVID-19 measures, such as handwashing and mask wearing, reducing the spread of all illnesses.

Figure 17: Mean days missed due to illness



As can be seen in Table 61 which reports the results of the difference-in-differences regression, these results are statistically significant for male students but not for female students. It is somewhat difficult to rationalize this result, as we would not expect the higher prevalence of natural disaster noted above to increase days off due to illness in the treatment group.

Table 61: Difference-in-differences, days missed due to illness

	Intervention			Comparison			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Days missed	1.6	0.5	-1.1	1.8	0.4	-1.4	0.3	0.06
Days missed female	1.8	0.5	-1.3	1.8	0.4	-1.4	0.2	0.48
Days missed male	1.4	0.5	-0.9	1.7	0.4	-1.3	0.4	0.03*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The indicator related to absences due to illness is McGovern-Dole Custom Outcome #16. The target for this indicator for year 3 was that 1.3 days would be missed monthly, equating to approximately 0.3 days weekly. Observing the mean days missed in the treatment group at midline (0.5 days overall), we can see that this target has not been met.

SCHOOL INFRASTRUCTURE

This section will analyze school infrastructure, which is identified in the results framework as a factor that may contribute to student attendance. Information on school infrastructure related to the school feeding program is analyzed in detail in the section “School Feeding Program;” we focus here on other infrastructure. Enumerators were asked to report whether the school has improved water, electricity, and a canteen. They were also asked to report how many toilets the school had. Summary scores are given below. Again, the first thing to note is the small sample size for toilets, food storage, and canteen

at baseline for the comparison group. As explained above, this poses issues for statistical significance and may introduce bias.

The difference-in-difference regression for all infrastructure components is reported in Table 62 below. An overall picture of quality of infrastructure between treatment and comparison is difficult to give, as in both groups some indicators see a deterioration while others see an improvement. Of particular note is the increase in schools in the intervention group with a toilet, improved water, and electricity, and the decrease in intervention schools with a canteen. This may point to success in the program's advocacy for infrastructure improvements and PTAs' support for and contributions to infrastructure.

Table 62: Change over time in school infrastructure

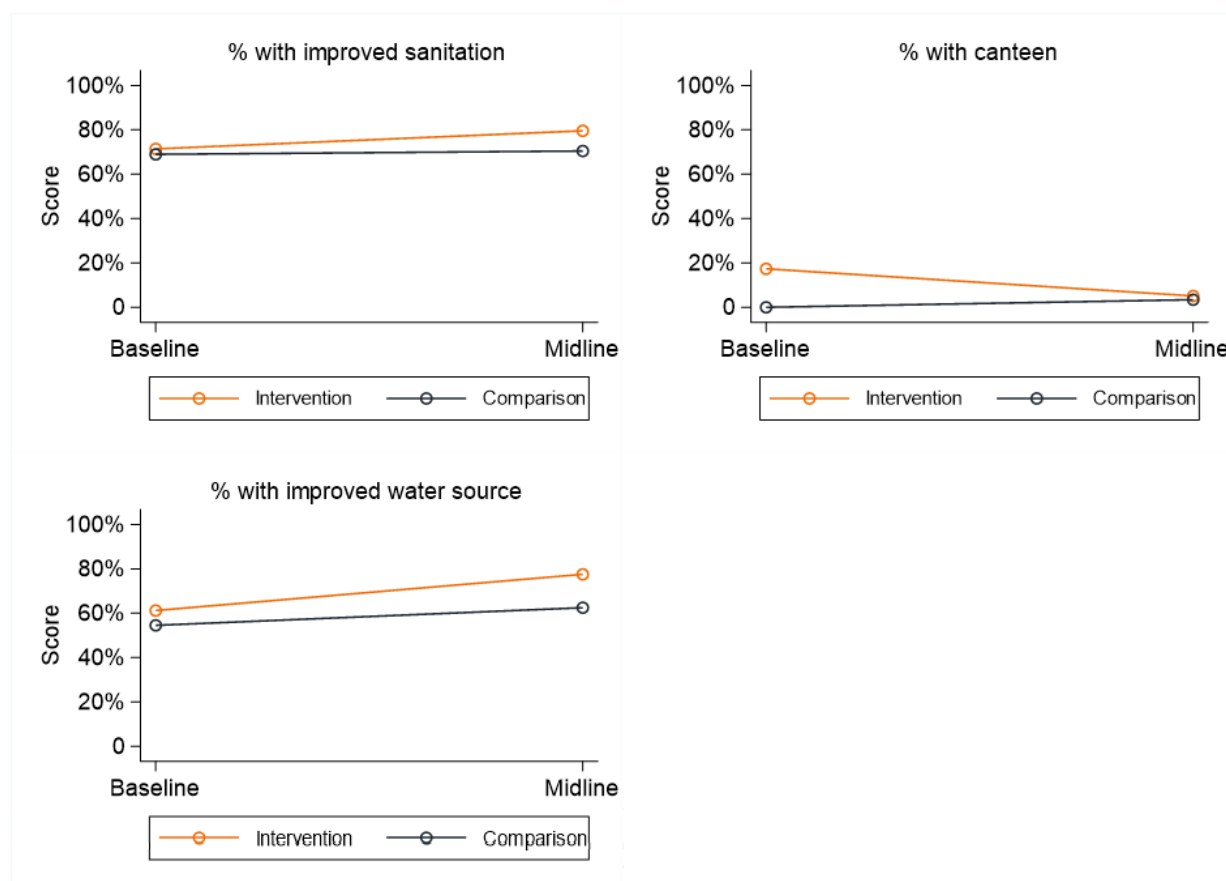
	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	88	88		98	98			
% with toilets	69.0% (n = 87)	70.6%	1.4	71.4%	79.6%	8.2	7.0	0.76
% with water	54.6%	62.5%	7.9	61.2%	77.6%	16.4	8.4	0.40
% with electricity	46.6%	47.7%	1.1	60.2%	66.3%	6.1	5.0	0.6
% with canteen	0.0% (n = 28)	3.4%	3.4	17.4%	5.1%	-12.3	-15.7	0.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

As mentioned earlier, the picture with respect to program impact is mixed. Some indicators show a negative difference-in-difference while others show a positive score. The is further complicated by the small sample size at baseline which may distort some results. None of the results are statistically significant aside from schools having a canteen, which shows a negative difference in difference score of 16 percentage points.

Three indicators are visualized below: Percentage with improved sanitation (defined as a school having at least one toilet), percentage with an improved water source, and percentage with a canteen. Percentage with improved sanitation and percentage with an improved water source had near 100% response rates at baseline and midline and showed a positive difference in difference. It can be clearly seen that the percentage with improved sanitation and with an improved water source grew faster in the intervention schools than comparison schools. The percentage of schools with a canteen dropped sharply for the intervention group; however, this finding is likely to have been due to enumerator error, as canteens are unlikely to have disappeared in schools unless the school was affected by a natural disaster. Canteens are rare in Timorese schools, and the presence of canteens may have been overestimated at baseline if enumerators did not fully understand what qualifies as a canteen.

Figure 18: Change over time in school infrastructure



The intervention group for these three indicators are disaggregated by municipality and rural/urban divide in Table 63. The indicators that improved at midline—percent with improved water and percent with improved sanitation—show a relatively even rise across all municipalities, suggesting that the improvement was not due to particular factors in one municipality and providing further evidence of program impact. Sanitation improved unevenly across location, with rural schools showing a 10 percentage point increase and urban schools showing no increase at all. It is also apparent that the decrease in schools with a canteen is entirely driven by the Ermera municipality, where the score fell from 32% to 0%. This further suggests that enumerator error may have affected the data on canteens.

Table 63: School infrastructure, disaggregated by municipality and location

	School Canteen		Water		Toilet	
	BL	ML	BL	ML	BL	ML
Municipality						
Ainaro	11.1%	11.1%	48.2%	63.0%	59.3%	59.3%
Ermera	31.7%	0.0%	68.3%	85.4%	68.3%	85.4%
Liquica	9.1%	9.1%	72.7%	90.9%	100%	100.0%
Manatuto	0.0%	5.3%	57.9%	73.7%	79.0%	84.2%
Location						
Rural	17.5%	3.8%	62.5%	77.5%	68.8%	78.8%
Urban	16.7%	11.1%	55.6%	77.8%	83.3%	83.3%

In Table 64 below the water source for the schools is reported. Intervention schools fared better over the study period, with increases in the percentage of schools with piped water either in the school or nearby and a decrease in those answering “no water available”.

Table 64: Water source of schools

	Comparison Schools		Intervention Schools	
	BL	ML	BL	ML
Number of respondents	88	88	98	98
Piped water at school	50.0%	50.0%	57.1%	63.3%
Piped water nearby	19.3%	15.9%	14.3%	17.4%
No water available/too far	15.9%	15.9%	16.3%	7.1%
Spring Water	2.3%	8.0%	7.1%	8.2%
Rainwater harvesting	3.4%	1.1%	3.1%	0.0%
Well or pump nearby	3.4%	4.6%	1.0%	3.1%
Well at school	3.4%	3.4%	1.0%	1.0%
Borehole at school	2.3%	1.1%	0.0%	0.0%

The final indicator of this section is the “infrastructure score,” which is an aggregate indicator giving an overall measure of the infrastructure at schools. It is reported on a scale of 1-6, where one point is assigned for the presence of six components of infrastructure: electricity, a kitchen, water, a canteen, toilets, and a food storage space. The table below shows the percentage of schools with “improved infrastructure,” defined as having at least 4 of the 6 elements. The percentage of schools meeting this standard improved for both the comparison and intervention groups, although again care must be taken in interpretation as the sample size at baseline is low.

Table 65: Change over time in schools with improved infrastructure

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	28	71		98	95			
Schools with improved infrastructure	21.4%	51.1%	29.7	62.2%	75.8%	13.6	-17.1	0.17

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

OTHER FACTORS

This section will analyze other factors that may affect school attendance, including factors related to school access and whether a student enjoys school or not. Caregivers were asked whether it was safe for their grade 2 child to walk to school and whether their child avoids or is afraid of school. At midline, the majority of caregivers with both male and female children reported that it was safe for their child to walk to school (90% of caregivers with male children and 89% of caregivers with female children). Caregivers of female children were somewhat less likely to report that their child didn't want to attend school (33%) than caregivers of male children (37%). Thirteen percent of caregivers of both male and female children reported that their child was sometimes afraid of going to school; this percentage is fairly high, suggesting that perhaps this issue could be made a greater target in the program.

Table 66 disaggregates results by comparison and intervention households at baseline and midline, and shows that intervention households saw a significant improvement in some indicators measuring access to public space and services. Caregivers of male children in intervention households were significantly less likely to report that their child avoided school or was afraid to go to school than would

be expected given the results of comparison households. Furthermore, caregivers of female children were significantly more likely to report that it was safe for their child to walk to school than expected from the results of comparison households.

Table 66: Change in access to public spaces and services

s	Comparison Households			Intervention Households			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Male								
n	186	321		248	358			
Safe to walk to school	95.2%	91.9%	-3.3	92.3%	88.3%	-4	-0.8	0.86
Avoids school	35.5%	36.1%	0.6	48.6%	38.5%	-10.1	-11.1	0.04*
Afraid of school	10.8%	15.3%	4.5	17.9%	11.1%	-6.8	-11.4	0.01*
Female								
n	192	298		233	370			
Safe to walk to school	97.9%	89.6%	-8.3	85.4%	89.2%	3.8	12.1	0.005**
Avoids school	31.9%	31.9%	0	41.0%	33.0%	-8	-8.0	0.22
Afraid of school	9.4%	11.8%	2.4	15.0%	14.2%	-0.8	-3.2	0.41

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

These results suggest that there may have been some improvement in access to public space and services for students in intervention areas as compared to comparison areas. This improvement appears to have been slightly more salient for boys than for girls (with improvement in two indicators for boys, as opposed to just one for girls); however, both male and female students appear to have benefited. Furthermore, at midline, female students do not appear to face more risks to accessing education than male students. These results are particularly notable given the impact of COVID-19, which, in many countries, has made students more afraid of going to school due to fear of catching the disease.

In Table 67 below two additional variables related to school access are summarized. The data is from the household survey and thus the analysis is cross-sectional. Parents were asked how long it takes the student to walk to school and whether the student needs to use any transportation. It is unlikely that these factors are affected in any substantial way by the program, but it is important to understand their effect as they may impact attendance rates.

Table 67: Summary scores for access to school

	Comparison Schools		Intervention Schools	
	BL	ML	BL	ML
School walk time				
n	372	602	475	676
Walk less than 30 minutes	72.3%	69.9%	65.7%	64.2%
Student transportation				
n	377	622	481	733
Takes transportation	4.8%	4.2%	4.4%	3.4%

The percentage of students within 30 minutes from the school ranged from just below 65% to just above 70%. This indicator decreased slightly across the study period for both treatment and comparison, although it is important to note that since this survey was cross-sectional we are not measuring the same students over time. The percentage needing public transport ranged from 3.4% to 4.8% and the percentage feeling safe on their walk to school ranged from 76% to 84%. Differences between treatment and comparison and changes across time were relatively modest.

Further disaggregating by urban/rural location, we find that while there was no significant difference in perceived safety for caregiver of female students across urban and rural locations, caregivers of male students were significantly more likely to consider the walk to school unsafe in urban areas than in rural areas (86% considered the walk safe in urban areas, compared to 95% in rural areas). This pattern was also found at baseline.

Caregivers of male students were significantly more likely to say that traffic makes their child's walk unsafe than caregivers of female students at midline; there were otherwise no significant differences in reasons for which the walk is unsafe across caregivers of male and female children. In general, at midline, the most common factors making a walk unsafe were long distance, heat or rain, poor roads, river crossings, and traffic. Harassment and abuse-related reasons were reported infrequently or not at all; for example, no caregivers of either male or female children reported that the risk of being sexually abused by adults or by other children made the walk to school unsafe.

Students are also more likely to attend school if they find it enjoyable; one driver of school enjoyment may be whether a student has friends at school. The percentage of students having difficulty making friends was low at between 3% and 7.4%. The percentage increased for students in the comparison group by around 3 percentage points and remained effectively the same in intervention schools from baseline to midline.

Table 68: Summary scores of school enjoyment

	Comparison Schools		Intervention Schools	
	BL	ML	BL	ML
n	373	619	480	732
% with difficulty making friends	3.2%	7.4%	4.6%	4.9%

Caregivers were additionally asked if their children have equal access to toilets at school. At baseline 96% of caregivers said that boys and girls have equal access, while at midline, 78% of caregivers said boys and girls have equal access, 6% said there were no toilets, and 15% said they did not know. Within the school survey, enumerators also asked how many functional toilets were available for students at each school, and how many of those toilets were available for female students. Nearly 25% of schools (30% comparison, 20% treatment) did not have any functional toilets and 17% of schools (19% comparison, 15% treatment) had only one toilet. Of the schools with more than one toilet, around half had an equal number of toilets for girls and boys. However, only 69% of the treatment schools and 51% of the comparison schools had toilets for girls' use only. These results suggest that there may remain some barriers to girls' equitable access to school infrastructure; however, they also point to low levels of school infrastructure overall.

The importance of these factors is explored below, by testing whether the mean number of days missed by the student was different when grouped by each variable. The differences found are in the direction expected for each variable—students within a 30-minute walk to school and those who felt safe on their walk missed fewer days of school on average, while students who relied on transportation, were afraid of going to school, or had difficulty making friends missed more days of school. All findings were statistically significant aside from those for transportation and difficulty making friends. Some of the differences were substantial as well, with students having to walk longer than 30 minutes to school missing an extra third of a day on average each week. The program does not include these factors as key targets, but the analysis suggests that they are significant determinants for student attendance and may in turn influence literacy scores.

Table 69: Mean days of school missed by key indicators

Indicator	Mean (answering yes)	Mean (answering no)	Difference	p-value
Within 30 minutes walk	0.9	1.2	-0.3	<0.001***
Uses transportation	1.3	1.0	0.4	0.1
Feels safe walking to school	1.0	1.1	-0.2	0.04*
Is afraid of school	1.2	1.0	0.3	0.009**
Has difficulty making friends	1.1	1.0	0.1	0.28
Gender	1.1 (male)	0.9 (female)	0.2	0.08

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

SCHOOL FEEDING PROGRAM

One of the HATUTAN program's objectives is to support the Government of Timor-Leste in the implementation of the school feeding program in all basic education and preschools throughout the school year. Four-hundred and forty schools have been targeted,¹⁴⁴ and up to 445¹⁴⁵ have been provided daily meals prepared with USDA commodities. From these schools, 220 schools in particularly disadvantaged circumstances received a broader intervention package, including supporting 220 parent and teachers' associations. This involves developing the capacity of teachers and administrators to ensure effective management of the SFP. Moreover, the program supports farmers through purchase of local produce for the SFP, which aims to boost local production for a sustainable source of nutritious food for local schools.

IMPLEMENTATION AND MANAGEMENT OF THE SFP

One of the goals of the project is to improve the implementation and management of the school feeding program. The school survey gathered information on meal provision, who is responsible for leading SFP implementation in each school, and food items included in the menu. The household survey provided supplementary information on the characteristics of the meals in terms of quantity, taste, and hygienic food preparation. The below analysis related to the implementation and management of the SFP primarily focuses on the cross-sectional sample; any analysis using panel data will be referenced as such.

A small number of schools (9%) in the baseline mentioned that there is a project supporting schools with meals. This increased at midline (43%), which indicates increased knowledge of the school feeding intervention. The survey also found that schools have become more familiar with the type of projects supporting the provision of meals to students. Most of the schools in the midline (79%) reported that there is a school feeding program through HATUTAN, compared to 6% that said the same in the baseline. Most schools (75% baseline, 88% midline) indicated that the school director or coordinator leads the implementation of the feeding program; there was no significant change in the person reported to have responsibility for the SFP among intervention schools at midline (see Table 70). Fifty-five percent of the treatment schools and 25% of the comparison schools have also affirmed that the PTA is responsible for the oversight of the program. Among those who responded other, at midline, 44% said that a service provider (*fornesedor*) is responsible for the feeding program.

¹⁴⁴ As an interim measure, the program imported USDA-provided food commodities of fortified rice, pinto beans and fortified vegetable oil to the 90,000 preschool and primary-aged children in 440 schools for three months in the first trimester of school years 2020-2022.

¹⁴⁵ In 2020 a total of 435 received the USDA commodities for their school meals. The number of schools increased to 445 in 2021.

Table 70: School feeding program responsibility, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	90	88		99	98			
Director or coordinator	81.1%	87.5%	6.4	68.7%	88.8%	20.1	13.7	0.08
Deputy director	6.7%	10.2%	3.6	8.1%	7.1%	-0.9	-4.5	0.43
PTA	22.2%	25.0%	2.8	38.4%	55.1%	16.7	13.9	0.14
Teachers	13.3%	22.7%	9.4	31.3%	41.8%	10.5	1.1	0.90
Other	28.9%	20.5%	-8.4	24.2%	16.3%	-7.9	0.5	0.95

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Schools reported on whether there was an ongoing feeding program in school on the day of the survey. Most comparison schools said that there was no feeding program. On the other hand, in treatment schools, while most said that they did not have school feeding for the day during baseline (99%), most were observed having meals prepared for the day during midline (88%) (Table 71). These findings are consistent with the findings from the household survey where more respondents in the treatment group reported that there was an ongoing feeding program in schools. There was an extremely large and significant difference-in-differences, which could be attributed to the HATUTAN program filling gaps in the government's implementation of the SFP, as per design. This is also in line with the finding that there has been increased familiarity about the feeding program. The difference-in-differences results are significant for both data provided in the school survey and for data reported by households; notably, households in intervention areas reported provision of school meals nearly 90 percentage points more at midline than at baseline, but households in comparison areas reported provision of school meals 30 percentage points less at midline.

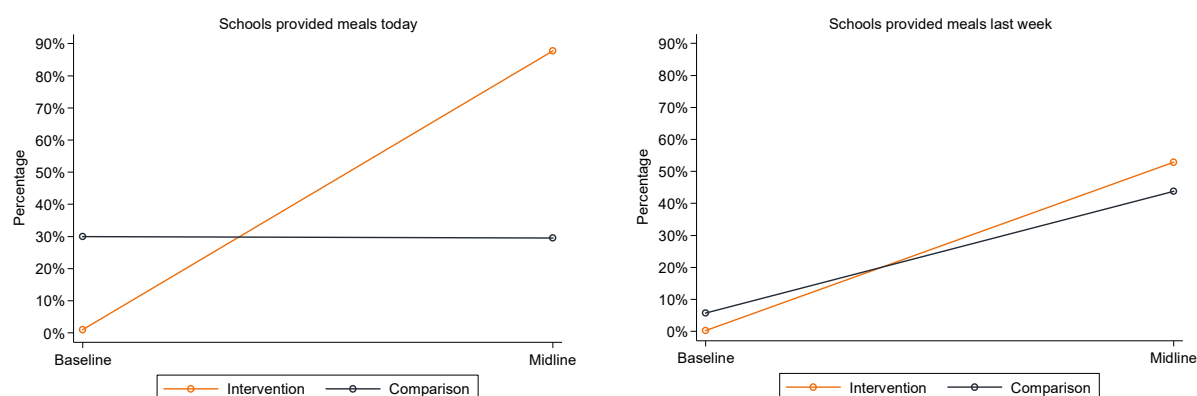
Table 71: Meals provided to students, by survey type and study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
School: Meals provided by school today								
n	90	88		99	98			
Yes	30.0%	29.6%	-0.4	1.0%	87.8%	86.8	87.2	<0.001***
Household: Meals provided by school last week								
n	344	599		461	701			
Yes	57.3%	26.9%	-30.4	2.0%	88.2%	86.2	116.6	<0.001***

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The responses of the treatment schools and treatment households are consistent in relation to the provision of meals to children in school. Moreover, the comparison group tends to report not having a feeding program in school on the day the school survey was conducted and the week prior to the household survey, which indicates poor implementation of the government SFP in comparison schools or poor knowledge of the program (see Figure 19).

Figure 19: Implementation of school feeding program



Within the school survey, most schools reported that they have a menu, indicating a level of preparedness and organization. Less schools reported following a menu during the midline (74%) compared to baseline (94%), especially among treatment schools (Table 72). This could be due to HATUTAN's programming efforts and training on flexible menus.

Table 72: School feeding program menu available, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	89	87		99	98			
Yes	94.4%	81.6%	-12.8	92.9%	67.4%	-25.6	-12.8	0.08

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Unfortunately, data collection issues at baseline resulted in school menu food items only being recorded in one intervention school; as a result of this extremely small sample size, we do not include a difference-in-differences analysis below, but report results for all baseline schools and midline intervention and comparison schools. Comparing portion sizes of each food item across schools would also be helpful to understand if children are provided with a healthy and balanced diet. However, data on food portions is not available.

In the midline, most schools consistently served carbohydrates (91%), legumes and nuts (64%), and dark green vegetables such as spinach, lettuce, and mustard greens (36%). This is consistent with the findings in the baseline. Overall, meals served to children in schools are lacking in fruit and meat. Midline schools in intervention areas were substantially more likely to report serving dark green vegetables and legumes, beans, and nuts than comparison schools. Comparison schools were much more likely to serve sweetened condensed milk (Table 73). Most schools reported serving one to six items from the list of items included in the menu, with an average of two to three items. In the comparison group, most schools mentioned two (41%) or three (37%) food items on the menu as part of children's meal for the day in school in the baseline. During midline, most comparison schools only identified two (81%) food items from the list. Among treatment schools, most mentioned two (48%) or three (41%) food items in the midline.

Table 73: Food items on the school feeding menu, by study group

	Baseline	Midline	
		Comparison	Intervention
n	28	26	86
Rice, maize, bread, and foods prepared with rice, maize, and wheat	89.3%	80.8%	94.2%
Pumpkin, carrot, purple sweet potato	25.0%	11.5%	16.3%
Potato, taro, yellow sweet potato, cassava, sago	21.4%	11.5%	7.0%
Dark green vegetables (e.g., spinach, lettuce, pumpkin leaves, cassava leaves)	35.7%	15.4%	41.9%
Other vegetables (e.g., cucumber, tomato, cabbage, eggplant)	0.0%	0.0%	0.0%
Yellow fruits (e.g., mango, papaya, honeydew melon, passionfruit)	0.0%	0.0%	2.3%
Other fruits (e.g., watermelon, tamarind, jackfruit)	0.0%	0.0%	0.0%
Meat from domesticated animals (beef, pork, sheep/goat meat, chicken, duck)	10.7%	11.5%	1.2%
Seafood (e.g., fresh or dry fish, shrimp)	7.1%	0.0%	0.0%
Legumes, beans, and nuts (e.g., beans, peas, soybeans or peanuts)	39.3%	15.4%	79.1%
Fresh milk	0.0%	7.7%	0.0%
Sweetened Condensed milk	3.6%	65.4%	0.0%
Other	35.7%	3.9%	8.1%
Eggs	10.7%	0.0%	0.0%

The food items mentioned are typically carbohydrates (e.g., rice, maize, bread), beans, legumes and nuts (e.g., beans, peas, peanuts), and vitamin A-rich vegetables (e.g., dark green vegetables including spinach and mustard greens). At midline, substantially more intervention schools reported serving beans, legumes, and nuts and vitamin A-rich fruits and vegetables than comparison schools. Very few schools at baseline or midline served dairy¹⁴⁶, eggs, and flesh foods in school lunches (Table 74).

Table 74: Food groups served to children in school, by study group

	Baseline	Midline		
	All	Comparison	Intervention	Difference
n	28	26	86	
Grain, roots, and tubers	92.9%	88.5%	94.2%	5.7
Beans, legumes and nuts	39.3%	15.4%	79.1%	63.7
Dairy	0.0%	7.7%	0.0%	-7.7
Eggs	10.7%	0.0%	0.0%	0
Vitamin A-rich fruits and vegetables	50.0%	19.2%	54.7%	35.5
Other fruits and vegetables	0.0%	0.0%	0.0%	0.0
Flesh foods	17.9%	11.5%	1.2%	-10.3

Schools reported serving food items that belonged to up to four food groups out of seven. At midline, meals in intervention schools were more likely to score a 2 or 3 on the dietary diversity scale compared to schools in comparison areas. Intervention schools were also far less likely to report serving meals with a dietary diversity score of 1 compared to comparison schools (Table 75).

¹⁴⁶ Milk is not usually consumed as part of the Timorese diet. Sweetened condensed milk is occasionally consumed as part of school meals, but its consumption is discouraged in intervention schools given the high levels of sugar and poor nutritional value.

Table 75: School menu dietary diversity score, by study group

	Baseline	Midline		
	All	Comparison	Intervention	Difference
n	28	26	86	
0	3.6%	0.0%	1.2%	1.2
1	28.6%	69.2%	7.0%	-62.2
2	28.6%	19.2%	53.5%	34.3
3	32.1%	11.5%	38.4%	26.9
4	7.1%	0.0%	0.0%	0.0

Information on the availability, quantity, preparation, and taste of meals served to children in school was also collected through the household survey. At both baseline and midline, most parents agreed that the food is available every day (92% baseline, 97% midline), quantity is sufficient (90% baseline, 91% midline), prepared in a hygienic manner (92% baseline, 97% midline), and tasty (90% in baseline and midline). The difference-in-differences analysis suggests that at midline, households in intervention areas were generally more likely to “agree completely” and less likely to “agree partially” with statements regarding school meals than would be expected given results in comparison areas (Table 76).

Table 76: Caregivers' reports of meal quality and availability, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
The food is available every day.								
n	322	153		438	604			
Agree completely	75.5%	73.2%	-2.3	68.5%	90.4%	21.9	24.2	<0.001***
Agree partially	17.4%	19.0%	1.6	22.4%	7.6%	-14.8	-16.3	<0.001***
Disagree partially	4.0%	2.0%	-2.1	5.0%	1.7%	-3.4	-1.3	0.51
Disagree completely	3.1%	5.9%	2.8	4.1%	0.3%	-3.8	-6.6	0.005**
The quantity of the food is sufficient.								
n	323	145		432	582			
Agree completely	72.1%	65.5%	-6.6	63.4%	73.7%	10.3	16.9	0.002**
Agree partially	17.7%	24.8%	7.2	26.4%	17.9%	-8.5	-15.7	0.002**
Disagree partially	7.4%	6.2%	-1.2	9.0%	4.5%	-4.6	-3.3	0.26
Disagree completely	2.8%	3.5%	0.7	1.2%	4.0%	2.8	2.1	0.29
The food is prepared in a hygienic manner.								
n	299	144		411	556			
Agree completely	76.9%	78.5%	1.6	69.8%	85.3%	15.4	13.9	0.006**
Agree partially	15.4%	18.1%	2.7	22.6%	11.9%	-10.8	-13.4	0.003**
Disagree partially	4.4%	2.8%	-1.6	5.6%	2.2%	-3.4	-1.9	0.40
Disagree completely	3.3%	0.7%	-2.7	2.0%	0.7%	-1.2	1.4	0.33

The food is tasty.								
n	284	137		385	542			
Agree completely	69.4%	68.6%	-0.8	62.6%	72.1%	9.5	10.3	0.07
Agree partially	18.3%	21.9%	3.6	29.1%	18.3%	-10.8	-14.4	0.005**
Disagree partially	6.7%	4.4%	-2.3	7.0%	6.6%	-0.4	1.9	0.50
Disagree completely	5.6%	5.1%	-0.5	1.3%	3.0%	1.7	2.2	0.39

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Based on qualitative data collected, school coordinators/directors are primarily responsible for the oversight of the school feeding program with support from teachers and the PTA. A school coordinator explained his role in the school feeding program:

In managing the school feeding program, I have to meet and discuss with [teachers and the PTA] clearly about transportation, food items to be delivered, where to store the food items, where to concentrate... so that when the food items are delivered to [administrative post name], we can organize the school feeding properly.¹⁴⁷

The FGDs also indicated that PTAs monitor the SFP to confirm that meals are served as planned, ensure that the kitchen and utensils are clean and appropriate for use, identify missing items or any food shortage, and determine budget needs. One of the school coordinators noted that to encourage parents and the PTAs' participation in the school feeding program they organized a meeting in December to discuss it and informed them about the meals that will be provided to students.¹⁴⁸

The PTA or one or two of its deputies go to the school to find out whether the school feeding items are utilized according to the calendar [and] if the school needs more money. Currently the Ministry of Education instructed that they should also [be involved] when the school needs to buy food items.¹⁴⁹

However, some respondents also mentioned that PTA involvement in school feeding was not very strong:

The PTA is established for less than a year. Talking about the lack of quality of the school meals I can say that it was caused by the lack of cooperation. For example, the newly established PTA must have a good coordination and cooperation with all parties, in terms of supervising the food items like rice, cooking oil and red beans. We all should work [together] to make sure student[s] consume the meals, and the food is sufficient for the whole month, but in reality, the PTA is not involved.¹⁵⁰

Children reported that they had a meal and parents expressed appreciation for the school meals provided through the government. Parents talked about the quality of food served to children in school based on their observations or what their children tell them when they return from school. Some of the recurring concerns that parents raised is the lack of diversity in meals, noting that children are often fed beans, and low food quality. Work is currently being done with schools to increase dietary diversity,

¹⁴⁷ FGD with school coordinators/directors, male, Manatuto municipality, Int. 126

¹⁴⁸ FGD with school coordinators/directors, male, Ermera municipality, Int. 136

¹⁴⁹ FGD with school coordinators/directors, male, Manatuto municipality, Int. 118

¹⁵⁰ FGD with fathers, male, Ermera municipality, Int. 132

and mobilization of the government budget may be able to further support schools to purchase the diverse and nutritious foods available locally.

*Sometimes the red beans were not cooked properly, with this the children did not want to finish their meal. The children wash their own dishes after meal. Older kids could wash theirs, but little kids could not. Washing dishes should be the responsibility of those who prepare the food. The beans were not cooked properly, and it has to be thrown away in the end.*¹⁵¹

*CARE only provides rice, red beans and cooking oil. The food is high quality, but in our observation all [that] the children eat every day is red beans and they may get [tired] of it. Some of the students did not want to eat the food, but the food is good and we are grateful for the contribution from CARE.*¹⁵²

*I think the program is good because the children would be diligent to come to school. But if we are talking about the quality, I do not think it has good quality as children would have to bring vegetables, firewood, and everyday they only eat rice. The vegetables only consist of beans, which we can say is healthy, but it is exactly what they already consume at home.*¹⁵³

A mother in Ainaro municipality additionally mentioned that the quantity of food was not sufficient, and that teachers sometimes ate food apportioned for students:

*Our children eat with a large spoon of rice, but the teachers eat good vegetables. When the children arrived at home, we asked about [their] portion of the meal, whether a lot or a little. They answered [that they] ate a large spoonful of rice and a piece of chicken thigh meat divided [among] four people, and a little vegetable. The teachers have a large portion. This is food aid for children, but the teachers also eat it.*¹⁵⁴

SCHOOL FEEDING PROGRAM HYGIENE AND RESOURCES

Most parents felt that food for their children in school is prepared in a hygienic manner, which is an important consideration to ensure that children do not get sick and miss classes. Among schools that reported having their own kitchen (86% baseline, 89% midline), more schools said that they have access to clean water to prepare meals in the midline (86%) compared to the baseline (64%). Most schools indicated keeping the kitchen clean by using detergent (79% baseline, 98% midline). Most schools also reported having storage spaces that are at least somewhat clean (78% baseline, 93% midline). Schools that have a kitchen all had a stove that used wood; none had electric or gas stoves. Less than half had a scale in the kitchen (35% treatment, 45% comparison), which represents a lower number of schools with a scale compared to the baseline (Table 77).

These results generally indicate that more schools are practicing hygienic food preparation over time. The difference-in-differences analysis suggests that some changes in hygienic food preparation have

¹⁵¹ FGD with mothers, female, Ermera municipality, Int. 101

¹⁵² FGD with fathers, male, Manatuto municipality, Int. 135

¹⁵³ FGD with fathers, male, Ermera municipality, Int. 105

¹⁵⁴ FGD with mothers, female, Ainaro municipality, Int. 108

occurred at similar rates in both intervention and comparison schools: there were no significant differences in whether the school had a kitchen or in improved usage of clean water or detergent at midline in intervention schools compared to comparison schools (Table 77). Furthermore, for the indicator measuring whether food storage is located at the school, comparison schools “caught up” at midline to intervention schools, improving by almost 45 percentage points from a relatively low level while intervention schools reported slightly lower rates of having an in-school storage space. In contrast, food storage areas in intervention schools were reported to be clean at higher rates than expected given the results of comparison schools at midline. Overall, these results suggest that the program may not have had a large impact on the use of hygienic food preparation practices in schools, but that changes over time may have occurred due to other factors that impact both intervention and comparison areas, including the increased access to water and detergent/soap due to the push for improved hygiene in schools as a result of the COVID-19 crisis.

Table 77: Hygienic preparation of food, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Kitchen at the school								
n	32	88		99	98			
Yes	71.9%	80.7%	8.8	93.9%	96.9%	3.0	-5.8	0.54
Clean water to prepare meals								
n	22	71		93	95			
Yes	45.5%	83.1%	37.7	68.8%	87.4%	18.6	-19.1	0.14
Use detergent to clean kitchen								
n	22	71		93	95			
No	22.7%	4.2%	-18.5	20.4%	1.1%	-19.4	-0.9	0.93
Sometimes	40.9%	29.6%	-11.3	34.4%	24.2%	-10.2	1.1	0.93
Often	0.0%	11.3%	11.3	6.5%	9.5%	3.0	-8.2	0.13
Every day	36.4%	54.9%	18.6	38.7%	65.3%	26.6	8.0	0.57
Storage located at the school								
n	14	40		69	74			
Yes	42.9%	87.5%	44.6	89.9%	85.1%	-4.7	-49.4	0.002**
Clean storage space								
n	30	40		99	74			
No	6.7%	7.5%	0.8	27.3%	6.8%	-20.5	-21.3	0.01**
Somewhat	40.0%	25.0%	-15.0	38.4%	24.3%	-14.1	0.9	0.94
Mainly	43.3%	20.0%	-23.3	13.1%	25.7%	12.6	35.9	0.005**
Yes	10.0%	47.5%	37.5	21.2%	43.2%	22.0	-15.5	0.20
Type of stove								
n	32	88		99	98			
Electricity	0.0%	0.0%	0.0	0.0%	0.0%	0.0	-	-
Gas	0.0%	0.0%	0.0	0.0%	0.0%	0.0	-	-
Wood	100.0%	100.0%	0.0	100.0%	100.0%	0.0	-	-
Scale in the kitchen								
n	32	88		99	98			
Yes	50.0%	45.1%	-4.9	57.0%	34.7%	-22.3	-17.3	0.22

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

For the indicators of school feeding facilities, the sample size at baseline was too small for a reliable panel analysis. We therefore narrow the sample to midline and compare the treatment and comparison groups. This analysis does not take into account the possibility that treatment and comparison schools differed before the program; we must therefore be more cautious in attributing differences to program impact.

Table 78: School facilities at midline

	Comparison	Treatment	Difference	p-value
School canteen	3.4%	5.1%	1.7	0.57
School kitchen	80.7%	96.9%	16.2	<0.001***
School kitchen water	83.1%	87.4%	4.3	0.44
School kitchen plates	81.8%	84.7%	2.9	0.60
School kitchen handwashing	32.4%	35.8%	3.4	0.65
School food storage	56.3%	77.9%	21.6	0.003**

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

For all facilities observed at midline, the treatment group had a higher percentage of schools with those facilities than the comparison group. Two of these were statistically significant – those replying they had a kitchen (16 percentage points higher in treatment schools) and those replying they had enough or some food storage (22 percentage points higher in treatment schools). Given the caveat above, we can cautiously interpret this as positive program impact in improving the effectiveness of school feeding programs through a better standard of facilities.

Data on storage spaces at or near each school were collected through the school survey and results are summarized in Table 79. Most of the schools (64% baseline, 69% midline) said they had some or enough storage space available. Treatment schools are more likely to have a storage space compared to comparison schools.

Table 79: Storage spaces, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Storage space								
n	30	71		99	95			
No	53.3%	43.7%	-9.7	30.3%	22.1%	-8.2	1.5	0.91
Yes, some	20.0%	36.6%	16.6	41.4%	45.3%	3.9	-12.8	0.28
Yes, enough	26.7%	19.7%	-7.0	28.3%	32.6%	4.4	11.3	0.33

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Most of the storage spaces had cement floors (74% treatment, 68% comparison) and brick walls (65% treatment, 60% comparison). It appears that the number of comparison schools with storage spaces with cement floors has declined between baseline and midline. This should be interpreted with caution due to the small sample of respondents. Nearly all of the treatment schools (96%) had an aluminum roof at the midline. About two in ten (22%) had storage spaces with a leaking roof; results show that the percentage of treatment schools with a leaking roof increased significantly at midline compared to comparison schools, where this percentage decreased; this may be an indication that resources should be allocated to maintaining the food storage spaces. A summary of these findings by study group and round can be found in Table 80.

Table 80: Storage space materials, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Storage space floor								
n	30	40		99	74			
Cement	86.7%	67.5%	-19.2	76.8%	74.3%	-2.5	16.7	0.16
Gravel	0.0%	0.0%	0.0	1.0%	1.4%	0.3	0.3	0.84
Mud	3.3%	2.5%	-0.8	7.1%	1.4%	-5.7	-4.9	0.34
Sand	3.3%	5.0%	1.7	3.0%	5.4%	2.4	0.7	0.90
Tile	3.3%	2.5%	-0.8	5.1%	4.1%	-1.0	-0.2	0.98
Wood or bamboo	3.3%	22.5%	19.2	7.1%	13.5%	6.4	-12.7	0.15
Storage space walls								
n	29	40		99	74			
Bricks	65.5%	60.0%	-5.5	55.6%	64.9%	9.3	14.8	0.29
Wood or bamboo	6.9%	25.0%	18.1	29.3%	14.9%	-14.4	-32.5	0.002**
Sand	0.0%	10.0%	10.0	5.1%	9.5%	4.4	-5.6	0.38
Tile	0.0%	2.5%	2.5	2.0%	0.0%	-2.0	-4.5	0.12
Mix	34.5%	22.5%	-12.0	20.2%	24.3%	4.1	16.1	0.21
Storage space roof								
n	0	40		3	74			
Aluminum sheet	-	87.5%	87.5	0.0%	96.0%	96.0	-	-
Bamboo	-	5.0%	5.0	66.7%	1.4%	-65.3	-	-
Mix	-	7.5%	7.5	33.3%	0.0%	-33.3	-	-
Branches and leaves	-	0.0%	0.0	0.0%	1.4%	1.4	-	-
Tile	-	0.0%	0.0	0.0%	1.4%	1.4	-	-
Leaking roof								
n	30	40		99	74			
No	93.3%	80.0%	-13.3	77.8%	86.5%	8.7	22.0	0.03*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The same number of schools (101) in the baseline and midline reported that their storage spaces have adequate ventilation. A quarter of the storage spaces (25%) did not have any method for raising food off the ground during baseline, and the number decreased during midline to only 3%. Most of the schools used pallets during baseline (53%). However, at midline, more schools reported raising food off the floor with methods other than using pallets or shelves (87%) (Table 81). The survey did not probe further on what these other methods or materials were. These findings are relevant to ensuring that food served to children are kept in a sanitary space to prevent potential sickness and therefore absence from school.

Table 81: Storage sanitation, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Ventilation in the storage space								
n	29	40		99	74			
Yes	79.3	95.0	15.7	78.8	85.1	6.3	-9.3	0.36
Method for raising food off the ground								
n	22	40		91	74			
Shelves	4.6	5.0	0.5	4.4	2.7	-1.7	-2.1	0.74
Pallets	59.1	15.0	-44.1	51.7	2.7	-49.0	-4.9	0.72
Other	22.7	75.0	52.3	16.5	93.2	76.8	24.5	0.05*
None	13.6	5.0	-8.6	27.5	1.4	-26.1	-17.5	0.07

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

One of the school directors explained during an FGD that the Technical Support Office (GAT, or *Gabinete de Apoio Técnico*) is in charge of the school feeding program.

*Whether it is about budget preparation or anything else, it is the GAT and the suppliers who handle the school feeding. They are the ones who know about it, what to eat or to drink it is the responsibility of the GAT and suppliers. I just control and oversee. If I find anything wrong, then I will complain and say this is not correct.*¹⁵⁵

Teachers take the lead in identifying the missing items for the school feeding program and share their findings with parents. However, one of the coordinators said that this is not mandatory.¹⁵⁶ School coordinators noted that parents contribute to the school feeding program by sending salt, onions, some vegetables, oil, and firewood when their children go to school when supplies are available.^{157, 158, 159} Some parents, however, refuse to get involved in the school feeding program due to lack of incentives,¹⁶⁰ and some of the school coordinators noted that making contributions requires financial investment from the side of the parents which make some reluctant to contribute. However, based on the FGDs, most of the parents are willing to do so to ensure that their children eat well-balanced meals in school.

Their contributions are the vegetables. Some bring vegetables one day, others don't. The little children each bring a vegetable... when they have them. Others don't bring vegetables but they bring pumpkins. Here it is like this. In Gleno, with 50 cents you can buy a good amount of vegetables, but here market is also far away. The parents want to provide some food [but] they can only give what they have. If they don't have it, everyone eats plain beans. How [can this] give health to the children? In the old days we used to buy [when] the government gave money, and we bought following a menu... to define what you can eat every day. For example, in a week how many times [will] you eat meat, eggs. That's because the government gave us money and we

¹⁵⁵ FGD with school coordinators/directors, female, Ainaro municipality, Int. 109

¹⁵⁶ FGD with school coordinators/directors, male, Ainaro municipality, Int. 122

¹⁵⁷ FGD with school coordinators/directors, female, Ermera municipality and Ermera municipality, Int. 103 & 129

¹⁵⁸ FGD with mothers, female, Manatuto municipality, Int. 117

¹⁵⁹ FGD with school coordinators/directors, male, Ainaro municipality, Int. 122

¹⁶⁰ FGD with school coordinators/directors, male, Ermera municipality, Int. 102

*bought according to a menu. But how are we, the teachers, supposed to buy food for the children if there is no money?*¹⁶¹

*I heard that the children did not want to eat papaya. They wanted to eat mustard greens, collard and cassava leaves. During the rainy season in the mountain, we had a lot of these vegetables. As parents, we could contribute, but we heard from them that they only needed cassava leaves, not papaya, not goat meat, not horse meat, but beef/cow meat. The children told us about this when they came home. So, they ate potatoes or eggs, milk, cake, green beans and red beans. So we told them that in the past they came and purchased some from us with the price around \$0.25 cents or \$0.50 cents or even \$1.00. They purchased from us. Sometimes if they needed salt and dish soap, we contributed. Sometimes we think of our children at this school [so] we also contribute.*¹⁶²

School coordinators/directors and parents also mentioned that access to adequate facilities, materials, and cooks was a challenge to school feeding. Some of the school coordinators/directors cited old kitchen facilities and limited storage space as challenges to implementing the school feeding program. They also highlighted the absence of some kitchen materials:

*We have no frying pans, dishes, and cutlery. The ones available here were supplied [in the year] 2000... they have turned black and people don't like to use them anymore. Other things are available, for example, water and pots but the latter is now being used for preparing vegetables because all the frying pans are broken.*¹⁶³

One of the mothers recalled raising this challenge during a PTA meeting, after which, immediate action was taken to address the problem.

*In the recent meeting, an issue was raised about the [poor] condition of the school kitchen. The [PTA] asked how the parents think about the kitchen's condition, [with] pillars about to fall apart. Most of parents supported to rebuild, but the school principal decided to give the responsibility to [the parents of] grade 9 students. They should be the ones responsible for the pillars, the parents will come and help with tearing down and reconstruct the kitchen. The [PTA] will contribute money for food.*¹⁶⁴

Some schools also experience a shortage of cooks, with one school coordinator citing that they have only one cook who has a high workload and does everything from preparing the vegetables, to cooking meals, to washing the dishes for hundreds of students per day.¹⁶⁵ Another school coordinator noted that in their school there are four cooks that share payment meant for two people.

The government now has a budget, they give two people a subsidy of \$100 per person, and since it is GAT who asks them to cook, they give \$25 each. And they say like this, if in the future the sisters change us we prefer not to continue cooking. It is very difficult

¹⁶¹ FGD with school coordinators/directors, female, Ermera municipality, Int. 103

¹⁶² FGD with mothers, female, Manatuto municipality, Int. 120

¹⁶³ FGD with school coordinators/directors, female, Ermera municipality, Int. 129

¹⁶⁴ FGD with mothers, female, Ermera municipality, Int. 101

¹⁶⁵ FGD with school coordinators/directors, male, Ainaro municipality, Int. 122; female, Ermera municipality, Int. 129

because they cook for many students and there are only four of them cooking. One of the difficulties is that the government pays \$100 for two people, they distribute for four.¹⁶⁶

Water supply for washing, preparing, and cooking food is another of the challenges that schools raised, which affects hygienic preparation of food and handwashing. Sometimes children bring their own water to school for drinking and handwashing. However, if they do not bring water, they eat without washing hands, which indicates poor adherence to healthy hygiene practices and may potentially lead to students missing school due to sickness.¹⁶⁷

We know very well that this water source is coming from Ainaro Town, but there is a schedule [for] water distribution. For example, we can get water today, tomorrow and the day after tomorrow, but [for] the next three days it won't be our schedule. [As a result,] the cook needs to get some water from the river to cook for the students.¹⁶⁸

This respondent added that there was an instance when meals were not prepared because there was no water.

Budget limitations were mentioned as a further challenge. Budget received from the government determines the schools' ability to purchase goods from local farmers. Some schools have arrangements with farmers, while others do not. Schools receive contributions from parents and the PTAs in terms of combustibles and food items. (This will further be discussed in the "Purchase of Local Food Items" sub-section below.)

Additionally, some parents reported that students are asked to bring vegetables or contributions to school, and that they are happy to contribute to ensure well-balanced meals for their children instead of consuming rice and beans only, especially when schools explain that there is no budget and that only rice and beans are available. Parents commented on the lack of diversity of meals served to students.

There has to be some sort of variation in their meals. They can't only eat meat. They also need eggs [and] vegetables. They need to vary each day. One day they can serve beans, the next day they should not eat beans again.¹⁶⁹

Everyday they (the students) would only eat (rice) porridge. If there are no vegetables, then [the school] would ask the parents to contribute cabbages and potatoes to put together to cook some vegetables to eat with the (rice) porridge.¹⁷⁰

What we want is that our children can have other good side dishes because so far in one year they only consumed red beans.¹⁷¹

Lastly, there appears to be an absence of a feedback mechanism on school feeding in many areas. Some of the parents said that they did not report concerns regarding the SFP to local or school

¹⁶⁶ FGD with school coordinators/directors, female, Ermera municipality, Int. 103

¹⁶⁷ FGD with mothers, female, Ermera municipality, Int. 138

¹⁶⁸ FGD with school coordinators/directors, male, Ainaro municipality, Int. 111

¹⁶⁹ FGD with mothers, female, Ainaro municipality, Int. 124

¹⁷⁰ FGD with fathers, male, Ainaro municipality, Int. 125

¹⁷¹ FGD with mothers, female, Manatuto municipality, Int. 120

authorities due to worries that it would create problems for their children in school, such as retaliation from teachers.

*We could share with other people but sometimes the teachers say that we are just being jealous of them. If we share it with others, the teachers could get angry with us. So, we just let them do things instead.*¹⁷²

*When our children inform us about it and we bring it up with the teachers, the teachers would get angry with the children instead, and said "if you just want a good food, it is probably better just stay home instead so that your parents prepare the good meals for you."*¹⁷³

ACCESS TO AND USE OF GOVERNMENT FUNDING FOR SFP

School feeding improves school enrollment and attendance, increases cognitive and academic performance, and contributes to gender equity in access to education.¹⁷⁴ The SFP is funded by the government of Timor-Leste but consistent implementation in all schools is hindered by budgetary and reporting delays from the government and school administrations, respectively, which reduces the program's reach and effectiveness.¹⁷⁵ The lack of resources (i.e., funding and rice) typically happens from January to March, which coincides with Timor-Leste's "hungry period."¹⁷⁶

As such, and to ensure sustainability of SFP-related activities upon the closure of the HATUTAN program, it is important that schools report consistent access to and use of government funding for school feeding. At midline, no quantitative data was collected on the source of funding for SFPs. However, in 2020, delays in approval of the national budget significantly delayed provision of government funding for school feeding, suggesting that challenges remain to access to and use of government funding for SFPs.

Most school coordinators/directors interviewed cited the lack of budget as one of the main challenges to the implementation of the school feeding program. Schools are able to purchase supplies from the community when they have access to funding but rely on contributions from parents in its absence.

*Before it was the government providing the school feeding. They would send us budget to buy the local products. We would order from the vendors and they bring it themselves. Parents buy local products like bananas, vegetables, cassava, chili, yams, potatoes. That they buy daily.*¹⁷⁷

We can't buy vegetables from the community. When the government gives us money then we'd buy, but now HATUTAN is the one who took responsibility, so the

¹⁷² FGD with mothers, female, Ermera municipality, Int. 138

¹⁷³ FGD with mothers, female, Ermera municipality, Int. 138

¹⁷⁴ Dongqing Wang and Wafaie W. Fawzi, "Impacts of school feeding on educational and health outcomes of school-age children and adolescents in low- and middle-income countries: Protocol for a systematic review and meta-analysis," *Systematic Reviews* 9, 55 (2020).

¹⁷⁵ A total of \$14.4M were allocated to the school feeding program under the 2018 National Budget (http://timor-leste.gov.tl/wpcontent/uploads/2018/08/2018_08_27_Discurso_orcamento_PT.pdf, accessed on May 24, 2019)

¹⁷⁶ The "hungry period" refers to the period when the previous year's subsistence staples have been consumed but the new season's produce is not yet ready. This usually occurs in November/December to January/February.

¹⁷⁷ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

*government stopped giving us school lunch money. How can we go to buy the community vegetables?*¹⁷⁸

*As for the vegetables, we can only buy them when the Ministry of Education provided some money, and we use the money to buy local produce such as pumpkin and cassava. But, with the HATUTAN program, we can't buy from the community because we don't have money. Instead, we ask the parents to contribute some vegetables so that we can cook with beans.*¹⁷⁹

*The school budget, or the school feeding budget, allocated \$0.25 per student per day. But what we purchase with this money are green mustards. Sometimes we buy what is recommended on the menu, meaning we have to cook/prepare vegetables with eggs or meat, following the menu that the school has established, and the cook always does this. On the rice, to my knowledge, in 2020 we didn't buy any rice because our rice was provided by CARE. There was always leftover. The same with oil, and so we continued using rice and oil from CARE. For example, last year we continued providing meals until the month of December because we had leftover rice and oil.*¹⁸⁰

PURCHASE OF LOCAL FOOD ITEMS

To boost local production and maintain a sustainable source of food for the SFP, schools are encouraged to purchase their produce locally. Most schools indicated during the baseline that they would purchase food locally sometimes (57%) or all the time (39%). Most school administrators stated they buy goods locally during midline. However, less than half of the schools (43%) at midline said that they do not buy produce from local farmers, most of which are intervention schools. Since only one intervention school provided meals at baseline, a difference-in-differences analysis cannot be conducted due to low sample size.

Among treatment schools, the primary reason for not purchasing produce from farmers at midline is not having the budget to buy produce (Table 82); during both baseline and midline, the government had not yet transferred money to schools for the SFP, the main reason for this finding. Purchasing patterns and dietary diversity continue to be constrained by the reality that the average cost of a nutritional diet using the most affordable and nutritious locally-available food items is 0.49 cents per student per day.¹⁸¹

¹⁷⁸ FGD with school coordinators/directors, female, Ermera municipality, Int. 103

¹⁷⁹ FGD with school coordinators/directors, male, Manatuto municipality, Int. 116

¹⁸⁰ FGD with school coordinators/directors, female, Manatuto municipality, Int. 133

¹⁸¹ "HATUTAN's Cost of the Diet Study: Understanding the availability and cost of nutritious food in four municipalities of Timor-Leste to inform school feeding," (March 2020).

Table 82: Schools buying local produce from farmers for the SFP, by study group

	Comparison		Intervention	
	BL	ML	BL	ML
The school buys local produce from farmers for the school feeding				
n	27	26	1	86
No	3.7%	3.9%	0.0%	54.7%
Yes, sometimes	59.3%	76.9%	0.0%	34.9%
Yes, all the time	37.0%	19.2%	100.0%	10.5%
Reason for not purchasing produce from farmers				
n	1	1	0	47
No budget to buy local produce	0.0%	0.0%	0.0%	74.5%
Farmers' produce is not sufficient	100.0%	100.0%	0.0%	6.4%
Farmers do not want to sell to the school	0.0%	0.0%	0.0%	0.0%
Poor quality of local produce	0.0%	0.0%	0.0%	0.0%
Inconsistent availability of produce	0.0%	100.0%	0.0%	2.1%
Local produce is not nutritious	0.0%	0.0%	0.0%	0.0%
Other	0.0%	0.0%	0.0%	21.3%

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Among schools that purchased local produce, the type of produce purchased appeared consistent. In the baseline, the main food items bought were dark green vegetables (89%), vitamin A-rich foods (e.g., pumpkin, carrot, and purple sweet potato) (70%), and starchy foods (e.g., potato, taro, yellow sweet potato, and cassava) (70%). In the midline, items typically bought were dark green vegetables (75%), starchy foods (56%), and vitamin A-rich foods (55%). This does not fully match what most schools reported to be part of the meals served to children in school, which are composed primarily of starchy food, beans, legumes and nuts, and dark green vegetables. This implies that food supplies purchased from farmers may not be a major component of school meals, which may be primarily composed of foods obtained from other sources. Low consumption of fruits may also be explained by the limited number of schools that purchased fruits locally.

At both comparison and intervention schools, many protein sources, including meat, eggs, and legumes, were purchased substantially less at midline. However, changes between baseline and midline for intervention schools should be interpreted with caution due to the large disparity in sample size, with data collected in only one intervention school at baseline (Table 83). In the case of rice, maize, and bread and legumes, beans, and nuts, it may be the case that the need to purchase these goods from local farmers was supplanted due to provision of goods by the HATUTAN program. In contrast, the relatively large percent of schools reporting purchases of pumpkin, carrot, and purple sweet potato; potato, taro, yellow sweet potato, and cassava; and dark green vegetables suggests that these types of produce are both available and desirable for school cooks.

Table 83: Local produce schools bought from farmers, by study group

	Comparison			Intervention		
	BL	ML	Difference	BL	ML	Difference
n	26	25		1	39	
Rice, maize, and bread	61.5	52.0	-9.5	0.0	23.1	23.1
Pumpkin, carrot, purple sweet potato	69.2	52.0	-17.2	100.0	56.4	-43.6
Potato, taro, yellow sweet potato, cassava	69.2	52.0	-17.2	100.0	59.0	-41.0

Dark green vegetables (e.g., spinach, mustard)	88.5	60.0	-28.5	100.0	84.6	-15.4
Other vegetables (e.g., cucumber, tomato, cabbage, eggplant)	38.5	4.0	-34.5	100.0	18.0	-82.1
Yellow fruits (e.g., mango, papaya, honeydew melon, passionfruit)	0.0	0.0	0.0	0.0	7.7	7.7
Other fruits (e.g., watermelon, tamarind, jackfruit)	3.9	0.0	-3.9	0.0	0.0	0.0
Meat from domesticated animals (e.g., beef, pork, sheep/goat meat, chicken, duck)	57.7	12.0	-45.7	100.0	12.8	-87.2
Seafood (e.g., fresh or dry fish, shrimp)	26.9	0.0	-26.9	100.0	5.1	-94.9
Legumes, beans, and nuts (e.g., beans, peas, soybeans or peanuts)	50.0	28.0	-22.0	0.0	15.4	15.4
Fresh milk	15.4	8.0	-7.4	0.0	0.0	0.0
Coconut oil	7.7	0.0	-7.7	0.0	2.6	2.6
Condiments	50.0	4.0	-46.0	0.0	2.6	2.6
Tofu or tempe	15.4	4.0	-11.4	0.0	0.0	0.0
Eggs	38.5	4.0	-34.5	0.0	2.6	2.6
Don't know	0.0	0.0	0.0	0.0	0.0	0.0

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Among comparison schools, more schools in Aileu and Covalima purchased goods from local farmers at midline than at baseline. However, fewer schools in all intervention municipalities purchased goods from local farmers at midline than at baseline; the decline in Ermera was particularly notable (Table 84). A possible contributing factor to this is the current food security context in Ermera. The Integrated Food Security Phase Classification (IPC) classified Ermera and Manufahi as under severe chronic food insecurity (IPC Phase 4) for 2018-2023. Low agricultural productivity, poor quality and quantity of food consumption are some of the contributing factors to chronic food insecurity in Timor-Leste.¹⁸²

Table 84: Schools buying local produce from farmers, by municipality and study group

	n	Group	Baseline	Midline	Difference	p
Aileu	23	Comparison	78.3%	95.7%	17.4	0
Ainaro	28	Intervention	100.0%	74.1%	-25.9	0.003*
Baucau	3	Comparison	100.0%	0.0%	-100.0	<0.001***
Bobonaro	32	Comparison	100.0%	93.6%	-6.4	0.15
Covalima	10	Comparison	70.0%	80.0%	10.0	0.63

¹⁸² IPC, *The First IPC Analysis Report on The Chronic Food Insecurity Situation in Timor-Leste: Evidence and Standards for Better Food Security and Nutrition Decisions*, National Directorate of Food Security and Cooperation, the Ministry of Agriculture and Fisheries (January 17, 2019). (http://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/3_IPC_Timor%20Leste_CFI_20182023_English.pdf, accessed on April 18, 2021)

Ermera	41	Intervention	95.1%	14.6%	-80.5	<0.001***
Liquica	11	Intervention	81.8%	72.7%	-9.1	1
Manatuto	19	Intervention	100.0%	57.9%	-42.1	0.001**
Manufahi	22	Comparison	90.9%	90.5%	-0.4	0.96

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

We note that this data cannot be further analyzed for only schools reporting that they have a feeding program, as sample size per municipality and baseline/midline round becomes too small for analysis.

When the SFP is operational, budget is available, and schools are able to purchase local goods, schools seek to purchase a variety of vegetables. Most of the school directors/coordinators interviewed noted that they did not have any formal agreement with farmers, although some do. One school coordinator said that they have an agreement with a farmer to supply local produce for the school feeding program, but there is no money/budget to purchase those supplies.¹⁸³

The school hasn't purchased any produce because the school doesn't have money to purchase them. The school only purchases them in April and May when it has budget [school concession budget]. But for the months of January, February and March we only use commodity from CARE.¹⁸⁴

So far, we don't have an agreement because all parents here are farmers. We can only make agreements with people in places where there are no farms, so we make a partnership and ask them to prepare vegetables only. But we have never done that as all community members here are creative and grow vegetables to supply what we need. The cooks know these people well. Today food items are supplied by several farmers, tomorrow will be the turn of other farmers to supply.¹⁸⁵

Parents either contribute goods/supplies or sell items at school.

In the first place, the school feeding has two supply sources: one is CARE-HATUTAN, and the other is the education department program that supplies rice. When money is available, we prioritize to buy our local products sold by parents. If they don't have the items, we buy from the shops, but parents always bring their products to sell at the school and we buy them.¹⁸⁶

Since 2018 to 2019, it was clearly instructed to buy all vegetables from parents, including buying the green couve/collard, shoot of pumpkin leaves, beans, cassava, banana, and papaya, as at that time government supported \$0.25 cents per student. But since 2020 up until now, we have not bought anything yet from the parents because the support we get from partners are only rice and cooking oil. We cannot take any decision to take that rice and exchange with vegetables or bring that cooking oil and exchange with that particular thing.¹⁸⁷

¹⁸³ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

¹⁸⁴ FGD with school coordinators/directors, female, Manatuto municipality, Int. 133

¹⁸⁵ FGD with school coordinators/directors, male, Manatuto municipality, Int. 118

¹⁸⁶ FGD with school coordinators/directors, male, Manatuto municipality, Int. 118

¹⁸⁷ FGD with school coordinators/directors, male, Ermera municipality, Int. 102

The typical items that the schools purchase for the feeding program include vegetables (e.g., cassava leaves, mustard green), eggs, and meat (e.g., chicken, beef).

Yes, what we normally buy from them such as papaya flowers, cassava leaves, kankung (water spinach), bitter mustard green and pumpkins. Recently because we have one vision [plan], that at enrollment, each student has \$2; but not all of them. So, with this cash that they have paid, then we can spend on buying the vegetables besides some support from parents. So, with this money, we the buy the vegetables that I described already.¹⁸⁸

In terms of rice, beans, we did not purchase because we have received them as assistance. This year and last year we received such assistance therefore we purchase the beans, potatoes other vegetables so we can cook. If we receive rice from the government, we can cook it along with chicken that we purchase from the market. We purchased eggs and chicken meats from local kiosk while the vegetables we bought from the community market.¹⁸⁹

From the farmers, we buy them every day, the vegetables, always every day. Sometimes meat. We purchase the beef from the market, but for the chicken, the local chicken, we buy them from the community.¹⁹⁰

School coordinators/directors also noted instances of inconsistent availability of local stock and prices of goods.

The problem we faced is the availability of local stock. For example, today we want the students to have the shoot of pumpkin leaves, and we buy them from a small place in the community. It of course will not be enough to feed these 700 students. So, we have to buy those vegetables from different places in the community. The availability of vegetables are not always enough for us to buy, in one place. There are many but in other places they are only few. That is the challenge.¹⁹¹

As we buy vegetables in large quantities, we don't buy them in this aldeia [hamlet] because farmers produce vegetables in small amounts sufficient only for their own consumption. We buy vegetables in large quantities because there are more than 200 students here, so [we] cannot buy in small quantities. Instead, we buy them in the market, not in the aldeia. If farmers from this aldeia sell their produce in the market, we could also buy from them and then cook them at the school. Here we usually buy vegetables, sometimes pumpkins, and pawpaw... We buy the food items occasionally. Sometimes it is supplied by the students because we asked parents to bring this or

¹⁸⁸ FGD with school coordinators/directors, male, Ainaro municipality, Int. 111

¹⁸⁹ FGD with school coordinators/directors, male, Ainaro municipality, Int. 122

¹⁹⁰ FGD with school coordinators/directors, female, Manatuto municipality, Int. 133

¹⁹¹ FGD with school coordinators/directors, male, Ermera municipality, Int. 102

*that. Only the school feeding provided by the Ministry of Education demands that vegetables must be bought on a daily basis.*¹⁹²

*The problems we faced include some farmers [providing] more items, others [offering] less for the same price, while some farmers charge a price equivalent to that of the capital city market. Some farmers do not do that because they are aware that the money they receive is big and it is quite difficult for them to earn. Farmers who have money provide less items for the same price and this is a problem.*¹⁹³

*One of the problems is this, it's [a] typical Timorese problem. When we purchase from the farm, the price and weight are different. When they sell these items in the market, sometimes they increase the amount to sell them faster. But if we purchase directly from them, things are never what we want. Sometimes we say the weight is 1 kg, but it [is always less than] 1 kg. More so with the produce they bundle themselves. But it's different when buy from the market. That's the issue.*¹⁹⁴

There is also a perception that students no longer want local goods which they typically eat at home. Thus, some schools wanted to serve goods such as bread and cookies but there is insufficient supply. Ultimately, they decided to serve rice and vegetables.

*In regard to local food, I just mentioned that the children are used to eating potatoes, cassava and bananas at home every day. So, if we buy local food items, they will refuse it because they say, "we eat these things at home every day, we don't want this anymore." We must buy bread and cookies so that they have the appetite to eat. But we thought although this school is able to buy bread or cookies, who is capable [of baking] bread or cookies [for] more than 200 pupils? So, we had to find other means and decided to buy rice and vegetables.*¹⁹⁵

CONTRIBUTIONS TO SFP

All schools providing basic education are supposed to be allocated 25 cents per day and 75-100g of unfortified rice per day for each child under the government SFP; however, schools are rarely provided with this funding in practice. Parents are to be involved in the selection and monitoring of cooks who prepare the rice and purchase other local produce required for meal preparation. This section discusses the contribution of schools and households to the school feeding program, including PTA involvement in the implementation. The school and household surveys provided information on the PTA's efforts to improve school feeding and participation of households in the PTA.

At both baseline and midline, most of the schools reported having a PTA that provides oversight of school feeding (79% baseline, 75% midline). At midline, PTA involvement in school feeding had improved significantly relative to comparison schools, which actually reported lower rates of PTA oversight of school feeding. There was also a substantial increase in the percent of respondents who felt that the PTA in their children's school was doing activities to improve school feeding (30% baseline, 72% midline). At midline, there was a significant increase in reported PTA improvement of school

¹⁹² FGD with school coordinators/directors, female, Ainaro municipality, Int. 109

¹⁹³ FGD with school coordinators/directors, male, Manatuto municipality, Int. 118

¹⁹⁴ FGD with school coordinators/directors, female, Manatuto municipality, Int. 133

¹⁹⁵ FGD with school coordinators/directors, female, Ainaro municipality, Int. 109

feeding in intervention areas as compared to comparison areas (Table 85), suggesting that PTAs in areas with HATUTAN programming may have had relatively more gains in effectiveness than those in areas not exposed to program activities.

Table 85: PTA involvement in school feeding, by study group

Comparison				Intervention			Difference in Differences	
BL		ML	Difference	BL	ML	Difference	DiD	p
PTA oversees school feeding								
n	85	86		94	98			
Yes	75.3%	57.0%	-18.3	81.9%	90.8%	8.9	27.2	0.002**
PTA improves school feeding								
n	378	208		482	353			
Yes	29.1%	62.5%	33.4	31.1%	77.3%	46.2	12.8	0.01**

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

PTA oversight of school feeding is related to the existence and level of activity of PTAs. Most schools reported having a PTA (95% baseline, 99% midline). At baseline, most comparison (66%) and treatment (53%) schools reported that the PTA did not meet during the year (Table 86). In the midline, there was an improvement in the frequency of PTA meetings in treatment schools, although most of them still reported that the PTA did not meet during the year (36%). Most comparison schools (70%) continued to report that they did not have a meeting for the year. Overall, at least a third of the schools reported their most recent PTA meeting to have been a month or more than a month earlier (37% baseline, 33% midline).

Table 86: Frequency of PTA meetings

	Comparison Schools		Intervention Schools	
	BL	ML	BL	ML
PTA meeting				
n	83	86	91	97
Last week	4.8%	4.7%	2.2%	23.7%
Last month	7.2%	12.8%	11.0%	17.5%
More than a month	21.7%	12.8%	34.1%	22.7%
Did not meet this year	66.3%	69.8%	52.8%	36.1%

However, the level of participation in the PTA remains low, indicating limited participation and potential influence in activities involving school feeding. Most households (65% baseline, 63% midline) do not have a member who participates in the PTA (Table 87). Furthermore, there was no significant difference in reported household participation rates across comparison and intervention groups from baseline to midline.

Table 87: Household participation in the PTA, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	208		482	353			
Yes	29.1%	34.6%	5.5	31.0%	37.4%	6.4	0.9	0.86

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Qualitative data indicate that there are varying levels of coordination between the PTA and parents in the schools. Moreover, school staff work with the PTA when there is a need for support in relation to

maintenance of facilities and contributions to the feeding program. A number of school coordinators/directors noted that men more actively participate in PTA meetings and other activities than women. However, in general, convening PTA meetings is a challenge due to low turnout attributed to lack of incentives.

This [PTA] thing... we are all Timorese. We call them every week. Sometimes they come, sometimes they don't. Here is the challenge. We Timorese are like that. They say, we will come but at least give us 5 dollars or a dollar. This is what it's all about. This is the challenge between our school and this [PTA]. They are constantly calling us, at least give us a bottle of aqua, right? So, we Timorese are already used to this kind of thing. But they also don't want to lose and they always look for ways to come back again.¹⁹⁶

School coordinators/directors shared that the PTA also monitor the school feeding program to ensure that there is sufficient supply of produce for meals, meals are prepared in a hygienic manner, and meals are served. However, as mentioned earlier, budget constraints and maintenance of facilities continue to be a challenge in some schools.

HEALTH AND NUTRITION

One of the key areas that HATUTAN focuses on is improving nutrition and WASH practices, and one of the McGovern-Dole strategic objectives is to increase the use of health, nutrition, and dietary practices. The results framework indicates that healthy practices decrease health-related absences in school, which improves students' school attendance and contributes to improved literacy. In-depth interviews and quantitative data from the household survey were used to gain an accurate understanding of health, nutrition, and dietary practices within households. Quantitative data from schools were also collected to provide insights on food preparation and storage in school.

This section will focus on the sub-factors identified in Table 88 as reasonable contributors to the increased use of health, nutrition, and dietary practices within households, including comparisons between study groups and findings in the baseline and midline studies. Data collected provide insight into knowledge and awareness of practices and access to tools related to health and nutrition. While knowledge does not demonstrate or imply healthy behaviors or health and nutritional status, it is a proxy for potential behavior. Further study on the gap between knowledge and behavior should be considered.

¹⁹⁶ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

Table 88: Factors contributing to health and nutrition per the results framework

Factors	Sub-Factors
Increased use of health, nutrition, and dietary practices	Knowledge of health and hygiene
	Knowledge of safe food prep and storage
	Knowledge of nutrition
	Access to clean water and sanitation
	Access to preventative health interventions
	Access to requisite food prep and storage tools

NUTRITION PRACTICES

The HATUTAN midline study collected data on the food consumption of caregivers and children under two. Women of child-bearing age play a fundamental role in their baby's development and nutrition through the variety of foods they consume in the household. Caregivers interviewed during the household survey reported on the types of food they consumed the day before. The nine food groups include grain, roots, and tubers (e.g., maize, rice, bread, cereals/porridge, other foods made from grains such as maize and wheat, white potatoes and yams, cassava, and other foods made from roots); legumes, beans, nuts, and seeds (e.g., food made from beans, peas, lentils, peanuts, pumpkin seeds, and sunflower seeds); vitamin A-rich dark leafy greens (e.g., dark green leafy vegetables such as spinach, and mustard greens); other vitamin A-rich vegetables and fruits (e.g., beta-carotene rich vegetables and fruits such as pumpkin, carrot, any dark yellow or orange-fleshed roots and tubers, ripe mangoes, melon, other fruits that are dark yellow or orange inside, and food made with red palm oil); other fruits and vegetables (e.g., cucumbers, tomatoes, cabbage, eggplant, watermelon, jackfruit, and any indigenous or wild fruits); flesh foods (e.g., meat such as beef, pork, goat, chicken, duck, wild animals, seafood, grubs, snails, and insects); organ meat (e.g., liver, kidney, heart, and other organ meats from domesticated animals and wild animals); dairy products (i.e. milk or food prepared with milk, excluding sweetened condensed milk); and eggs.

Table 89 shows that the most common food consumed by women between the ages of 15 and 49 at the midline was grains such as rice, maize, or bread (99%), followed by vitamin A-rich dark leafy greens (73%) and other vitamin A-rich vegetables and fruits (58%). Women in the comparison group tend to consume slightly more grains, roots, and tubers, and vitamin A-rich dark leafy greens than women in the treatment group, at both baseline and midline. However, the difference-in-differences analysis suggests that there was no significant change in food consumption habits in treatment groups as compared to comparison groups. Overall, the data suggests that most of the households live on a carbohydrate-based diet with little access to protein sources. Moreover, there has been a decrease in the percentage of females consuming each food group in general between baseline and midline. This is consistent with the findings from the 2020 Timor-Leste Food and Nutrition Survey (TLFNS), which shows that there are few shifts in dietary diversity in Timor-Leste with carbohydrates and green leafy vegetables being the predominant in their diets.¹⁹⁷

¹⁹⁷ Timor-Leste Food and Nutrition Survey 2020, Government of Timor-Leste. Preliminary report was launched in November 2020, but final report has yet to be released; see <https://www.unicef.org/timorleste/press-releases/timor-leste-food-and-nutrition-survey-2020-preliminary-results-steady-progress-made>.

Table 89: Food consumed by women of childbearing age (15-49), by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	304	518		378	607			
Grain, roots and tubers	100.0%	98.8%	-1.2	99.7%	98.7%	-1.1	0.1	0.88
Vitamin A-rich dark leafy greens	79.9%	76.8%	-3.1	78.3%	69.7%	-8.6	-5.5	0.18
Other vitamin A-rich vegetables and fruits	86.8%	56.0%	-30.9	84.7%	58.8%	-25.9	5.0	0.21
Legumes, beans, nuts and seeds	8.2%	8.3%	0.1	16.1%	11.2%	-4.9	-5.0	0.10
Other fruits and vegetables	24.0%	12.2%	-11.9	23.0%	11.2%	-11.8	0.0	0.99
Eggs	10.5%	5.4%	-5.1	11.9%	9.1%	-2.8	2.3	0.43
Flesh foods	12.5%	8.1%	-4.4	11.4%	8.1%	-3.3	1.1	0.72
Dairy products¹⁹⁸	0.7%	0.6%	-0.1	2.9%	2.5%	-0.4	-0.4	0.77
Organ meat	4.9%	2.1%	-2.8	4.5%	2.5%	-2.0	0.8	0.67

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Mothers of children under two years of age reported the foods consumed by their children between 6 and 23 months old the day prior to data collection. At midline, most of these children consumed grains, roots, or tubers (93%), at a slightly lower rate than at the baseline (96%) (Table 90). Similar to the food consumption trend of women above, there has also been a decrease in the percentage of children consuming each food group between baseline and midline, except for dairy products where food consumption doubled (from 14% at baseline to 30% at midline). However, the decline in the consumption of most food types among children is not as sharp as for women. The decrease in consumption of dairy products among women but increase in consumption among children, as well as their limited consumption of such products to begin with, may be an indication that these are preferably given to children than consumed by their caregivers. The difference-in-differences analysis presented in Table 90 also shows that there was little significant difference in food consumption between treatment and comparison groups from baseline to midline, except for vitamin A-rich vegetables and fruits. For this food group, consumption in the comparison group dropped by 20 percentage points at midline, but consumption in the treatment group remained relatively stable. This suggests that program activities (or some other factor affecting only treatment groups) may have helped stabilize consumption of vitamin-A rich foods for children in intervention areas.

¹⁹⁸ Dairy products are not commonly consumed as part of the Timorese diet.

Table 90: Food consumed by children between the ages of 6 and 23 months, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	77	147		98	212			
Grain, roots, and tubers	97.4%	92.5%	-4.9	94.9%	93.9%	-1.0	3.9	0.33
Vitamin A-rich vegetables and fruits	50.7%	30.6%	-20.0	44.9%	44.8%	-0.1	20.0	0.03*
Legumes and nuts	2.6%	0.7%	-1.9	7.1%	4.7%	-2.4	-0.5	0.89
Other fruits and vegetables	3.9%	3.4%	-0.5	6.1%	2.4%	-3.8	-3.3	0.39
Eggs	15.6%	6.8%	-8.8	10.2%	12.3%	2.1	10.8	0.07
Flesh foods	6.5%	4.1%	-2.4	4.1%	5.7%	1.6	4.0	0.34
Dairy products	14.3%	34.0%	19.7	14.3%	26.4%	12.1	-7.6	0.30

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Dietary diversity scores of women and children were calculated, and mean scores are reported in Table 91. The women's dietary diversity score (WDDS) reflects the diversity of a woman's diet based on the number of food groups (up to nine) that they consumed. All respondents mentioned that they consumed at least one of the food groups. Women in the comparison group mentioned a maximum of seven food groups at the baseline and six at midline. Overall, most of them mentioned consuming three (47%) to four (21%) food groups the previous day at the baseline, and two (35%) to three (37%) food groups at midline. Thirty-three percent of women at baseline and 18% at midline consumed four to eight food groups the previous day. In the baseline study, the average number of food groups that women consumed the previous day was similar for both treatment (3.33) and comparison (3.28). A decrease in score was observed at midline where women consumed an average of 2.68 out of nine food groups in the comparison group and 2.72 in the treatment group. This appears to be primarily driven by the decline in the consumption of other vitamin A-rich vegetables and fruits. This decline was consistent across both comparison and treatment groups, suggesting that program activities had no impact on dietary diversity for women in intervention areas. The limited dietary diversity may indicate poor nutrition practices, which may be an effect of the "hungry season"¹⁹⁹ or limitations to financial access to purchase food during COVID-19.²⁰⁰

In the comparison group, most of the respondents in Aileu and Covalima consistently mentioned three food groups at baseline and midline, whereas the number of food groups mentioned by most of the respondents in Baucau, Bobonaro, and Manufahi dropped to two. Among respondents in the treatment group, most respondents in all four municipalities mentioned three food groups at the baseline (52% in Ainaro, 50% in Ermera, 34% in Liquica, and 31% in Manatuto). However, at the midline, a higher percentage of respondents in Liquica (45%) and Manatuto (46%) mentioned two food groups. While most respondents in Ainaro and Ermera still mentioned three, the percentage of people that said so decreased to 38% and 37%, respectively.

¹⁹⁹ During the 'hungry season', households' food stock may have been depleted before the pre-harvest period and families tend to resort to coping strategies such as limiting food consumption.

²⁰⁰ Based on the program's semi-annual report, as of October 2020, 28% of the households reported having experienced food insecurity (having no food or insufficient food for 10 or more days) since the lockdown started. Annex 3: HATUTAN Education and Nutrition Program, Semi-Annual Report FY2020 (April to September 2020), CARE.

Table 91: Mean dietary diversity score (DDS) of women (15-49 years old) and children (6-23 months old), by study group

	Comparison			Treatment		
	BL	ML	Difference	BL	ML	Difference
n (women)	304	518		378	607	
Women's DDS	3.3	2.7	-0.6	3.3	2.7	-0.6
n (children)	77	147		98	212	
Children's DDS	1.9	1.7	-0.2	1.8	1.9	0.1

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Children's dietary diversity scores were also calculated to determine if they meet the minimum acceptable diet (MAD) for children ages 6 to 23 months, following the same method as for WDDS but with seven instead of nine food groups.²⁰¹ To meet the MAD requirement for children ages 6 to 23 months, children should have consumed at least four of the seven food groups during the previous day.²⁰² Only 6% of children met the MAD requirement at baseline and 4% at midline. This is an indication of the declining quality of nutrition provided to children. There was, however, no substantial difference in the average dietary diversity within intervention or comparison groups.

At the midline, in comparison municipalities, most caregivers in Aileu (48%) and Covalima (52%) fed their children with food from two food groups at the midline and most caregivers in the other three municipalities mentioned one, which is consistent with the trend among women where those residing in Aileu and Covalima were more likely to mention more food groups. On the other hand, most of the respondents in all four treatment municipalities mentioned two food groups. However, these results should be interpreted with caution noting the smaller sample size of children compared to that of women. Thus, there is a lower level of confidence in the findings on dietary diversity of children than women.

At both baseline and midline, most mothers reported that their children under six months were exclusively breastfed the previous day: 79% of mothers at baseline and 76% of mothers at midline reported that their child ate and drank only breastmilk. Rates of exclusive breastfeeding were slightly higher in treatment areas than comparison areas at both baseline and midline (75% in comparison and 81% in treatment at baseline, and 70% in comparison and 79% in treatment at midline). Mothers of children ages 6-23 months were much less likely to report exclusive breastfeeding; 46% of mothers reported that their child ate and drank only breastmilk at baseline, and only 23% of mothers at midline. At baseline, mothers in treatment areas were more likely to report exclusively breastfeeding their child age 6-23 months than mothers in comparison areas, but at midline, mothers in these areas were less likely exclusively breastfeed than mothers in comparison areas. However, it is worth noting that at midline, the average age of children was slightly older in treatment areas and slightly younger in comparison areas; some of the relative decrease in exclusive breastfeeding may be explained by this difference.

Mothers in the comparison group were more likely to give their children formula milk at the baseline (14%) than mothers in the treatment group (9%). Reported consumption of formula milk declined for both the comparison and treatment groups at the midline (4% comparison, 8% treatment), but more mothers in the treatment group reported giving their children ages 6 to 23 months formula milk.

Food consumption may be influenced by traditional nutrition practices for children, infants, and pregnant and lactating women in Timor-Leste. Previous research in Liquica and Ermera on traditional nutrition practices found that there is a preference for feeding children with rice instead of maize and

²⁰¹ Breastfed and non-breastfed children were included in the same analysis since very few of the non-breastfed children consumed milk.

²⁰² WHO indicator: Children aged 6-23 months who receive a minimum acceptable diet (MAD), defined as the percentage of children 6-23 months of age who received foods from ≥ 4 (out of 7) food groups during the previous day.

a range of taboos for pregnant women, such as consumption of cold water, coconut water, goat meat, fish, cassava leaves, and “hard foods.”²⁰³

Based on qualitative data, parents demonstrated awareness of healthy nutrition practices. Parents said that they prepare a nutritious breakfast for their children before they go to school and dinner in the evening before they go to bed. The types of food served in the household include a variety of vegetables (e.g., spinach, passion fruit leaves, potato leaves, cassava leaves, cabbage, pumpkin, mustard greens), cassava, beans, eggs, meat (e.g., chicken, beef), and rice. However, responses indicate that provision of healthy and diverse meals at home is dependent on availability of financial resources.

*We have cassava leaves, Chinese cabbage, and broiler chicken meat. If we have money, we buy buffalo meat for ourselves. If we don't have money, we don't [buy].*²⁰⁴

*If we have money, we will buy eggs and other side dishes for our children. This includes local food [such as] cassava, shoot of pumpkin leaves.*²⁰⁵

*The children eat rice twice a day. They eat potato leaves for lunch and cassava leaves for dinner. We don't always give the same thing.*²⁰⁶

*Firstly, for our daily consumption we have vegetables; secondly, we have eggs when they are available; and thirdly, we also depend on availability of money. If money is available, we can buy meat for consumption of the household and in one side we also consume red beans because it benefits our health. Sometimes we also have tofu and tempeh. We buy them from vendors [who] came to sell here and this is available because usually vendors bring them from Buarahun... From Manatuto they bring fresh fish, so we buy fish, tofu and tempeh, but this does not happen every day. Our main consumption for every day is vegetables like mustard for one day and other vegetables for another day. When fruits are available, we also consume them. We cannot consume meat every day but rarely.*²⁰⁷

One of the fathers also recalled information shared by CARE on maintaining healthy diets.

*Last year CARE disseminated information about giving fruits to children and pregnant mothers who are about to give birth. [They] also reminded to not just eat like this, eat also this. We listened so we understand that in order to get healthy we must also consume fruits.*²⁰⁸

KNOWLEDGE OF HEALTH AND HYGIENE

Knowledge of health and hygiene practices is one of the six factors that contribute to increased use of health, nutrition, and dietary practices included in the results framework. Images representing healthy

²⁰³ Castro, A.F. (2013) An Approach to the Field Habits of Three Communities in Timor-Leste.

²⁰⁴ FGD with mothers, female, Ainaro municipality, Int. 108

²⁰⁵ FGD with mothers, female, Manatuto municipality, Int. 120

²⁰⁶ FGD with mothers, female, Ermera municipality, Int. 138

²⁰⁷ FGD with fathers, male, Manatuto municipality, Int. 121

²⁰⁸ FGD with fathers, male, Manatuto municipality, Int. 128

and unhealthy hygiene practices were shown to caregivers to demonstrate their knowledge of these practices. Most of them were able to correctly identify the healthy hygiene practices.

The general trend observed is that there is an improvement in the percentage of mothers that were able to identify healthy hygiene practices. Notable improvements are in the identification of burying trash, cleaning a runny nose, wearing sandals/shoes, and using the toilet to defecate or urinate as healthy hygiene behaviors (Table 92). However, it should be noted that a change in knowledge does not mean change in behavior. Further data collection and analysis would help to understand how these healthy behaviors identified by mothers are practiced in the household. Furthermore, the difference-in-differences analysis suggests that there was little change in the intervention group as compared to the comparison group at midline, suggesting that program activities did not have a large impact on knowledge of healthy hygiene practices because rates were already fairly high.

Table 92: Knowledge of healthy hygiene practices, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	362	625		163	982			
Wash hands after caring for animals	93.9%	94.1%	0.2	92.0%	95.2%	3.2	3.6	0.19
Wash hands before caring for children	94.8%	95.8%	1.1	89.6%	95.2%	5.6	4.8	0.09
Wash hands after cleaning children	95.3%	97.1%	1.8	96.9%	95.7%	-1.2	-2.7	0.17
Wash hands before eating	98.6%	98.4%	-0.2	97.6%	97.3%	-0.3	0.6	0.68
Wash hands before feeding children	98.1%	98.6%	0.5	98.2%	97.2%	-1.0	-1.1	0.43
Wash hands before preparing food	97.0%	97.3%	0.3	95.1%	97.4%	2.3	2.3	0.26
Wash hands after using the toilet	95.9%	98.1%	2.2	95.7%	96.1%	0.4	-1.7	0.40
Wash hands after picking the trash	93.9%	93.8%	-0.2	92.6%	94.7%	2.1	2.6	0.33
Drink boiled water	97.2%	96.5%	-0.8	96.3%	98.2%	1.9	2.6	0.17
Brush the teeth	95.6%	99.0%	3.5	95.1%	98.5%	3.4	-0.1	0.97
Bury the trash	91.2%	95.5%	4.4	89.0%	94.0%	5.0	0.7	0.83
Cover the food after preparing it	95.9%	98.6%	2.7	93.9%	98.3%	4.4	1.7	0.45
Trim the nails	96.4%	98.9%	2.5	93.3%	97.1%	3.8	1.3	0.57
Clean a runny nose	89.0%	94.7%	5.8	85.9%	93.1%	7.2	1.4	0.68
Use sandals/shoes	93.9%	98.1%	4.2	87.7%	96.2%	8.5	4.3	0.15
Use the toilet to urinate/defecate	96.4%	98.2%	1.8	89.6%	98.2%	8.6	6.8	0.01*
Wash clothes	95.0%	98.1%	3.1	90.8%	97.8%	7.0	3.9	0.14
Wash the dishes	96.1%	98.9%	2.8	90.8%	97.6%	6.8	4.0	0.12
Wash fruits and vegetables	95.0%	97.6%	2.6	96.3%	96.5%	0.2	-2.4	0.25

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The McGovern-Dole Custom Outcome #21 asked that caregivers identify at least 17 out of 19 healthy hygiene practices. This improved between baseline to midline where more caregivers achieved the standard in both comparison (94%) and treatment areas (93%) (Table 93). This is an indication that more caregivers are becoming aware of important hygiene/sanitation practices. Achievement of this

outcome increased by a greater percentage in intervention areas (12 percentage points) than comparison areas (6 percentage points) from baseline to midline; although this difference was not significant and thus results are not conclusive, this suggests that program activities may have had some impact on achievement of this outcome in treatment areas.

Table 93: Custom outcome #21: Percentage of participants who can identify at least 17 out of 19 important hygiene/sanitation practices, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	362	625		163	982			
Achieved	88.7%	94.4%	5.7	80.4%	92.6%	12.2	6.5	0.08

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Despite the mothers' abilities to identify healthy hygiene practices, some unhealthy hygiene practices were also identified as hygienic. Table 94 shows that there was an improvement between baseline and midline regarding the practice of keeping livestock and animals in the house or the kitchen. However, the percentage of caregivers who said that throwing trash outside is a healthy hygiene practice increased in both treatment and comparison groups. The percentage of caregivers that said that throwing trash outside is a healthy behavior more than doubled during the midline (from 21% to 52% in the treatment and from 23% to 54% in the comparison group). Identification (accurately or inaccurately) of these practices changed relatively uniformly across treatment and comparison groups from baseline to midline, suggesting that the change was not as a result of program activities. However, it is not clear if the image shown to mother about throwing trash outside might have been misinterpreted, as disposing trash from the house to an outside area may be considered a healthy hygiene behavior because it prevents pests from entering the house and contributes to maintaining sanitation inside the house.

Table 94: Unhealthy hygiene practices identified as healthy hygiene practices, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	362	625		163	982			
Keep livestock/animals in the house or kitchen	69.1%	27.2%	-41.9	65.6%	22.1%	-43.5	-1.7	0.74
Throw trash outside	23.2%	55.0%	31.8	20.9%	52.4%	31.6	-0.3	0.96

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

During the midline, data on knowledge of COVID-19 prevention measures was gathered. Caregivers were asked what behaviors help prevent COVID-19 and which ones they practice. Table 95 shows the list of COVID-19 measures that respondents were aware of and which ones they practice. More respondents from comparison households are aware of staying home as a prevention measure and comparison households are significantly more likely than treatment households to practice this behavior. There were no other significant differences between treatment and comparison households. Overall, most households at midline had knowledge of mask wearing, handwashing, and social distancing as safety measures. However, there is still a minority of respondents that were unaware of such preventative practices, which may indicate an urgent need to reinforce them. Furthermore, in general, practices were reported less frequently than knowledge, suggesting that knowledge does not also translate into a healthy behavior. It should be noted, however, that COVID-19 practices that respondents applied were self-reported and not observed.

Table 95: Knowledge of COVID-19 prevention behaviors, by study group

	Knowledge		Difference in Means		Practice		Difference in Means	
	Comparison	Treatment	Diff.	p	Comparison	Treatment	Diff.	p
n	625	734			625	734		
Wearing a mask when leaving the house	58.2	60.0	-1.7	0.52	53.0	52.5	0.5	0.85
Handwashing with soap	78.4	80.1	-1.7	0.44	71.0	71.5	-0.5	0.84
Staying at home	19.7	14.0	5.6	0.005**	22.9	15.9	6.9	0.001***
Maintaining social distance	32.2	32.8	-0.7	0.79	27.5	27.0	0.5	0.82
Avoiding gatherings	6.4	5.2	1.2	0.33	5.6	4.8	0.8	0.49
Staying away from sick people	4.6	5.0	-0.4	0.73	4.2	4.1	0.1	0.95
Not spitting in public spaces	1.8	3.4	-1.6	0.06	1.8	2.5	-0.7	0.38
If experiencing fever, coughing, sneezing or difficulty breathing, seek treatment immediately	3.2	2.6	0.6	0.50	3.7	2.6	1.1	0.25
Cover your sneeze or cough using a tissue or your elbow	4.3	6.3	-1.9	0.11	3.4	4.4	-1.0	0.34
Throw used tissues in the trash	3.5	5.3	-1.8	0.11	2.2	3.8	-1.6	0.10
Do not know	15.2	13.8	1.4	0.45	-	-	-	-

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Qualitative data, consistent with findings from the household survey, indicated that households are familiar with healthy hygiene practices, particularly relating to handwashing practices before eating, after playing, and after using the toilet. Most parents are also promoting the use of soap during handwashing, especially during the COVID-19 pandemic.

*Easy for my children, we put 5 liters of water into a jerry can, punch on the lower part of the jerry can with unused pen covering, and then put hand soap near a door. When they finish using the toilets, they wash their hands.*²⁰⁹

*[Children] would wash their hands this way in front of us. Now to ensure whether they wash their hands or not, we would have to stay home all day. For example, when we are having meals, we tell them to wash their hands. When we are not around, sometimes they would only care if they eat till they are full.*²¹⁰

We wanted our children to be healthier with these foods, even if only vegetables there would be no problem, as long as they are well-washed - washed three or four times so they are clean before cooking and should be well cooked. Only then can [vegetables]

²⁰⁹ FGD with mothers, female, Ermera municipality, Int. 101

²¹⁰ FGD with fathers, male, Ermera municipality, Int. 105

*be given to the children, so that the coronavirus disease that is out there cannot affect them.*²¹¹

*In my house, when my children take a bath, they have to brush their teeth too, cut their nails, and that's good for their health. When they start to eat, they need to wash their hands with soap first, and they need to eat some clean food.*²¹²

Some parents, however, mentioned that there were challenges to practicing handwashing, including reluctance of children and lack of appropriate handwashing facilities.

*Well, firstly, normally before the spread of the COVID, we are not accustomed with washing hands with soap before having our meals or doing anything. Secondly, when COVID emerged, the health department announced that we must wash hands with soap before having our meals, but we don't do it most of the time because of the failure from parents who sometimes do not prepare water and soap for their children because what we have prepared [by the side of the] door can be spoiled. We must always pay attention to them. [Whether you] like it or not, they must practice this behavior, but when parents don't control them even 5 liters of water [that] is prepared [for handwashing] they will not finish in a week. All these things depend on the parents because they are the guides for their children.*²¹³

*It is difficult because we do not have bathroom. It is different here compared to Maubisse or in Dili. We will use our pigsty as toilet, so it is difficult to wash our hands after we use the toilet. This will affect our health.*²¹⁴

KNOWLEDGE OF NUTRITION

Caregivers were asked to provide examples of important maternal childcare practices during the household survey to gauge their knowledge of healthy nutrition practices. Table 96 summarizes the caregivers' responses. During the midline, the nutrition practices identified most commonly by caregivers included using a variety of nutritious local foods, exclusively breastfeeding for six months, and introducing appropriate, safe and adequate complementary foods for children aged six months to two years and beyond. The difference-in-differences analysis finds that caregivers in intervention areas were significantly more likely to identify immediate initiation of breastfeeding, exclusive breastfeeding, continuing to breastfeed for 1 to 2 years, immunizing children, continuing to breastfeed when a child is sick, and providing expressed breastmilk to a child unable to suckle as healthy practices at midline than expected given the change in comparison areas from baseline to midline. However, caregivers in intervention areas were significantly less likely to identify using a variety of nutritious foods as a healthy practice at midline than expected: 6 percentage points fewer caregivers identified this practice at midline than at baseline, while 6 percentage points *more* caregivers identified the practice in comparison areas. Overall, these results point to generally, but not uniformly, positive program impact on identification of child nutrition practices. However, there remain many practices for which identification remains low.

²¹¹ FGD with mothers, female, Ermera municipality, Int. 106

²¹² FGD with mothers, female, Ainaro municipality, Int. 113

²¹³ FGD with fathers, male, Manatuto municipality, Int. 121

²¹⁴ FGD with fathers, male, Ainaro municipality, Int. 125

Table 96: Caregivers' knowledge of child nutrition practices, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	625		482	982			
Use a variety of nutritious local foods for infants and young children	49.7%	55.4%	5.6	56.6%	50.5%	-6.1	-11.8	0.006**
Initiate breastfeeding within one hour of delivery	42.6%	39.2%	-3.4	31.5%	39.9%	8.4	11.8	0.005**
Exclusively breastfeed for six months	36.2%	48.2%	11.9	25.5%	45.4%	19.9	8.0	0.05*
Introduction of appropriate, safe and adequate complementary foods at 6 months up to 2 years and beyond	32.5%	46.4%	13.9	24.7%	42.0%	17.3	3.4	0.40
Feed foods rich in iron	27.0%	24.8%	-2.2	28.8%	24.5%	-4.3	-2.1	0.58
Feeding frequent meals and snacks to child	28.8%	14.4%	-14.4	26.8%	16.2%	-10.6	3.9	0.28
Breastfeed frequently on demand, both day and night	29.4%	39.4%	10.0	25.7%	37.5%	11.7	1.8	0.66
Take children to health promotion sessions or health facility	23.8%	35.4%	11.6	20.3%	34.9%	14.6	3.0	0.42
Continue breastfeeding for 1 to 2 years	19.8%	27.0%	7.2	15.8%	29.8%	14.1	6.9	0.05*
Ensure timely immunizations	19.3%	20.3%	1.0	15.4%	26.1%	10.7	9.7	0.004**
Maintains health card to monitor growth and development of child	16.4%	12.6%	-3.8	16.2%	17.0%	0.8	4.6	0.14
Ensure child sleeps under treated mosquito net	16.1%	18.7%	2.6	15.4%	16.5%	1.2	-1.4	0.65
Appropriate care for pregnant and lactating women, including adequate quantities of nutritious food	11.9%	10.2%	-1.7	8.9%	13.2%	4.3	6.0	0.03*
Continue or increase breastfeeding when mother or child is sick	8.2%	7.2%	-1.0	7.3%	11.3%	4.0	5.0	0.03*
When infant unable to suckle, provide expressed breastmilk in cup or tube	6.4%	4.5%	-1.9	2.7%	5.0%	2.3	4.2	0.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

McGovern-Dole Custom Outcome #23 on nutrition knowledge asks that mothers identify at least three important nutrition or dietary recommendations to meet this standard. At baseline, about half of caregivers (49%) achieved this outcome; the percentage increased at midline (66%) (Table 97). Consistent with the baseline study, more caregivers in the comparison group achieved this outcome compared to caregivers in the treatment group. However, there was greater improvement in the intervention group than the comparison group—around 19 percentage point compared to 15 percentage points respectively—although this difference was not significant.

Table 97: Custom outcome #23: Percentage of participants in program target groups who can identify at least three important nutrition/dietary recommendations, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	366	596		482	982			
Achieved	53.8%	68.6%	14.8	45.8%	64.7%	18.9	4.1	0.36

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ACCESS TO CLEAN WATER AND SANITATION IN THE HOUSEHOLD AND IN SCHOOL

The results framework identifies access to clean water and sanitation as a sub-factor that contributes to health, nutrition, and dietary practices. One of the intermediate results expected from the program is increased access to clean water and sanitation services, which includes access to an improved water source and improved sanitation facilities. Households' access to an improved water source has declined, while reported sanitation services in schools have improved. The presence of handwashing stations with soap in schools was mandatory during the reopening in July 2020. This process was facilitated by HATUTAN in close collaboration with PTAs in intervention schools.

Respondents to the household survey were asked if there was a time of the year when drinking water becomes unavailable, as well as if their household had a toilet of any kind. Almost six in ten respondents (58% baseline, 56% midline) said drinking water is not available the whole year. Less than half of the respondents at the baseline (46%) said that drinking water was unavailable on some days, but at the midline almost half (49%; 42% treatment, 54% comparison) of those who did not have reliable access to water said that water was unavailable for more than a month (see Table 98). Almost six in ten in both rural (58%) and urban (59%) treatment locations said that there are times during the year that drinking water is not available.

While access to a source of drinking water has slightly improved between baseline and midline, it appears that availability has been intermittent in some locations. Among treatment municipalities, most respondents in each municipality at the baseline (65%) said that there is a time of the year that drinking water becomes unavailable, but this decreased during the midline. At the midline, most of the respondents in Ainaro (54%) said that water was available the whole year.

In relation to access to toilets, eight in ten (82%) said they had a toilet of some kind at baseline and a slight increase in the number of respondents that said the same at midline (83%). Possible responses for type of toilet available at home include covered and uncovered pit latrines, improved pit latrines, composting latrines, and flushing toilets. Most households have covered pit latrines (40% baseline, 35% midline), uncovered pit latrines (26% baseline, 19% midline), and improved pit latrines (10% baseline, 28% midline) (see Table 98). These are the three main types of toilets used in both comparison and treatment municipalities, regardless of whether the treatment location (i.e., rural or urban).

Table 98: Access to drinking water and toilets, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	625		482	982			
Drinking water is available the whole year								
Yes	51.3%	41.9%	-9.4	34.7%	44.9%	10.2	19.6	<0.001**
Household has a toilet								
No toilet	13.8%	17.3%	3.5	21.2%	16.3%	-4.9	-8.4	0.009**
Pit latrine, uncovered	25.7%	18.6%	-7.1	26.4%	19.5%	-6.9	0.2	0.96
Pit latrine with a slab	46.6%	30.9%	-15.7	35.1%	37.4%	2.3	18.0	<0.001***
Improved pit latrine	7.7%	32.2%	24.5	11.2%	25.9%	14.7	-9.8	0.001***
Composting latrine	1.9%	0.2%	-1.7	1.3%	0.1%	-1.2	0.6	0.53
Flushing toilet	4.5%	1.0%	-3.5	5.0%	0.9%	-4.1	-0.5	0.73

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 98 also shows that there was a significant change in access to drinking water and some types of toilets in intervention areas as compared to comparison areas. At midline, access to drinking water increased in treatment areas while decreasing in comparison areas, suggesting that program activities (or other factors applying to only intervention municipalities) may have significantly improved year-round access to drinking water for affected households. Similarly, the number of intervention households that did not have a toilet declined substantially compared to comparison households—in which more households reported not having a toilet at midline—and the number of intervention households with a pit latrine with a slab significantly increased relative to comparison areas. While a higher percent of intervention households also reported having an improved pit latrine at midline than at baseline, access to these types of toilets increased by a significantly greater amount in comparison areas. However, overall, these results suggest that program activities may have had some impact on access to water and toilets in intervention households.

Table 99 shows that the majority of the households in treatment and comparison groups obtained water from a public tap (36% baseline, 34% midline), followed by pipes connected to their yard/plot (24% baseline, 22% midline), spring water (15% baseline, 12% midline), and pipes in their houses (10% baseline, 14% midline). A statistically significant change in the use of public taps was observed. More treatment households were collecting water from a public tap compared to the baseline, but the opposite is true for the comparison households where collection of water from a public tap declined between baseline and midline. There was also a statistically significant change in collection of spring water, with a slight increase in the collection of spring water among comparison households and a decrease among treatment households. Rainwater harvesting was practiced among respondents in the comparison group during the midline but not in the baseline, and a slight increase in the use of this practice was observed among treatment households, although not statistically significant. At the baseline, rural treatment households are more likely to drink water from a public tap (34%) or water piped into their yard/plot (27%). Those in urban areas are more likely to drink water piped into their yard/plot (26%) or drink spring water through bamboo pipes (26%). However, at the midline, most rural and urban households collected drinking water from public taps (36% rural, 32% urban) or pipes connected to their yard/plot (22% rural, 25% urban).

Table 99: Main source of household drinking water, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	625		482	982			
Piped into dwelling	9.0%	16.0%	7.0	11.0%	13.2%	2.2	-4.9	0.08
Piped into the yard/plot	19.6%	20.2%	0.6	27.0%	22.6%	-4.3	-4.9	0.16
Public tap	41.0%	32.0%	-9.0	32.0%	34.8%	2.8	11.8	0.004**
Borehole	6.9%	4.3%	-2.6	3.1%	3.7%	0.6	3.1	0.09
Dug well (protected)	4.8%	2.6%	-2.2	2.1%	2.6%	0.5	2.7	0.08
Dug well (unprotected)	1.6%	4.3%	2.7	2.3%	2.6%	0.3	-2.5	0.07
River/lake	7.1%	9.1%	2.0	2.9%	6.7%	3.8	1.8	0.38
Rainwater harvesting	0.0%	0.5%	0.5	0.6%	0.7%	0.1	-0.4	0.46
Spring/spring water through bamboo pipes	10.1%	11.0%	1.0	19.1%	13.3%	-5.8	-6.8	0.02*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Drinking water supply can be categorized as coming from unimproved or improved drinking water sources.²¹⁵ Table 100 shows the change in use of improved drinking water sources between baseline and midline and the difference between the two study groups. Use of improved water sources declined in both intervention and comparison areas at midline, although the majority of households in both areas still use improved sources.

Table 100: Collection of water from improved drinking water sources, by study group

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	340	556		390	762			
Yes	90.3%	84.9%	-5.4	93.6%	89.5%	-4.1	1.3	0.68

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 101 shows that households in urban areas are more likely to collect drinking water from improved sources than rural households. Only households in the treatment group were grouped into rural and urban locations. Findings show that there has been a decline in the collection of drinking water from improved sources in both study groups and in both rural and urban areas.

²¹⁵ World Health Organization indicator: Population using improved drinking water sources. "An improved drinking water source, by nature of its construction and design, is likely to protect the source from outside contamination, in particular from fecal matter. Improved drinking water sources include: Piped water into dwelling, plot or yard, public tap/stand pipe, tube well/borehole, protected dug well, protected spring, and rainwater collection. On the other hand, unimproved drinking water sources are unprotected dug well, unprotected spring, cart with small tank/drum, tanker truck, surface water (river, dam, lake, pond, stream, canal, irrigation channel and any other surface water), and bottled water (if it is not accompanied by another improved source)."

Table 101: Collection of water from improved drinking water sources among treatment households, by rural or urban location

	Intervention		
	BL	ML	Difference
Yes	93.6% (n = 390)	89.5% (n = 762)	-4.1
Rural	92.6% (n = 323)	89.2% (n = 618)	-3.4
Urban	98.5% (n = 67)	91.0% (n = 144)	-7.5

The school survey included a number of items related to access to clean water for cooking and sanitation. As noted in the baseline study, more than 60 of the comparison schools did not provide responses to these questions and results in Table 102 below should be interpreted with caution. Overall, about six in ten schools (64%) had clean water available to prepare meals at baseline; this increased (86%) at midline in both comparison and treatment schools. About a third (37% baseline, 34% midline) of schools had a handwashing station in their kitchen; access to a handwashing station in the kitchen improved in comparison schools but decreased in intervention schools from baseline to midline, although this difference was not significant. At midline, 86% of intervention schools had handwashing stations (76% with soap) and 88% of comparison schools had handwashing stations (80% with soap). In addition, less than half of schools (43%) reported using detergent often or everyday to clean the kitchen during baseline, and this increased substantially during midline (71%), particularly in treatment schools. However, findings should be interpreted with caution due to the low number of comparison schools that responded to questions on sanitation during baseline. The number of comparison and treatment schools that responded to questions on sanitation was more balanced during midline. Results indicate that the program contributed to the improvement of sanitation in schools in terms of access to clean water for meal preparation and regular cleaning of the kitchen; however, the difference-in-difference analysis found no significant differences across treatment or comparison groups.

Table 102: Sanitation in school, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Clean water to prepare meals								
n	22	71		93	95			
Yes	45.5%	83.1%	37.7	68.8%	87.4%	18.6	-19.1	0.14
Handwashing station in the kitchen								
n	30	71		99	95			
Yes	23.3%	32.4%	9.1	41.4%	35.8%	-5.6	-14.7	0.22
Handwashing station at the school								
n	-	88		-	98		-	-
Yes	-	87.5%	-	-	85.7%	-	-	-
Use detergent to clean kitchen								
n	22	71		93	95			
No	22.7%	4.2%	-18.5	20.4%	1.1%	-19.4	-0.9	0.93
Sometimes	40.9%	29.6%	-11.3	34.4%	24.2%	-10.2	1.1	0.93
Often	0.0%	11.3%	11.3	6.5%	9.5%	3.0	-8.2	0.13
Every day	36.4%	54.9%	18.6	38.7%	65.3%	26.6	8.0	0.57

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

One of the hygiene challenges that most school directors/coordinators raised during the focus group discussions was access to water. One of the school coordinators noted, “When there is no water it is like we are talking about life and death. [Each child] when they come every day has to bring at least one bottle of [water]. When it is the rainy season, we make use of the rainwater.”²¹⁶ Another cited that they have to fetch water from at least half a kilometer away from the school.²¹⁷ A school coordinator in Manatuto municipality emphasized the need for access to water to maintain proper hygiene practices in school.

*We have said it quite often. To practice it daily, we need water which is difficult to get. You can see it here. This new jerry can is empty. Sometimes we put water in that one container for the students to use. It doesn't mean that we only have one container. We have many containers. We have mostly everything, and they are enough, except water. It's very difficult to get water... I have informed them (the PTA) [about challenges regarding access to water] and we have had a Community Score Card (CSC) meeting where they supported us by providing water pipes. The difficult part is the canalization. So, we still just keep the materials here while we go find a place (water source) to pipe the water from. But even then, it makes not much difference because they use a solar-based system so sometimes we don't get water when the sunlight is not intense or bright.*²¹⁸

Not all of the schools appear to be facing challenges with water access. A school director noted that CARE provided them with tools for keeping water when it is available, as well as soap: “CARE provided a complete set of items. They provided ‘klin’ [detergent], a water tank for washing hands in this school, including soap, so there are no difficulties.”²¹⁹

One of the parents raised a concern regarding hygiene and sanitation during an FGD, noting that they were not informed earlier that the students go into the bushes when they need a toilet and do not use water to clean up after themselves, although sometimes they put water and soap.²²⁰

ACCESS TO PREVENTATIVE HEALTH INTERVENTIONS

The household survey covered questions on access to healthcare and use of savings and loans for healthcare.²²¹ Respondents were asked if they can afford the cost of taking someone in the household to the doctor or clinic. At midline, an additional option was added to the survey: “No need to pay.” While this option is useful for improving our understanding of access to healthcare, its inclusion at only midline means that results are not directly comparable. As such, we do not include a difference-in-differences analysis below.

The percentage of caregivers that said they could not afford to take someone to the doctor or a clinic decreased between baseline (35%) and midline (16%). However, most of the respondents in the midline said that they do not need to pay (54%), which was not an option in the baseline survey. This likely explains the lower number of respondents who chose the other four options (all the time, most of the time, sometimes, and cannot pay). Table 103 shows the distribution of responses by study group and round. At the baseline, most of the respondents in the treatment households responded that they can sometimes pay or cannot pay at all. However, urban households were more likely to say that they

²¹⁶ FGD with school coordinators/directors, male, Ermera municipality, Int. 107

²¹⁷ FGD with school coordinators/directors, male, Ermera municipality, Int. 136

²¹⁸ FGD with school coordinators/directors, male, Manatuto municipality, Int. 133

²¹⁹ FGD with school coordinators/directors, female, Ainaro municipality, Int. 109

²²⁰ FGD with mothers, female, Ermera municipality, Int. 138

²²¹ Use of savings and loans is addressed in more detail in the section “Economic Empowerment.”

can sometimes pay (42%) while rural households are more likely to say that they cannot pay (38%), suggesting that rural households may have faced more barriers to accessing healthcare. At the midline, most treatment households (56%) in both rural and urban areas said that they do not need to pay (58% and 47%, respectively) and the percentage of respondents that said they cannot pay drastically declined (14% rural, 21% urban). However, outside of just affordability, distance of health facilities from remote areas, access to transportation, and road accessibility, especially during the rainy season, may affect access to healthcare. In future evaluations, inclusion of questions referring to households' behaviors towards visiting a doctor or the clinic and barriers to access may also be good indicators of access to healthcare.

Table 103: Access to healthcare, by study group

	Comparison			Treatment		
	BL	ML	Difference	BL	ML	Difference
n	378	625		482	982	
All the time	32.3%	12.5%	-19.8	23.2%	12.2%	-11.0
Most of the time	6.9%	1.8%	-5.1	4.8%	3.2%	-1.6
Sometimes	28.0%	14.9%	-13.2	34.9%	13.9%	-21.0
Cannot pay	32.8%	18.1%	-14.7	37.1%	15.5%	-21.7
No need to pay	0.0%	52.8%	52.8	0.0%	55.3%	55.3

Heads of households reported whether anyone in the home had savings or a loan, as well as whether that money was used for healthcare. Table 104 shows that at both baseline and midline, less than one-fifth of respondents reported using savings or loans for healthcare expenses. Use of loans for healthcare decreased substantially at midline, by 12 percentage points in comparison areas and 11 percentage points in treatment areas. Use of savings for healthcare, in contrast, was relatively more stable from baseline to midline.

Table 104: Use of savings and loans, by study group

	Comparison			Treatment			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Use of savings for health care	19.3% (n = 202)	16.5% (n = 377)	-2.9	15.0% (n = 233)	19.7% (n = 646)	4.6	7.5	0.09
Use of loan for health care	18.0% (n = 78)	6.0% (n = 50)	-12.0	14.9% (n = 134)	3.9% (n = 178)	-11.0	1.0	0.88

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Medical care is free in Timor-Leste, and a related expense is usually transportation. Figure 20 shows that the use of savings for healthcare increased among households in the treatment group, while households in the comparison group allocated their savings for other purposes, as can be seen in the decline between baseline and midline. The use of loans for healthcare also declined in both treatment and comparison groups. Within intervention areas, the decline in the number of households that used loans for healthcare may be due to the increase in the number of households that are using their savings instead. This may imply changes in income or spending habits, which may have effects on health, nutrition, and dietary outcomes. Among treatment households, urban households are more likely to use their savings and loans for healthcare than rural households; Figure 21 illustrates this.

Figure 20: Use of savings and loans for healthcare, by study group

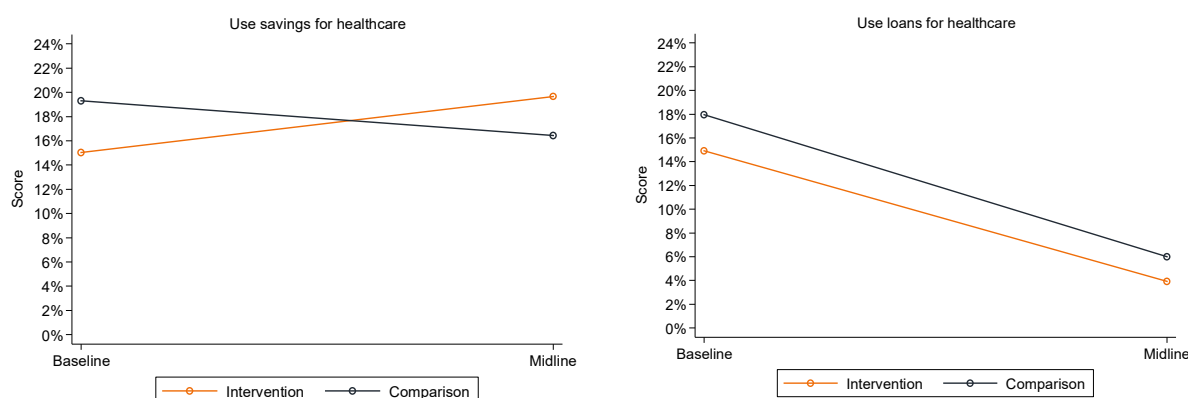
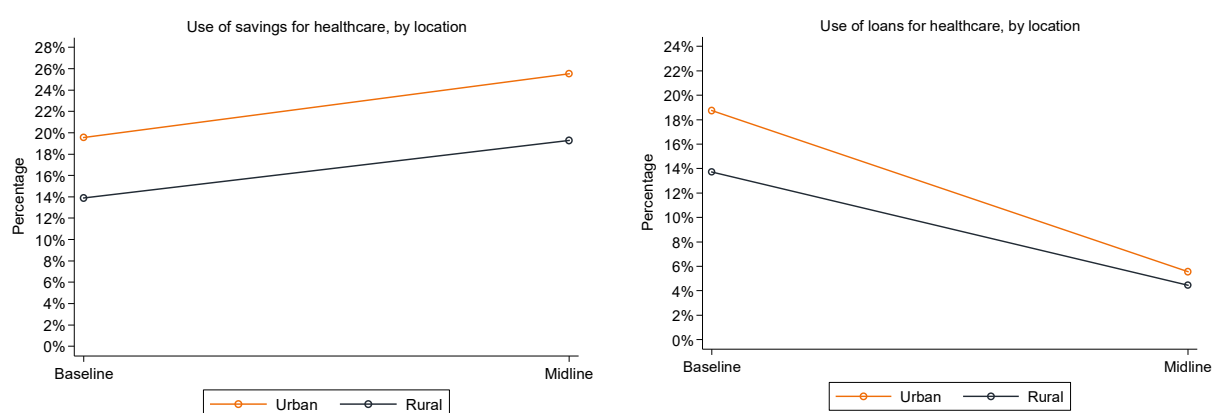


Figure 21: Use of savings and loans for healthcare among treatment households, by urban and rural location



Some of the parents indicated during the FGDs that they visit clinics to get their children checked or to ensure good health during pregnancy.

We go to a clinic [to] buy some vitamins and give it to the children. After consuming it, their body will recover. We must not find other medication. We must go to a clinic. That's what I think. If the children are losing weight, we have to see a doctor and ask for medicines. We should inform the doctor that our children are losing appetite and they give us the proper medicine to regain their appetite and increase their body weight.²²²

In the past we provided local food including cassava and bananas and that's what make them healthy, but now the children refuse to consume local food because this food is not good and [they] consume eggs, porridge, and meat instead. Meanwhile mothers must visit clinics in order to deliver their babies, and they also visit clinics so often to consult with the doctors so that their baby are healthy.²²³

²²² FGD with mothers, female, Hatu-Emera, Int. 117

²²³ FGD with mothers, female, Ermera municipality, Int. 131

ECONOMIC EMPOWERMENT

In this section different aspects of economic empowerment will be explored, with a focus both on differences by whether the respondent was in the treatment group or not, and differences by whether the respondent was part of a village savings and loan association. There are several possible ways in which the program is expected to be beneficial for economic empowerment. One is from positive spill-over effects from other parts of the HATUTAN program. The results framework includes in its “foundational results,” for example, increased capacity of government institutions and increased government support, which if realized could improve the economic condition of households. Another is through trainings designed to promote gender equality which could equalize intra-household decision-making and boost female economic empowerment. A final possible method is through the VSLA—the program aims to either boost capacity of or establish new VSLAs.

We will initially analyze savings, including how many households had savings, where their savings were kept, and what the main uses of savings were. This will be followed by a specific section on VSLAs, where we analyze the effect of VSLA membership and the HATUTAN program on educational and nutritional outcomes. Throughout the analysis the sample used is the cross-sectional sample, although for many of the indicators the data was only collected at midline. Questions on savings were asked in both the household survey and, at midline, in the farmer’s group booster survey. We note that the farmer’s group sample is demographically different than the sample from the household survey; as such, we present results both excluding and including this group.

SAVINGS

The summary statistics for the number of households with savings are presented in Table 105. Overall, the proportion of households with savings was just over half, with the comparison group having a slightly higher percentage than the intervention group. For the sample of just household survey respondents, the percentage for both groups grew by a similar amount over the study period. The difference in differences is correspondingly small and not significant, providing limited evidence of any program impact for this outcome. However, when respondents to the farmer’s group survey are included, we find a positive and significant increase in savings in intervention areas at midline as compared to comparison areas.

Table 105: Households with savings, difference-in-differences

	Comparison			Intervention			Difference in Differences	
Household survey	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	621		482	733			
Households where at least one member has savings	53.4%	60.2%	6.8	48.3%	54.7%	6.4	-0.4	.92
Household + farmers								
n	378	621		482	981			
Households where at least one member has savings	53.4%	60.2%	6.8	48.3%	65.9%	17.6	10.7	0.01*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 106 shows where households kept their savings, for those who answered positively to the question above. More than one answer was possible. For this question we only have data at midline, meaning we are unable to study change over time. For the sample of households only, the most common place to keep savings was at home, with nearly three-quarters of both groups keeping all or some of their savings at home. The next most common answer was a VSLA, with the comparison group having a higher score; the difference between comparison and intervention groups for this response was statistically significant. Most of the rest of the households kept their savings either in a microfinance group or a bank. Overall, while most kept their savings at home, the comparison group comparatively preferred VSLAs while the intervention group tended to use microfinance groups and

banks. Because we cannot assess changes over time, however, it is difficult to draw conclusions with respect to program impact—the treatment and comparison group could differ for reasons other than the program.

We do note interesting differences when respondents from the farmer's group survey are included: This group was much less likely to keep their savings at home and much more likely to keep savings in a VSLA, as expected since the sampling framework for this group were participants trained on VSLA.

Table 106: Location of savings at midline

	Households Only		Households and Farmers	
	Comparison	Intervention	Comparison	Intervention
n	374	401	374	646
VSLA	25.9%	14.5%	25.9%	43.3%
Microfinance group	1.9%	3.5%	1.9%	2.6%
Kept at home	71.4%	74.6%	71.4%	53.3%
Bank	7.0%	11.5%	7.0%	7.1%
Other	0.5%	0.5%	0.5%	0.3%

Table 107 shows the main uses households had for their savings; we restrict this analysis to respondents from the household survey only for comparability purposes. The percentage using their savings for food was high in both groups and rose by 7% to 9% from baseline to midline. The next most common use was education, and again this rose in both groups by a relatively similar amount across the study period. For all other categories the score was approximately twenty or below twenty percent. An interesting pattern is the decrease in percentage spending savings on a business and an increase in spending savings on debt. This could be interpreted as a possible deterioration in households' financial health overall.

Table 107: Change over time in use of savings

	Comparison			Intervention			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	202	374		233	401			
Food	86.6%	94.1%	7.5	84.6%	93.0%	8.4	0.9	0.78
Healthcare	19.3%	16.6%	-2.7	15.0%	20.0%	5.0	7.7	0.10
Education	86.1%	91.2%	5.1	85.4%	90.5%	5.1	0.08	0.98
Agriculture	26.2%	20.6%	-5.6	28.3%	23.2%	-5.1	0.5	0.92
Invest in business	28.7%	20.1%	-8.6	24.0%	17.2%	-6.8	1.8	0.71
Funeral/wedding/other ceremony	-	42.8%	-	-	40.7%	-		
Pay debt	1.0%	8.6%	7.6	6.4%	12.0%	5.6	-2.0	0.53
Invest in household asset	13.9%	9.6%	-4.3	10.3%	12.0%	1.7	5.9	0.12

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The difference in difference is positive for all uses except for paying debt (suggesting the percentage in the intervention group rose by a greater amount or fell by a lesser amount than the comparison), while the scores were negative for use on debt. However, the figure was not statistically significant for any indicator, suggesting that changes in economic empowerment occurring between baseline and midline were not driven by program activities.

For the sample of farmers in intervention areas at midline, respondents reported similar rates of spending on healthcare (19%) and agriculture (23%) as the comparable group of household respondents in intervention areas at midline. Respondents were less likely to report using their savings

for food (87%), education expenses (82%), ceremonies (25%), to pay debt (6%), or to invest in a household asset (6%). They were, however, substantially more likely to report using savings to invest in a business (31%). These results may suggest somewhat higher levels of food security and general economic well-being among respondents in farmers' groups, although expenditure on basic needs—food and education—remains high. However, the relatively higher rate of investment in businesses suggests that respondents in farmers' groups may be somewhat more able to use their savings for long-term goals when compared to respondents from the household survey. It is, however, worth noting that some of these findings may be driven by demographic differences between farmers' survey and household survey respondents, rather than by program impact.

In Table 108 we disaggregate the treatment group by municipality for some of the key indicators—whether a household had any savings, whether they used savings on education, and whether they used savings on debt. The aggregated increase in savings we saw earlier was mostly driven by improvements in the Liquica municipality, where the figure rose from 44% to greater than 80%. The percentage using savings on education was uneven across municipalities, with Ainaro and Liquica dropping slightly and Ermera and Manatuto both increasing. We can also see that the increase in households using savings to pay off debts is mostly driven by a large increase in Liquica, from 0% to around 23% for just household respondents and 17% for household and farmers' group respondents. This is further evidence that any changes in the indicators were not driven by the program but other factors, including municipality-specific factors.

Table 108: Savings indicators disaggregated by municipality

	Savings		Use on Education		Use on Debt	
	BL	ML	BL	ML	BL	ML
Households only						
Ainaro	49.2%	46.6%	90.5%	87.5%	6.4%	8.0%
Ermera	51.0%	51.7%	80.2%	90.4%	6.6%	14.5%
Liquica	43.6%	82.6%	95.8%	93.0%	0.0%	22.5%
Manatuto	44.0%	55.5%	85.0%	92.1%	10.0%	1.3%
Households and farmers						
Ainaro	49.2%	62.9%	90.5%	87.3%	6.4%	7.5%
Ermera	51.0%	59.3%	80.2%	88.1%	6.6%	12.3%
Liquica	43.6%	89.4%	95.8%	88.1%	0.0%	16.7%
Manatuto	44.0%	65.9%	85.0%	85.8%	10.0%	0.8%

VSLAS

In this subsection we focus on households who participated in a VSLA. Following on from the previous section, Table 109 below compares spending on savings at midline disaggregated by where that household kept their savings. The table includes both respondents to the household and farmer's surveys. VSLA members were less likely to spend savings on food, a ceremony, and debt, and more likely to spend savings on business. Of particular interest for the program is spending on education; the results suggest that VSLA members were slightly less likely to spend on education, although the inclusion of farmers in this table—instead of just households with a second grade child—affects these results. It is also important to note here that this variable only ask for *savings* spent on education and not overall spending on education.

Table 109: Savings use by VSLA participation, midline

	Non-VSLA	VSLA
n	645	378
Food	95.2%	86.5%
Healthcare	18.3%	18.8%
Education	90.2%	86.5%
Agriculture	22.5%	22.0%
Invest in business	17.1%	29.4%
Funeral/wedding/other ceremony	44.0%	26.5%
Pay debt	11.0%	6.4%
Invest in household asset	10.5%	8.5%

Below, the key indicators relating to VSLA participants are summarized according to treatment and intervention group. In each household with savings in the VSLA there is a mean of 1.6-1.7 VSLA members. Members are more likely to be female than male. In the intervention group more members had taken out a loan in the past three months; this difference was significant. There was a slightly higher number of mean VSLA meetings in the intervention group which could possibly suggest success on the part of the program's efforts to improve frequency of VSLA meetings. However, the difference was not significant. Overall, there seems to be little difference in VSLA characteristics, membership, and frequency of meetings based on the program's effects, other than loan rates, which were substantially higher in the intervention group than the comparison group. However, small sample size and potential comparability issues between households and farmers are a constraining factor in making rigorous comparisons for VSLA indicators.

Table 110: Summary scores for VSLA indicators

	Comparison	Intervention
Participants		
n	95	291
Mean number of VSLA participants	1.6	1.7
Female participants		
n	95	290
Mean number of female VSLA participants	1.0	1.0
Borrowed past 3 months		
n	473	772
% yes	26.9%	40.4%
Meeting frequency		
n	95	284
Mean meetings last month	1.7	1.9

We briefly analyze gendered decision-making power over what to do with VSLA loans below (for more discussion of gendered decision-making, see "Gender and Power"). All respondents to this question in comparison areas were female; the farmer's survey, however, included some male respondents. For female respondents, most women reported making decisions of VSLA loans jointly with their spouse. Male respondents were somewhat more likely to make decisions themselves. Across both comparison and intervention it was more common that the respondent had sole responsibility rather than his or her spouse having sole responsibility; this is likely reflective of the fact that respondents were the participants in the VSLA, rather than their spouses. For women, respondents in treatment areas were substantially less likely to report making decisions with their spouse, and more likely to state that either they or their spouse alone made decisions.

Table 111: VSLA loan decision-making by gender

Decision on VSLA loan	Female		Male	
	Comparison	Intervention	Comparison	Intervention
n	49	126	0	52
Myself	24.5%	31.0%	-	46.2%
My spouse	10.2%	18.3%	-	1.9%
Me and my spouse	61.2%	47.6%	-	48.1%
Other	4.1%	3.1%	-	3.8%

In Table 112 we observe the main uses of VSLA loans by comparison and treatment for just respondents to the household survey and for both the household and farmers' surveys. More than one answer was allowed. The most common use for loans was education, followed by food, while the least common were debt and agriculture. Again, there were few meaningful differences between treatment and control. The intervention group spent a slightly lower percentage on education, but the difference was not statistically significant. The difference in spending loans on agriculture—which the intervention group reported almost 10 percentage points more than the comparison group for just households—was significant at the 5% level, as were the differences in loans on investing in business and paying debt. As we cannot measure changes, however, we must be careful in attributing these differences to the impact of the HATUTAN program.

Table 112: VSLA loan use, midline

	Household Only		Household and Farmers	
	Comparison	Intervention	Comparison	Intervention
n	127	174	127	312
Food	59.1%	62.1%	59.1%	62.2%
Healthcare	13.4%	13.2%	13.4%	8.7%
Education	74.0%	67.8%	74.0%	69.2%
Agriculture	9.5%	19.0%	9.5%	14.4%
Invest in business	22.8%	32.8%	22.8%	32.1%
Funeral/wedding/other ceremony	34.7%	31.0%	34.7%	18.9%
Pay debt	3.9%	8.6%	3.9%	5.8%
Invest in household asset	11.8%	5.8%	11.8%	3.9%

Table 113 shows the main perceived benefits of being in a VSLA. Across both groups the most common benefits listed were obtaining money for family needs, solidarity, saving money for future needs, and obtaining capital for business. Few respondents mentioned information and learning as a benefit. There are some differences between treatment and comparison groups but given the small sample size, discrepancies are expected. The difference in percentage of respondents citing obtaining capital for a business is quite large, however, with the intervention group scoring 24 percentage points higher, a significant difference. Again, it is difficult to attribute this effect to the program, and it is difficult to rationalize why it might be the case that those in the treatment group would value capital for business higher. Notably, when respondents to the farmer's group survey are included, the percent of total respondents listing obtaining capital for business as a benefit decreases by almost 11 percentage points. This suggests that the household sample in intervention areas may, by chance, have had greater need for capital for businesses. Alternatively, because respondents to the farmer's group survey frequently reported using savings to invest in a business (see Table 106), these respondents

may consider investing in a business to be a core component of their savings plan, rather than a benefit of VSLA participation.

Table 113: Perceived benefits of VSLA participation

	Household Only		Household and Farmers	
	Comparison	Intervention	Comparison	Intervention
n	95	65	95	291
Solidarity	41.1%	33.9%	41.1%	37.5%
Obtain capital for business	15.8%	40.0%	15.8%	29.2%
Obtain money for family needs	62.1%	56.9%	62.1%	59.1%
Information/learning	4.2%	4.6%	4.2%	10.0%
Save money for future needs	36.8%	29.2%	36.8%	34.0%
Information about health/nutrition	5.3%	1.5%	5.3%	2.4%
Obtain information about agriculture	2.1%	3.1%	2.1%	4.5%
Other	13.7%	7.7%	13.7%	15.1%

Below we report a series of regressions to test the effect of being in a VSLA and in the treatment group on the key outcomes of interest for the HATUTAN program. Each regression includes an interaction term for being in a VSLA and in the treatment, which will capture the effect from belonging to both of these groups at the same time over belonging in just one. If, for example, being in a VSLA is not beneficial for nutrition outcomes but being in a VSLA and being in the treatment group is beneficial for nutritional outcomes (possibly due to a higher functioning VSLA or nutritional trainings through the group), we would expect the main effects term not to be significant but the interaction term to be significant and positive. This method should allow us to gain an initial understanding of whether VSLAs have been improved in some respect as regards reaching key program outcomes by interacting with HATUTAN.

Table 114: Impact of VSLA membership on nutrition and literacy outcomes

	Main effect: VSLA		Main effect: treat		Interaction	
	Coefficient	p	Coefficient	p	Coefficient	p
Student eaten on test day (%)	-1.5	0.63	1.1	0.25	2.4	0.62
Household not eaten (%)	-5.8	0.12	2.9	0.15	-2.6	0.62
Literacy score (0 - 1)	-0.13	<0.001***	-0.0064	0.41	-0.034	0.42
Working memory score (0 - 100)	-9.1	<0.001***	-3.05	<0.001***	3.55	0.29

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

For the nutritional variables, none of the predictor variables were significant, suggesting that being in the intervention group (for this small sample) or being in a VSLA was not associated with nutritional outcomes. For literacy score and working memory the main effects for VSLAs were statistically significant, negative, and quite large. Due to the interaction term, this figure is interpreted as the effect of being in a VSLA only for those in the comparison group. It is interesting that these are negative and significant—this might be due to another variable correlated with both education scores and VSLA membership, rather than a causal relationship between the two. The key point from this analysis is that none of the interaction terms were significant. We can infer that the effect that being in a VSLA has on educational and nutritional outcomes is not changed by the HATUTAN program, although again it is important to keep in mind the small sample size and the effect this has on drawing inference.

AGRICULTURAL PRACTICES

In this section we focus on agriculture. The main focus is on two agricultural practices—keyhole gardens and permagardens—and we aim to understand the characteristics of each of these techniques and the effects they may have on economic and nutritional outcomes. The HATUTAN program aims to support farmers in adopting these techniques to boost yields and create sustainable sources of foods, including through farmer trainings.

Most of the data for this section was collected at midline and the sample comprised households of grade 2 students that had farmers as well as a separate group of farmers sampled independently of the household survey. This means we are mostly unable to assess changes over time and gain difference-in-differences estimates, and the analyses possible are limited by the small sample size.

OVERVIEW OF AGRICULTURAL PRACTICES

In Table 115 below we summarize the overall indicators for agricultural techniques and training, disaggregated by treatment and control. The vast majority of respondents had received no agricultural training, although more in the intervention area had received training than in the comparison. Most of the training received covered both keyhole gardens and permagardens, and a higher proportion of the intervention group received training in keyhole gardens. Cultivating keyhole gardens was more popular than permagardens for both groups. Interestingly, although a higher proportion of the intervention group had received training covering these agricultural techniques, a lower proportion of respondents were actually cultivating such gardens than the comparison group, potentially reflecting the impact of other interventions in the comparison municipalities. It is important to note here, however, that the sample size for the comparison group in the farmer survey was very low.

Table 115: Summary scores for main agricultural indicators

	Comparison	Intervention Farmer's survey only	Intervention Household survey only
Received training			
n	496	248	734
No training received	76.0%	63.7%	73.6%
Keyhole garden only	0.5%	10.1%	1.0%
Permagarden only	1.0%	5.2%	0.7%
Both keyhole and permagarden	1.9%	16.9%	1.6%
Don't know	20.6%	4.0%	23.2%
Currently cultivating keyhole			
n	15	67	86
Yes	73.3%	56.7%	59.3%
Currently cultivating permagarden			
n	18	55	72
Yes	61.1%	38.2%	41.7%

Below in Table 116 we summarize the main crops cultivated for farmers with keyhole gardens. A diverse range of crops were grown overall, but most farmers cultivated only a small selection—53% only cultivated one or two crops while 73% only cultivated three or fewer crops. By far the most popular crop was mustard greens, with 86% cultivating this crop. There was also a relatively high proportion who grew onions, carrots, kangkong, and lettuce, each between 25% to 40%. Due to the small sample size we did not disaggregate by treatment group.

Table 116: Crops grown in keyhole gardens

	Total
Keyhole crops cultivated	
n	62
Carrots	27.4%
Onion	37.1%
Garlic	14.5%
Spinach	11.3%
Mustard green	85.5%
Lettuce	25.8%
Kangkung/morning glory	27.4%
Collard greens	11.3%
Other	38.7%

Below we report the crops grown by those with permagardens. A wider range of crops are grown than with keyhole gardens. Only 27% grew one or two crops, while 46% grew at least four crops. This is likely at least in part due to the wider range of possible crops to list, but the “other” option has only fallen by about 10 percentage points after the inclusion of these additional options, suggesting part of the increased diversity is real. Again the most popular crop was mustard greens, with lettuce, bok choy, and tomato also grown by over 30% of respondents.

Table 117: Crops grown in permagardens

	Total
Permagarden crops	
n	41
Carrot	17.1%
Peanut	9.8%
Mustard greens	58.5%
Kangkung/morning glory	12.2%
Lettuce	34.2%
Chilli	39.0%
Cucumber	12.2%
Parsley	12.2%
Cabbage	24.4%
Bok choy	39.0%
Onion	22.0%
Garlic	19.5%
Tomato	31.7%
Eggplant	26.8%
Orange sweet potatoes	17.1%
Pumpkin	24.4%
Spinach	9.8%
Soybean	2.4%
Mung bean	14.6%
Other	26.8%

Indicators relating to sales of home-grown produce are summarized below. Most respondents in the farmer survey grew vegetables for both household consumption and sale rather than household consumption alone. However, for both treatment and comparison, most vegetables that they grew were kept for household consumption rather than sale. The majority of respondents in both areas report making a profit from selling produce. There are some differences between treatment and control groups, but given that the magnitude of differences are overall quite small and the sample sizes are very small (as low as 5 for the comparison group) we will not attempt to attribute this to the effect of the intervention.

Table 118: Summary scores for sales of homegrown produce

	Comparison	Intervention
For sale or household consumption		
n	10	30
Household consumption only	50.0%	30.0%
Both household consumption and sale	50.0%	70.0%
Proportion of vegetables grown for sale		
n	5	21
Proportion	29.0%	19.0%
Made profit from selling produce		
n	5	21
Yes	80.0%	95.2%

We also report the main challenges faced by users of keyhole gardens and permagardens. Again, the sample size for the comparison group was very small so we did not disaggregate by intervention/comparison status, as it added little value to the analysis. Just over half of the farmers reported facing any challenges. Of those challenges listed by respondents, the main challenges were, in order, lack of tools, materials, and seeds; natural disaster; pests; and lack of technical support. These are possible areas that could be improved by training or by direct support from the program. However, by far the most common choice was “other,” suggesting there were challenges not measured by the survey.

Table 119: Main challenges faced by farmers

	Total
Faced any challenges	
n	69
Yes	52.2%
Type of challenges	
n	44
Other	47.7%
Lack of tools, materials, seeds	22.7%
Natural disaster	18.2%
Pests	18.2%
Lack of technical support	13.6%
Limited production	11.4%
Lack of money to invest	9.1%
Poor quality of seeds	6.8%
Unable to sell	6.8%
Personal issues	6.8%
Damaged/stolen produce	4.6%

Limited amount of land	4.6%
Poor quality of produce	2.3%

Finally, it is also of interest to understand other sources of agricultural support, including access to extension services and sources of information. These are summarized below in Table 120; we note that the sample sizes are quite large as they include households outside of the farmer survey. Around 17% of households had received extension support, and this did not differ greatly between treatment and comparison groups. Most respondents did not receive information about agriculture from any source, and for the sources listed there was little difference between treatment and control overall. However, when analyzing data from just the farmers' group survey, respondents were substantially less likely to report receiving no information, and more likely to report receiving information from training or an NGO worker. While these results are to be expected, given that the survey was specifically targeted towards farmers who had received training, they do point to some level of program impact.

Table 120: Access to AES and sources of information

	Comparison	Intervention Households + farmers	Intervention Farmers only
Received extension support			
n	415	723	248
Yes	17.8%	16.7%	14.9%
Source of agricultural information			
n	625	982	248
Received no information	63.7%	62.5%	52.8%
Training	5.4%	5.7%	16.1%
Extension services	13.6%	14.0%	14.5%
NGO worker	8.3%	9.3%	18.2%

RELATIONSHIP BETWEEN AGRICULTURAL PRACTICES AND ECONOMIC AND NUTRITIONAL OUTCOMES

As mentioned in the introduction, adoption of improved agricultural practices is expected to impact nutritional and economic outcomes, and farmer trainings through the program are intended to either introduce these techniques or improve their implementation. The relationship between these outcomes, agricultural practice, and participation is analyzed in this section with a regression including interaction effects. This allows us to compare, for example, the impact of having undergone agricultural training for someone in the comparison group with the impact of having undergone training for someone in the treatment group. If farmers have benefited from HATUTAN trainings on agriculture, we would expect the interaction term to be positive and significant.

There are three nutritional outcome variables of interest: whether a student had eaten on the day of the test, whether anyone in the household had gone a day without eating in the past 30 days, and whether the household consumed more green vegetables. There are three economic indicators: whether a household had any savings, whether a household used those savings to invest in a business (used here as an indicator of good financial health) and whether a household had used savings to pay debt (used as an indicator of poor financial health). The predictor variable is whether the household had taken agricultural training.

Table 121 below summarizes the regression testing the effect of agricultural training. The main coefficient for training (capturing the effect of undertaking training for those in the comparison group) was positive and substantial for the percentage of households where someone had not eaten in the

past 30 days, while the coefficient for the interaction term (capturing the effect for those in the treatment group) was negative and substantial, though not significant—a 15 percentage point difference between groups. Interestingly, this suggests there were worse nutritional outcomes for those who took training and were in the comparison group, and better nutritional outcomes when a household took training and was in the treatment group. This could be interpreted as a benefit from the HATUTAN program, although we must be cautious in assigning causality. For consumption of green vegetables, the coefficient was negative and significant for training and treatment; in other words, undertaking a training was associated with a lower chance of consuming green vegetables, as was being in the treatment group. The interaction term, however, was positive, suggesting that households that had both taken training and were in the treatment group were substantially more likely to have consumed green vegetables. When household savings was the outcome the coefficients were positive for training and treatment but not for the interaction. This suggests that undertaking a training was associated with a higher chance of having savings, as was being in the treatment group, but there was no interaction between taking training and being in the treatment group at the same time.

Table 121: Relation between agricultural trainings, health, and economic status

Outcome variable	Main effect: training		Main effect: treat		Interaction	
	Coefficient	p	Coefficient	p	Coefficient	p
Student eaten on test day (n = 1,059)	0.58	0.93	-0.09	0.96	1.3	0.89
Household not eaten (n = 1,285)	12.8	0.08	2.1	0.28	-15.3	0.06
Consumed green vegetables (n = 1,298)	-24.3	0.02*	-7.6	0.004**	31.2	0.005**
Household has savings (n = 1,297)	22.1	0.03*	6.3	0.02*	6.9	0.55
Uses savings on investment (household or business) (n = 849)	-5.0	0.44	4.4	0.02*	-0.3	0.96
Uses savings on debt (household or business) (n = 849)	0.2	0.98	4.4	0.19	9.8	0.42

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

GENDER AND POWER

The midline study was also designed to allow for an analysis of gender and power dynamics. At baseline, key areas of inquiry around gender and power were defined by CARE based on extensive field research. The key areas of inquiry correspond to areas where the characteristics and dynamics of gender and power relations are typically negotiated. Overall, the gender analysis seeks to analyze information on the different roles of women and men across both the public and private spheres in order to better understand the different priorities, needs, activities, and responsibilities of men and women (and boys and girls). This analysis will help assess and inform HATUTAN programming to allow it to better transform gender dynamics and power, promote inclusiveness and equality, remain accountable to beneficiaries of all genders (and other demographic characteristics), and build an evidence base to contribute to broader advocacy and social movement for all genders. Additionally, the gender analysis further explores how the intersectionality of marginalization—such as through overlaps in age or disability status—can give rise to discrimination or exclusion in society.²²⁴

²²⁴ See CARE International, *Good Practices Framework: Gender Analysis* (Geneva: CARE International, 2012).

DIVISION OF LABOR

Within the household survey, caregivers were asked to mention household tasks that their grade 2 child performs, including caregiving, housework (e.g., cooking or cleaning), fetching water or firewood, agricultural work, and helping with a family business. They were also asked to estimate the amount of time it takes their child to perform those tasks and whether household responsibilities ever cause their child to arrive late to school or reduce the amount of time the child spends studying or doing homework.

At midline, both male and female students were less likely to perform all household tasks than at baseline. However, female students remained significantly more likely than male students to participate in caregiving, which 78% of female students performed compared to 71% of male students, and to do housework, which 62% of female students performed compared to 39% of male students—a substantial gap of over 23 percentage points. In contrast, male students were significantly more likely to assist with agricultural work at both baseline and midline, with 38% of male students performing this task at midline compared to 25% of female students. There were no significant differences in the gendered division of fetching and helping with a family business; overall, fetching water was the most common task performed by both boys and girls and assisting with a family business was the least common.

Table 122: Gendered division of labor within household

	Baseline			Midline		
	Male	Female	p	Male	Female	p
n	434 ²²⁵	426 ²²⁶		682	677	
Caregiving	82.5%	89.6%	0.02*	71.0%	78.0%	0.002**
Housework	50.5%	75.1%	<0.001***	38.6%	62.0%	<0.001***
Fetching water	85.9%	88.0%	0.33	79.9%	83.3%	0.14
Agriculture	54.8%	42.7%	0.002**	37.8%	24.7%	<0.001***
Business	24.0%	22.3%	0.53	10.1%	13.0%	0.07

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The difference-in-differences regression analysis suggests that there were no significant changes in the gendered division of labor in treatment households at midline when compared to the changes seen in comparison households for either male or female students (Table 123). Rather, participation in household tasks decreased fairly consistently across comparison and treatment households for both male and female students. The mechanism driving this consistent decrease in participation in household tasks is not entirely clear; however, this analysis suggests that the HATUTAN program likely did not have a large effect on changing the gendered division of labor within households.

²²⁵ n = 433 for caregiving and assisting with a business.

²²⁶ n = 424 for caregiving and n = 421 for assisting with a business.

Table 123: Change in gendered division of labor within household

	Comparison Households			Intervention Households			Difference in Differences	
Male	BL	ML	Difference	BL	ML	Difference	DiD	p
n	186	321		248	361			
Caregiving	83.3%	69.5%	-13.8	81.8%	72.5%	-9.3	4.6	0.51
Housework	46.8%	35.8%	-11.0	53.2%	41.0%	-12.2	-1.3	0.88
Fetching water	83.9%	78.2%	-5.7	87.5%	81.4%	-6.1	-0.4	0.94
Agriculture	52.7%	38.0%	-14.7	56.5%	37.7%	-18.8	-4.1	0.64
Business	23.2%	10.3%	-12.9	24.6%	10.0%	-14.6	-1.6	0.86
Female	BL	ML	Difference	BL	ML	Difference	DiD	p
n	192	304		234	373			
Caregiving	85.8%	75.7%	-10.1	92.7%	79.9%	-12.8	-2.7	0.60
Housework	72.9%	58.6%	-14.3	76.9%	64.9%	-12.0	2.3	0.76
Fetching water	86.5%	80.3%	-6.2	89.3%	85.8%	-3.5	2.7	0.61
Agriculture	39.1%	20.4%	-18.7	45.7%	28.2%	-17.5	1.1	0.90
Business	22.8%	11.8%	-11.0	22.0%	13.9%	-8.1	2.9	0.77

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In general, there were few significant differences in caregivers' perceptions of how much time male and female students spent on daily tasks.²²⁷ At midline, caregivers were significantly less likely to report that female students spent their whole day doing work compared to male students; however, this absolute difference was small, at only one percentage point. Otherwise, caregivers generally reported very similar amounts of time spent on tasks for students of both genders. This is similar to the finding at baseline in which caregivers reported that girls were significantly less likely not to do chores, but otherwise reported relatively similar student time dedicated to daily tasks for both girls and boys.

Table 124: Caregivers' perceptions of student time spent on daily tasks

	Baseline			Midline		
	Male	Female	p	Male	Female	p
n	425	414		625	642	
Whole day	0.7%	0.0%	0.07	1.4%	0.3%	0.03*
Half day	15.5%	18.1%	0.39	7.2%	8.9%	0.2
Quarter day	59.5%	59.2%	0.94	63.5%	62.9%	0.82
An hour a day	14.1%	17.4%	0.26	21.8%	22.7%	0.7
Does not do chores	10.1%	5.3%	0.03*	5.8%	4.5%	0.34

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At both midline and baseline, there were neither substantial nor significant differences in the perceptions of caregivers on whether tasks made their male or female children late for school. At baseline, caregivers with male children were moderately, though insignificantly, more likely to report that their children had less time to study or do homework due to household tasks than caregivers with

²²⁷ We note that answers to this question were influenced by social desirability bias: In some households, when asked how much time their child spends on tasks, the respondent stated that the child does not do chores, but elsewhere in the survey, also affirmed that the child does at least one task of caregiving, housework, fetching, agriculture, or helping with a business. As such, these results should be taken as suggestive, but not definitive.

female children. At midline, this gap between male and female children decreased, with caregivers of male and female children reporting approximately equivalent rates at which household tasks took time from studying. In general, at midline, caregivers of both male and female grade 2 students were less likely to report that household tasks made their children late to school or took away time from studying.

Table 125: Caregiver perceptions on tasks and time for education

	Baseline			Midline		
	Male	Female	p	Male	Female	p
n	427	420		625	642	
Tasks make student late often	1.2%	1.2%	0.98	0.6%	0.9%	0.45
Tasks make student late sometimes	14.3%	13.1%	0.58	8.5%	8.9%	0.82
Tasks do not make student late	84.5%	85.7%	0.60	90.4%	90.2%	0.91
Student has less time to study	18.4%	13.9%	0.12	10.1%	9.2%	0.63

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Similarly, we analyze whether there is a correlation between the time spent on tasks and overall literacy scores, days of school missed, or whether a student had ever dropped out of school. We find that while the regression results are positive, they are not significant. We note some limitations to this data, however; some caregivers reported that their child “does not do” chores, but when asked about specific tasks, such as housework or fetching, later in the survey, answered that the child does perform that task. Accordingly, we are likely to underestimate the time spent by students on tasks due to social desirability bias which influences caregivers to report that children do not spend much time on chores.

A difference-in-differences analysis suggests that at midline, there was little significant change in the time students spent on tasks or whether those tasks detracted from their schooling in intervention areas as compared to control areas (Table 126). The analysis finds a significant difference only for the percent of male students reported to spend the whole day on tasks, which decreased slightly in intervention households but increased in comparison households, and whether household tasks took time from female students to study and do homework, which decreased substantially in intervention households but increased slightly in comparison households. Overall, while these results suggest that there may have been some minor changes to children’s participation in household labor in intervention households as compared to control households, there appears to have been little significant effect on either the total amount of labor performed by children or the gendered division of that labor. It is worth noting, however, that the division of labor within a household is based on a variety of factors which do not tend to change drastically over short periods of time, such as social norms and economic status. We thus do not expect to see a large variation in these indicators over just a year.

Table 126: Time taken for household tasks and impact on schooling, difference-in-differences analysis

s	Comparison Households			Intervention Households			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	181	286		244	339			
Whole day	0.6%	2.8%	2.2	0.8%	0.3%	-0.5	-2.8	.003**
Half day	14.4%	9.8%	-4.6	16.4%	5.0%	-11.4	-6.8	0.15
Quarter day	58.6%	62.2%	3.6	60.2%	64.6%	4.4	0.7	0.93
An hour a day	12.7%	15.2%	2.5	19.9%	23.3%	3.4	0.9	0.90
Does not do chores	13.8%	5.2%	-8.6	7.4%	6.2%	-1.2	7.4	0.10
Late often	2.2%	0.7%	-1.5	0.4%	0.6%	0.2	1.7	0.14

Late sometimes	14.9%	10.1%	-4.8	13.8%	7.1%	-6.7	-2.0	0.60
Never late	82.9%	88.5%	5.6	85.8%	92.0%	6.2	0.7	0.87
Limits studying	18.6%	13.0%	-5.6	18.3%	7.8%	-10.5	-4.9	0.36
Female								
n	185	284		229	358			
Whole day	0.0%	0.4%	0.4	0.0%	0.3%	0.3	-0.07	0.87
Half day	21.1%	9.5%	-11.6	15.7%	8.4%	-7.3	4.2	0.44
Quarter day	55.1%	58.1%	3.0	62.4%	66.8%	4.4	1.4	0.83
An hour a day	17.3%	17.5%	0.2	25.4%	20.7%	-4.7	-4.9	0.46
Does not do chores	6.5%	6.0%	-0.5	4.4%	3.4%	-1.0	-0.5	0.86
Late often	1.1%	1.1%	0.0	1.3%	0.8%	-0.5	-0.4	0.71
Late sometimes	9.0%	10.2%	1.2	16.4%	7.8%	-8.6	-9.7	0.08
Never late	89.9%	88.7%	-1.2	82.3%	91.3%	9.0	10.2	0.07
Limits studying	9.3%	9.5%	0.2	17.5%	9.0%	-8.5	-8.8	0.04*

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At baseline, the report further analyzed whether perceptions of the division of labor between girls and boys differ by the gender of the caregiver. This analysis found that, in general, male caregivers viewed boys as doing most household tasks at higher rates than claimed by female caregivers; male caregivers viewed boys as spending more time on daily tasks, while female caregivers viewed girls as spending more time; and male caregivers were more likely to report that household tasks made their child late to school if that child was a boy, while female caregivers' perceptions were similar for boys and girls. Unfortunately, at midline, only 32 caregivers were male (2% of all caregivers), and only seven male caregivers had female grade 2 children. Due to this very low sample size, the midline analysis of the gendered division of labor could not be further disaggregated by the gender of the caregiver, as any results would be highly susceptible to bias from low sample size.

Within the qualitative data, parents mentioned that chores detracted from studying time for both boys and girls. However, both mothers and fathers frequently mentioned that girls often had less time to study due to chores than boys. For example, a mother from Ermera municipality stated:

*At home, girls and boys are different. Sometimes the boys focus more on their study because they are relaxed and free as they don't have work to do in the house. However, the girls have to do a lot of work in the house, for instance, prepare vegetables, do dishes and other [chores], so they can only study in the evening.*²²⁸

The qualitative data does not suggest a major gap in perceptions of the time girls and boys spend on chores by the gender of the respondent, although these findings are not definitive.

HOUSEHOLD DECISION-MAKING

Caregivers in the household survey and the farmer's group booster survey were asked who makes decision regarding children's eating and hygiene practices. At baseline, the survey only included options for "Myself," "Myself and my spouse," "Myself and/or my spouse in consultation with elders,"

²²⁸ FGD with mothers, female, Ermera municipality, Int. 138

and “Elders/grandparents.” At midline, these options were expanded to include “My husband [spouse] alone.” While this expanded list of choice options provides for better understanding of the gender- and age-related dynamics of household decision-making, the change in choices from baseline to midline precludes statistically robust analysis of change in these indicators over time. As such, we present a summary of midline findings below, but do not conduct a difference-in-differences analysis.

At midline, the vast majority of respondents reported that either they themselves or they in conjunction with their spouse made decisions on their child’s or baby’s eating and hygiene practices (Table 127). Only a small minority of respondents reported that decisions were made with elders involved or that their spouse alone made decisions about the child’s or baby’s eating and hygiene practices. Similar results were found at baseline, with over 95% of caregivers reporting that either they alone or they and their spouse were involved in household decision-making.

Table 127: Household decision-making at midline

	Comparison Households	Intervention Households
Decision on child's eating practices		
n	625	982
Myself alone	48.8%	46.3%
Myself and my spouse	46.2%	45.4%
Myself and/or spouse with elders	1.4%	2.0%
Elders	1.8%	0.9%
Spouse alone	1.8%	3.3%
Decision on child's hygiene practices		
n	625	982
Myself alone	51.4%	60.9%
Myself and my spouse	45.6%	40.7%
Myself and/or spouse with elders	1.6%	1.7%
Elders	0.8%	0.9%
Spouse alone	0.6%	0.9%
Decision on baby's eating practices		
n	50	87
Myself alone	60.0%	70.1%
Myself and my spouse	36.0%	27.6%
Myself and/or spouse with elders	2.0%	1.1%
Elders	0.0%	1.1%
Spouse alone	2.0%	0.0%

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The farmer’s group booster sample included more male respondents than the household survey, allowing for more substantial analysis by respondent gender. However, we note that within the farmer’s survey, respondents were not necessarily caregivers, and thus may be less likely to be involved in the care of children due to their role in the household. Disaggregating by caregiver gender within the household survey alone suggests that at midline, male caregivers were slightly more likely to report making household decisions alone than female caregivers: 53% of male caregivers and 48% of female caregivers reported making decisions on their child’s eating practices alone, and 56% of male caregivers and 54% of female caregivers reported making decisions on their child’s hygiene practices alone. However, when respondents from the farmer’s group booster sample are also included in the analysis, male respondents were significantly less likely to make household decisions than female

respondents: Only 31% of male respondents reported making decisions about what children should eat alone compared to 49% of female respondents, and only 30% of male respondents reported making decisions about children's hygiene practices alone compared to 55% of female respondents. These results—along with the overall low number of men who are caregivers to children—suggest that there is still a strong gendered component to caregiving. In most households, women appear to be primarily responsible for the care of children, including feeding and hygiene, while men either take a supplementary role where they are consulted about decisions by their partner or are not involved in caregiving at all.

CONTROL OF PRODUCTIVE ASSETS

Caregivers were asked who controls decision-making related to productive assets, including large and small household purchases, loans, gardens, produce²²⁹ and livestock, and household businesses. At midline, in general, it was more likely for caregivers to report making decisions with limited monetary impact—such as making a small household purchase or selling a chicken—themselves, and to make decisions with greater monetary impact—such as a large purchase or selling livestock—jointly with their spouse. Compared to baseline, for most productive assets, caregivers were somewhat more likely to report making decisions jointly with their spouse at midline and slightly less likely to report making decisions themselves.

Table 128 shows, however, that there was little significant change in control of productive assets, as measured by caregivers' reported involvement in decision-making, between baseline and midline across intervention and comparison groups. The only significant change in decision-making occurred for small businesses, for which intervention households were 9 percentage points less likely than comparison households to report that their spouse made the decision.

Table 128: Control of productive assets, difference-in-differences analysis

s	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Large purchases								
n	378	625		482	982			
Myself	10.1%	7.2%	-2.9	11.6%	12.3%	0.7	3.6	0.32
My spouse	25.4%	18.2%	-7.2	28.6%	24.2%	-4.4	2.8	0.63
Myself and my spouse	57.9%	63.5%	5.6	50.6%	53.1%	2.5	-3.2	0.59
Others	6.6%	11.1%	4.5	9.2%	10.4%	1.2	-3.2	0.4
Small purchases								
n	378	625		482	982			
Myself	52.1%	50.9%	-1.2	50.2%	56.5%	6.3	7.5	0.34
My spouse	6.9%	4.2%	-2.7	11.6%	8.8%	-2.8	-0.1	0.95
Myself and my spouse	38.1%	41.3%	3.2	33.8%	30.0%	-3.8	-7.0	0.34
Others	2.9%	3.6%	0.7	4.4%	4.7%	0.3	-0.4	0.85
Sell chicken								
n	363	625		448	982			
Myself	21.2%	21.3%	0.1	25.2%	22.8%	-2.4	-2.5	0.56
My spouse	15.4%	14.6%	-0.8	17.4%	15.4%	-2.0	-1.2	0.78

²²⁹ In the household survey, only 26 respondents were asked who makes decisions regarding the sale of produce at midline, so this indicator is not included in the analysis.

Myself and my spouse	59.5%	57.9%	-1.6	52.7%	52.8%	0.1	1.7	0.72
Others	3.9%	6.2%	2.3	4.7%	9.0%	4.3	2.0	0.46
Sell livestock	BL	ML	Difference	BL	ML	Difference	DiD	p
n	370	625		447	982			
Myself	7.3%	6.6%	-0.7	10.3%	10.8%	0.5	1.2	0.68
My spouse	18.4%	18.2%	-0.2	22.4%	22.5%	0.1	0.3	0.95
Myself and my spouse	66.5%	65.1%	-1.4	58.6%	56.0%	-2.6	-1.2	0.84
Others	7.8%	10.1%	2.3	8.7%	10.7%	2	-0.3	0.94
Start business	BL	ML	Difference	BL	ML	Difference	DiD	p
n	335	625		401	982			
Myself	14.3%	14.6%	0.3	21.0%	22.5%	1.5	1.3	0.74
My spouse	9.0%	12.3%	3.3	16.5%	11.2%	-5.3	-8.6	0.03*
Myself and my spouse	73.1%	60.6%	-12.5	60.0%	50.0%	-10.0	2.6	0.6
Others	3.6%	12.5%	8.9	2.5%	16.3%	13.8	4.7	0.3

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In addition to the questions detailed in Table 128, respondents were also asked about garden-related decision-making. At baseline, all caregivers were asked to state who make decisions about large and small gardens. At midline, however, farmers who had received training on keyhole gardens and permagardens were asked who makes decisions about these gardens. As a result, the baseline and midline data are not directly comparable. However, overall, Table 129 shows that at both baseline and midline, the majority of respondents stated that both they and their spouse made decisions related to the garden, followed by just the respondent themselves or just the respondent's spouse.

Table 129: Garden-related decision-making

	Baseline		Midline	
	Small garden	Large garden	Keyhole garden	Permagarden
n	775	751	62	41
Myself	23.1%	13.7%	22.6%	19.5%
My spouse	18.1%	20.5%	19.4%	9.8%
Myself and my spouse	54.2%	59.7%	46.8%	53.7%
Others	4.6%	6.1%	11.2%	17.0%

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At midline, respondents who participated in VSLAs were also asked who in their family made the final decision about how to use the VSLA loan. Consistent with the results thus far, the majority of respondents (51%) stated that they and their spouse together had made the decision, followed by just the respondent themselves (33%) or just the respondent's spouse (13%).

Disaggregating by caregiver gender suggests that at midline, male caregivers were much more likely to report making decisions about, for example, large purchases or the sale of livestock by themselves than female caregivers.²³⁰ Forty-nine percent of male respondents reported that they themselves have

²³⁰ As above, we note some comparability issues across the farmer's group booster sample and household survey: Respondents to these questions in the household survey were caregivers, while respondents in the farmer's survey were not necessarily caregivers.

the final say on large household purchases compared to only 7% of female respondents, 44% of male respondents reported themselves to have the final say on the sale or consumption of livestock compared to only 6% of female respondents, 39% of male respondents reported themselves to have the final say on the sale or consumption of chickens compared to 21% of female respondents, and 43% of male respondents reported themselves to have the final say on starting a business compared to 18% of female respondents. This gap was reversed for decisions regarding small household purchases, for which 36% of male respondents and 56% of female respondents reported having sole decision-making power. These results are suggestive of a gendered gap in decision-making power over productive assets, whereby men often make the primary decisions that have major implications for household finances or food security, and women are often limited to making decisions that have smaller financial implications.

For female caregivers, we further analyze results by the age of the caregiver to better understand the potential intersection between age and gender as it relates to the control of productive assets.²³¹ We find that among all female midline respondents, older respondents were significantly more likely than younger respondents to make decisions about the sale/consumption of a chicken or livestock by themselves. Furthermore, among respondents in intervention municipalities, older respondents were significantly less likely to report that their spouse made the decision to sell or consume livestock by themselves and significantly more likely to report making decisions about large purchases or businesses themselves. These results suggest that it may be more beneficial to target young women (or households with young, female caregivers) for activities seeking to improve women's control of productive assets, as young women may face the greatest deficits in decision-making power.

CLAIMING RIGHTS AND MEANINGFUL PARTICIPATION IN PUBLIC DECISION-MAKING

At midline, heads of household were asked about household members' participation in VSLAs and caregivers were asked about whether they had received training on keyhole gardens or permagardens and whether they received support from agriculture extension services (AES). We note that at baseline, questions were phrased differently as program activities had not yet taken place—for example, respondents were asked if they would be able to attend an agricultural training, rather than whether they had received training, and were asked if they have access to AES, rather than if they receive support from AES (a more difficult criteria to meet). Additionally, at baseline, respondents were not asked about participation in VSLAs. As such, we do not make comparisons over time for these indicators, but provide an analysis of findings for midline respondents.

At midline, among all respondents, men reported receiving support from AES at slightly higher rates than women—19% compared to 15% respectively. However, both men and women were equally likely to state that they had not received support from AES (75% for both genders), but women were more likely to respond that they did not know if they had received AES support. Similarly, for garden training, men and women were equally likely to state that they had not received training (73% for both genders), but men were more likely to respond that they had received training (20%) while women were more likely to state that they did not know if they had received training (20% responded “don't know” and 7% responded that they had received training). In comparison, at baseline, men also reported

This issue is less salient for control of productive assets than for household decision-making, but may still affect results, as respondents to the farmer's survey may be more likely to be heads of household.

²³¹ We control for caregiver education in this analysis, as older respondents may be more likely to have higher levels of education which may affect results.

somewhat higher rates of access to AES, membership in a farmer's group,²³² and ability to attend an agricultural training.

Table 130 disaggregates results by comparison and intervention groups for respondents who stated affirmatively that they had received support or training. While there were no statistically significant differences in support from AES for male and female respondents, female respondents in intervention households were significantly less likely to have received garden training than male respondents. This result points to a potential gender imbalance in provision of training, which may be more frequently provided to male farmers than female farmers.

Table 130: Access to AES and garden trainings by gender, midline respondents

	Comparison Households			Intervention Households		
	Male	Female	p	Male	Female	p
Support from AES	14.3% (n = 14)	16.4% (n = 438)	0.70	20.0% (n = 105)	14.5% (n = 690)	0.31
Garden training	6.7% (n = 15)	1.8% (n = 610)	0.60	21.5% (n = 107)	9.3% (n = 875)	.003**

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Respondents were also asked about the number of total and female household members participating in a VSLA. Around 38% of households kept their savings in a VSLA; intervention households were around twice as likely to participate in VSLAs as comparison households. Among households with members participating in VSLAs, on average, one to two household members participated in a VSLA. Women were more likely to participate in a VSLA than men; on average, midline respondents reported that 63% of the household members participating in a VSLA were female. Furthermore, 44% of households reported that all VSLA members were female, while only 19% of households reported that all VSLA members were male. These results suggest that VSLAs may be effectively targeted towards women as compared to men; this dynamic may help increase women's financial independence and ability to claim rights.

CONTROL OVER ONE'S BODY

Caregivers and respondents to the farmer's group booster survey were asked whether a husband is justified in beating his wife given four circumstances—if she goes out without telling him, if she neglects the children, if she argues with him, and if she burns the food. Among respondents who were asked about at least one scenario,²³³ 41% believed that a husband was justified in beating his wife in at least one scenario at midline, compared to 55% at baseline. At baseline, comparison households were slightly less likely to say that spousal violence was justified in at least one situation; at midline, in contrast, comparison households were slightly more likely to say that violence was justified in at least one situation than intervention households.

Looking at the difference-in-differences analysis (Table 131), intervention households saw a slight but insignificant improvement in attitudes towards spousal violence compared to comparison households for some scenarios. At midline, intervention households responded less positively towards statements that a husband was justified in beating his wife if she goes out without telling him or if she burns the food than would be expected given the results of comparison households. They also were less likely

²³² At midline, respondents were not asked if they were a member of a farmers' group; the HATUTAN program refers to these groups as VSLAs, not farmers' groups. Responses to questions asking about membership in farmers' groups and VSLAs are not directly comparable across baseline and midline, however, and we therefore only analyze changes in support from AES and garden training.

²³³ Respondents were not asked these questions if another person was present; 564 respondents were asked at least one question at midline and 860 at baseline. Only 29 male caregivers/respondents were asked at least one of these questions at midline; as such, answers are not disaggregated by caregiver/respondent gender.

to respond positively to at least one statement about spousal violence than would be expected given the results of comparison households. However, intervention households were slightly—though again, insignificantly—more likely to agree that a husband was justified in beating his wife if she argues with him at midline compared to the change over time in comparison households.

Table 131: Scenarios in which beating one's wife is justified, difference-in-difference analysis

s	Comparison Households			Intervention Households			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD ²³⁴	p
She goes out without telling him	42% (n = 226)	30.7% (n = 202)	-11.3	40.2% (n = 271)	22.4% (n = 348)	-17.8	-4.7	0.6
She neglects the children	47.6% (n = 225)	30.9% (n = 204)	-16.7	49.1% (n = 277)	30.0% (n = 353)	-19.1	0.0	1.00
She argues with him	31.3% (n = 227)	31.7% (n = 202)	0.4	24.3% (n = 272)	25.2% (n = 349)	0.9	2.6	0.62
She burns the food	17.9% (n = 223)	16.3% (n = 202)	-1.6	19.1% (n = 273)	11.8% (n = 347)	-7.3	-4.9	0.27
At least one of the above	53.0% (n = 232)	43.9% (n = 205)	-9.1	56.1% (n = 287)	39.3% (n = 359)	-16.8	-7.7	0.35

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Overall, these quantitative results suggest that fewer respondents believed that a husband was justified in beating his wife in most scenarios at midline, except if she argues with him. However, the similarity of results across comparison and intervention households suggests that this improvement in attitudes towards spousal violence is the result of a broader change in social norms or attitudes towards gender-based violence, rather than a result of any specific program interventions. Further analysis by the age of the caregiver finds no significant relationships between age and attitudes towards gender-based violence among midline respondents; this suggests that any change in attitudes may not be generational, but rather a change occurring among respondents of all ages.

Within the qualitative data, respondents who acknowledged that gender-based violence did occur in some households stated that it often occurred because of food insecurity. For example, a father in Ermera municipality stated:

Domestic violence happens when food is not available and can trigger fights within the household; our children cry because they have no food, we may beat them and spouses can also beat one another.²³⁵

A mother in Ermera municipality similarly described how food insecurity and poverty could lead to violence:

[Violence] could potentially happen here. Because when you have unmet essential needs, it could happen. People are stressed because their local produce are not sold

²³⁴ A relatively large percent of respondents said they “don’t know” to questions about spousal violence. As these “don’t know” answers are included in the calculation of percentages but not included in the regression analysis, the difference-in-differences figures may not exactly match the percentages reported on the left of the table. Notably, at midline, around 8% of respondents to all questions stated that they “don’t know” if violence is justified in the given scenario; at baseline, respondents only answered “yes” or “no.”

²³⁵ FGD with fathers, male, Ermera municipality, Int. 132

*out, they don't have food. So it can cause rift among husbands and wives and lead them to arguments or fights.*²³⁶

However, many respondents denied that violence occurred within households, or stated that it only occurred with parents beating their children, rather than spousal violence.

VIOLENCE AND RESTORATIVE JUSTICE

Caregivers were asked about the ways in which teachers handle misbehavior in class. They were asked about both positive or neutral forms of discipline, including giving a verbal warning, informing parents, and having a conversation with the child, and negative forms of discipline, including shouting at the child, using corporal punishment, and assigning chores. At midline, 49% of caregivers reported that teachers gave verbal warnings, 46% that teachers had a conversation with the child, 42% that teachers shouted at the child, 31% that teachers used corporal punishment, 22% that teachers informed parents, and 15% that teachers assigned chores to the child. Overall, among all respondents at midline, caregivers were slightly less likely to state that teachers use corporal punishment or have a conversation with the child compared to baseline, substantially less likely to state that teachers assign chores to the child or inform the parents compared to baseline, and slightly more likely to state that teachers give a verbal warning or shout at the child compared to baseline.

Table 132 shows that at both baseline and midline, caregivers of male children in grade 2 were significantly more likely to say that teachers use corporal punishment than caregivers of female children. Accordingly, at both baseline and midline and among both treatment and comparison groups, caregivers of both male and female children were more likely to say that corporal punishment was justified against boys than against girls. Additionally, Table 132 shows that at midline, caregivers of male children were also significantly more likely to say that teachers shout at students than caregivers of female children. Otherwise, there were no significant differences in the discipline practices reported by caregivers of male and female children.

Table 132: Teachers' use of positive and negative discipline practices, by gender of student

	Baseline			Midline		
	Male	Female	p	Male	Female	p
n	434	426		682	677	
Gives verbal warning	49.3%	44.4%	0.12	48.1%	49.0%	0.7
Informs parents	32.3%	27.9%	0.15	22.1%	21.6%	0.80
Has conversation with child	48.6%	45.5%	0.35	44.4%	47.4%	0.29
Shouts at the child	39.9%	39.9%	0.99	44.7%	38.7%	0.03*
Uses corporal punishment	36.9%	30.3%	0.04*	33.6%	28.4%	0.04*
Assigns chores	28.6%	26.3%	0.50	14.8%	14.3%	0.79

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 133 displays the difference-in-differences regression results and shows that intervention households were significantly more likely to state that teachers give a verbal warning at midline compared to the results expected given comparison households. Indeed, 4 percentage points fewer comparison households stated that teachers gave students verbal warnings at midline than at baseline, while 6 percentage points more intervention households stated that verbal warnings were used at midline than at baseline. There were no other significant changes in the discipline practices reported by intervention households compared to comparison households.

²³⁶ FGD with mothers, female, Ermera municipality, Int. 138

Table 133: Change in teachers' positive and negative discipline practices

s	Comparison Households			Intervention Households			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
n	378	625		482	734			
Gives verbal warning	51.6%	47.5%	-4.1	43.2%	49.5%	6.3	10.4	0.04*
Inform parents	29.9%	22.2%	-7.7	30.3%	21.5%	-8.8	-1.1	0.88
Has conversation with child	48.2%	46.2%	-2	46.3%	45.6%	-0.7	1.3	0.87
Shouts at the child	39.4%	44.3%	4.9	40.3%	39.5%	-0.8	-5.6	0.31
Uses corporal punishment	30.4%	30.1%	-0.3	36.1%	31.7%	-4.4	-4.0	0.50
Assigns chores	28.8%	13.9%	-14.9	26.4%	15.1%	-11.3	3.7	0.39

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

We further analyze the reported use of discipline practices by whether the household reports their second grade child to have a physical or mental disability. We find no significant correlation between physical disability and discipline practices at midline, though we note that few households reported their child to have a physical disability. However, we do find significant relationships between mental disability and reported discipline practices.²³⁷ In intervention households that reported their child to have at least one mental/emotional disability at midline, caregivers were significantly more likely to report that teachers shouted or assigned chores to the student and less likely to report that teachers had a conversation with the student. In control households, caregivers of mentally disabled students were more likely to report the use of corporal punishment against their child. These results suggest that disabled students may face amplified challenges in schools; it is worth further considering the intersection between disability, gender, and power in program activities that seek to address violence and restorative justice.

These findings are particularly pertinent because caregivers who report that teachers use corporal punishment or shout at children are also significantly more likely to report that their child is afraid to go to school or avoids school. Controlling for student gender and school fixed effects, reported use of corporal punishment is associated with a 10 percentage point greater likelihood that the student is afraid to go to school and a 5 percentage point greater likelihood that the student avoids school. Teachers shouting at students is associated with a 9 percentage point greater likelihood that the student is afraid to go to school and a 5 percentage point greater likelihood that the student avoids school. This suggests that the use of negative discipline practices may affect students' desire to attend schools.

Within the household survey, caregivers were also asked who they could report abuse to if their grade 2 child was abused or harassed at school. Table 134 shows that the majority of caregivers stated that they would report the abuse to the head teacher; a large portion also answered "other," of which the majority of respondents specified that they would report to the teachers. Notably, at baseline, 9% of respondents answered that they would not be able to report abuse; this percentage increased at midline, to almost 15%. This suggests that there may remain barriers to accessing restorative justice for children abused at school.

In addition to the household survey, school directors were asked about avenues for reporting abuse at the toilets or abuse by a teacher within the school survey. The majority of directors at both baseline

²³⁷ Students are classified as having a mental/emotional disability if they were reported to have an issue with at least one of memory, self care, communication, anxiety (monthly or more frequently), or depression (monthly or more frequently).

and midline stated that abuse could be reported to the director/coordinator of the school or to a teacher (Table 134). Only a very small percent of directors stated that abuse could not be reported. The contrast in results across the household and school survey suggests that while directors may believe that there are effective avenues for reporting abuse and obtaining justice, caregivers do not necessarily agree or are not necessarily aware of these avenues for reporting.

Table 134: Reporting of abuse at school

	Baseline	Midline
Household survey: Reporting abuse of child at school		
n	860	1,359
Head teacher	79.3%	67.2%
Police	1.9%	1.8%
Social services	0.5%	0.2%
Local authorities	3.4%	2.3%
Cannot report	9.4%	14.9%
Other	9.8%	19.8%
School survey: Reporting abuse of child in toilets		
n	185	186
Family/relatives	10.3%	6.5%
Director/coordinator	45.4%	46.8%
Teacher	33.0%	26.3%
PTA	2.7%	9.1%
Cannot report	0.5%	0.0%
Other	8.1%	8.0%
Don't know or did not respond	0.0%	3.3%
School survey: Reporting abuse of child by teacher		
n	184	186
Family/relatives	16.9%	7.0%
Director/coordinator	44.6%	50.0%
Teacher	19.6%	22.6%
PTA	1.6%	7.5%
Cannot report	0.5%	0.5%
Other	16.8%	8.6%
Don't know or did not respond	0.0%	3.8%

ASPIRATIONS AND STRATEGIC INTERESTS

Caregivers were asked about whether boys and girls were equally skilled at math and reading as well as a variety of question about whether girls and boys had equitable experiences at schools. At midline, caregivers were given the response option “don’t know,” while this option was not included in the baseline. “Don’t know” was selected quite frequently at midline, comprising over 20% of all responses. Due to the inclusion and frequency of selection of this response at midline, baseline and midline results are not directly comparable. We thus do not include a difference-in-differences analysis of change over time; we present only an analysis of midline results disaggregated by the gender of the student.

At midline, the majority of caregivers stated that they thought boys and girls had the same capacity for reading and writing and for math. However, caregivers were more likely to state that girls have more capacity for reading and writing or math than boys: 18% of respondents stated that girls have more capacity than boys for reading and writing and 15% stated that girls had more capacity for math.

In general, the vast majority of respondents either believed that schools treated students equitably by gender (as measured by encouragement of students and students’ abilities to ask questions or get

help in class) or stated that they were unsure; on average, these two responses accounted for around 90% of all responses to questions about who was encouraged to participate, able to ask questions, and able to ask for help in class at midline. Table 135 disaggregates by the gender of the caregiver's child, and shows that in general, caregivers of male and female grade 2 students had similar views on the capacity of girls and boys. At baseline, caregivers of female students were significantly less likely to state that girls and boys had equal capacity for reading and writing and significantly more likely to state that boys had more capacity for reading and writing; there were no other significant differences.

Table 135: Caregivers' beliefs in the capacity of male and female students and school equity, by gender of child

s	Baseline			Midline		
Reading skill	Male	Female	p	Male	Female	p
n	377	375		682	677	
Equal	75.3%	67.7%	.03*	56.6%	52.7%	0.19
Girls better	18.3%	22.9%	0.20	16.0%	20.2%	0.06
Boys better	6.4%	9.3%	.04*	6.2%	5.2%	0.44
Neither	0.0%	0.0%	-	0.3%	0.0%	0.17
Don't know	-	-	-	21.0%	21.9%	0.66
Math skill	Male	Female	p	Male	Female	p
n	376	373		682	677	
Equal	73.9%	72.1%	0.45	54.0%	51.1%	0.20
Girls better	16.8%	17.7%	0.73	14.4%	15.5%	0.55
Boys better	9.0%	9.9%	0.68	8.2%	8.1%	0.95
Neither	0.3%	0.3%	1.0	1.0%	0.9%	0.61
Don't know	-	-	-	22.4%	24.4%	0.29
Encouraged to participate	Male	Female	p	Male	Female	p
n	376	361		682	677	
Equal	93.4%	93.9%	0.79	64.5%	65.1%	0.81
Girls more	2.7%	3.9%	0.39	4.0%	4.0%	0.98
Boys more	2.9%	1.9%	0.44	3.2%	3.1%	0.89
Neither	1.1%	0.3%	0.29	0.9%	1.2%	0.36
Don't know	-	-	-	27.4%	26.6%	0.66
Able to ask questions	Male	Female	p	Male	Female	p
n	350	348		682	677	
Equal	90.9%	87.6%	0.18	55.1%	55.8%	0.79
Girls more	5.7%	9.5%	0.06	5.7%	5.8%	0.97
Boys more	2.9%	2.6%	0.75	3.7%	2.4%	0.27
Neither	0.6%	0.3%	0.58	1.8%	1.6%	0.89
Don't know	-	-	-	33.7%	34.4%	0.74
Able to get help	Male	Female	p	Male	Female	p
n	360	351		682	677	
Equal	92.2%	92.3%	0.96	62.5%	62.8%	0.89
Girls more	4.2%	4.8%	0.54	4.4%	4.4%	0.98
Boys more	3.1%	2.0%	0.30	2.6%	2.8%	0.80
Neither	0.6%	0.9%	0.64	1.5%	0.9%	0.40
Don't know	-	-	-	29.0%	29.1%	0.97

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At baseline, it was found that a greater percentage of male caregivers believed that girls were more skilled at reading, writing, and math than boys, while female caregivers were more likely to say that boys were more skilled in reading, writing, and math. We find broadly similar results at midline: A higher percentage of female caregivers said boys are more skilled at reading, writing, and math than male caregivers, and a smaller percentage of female caregivers said girls are more skilled at math than male caregivers. However, in contrast to baseline findings, a higher percentage of female caregivers than male caregivers also said that girls are more skilled at reading and writing. In other words, at midline, female caregivers were more likely than male caregivers to believe that one gender—either boys or girls—was innately more skilled at reading, writing, or math.

Additionally, at baseline, male and female caregivers generally perceived boys' and girls' experiences at school to be equal. In contrast, at midline, male caregivers were more likely to believe that girls were more able to ask questions and ask for help in class than boys. However, we note as above that due to very low sample size for male caregivers at midline, these results should not be taken as conclusive.

Qualitative data suggests that many respondents view female and male students differently, both in terms of their capacity to learn and their behavior within the classroom. While respondents generally said that both boys and girls may struggle to learn in school, multiple respondents stated that girls are more focused on their learning materials, that girls learn faster than boys, or that girls spend more time studying at home while boys spend more time playing. For example, a school coordinator stated:

*The boys' capacity to absorb lessons is less than that of the girls. The girls absorb lessons faster than boys when we teach them. When we write a letter in the blackboard and ask them the letter's name or ask them to read they can do it correctly, better than boys. The boys are more interested in playing around, chasing each other back and forth; while the girls always listen when we talk.*²³⁸

A mother in Ermera municipality similarly stated:

*Girls and boys are different because the girls read more than the boys. It is because the boys spend more time playing. The girls are focused more on their learning materials that teachers gave to them, however the boys take their leisure times more and return home in the late evening. We ask the boys to read, they don't listen.*²³⁹

However, multiple respondents also mentioned that girls tended to be more shy than boys, which could make them difficult to teach:

*Some of the boys, not all, are smart and easy to teach. But girls tend to be quiet, so it is difficult for us to teach them if they are always quiet. We don't know if they know or not.*²⁴⁰

Boys, in contrast, were frequently described as “naughty,” poorly behaved, or only interested in playing, rather than studying:

Most of the time boys are hard-heads, we can say they are roguish or naughty while girls are interested to learn because they are sensitive, we know everywhere girls are

²³⁸ FGD with school coordinators, female, Ermera municipality, Int. 129

²³⁹ FGD with mothers, female, Ermera municipality, Int. 138

²⁴⁰ FGD with mothers, female, Ainaro municipality, Int. 124

*always diligent. But boys are hard-headed, they just mingle together with their peers. What could we do, some listen to us, but after one hour they started misbehaving again by playing with rubber bands, slingshot, and other games. Girls, as I just said, I always accompany them. Girls usually study in groups together, read together.*²⁴¹

While respondents did not explicitly mention a difference in the engaging teaching practices used with girls and boys—for instance, encouraging girls more than boys—it is worth noting that many respondents mentioned that it was important to engage and encourage shy students. Given that respondents tend to view girls as more shy than boys, this dynamic may mean that some teachers may tend to encourage girls more in class than boys.

Caregivers were also asked the maximum level of education they could support for their grade 2 child. At both baseline and midline, the majority of caregivers stated that they could support their boys and girls through university (at baseline, 80% and 77% respectively; at midline, 73% and 70% respectively). These responses likely reflect social desirability bias rather than caregivers' actual ability or intent to support their children's education; as a result, this is not necessarily a true picture of the differences in households' abilities and desires to support and girls through school.

PREDICTIVE ANALYSIS

PREDICTORS OF LITERACY

In this section, we test the relationship between various student- and school-level characteristics and the overall literacy score. We first look at individual-level variables, such as attendance and age, and then analyze variables that are measured at the school level, such as class size. As at baseline, we use hierarchical linear modeling (HLM) to determine the extent to which student-level and school-level characteristics determine variability in reading scores. HLM is a form of ordinary least squares regression that is used to analyze variance in outcome variables—in this case, literacy scores—when the predictor variables (our student- and school-level characteristics) are at varying hierarchical levels.²⁴² In our case, because the literacy scores of students in a classroom vary according to their common teacher, classroom, and school, HLM is an appropriate approach. Indeed, using HLM, it was determined that at midline, 11% of the variation in students' overall literacy scores was a result of factors at the school level, while at baseline, 26% of the variability in scores was a result of their school. These findings indicate, as expected, that students' literacy scores are in part related to school-level factors, such as school resources or teacher quality.

At baseline, the model included eight student-level predictors identified as being statistically significant predictors of literacy scores using stepwise regression: age, gender, working memory, caregiver education level, preschool attendance, school absences, whether the student reads at home, and whether a toilet is available at home (a proxy for student health and hygiene). At midline, the stepwise regression analysis identified five of these variables as still significant: age, gender, working memory, caregiver education level, and school absences. In addition, whether a student had studied at home during COVID-19 school closures and the caregiver's nutrition level (a proxy for household nutrition status) were also identified as significant.²⁴³ In the analysis below, we include all ten predictors

²⁴¹ FGD with teachers, male, Ermera municipality, Int. 110

²⁴² See Heather Woltman et al., "An introduction to hierarchical linear modeling," *Tutorials in Quantitative Methods for Psychology* 8 (1): 2012, 52-69.

²⁴³ The variables included in the stepwise regression analysis were student age, student gender, student language, working memory, caregiver education level, number of household members, preschool attendance, school absences, whether the student

identified in either the baseline or midline as statistically significant in the model.²⁴⁴ In our first model, which seeks to identify individual-level predictors of literacy, we also include school fixed effects—an HLM technique. These variables control for any variation which occurs at the school level and which does not change, or changes at a constant rate, over time. This model does not allow us to understand which specific school-level factors predict literacy; however, it is the most robust model available to understand individual-level predictors. We run our models using data from all students assessed using the household survey at baseline and midline.

Figure 22 shows the results of the regression analysis with school fixed effects (Table 136 presents a summary of all regression results). Age and gender were both significant predictors of literacy scores with substantial effect sizes at both baseline and midline, as well as in the model including all cross-sectional students: Older students and female students tend to have better overall literacy scores. For all three models, working memory scores and caregiver education were also significant predictors of overall literacy scores, though with smaller effect sizes: Better working memory and a more highly-educated caregiver are correlated with higher overall literacy scores. At baseline, whether the student reads at home was also a significant predictor of literacy scores, with a positive and substantial effect size. At midline, preschool attendance was a significant predictor of literacy, with a positive though small effect size for students who had attended preschool. For the model including all students in the cross-sectional cohort, preschool attendance, student attendance,²⁴⁵ toilet availability at home, and dietary diversity of the caregiver were all significant and positive predictors of overall literacy, though all had very small effect sizes.

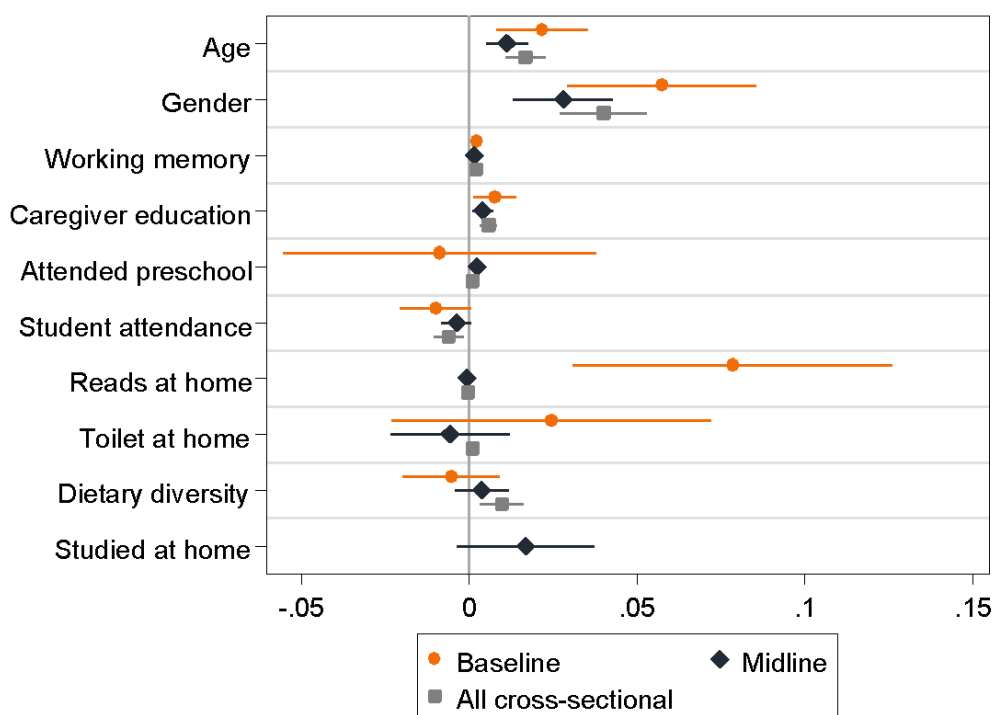
Overall, among the cross-sectional cohort, we find similar effect sizes for most variables—i.e., the coefficients of most variables in the baseline, midline, and full cross-sectional models fall within the confidence intervals of the coefficients in each other model. The one variable for which this is not true is whether the student reads at home. This variable had a large and significant effect on literacy scores at baseline but was not significant for the models run for just midline or all cross-sectional students. The reason for this change is unclear; it is possible that if students gained greater access to reading materials outside of the home at midline, the importance of reading at home for overall literacy scores may have been reduced. It is also possible that the COVID-19 pandemic reduced the impact of reading at home on overall literacy; because of school closures, students' literacy skills may have been too weak to benefit significantly from reading at home.

reads at home, whether the student has a physical disability, whether the student has a mental disability, whether a toilet is available at home, proximity to school, whether someone helps the student read at home, caregiver nutrition, and whether the student studied at home during COVID-19 school closures.

²⁴⁴ The variable for whether a student studied at home is only included in the model for midline students, as it only affected students at midline.

²⁴⁵ Note that the student attendance variable was measured by recording the number of absences of the student per week; a coefficient below zero thus implies that students with more absences have worse overall literacy scores, or, conversely, that student attendance and overall literacy scores are positively correlated.

Figure 22: Predictors of literacy, individual level

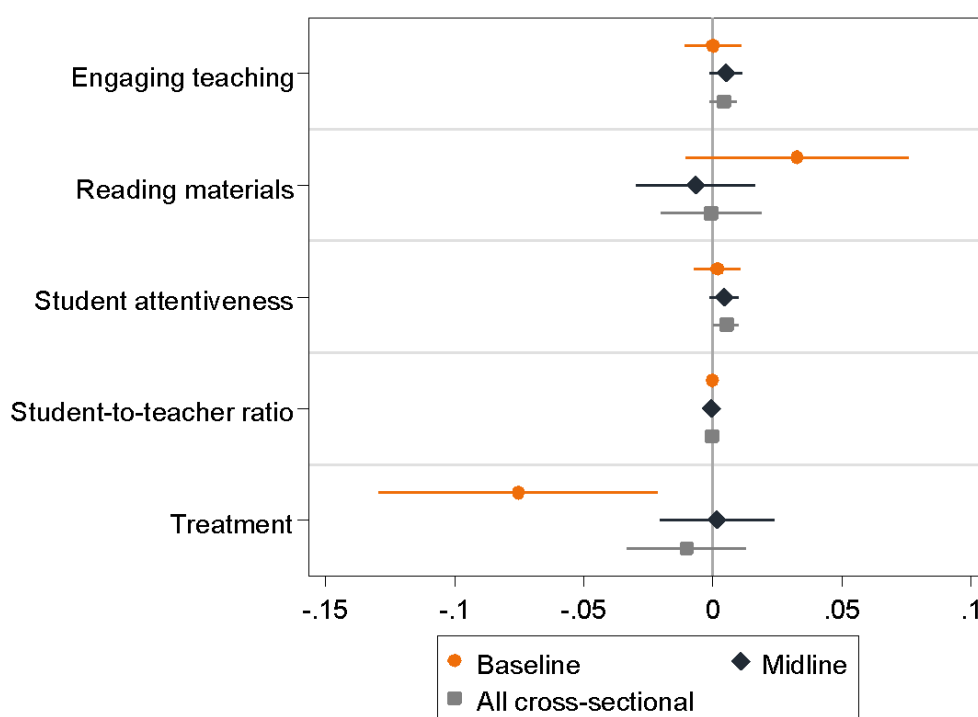


In order to better understand the specific school-level factors which may predict literacy scores, we now run a model which includes four school-level variables: availability of books/magazines, student attentiveness, engaging teacher practices, and grade 2 student-teacher ratio.²⁴⁶ The model further controls for intervention/control group status. This model still includes individual-level variables to control for student-specific variation that influences literacy scores, but does not include school fixed effects, as these would be colinear with the school-level predictors of interest. As a result, this model is less rigorous for determining individual-level predictors; we thus only report the values of school-level variables below.

Figure 23 shows that, in general, most school-specific predictors were not found to be significant at baseline, midline, or for all baseline and midline cross-sectional students. At baseline, students in the treatment group were, on average, less likely to score well on overall literacy than students in the comparison group; at midline and among all cross-sectional students, there were no longer any significant differences between intervention and comparison groups' scores. Among all cross-sectional students, student attentiveness was also found to be a significant and positive predictor of overall literacy, although the effect size was small. Otherwise, no significant school-level predictors were found.

²⁴⁶ At baseline, a variable was included for grade 2 enrollment. This is replaced by grade 2 student-teacher ratio for the midline analysis, as this variable—a proxy for class size—is a more relevant predictor of literacy scores.

Figure 23: Predictors of literacy, school level



We also run a predictive analysis for only grade 3 and 4 students—those in the panel sample at midline. Grade 2 students, and particularly those in the cross-sectional sample who were affected by COVID-19 school closures, have limited school exposure and literacy abilities; this may thus affect the impact of some predictors. For example, access to reading materials will not have an effect on the reading ability of a student who has not attended school long enough to identify letters, but may have a more significant impact on students in later grades with more mastery of foundational skills. At midline, the household survey was not conducted for students in the panel sample, and classroom observations were only conducted in grade 2 classrooms. We therefore use an adjusted set of variables collected for grade 3 and 4 students. Individual-level predictors include student age, gender, working memory score, native language (Tetum or non-Tetum), while school-level predictors include school-wide student-to-teacher ratio, whether the school has a literacy project, whether students at the school can borrow books, and whether the school provided meals to students the day of data collection.

Controlling for school fixed effects, we find all four individual-level predictors to be significant. Older students, female students, students with higher working memory scores, and students whose native language is Tetum have significantly higher overall literacy scores, all else held constant. Looking at school-specific predictors, we also find that the existence of a literacy project and the ability of students to borrow books have significant and positive impacts on overall literacy scores. These results suggest that access to reading materials and literacy projects may indeed have a more substantial impact for students with a stronger grasp of foundational skills.

Table 136 below presents a summary of all regression results for predictors of literacy. As described above, the most substantial predictor of literacy scores is gender; among all cross-sectional students included in the regression analysis, on average, female students scored 4 points higher on the overall literacy assessment than male students, all else held constant. Student age was also a substantive predictor; one additional year of age corresponds to a 1.7 point higher overall literacy score, all else held constant. These results are both consistent with expectations: Within Timor-Leste, female

students tend to outperform male students, and in general, older students tend to perform better in school due to additional exposure to learning environments, either at home or at school. While many other variables—particularly at the individual level—were also significant predictors of literacy, none had an effect size near that of age or gender.

Table 136: Summary of predictors of literacy results

	Baseline		Midline		All cross-sectional		Grade 3 and 4	
	Individual-level variables							
	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p
n	685		1,228		1,924		1,984	
Age	2.2	.002**	1.1	.001**	1.7	<0.001***	1.7	<0.001***
Gender	5.7	<0.001***	2.8	<0.001***	4.0	<0.001***	11.9	<0.001***
Working memory	0.2	<0.001***	0.1	<0.001***	0.2	<0.001***	0.4	<0.001***
Caregiver education	0.8	.02*	0.4	.01**	0.6	<0.001***	-	-
Attended preschool	-0.9	0.71	0.2	<0.001***	0.1	.003**	-	-
Student attendance	-1	0.07	-0.4	0.11	-0.6	.009**	-	-
Reads at home	7.8	.001**	-0.08	0.18	-0.03	0.46	-	-
Toilet at home	2.4	0.32	-0.6	0.53	0.1	<0.001***	-	-
Dietary diversity	-0.5	0.47	0.4	0.34	1.0	.004**	-	-
Studied at home	-	-	1.7	0.11	-	-	-	-
Native language	-	-	-	-	-	-	3.7	0.02*
	School-level variables							
	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p
n	425		1,231		1,656		1,944	
Engaging teaching	0.02	0.97	0.5	0.11	0.4	0.11	-	-
Reading materials	3.3	0.13	-0.6	0.58	-0.05	0.96	-	-
Student attentiveness	0.2	0.68	0.5	0.1	0.5	.03*	-	-
Grade 2 student-to-teacher ratio	-0.004	0.94	-0.04	0.3	-0.04	0.25	-	-
Treatment group	-7.5	.007**	0.2	0.88	-1.0	0.39	-	-
School student-to-teacher ratio	-	-	-	-	-	-	-0.2	0.12
Literacy project	-	-	-	-	-	-	4.6	0.03*
Lend books	-	-	-	-	-	-	0.07	0.004**
School feeding	-	-	-	-	-	-	-4.2	0.06

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

PREDICTORS OF ENGAGING TEACHING PRACTICES

As at baseline, a linear regression model was used to determine predictors of the use of engaging teaching practices. The outcome variable measured the number of engaging teaching practices used during the classroom observation (a maximum of nine, as described in the section “Quality of Instruction”). Predictor variables included teacher gender, teacher education (secondary school or greater than secondary school), teacher experience (in years), class size, availability of reading materials, whether the school had electricity, whether the school had a PTA, and whether the director provides coaching to teachers.

Table 137 shows the results of the regression analysis for just baseline data, just midline data, and baseline and midline data combined. At baseline, class size and availability of reading materials at

school were both significant predictors of engaging teaching practices. Larger class sizes were correlated with less use of engaging teaching practices, while access to reading materials was correlated with more engaging teaching practices. At midline, there were no significant predictors of engaging teaching practices. However, in the regression including all baseline and midline data combined, class size was found to be a significant predictor of engaging teaching practices, again with a negative coefficient.

Table 137: Predictors of engaging teaching practices

	Baseline		Midline		All data	
	Coefficient	p	Coefficient	p	Coefficient	p
n	138		185		323	
Gender	0.52	0.16	0.40	0.22	0.45	0.07
Education	0.81	0.05	0.25	0.44	0.28	0.25
Experience	-0.02	0.66	0.03	0.41	0.008	0.76
Class size	-0.03	.03*	-0.01	0.32	-0.02	.03*
Electricity	0.46	0.23	0.27	0.41	0.40	0.11
Reading materials	1.1	.004**	-0.38	0.25	0.18	0.45
PTA	-0.30	0.67	1.1	0.12	-0.05	0.94
Director coaching	0.23	0.66	-0.35	0.38	0.14	0.65

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

These findings suggest that teachers of large classes find it more difficult to use engaging teaching practices. To further analyze this finding, we run a simple regression on the use of each of the nine engaging teaching practices and class size. We find that class size is significantly and negatively correlated with teachers engaging students who are not participating in class and with teachers soliciting student opinions. In large classes, teachers may find it more difficult to engage quiet or inattentive students or solicit students' opinions due to the difficulty of managing large numbers of students and tracking which students are or are not participating.

We also analyze the specific engaging teaching practices correlated with availability of reading materials and find that access to reading materials is significantly and positively correlated with the use of a reading corner in class. This result is unsurprising; schools without reading materials would be expected not to have reading corners, and teachers in these schools would therefore be unable to use a reading corner to engage students.

PREDICTORS OF STUDENT ATTENTIVENESS

A linear regression analysis was used to test factors that predict student attentiveness. Separate regressions were run for each student attentiveness indicator (as described in the "Student Attentiveness" section). As well as the variables listed in the tables, the model included treatment group and round as controls. Each table shows the coefficient and p-value for each variable included.

Table 138 shows the results of the regression with self-reported student attentiveness as the outcome variable. The only significant predictor was whether the student had eaten that day, which had a positive association with attentiveness. This is expected and indeed confirms the assumption in the log frame: A student that is hungry is likely to have more difficulty concentrating and paying attention.

Table 138: Predictors of self-reported attentiveness

	Coefficient	p
School served food today	0.01	0.25
School purchased local produce	-0.006	0.62
PTA responsible for school feeding	0.006	0.58
Student had eaten today	0.08	<0.001***
Member of student's household went without eating	0.003	0.81

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

Table 139 shows the results with working memory score as the outcome variable. The only significant predictor was whether a school had purchased local produce or not, which indicates a positive association with working memory score. Causality for this relationship is unclear; this could be another proxy indicator for school meals, and thus for student hunger, an indicator for greater availability of local vegetables and thus better dietary quality, or a proxy for the economic status of a community, as families in wealthier communities may have been more likely to contribute money to the SFP thus enabling those schools to purchase local produce. Additionally, whether the student lived in household where someone had gone a day without eating in the past 30 days (an indicator of inadequate food intake) predicted a lower memory score and was significant at the 10% level.

Table 139: Predictors of working memory scores

	Coefficient	p
School served food today	-1.2	0.38
School purchased local produce	3.5	0.009**
PTA responsible for school feeding	-1.5	0.17
Student had eaten today	2.1	0.14
Member of student's household went without eating	-2.2	0.09

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

Table 140 below shows the results of the regression using observed student attention as the outcome variable. There were three significant predictors: More students paying attention was predicted by whether students were reading to themselves, working in groups, or playing games. In contrast, some practices classified as strong teaching practices, such as asking open questions or asking for opinions, did not have an effect on the number of students paying attention. However, traditional teaching practices—including students copying from the board and repeating the teacher—also had no significant relationship with the number of students paying attention in class. Overall, these results suggest that engaging teaching practices do tend to increase student attentiveness, but that teaching practices are not the only factor that affects attentiveness in class.

Table 140: Predictors of observed student attentiveness

	Coefficient	p
Teacher gender	-0.7	0.09
Teacher education higher than secondary	-0.4	0.28
Teacher experience in school	-0.04	0.35
Teacher total experience	0.09	0.06
Students copy from board	0.3	0.45
Students repeat after teacher	-0.3	0.50
Students read by themselves	1.3	0.005**
Students participate in group work	1.0	0.02*

Teacher reading aloud	0.0	1.0
Teachers using reading corner	0.6	0.23
Teacher uses games	1.4	0.003**
Teacher asks open questions	0.1	0.83
Teacher asks for opinion	-0.1	0.78
Teacher engages non-participating students	-0.4	0.45
Teacher encourages boys	-0.8	0.18
Teacher encourages girls	0.03	0.96
Teacher questions boys	0.04	0.95
Teacher questions girls	1.0	0.18
Teacher gets angry at boys	-0.6	0.28
Teacher gets angry at girls	-0.3	0.71

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

PREDICTORS OF STUDENT ATTENDANCE

A linear regression analysis was used to test factors that predict school absences. As well as the variables listed in the table below, the model included treatment group and round as controls. The outcome variable was the number of absences the previous week, as reported by the primary caregiver at the household survey. The table below shows the coefficient and p-value for each variable included.

Table 141: Predictors of student absences

	Coefficient	p
Student age	0.026	0.37
Student gender	-0.061	0.37
Spent at least half a day on chores	0.23	0.03*
Suffered from serious illness	0.13	0.31
Had difficulty with self-care	0.19	0.01*
Within 30-minute walk	-0.15	0.05
Walk to school was safe	-0.087	0.33
Improved sanitation in school	0.045	0.57
Improved sanitation in household	-0.16	0.02*
Experienced anxiety	0.15	0.03*
Experienced depression	0.13	0.07

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

Within the household survey, respondents were asked how much time the student was required to spend on household chores. The options were (1) the student does no chores, (2) a whole day, (3) a half-day, (4) a quarter day, or (5) a little time. Spending a whole day or half-day counted as spending a large amount of time on chores, while spending less was classified as not spending a long time on chores. The analysis finds that spending at least half a day on chores is a significant predictor of student absences, with students who spend a long time on chores missing on average 0.23 more days a week than those who do not, all else held constant. This result suggests that children's participation in household labor may detract from their ability to succeed in school; this is particularly problematic given that household labor has gendered dynamics, with girls tending to have more household responsibilities than boys. Notably, this gendered division of labor may not only affect academic achievement, but also leadership and organizational skills that children obtain through play and free

time, as well as health from higher stress levels due to household responsibilities and the need to manage time.

Several mental/emotional disabilities are also significantly correlated with school absences. Parents/caregivers were asked to rate on a scale of 1-4 how much difficulty the student had with self-care such as washing themselves, with 1 meaning no difficulty and 4 meaning they cannot do so at all. Any answer other than 1 was classified as a student having difficulty with self-care. This described approximately 30% of students, and the analysis here suggests they miss 0.19 more days of school per week. However, students showing anxiety at any time of year missed 0.15 fewer days per week than students who did not show anxiety; there may be some reverse causality driving this result, with students who have to balance both schoolwork and (potentially) housework having more anxiety due to higher levels of responsibility. Overall, mental health may play a significant role in school attendance, and thus also in educational outcomes. Provision of psychosocial support to vulnerable households or students with disabilities may therefore be an effective way to improve attendance and literacy.

Lastly, a household was classified as having improved sanitation if it had a pit latrine with a slab or with ventilation, a composting latrine, or a flush toilet. No toilet or an uncovered pit latrine was classified as not having improved sanitation. Households with improved sanitation missed 0.16 fewer days on average. This result is likely due to the relationship between improved sanitation and health: Students in households with improved sanitation are likely to have fewer illnesses and thus better attendance.

PREDICTORS OF HEALTH KNOWLEDGE

This section analyzes the relationship between various characteristics of households pertaining to health, hygiene, and nutrition and their hygiene and nutrition knowledge, dietary diversity scores, and COVID-19 prevention behavior. The results framework identifies “increased use of health, nutrition and dietary practices” as a McGovern-Dole priority outcome. However, the household survey mostly collected information on health, hygiene, and nutrition knowledge rather than health behaviors. COVID-19-related behaviors were collected in the midline, but behavior change cannot be determined because COVID-19 was not present during the baseline. It is important to note that change in behavior is not measured during the analysis because change in knowledge does not imply change in behavior. However, given the available data, hygiene and nutrition knowledge are used as a proxy.

Linear regression analysis was conducted to identify predictors of hygiene and nutrition knowledge, dietary diversity, and COVID-19 prevention behavior. Moreover, predictors of diet diversity within households were also identified through the same analysis. The tables presented in each sub-section show the coefficients and p-values for each predictor. Caregiver’s highest level of education is consistently a statistically significant predictor of hygiene and nutrition knowledge and dietary diversity in the baseline, but only for dietary diversity in the midline.

PREDICTORS OF HYGIENE KNOWLEDGE

The regression for hygiene knowledge in the midline included five variables at both baseline and midline: caregiver’s gender, caregiver’s highest level of education, toilet available at home, handwashing station availability, and handwashing with soap. An additional COVID-19-related variable was included at midline.²⁴⁷ Hygiene knowledge scores range from 0 to 19, which corresponds with the number of healthy hygiene practices the respondent identified during the household survey. While the caregiver’s level of education was found to be a significant predictor at baseline in the preliminary stepwise regression analysis, in the midline analysis, which included additional COVID-

²⁴⁷ When running stepwise regression on Stata with the full model for the midline (i.e., including COVID-19-related predictors), caregiver level of education was omitted from the model, while decision-maker on child hygiene practices was omitted in both baseline and midline models.

related variables, it was no longer found to be a significant predictor in either the preliminary or final models. A potential explanation for this is that caregivers with higher levels of education had more knowledge of healthy hygiene practices prior to the implementation of the program and program activities may have equalized hygiene knowledge, where those with a lower level of education improved their knowledge of hygiene practices.

Table 142: Predictors of hygiene knowledge

	Baseline		Midline	
	Coefficient	p	Coefficient	p
n	525		1,359	
Caregiver highest level of education	0.09	0.009**	0.01	0.38
Caregiver gender	0.59	0.09	1.44	<0.001***
Toilet available at home	0.41	0.13	0.22	0.04*
Handwashing station available at home	-	-	0.23	0.003**
Handwashing with soap to prevent spread of COVID-19	-	-	0.13	0.13

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

In the baseline, there was a statistically significant difference in the ability to name healthy hygiene practices between caregivers with higher levels of education compared to those with lower levels of education. The effect size between education levels was small, while controlling for all variables, with an even smaller and non-statistically significant effect in the midline. At baseline, as caregivers gain an additional level of education,²⁴⁸ for example from no education to primary education, their hygiene knowledge score increases by an average of 0.09 points, all else held constant.

Caregiver's gender may influence hygiene knowledge because mothers are more likely to take the lead on decision-making on child hygiene practices. However, we caution against overinterpretation of these results due to the low sample size of male caregivers at midline.

In the midline, availability of toilets and handwashing stations was a statistically significant predictor of the number of healthy hygiene practices that respondents can identify. There was a statistically significant effect on the ability to name healthy hygiene practices between households with a toilet and a handwashing station at home and those without. This could be explained by households with a toilet or handwashing station having more access to resources and knowledge of hygiene and being aware of the importance of toilets and handwashing stations in maintaining good hygiene.

PREDICTORS OF NUTRITION KNOWLEDGE

The regression for nutrition knowledge included four variables:²⁴⁹ caregiver's highest level of education, number of household members, household dietary diversity score, and use of loans for food. Nutrition knowledge scores range from 0 to 15, which is the number of nutrition or dietary recommendations that respondents identified. Caregiver's gender was added to the stepwise regression but was removed from the final regression due to lack of significance and co-linearity at midline, despite appearing as a significant predictor of nutrition knowledge at baseline.

Table 143: Predictors of nutrition knowledge

	Baseline		Midline	
	Coefficient	p	Coefficient	p
n	126		90	

²⁴⁸ Levels of education were divided into no education, primary education, pre-secondary education, secondary education, and university.

²⁴⁹ Using the stepwise regression, variable on using savings for food was omitted in both baseline and midline models.

Caregiver highest level of education	0.24	0.03*	0.17	0.23
Number of household members	-0.29	0.01*	0.32	0.02*
Household dietary diversity	0.58	0.03*	-0.59	0.07
Use loans for food	0.89	0.13	1.2	0.09

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

At the baseline, all of the variables except using loans for food were statistically significant predictors of nutrition knowledge. The effect size of varying education levels was small in the baseline and even smaller in the midline, holding all variables constant. Analysis also shows that household size statistically significantly and negatively affects nutrition practices at the baseline, implying that prior to exposure to program activities, the smaller the household size the better knowledge the respondents have about nutrition practices. However, at the midline, household size statistically significantly and positively affects nutrition practices. Household size positively affects nutrition knowledge in both comparison and treatment groups at the midline, which implies that treatment households may already have existing knowledge of healthy nutrition practices prior to the implementation of program activities. At the baseline, household size of comparison households negatively and significantly affects nutrition knowledge; however, in treatment households, the effect was positive but not significant.

There is also a statistically significant difference in the ability to identify nutrition practices between households with varying dietary diversity at the baseline. Dietary diversity is determined by the number of food types that the household consumed the previous day. The more food types the household consumed, the more nutrition or dietary recommendations they identified, which may be an indication that they have seen the positive nutrition outcomes of maintaining a diverse and balanced diet. However, the opposite is true at the midline. The dietary diversity scores of households, particularly of women of childbearing age and infants 6-23 months old, decreased between baseline and midline. As a result, it may have become necessary for households to practice nutrition recommendations that they are aware of to make up for nutrition loss.

PREDICTORS OF DIETARY DIVERSITY

The regression for household dietary diversity included three variables:²⁵⁰ caregiver's highest level of education, number of household members, and number of food groups consumed the previous day. The household dietary diversity scores range from 0 to 9 and refer to the number of food groups that the household consumed the previous day.

Table 144: Predictors of dietary diversity

	Baseline		Midline	
	Coefficient	p	Coefficient	p
n	860		1,359	
Caregiver highest level of education	0.06	<0.001***	0.05	<0.001***
Number of household members	0.009	0.52	0.02	0.04*
Number of dietary or nutrition recommendations identified	-0.02	0.16	0.04	<0.001***

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Caregiver level of education statistically significantly predicted a household's dietary diversity, with an effect size of 0.06 at baseline and 0.05 at midline. Household size and knowledge of nutrition practices are also statistically significant predictors of dietary diversity in the midline. A plausible assumption would be that dietary diversity would be lower as household size increases because resources will have to be spread more thinly over larger households. However, the resulting regression model counters that assumption. On the other hand, the better knowledge a household has about nutrition

²⁵⁰ Using the stepwise regression, the variable referring to a household not having food for at least a day in the past 30 days was omitted.

practices, the more food types they consume, which may imply that they are somewhat aware of how to maintain a diverse and balanced diet. This indicates that households with better nutrition knowledge translate that into better nutrition practices through higher dietary diversity, demonstrating a link between knowledge and behavior.

PREDICTORS OF COVID-19 PREVENTION BEHAVIOR

The regression for COVID-19 prevention behavior included four variables: caregiver's highest level of education, number of COVID-19 prevention practices known, number of healthy hygiene practices identified, and availability of a handwashing station at home. Respondents may identify up to ten COVID-19 prevention practices. COVID-19 prevention behavior refers to the prevention measures that respondents said they practice. This is only applicable to the midline.

Table 145: Predictors of COVID-19 prevention behavior

	Midline	
	Coefficient	p
n	1,359	
Caregiver highest level of education	0.005	0.55
Number of COVID-19 prevention practices known	0.87	<0.001***
Number of healthy hygiene practices identified	0.001	0.97
Handwashing station available at home	0.19	<0.001***

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The number of COVID-19 prevention practices known and availability of a handwashing station at home statistically significantly predict people's COVID-19 prevention behavior. This implies that they have been exposed to some form of communication or information related to COVID-19, particularly how to limit its transmission. People are more likely to practice COVID-19 prevention behaviors if they are given information of COVID-19 prevention precautionary measures. Having a handwashing station at the household level also encourages people to wash their hands frequently; most respondents (64%) in the midline said that they wash their hands most times or always. Caregiver's level of education and knowledge of healthy hygiene practices have very limited and statistically insignificant influence on people's COVID-19 behavior.

LEARNING AGENDA DISCUSSION

The McGovern-Dole Learning Agenda aims to answer questions related to school meal program implementation and education. The HATUTAN program-specific learning agenda also focuses on questions related to literacy, health, nutrition, agriculture, gender-based violence, and sustainability. Below, we discuss the learning agenda for education (including literacy), health, nutrition, agriculture, and gender-based violence. Sustainability is addressed in the later section on program implications. Findings are focused on the results of the treatment group in order to best inform program implementation.

Throughout this report, we have noted the substantial impact that COVID-19 had on outcomes of interest, particularly education, nutrition and food security, health, and gender-based violence. Lessons learned at midline have also been strongly affected by the pandemic, as program activities have adapted to new conditions, new activities have been added to the program, and some activities have been delayed.

EDUCATION AND LITERACY

Both the McGovern-Dole and HATUTAN program learning agendas seek to understand the effects of school meal interventions on student literacy and education. Key learning agenda questions include the following:

1. In what ways do the combination of school meal interventions and educational interventions improve education and literacy levels? How can these combinations improve cost-effectiveness?
2. In what ways do school meal interventions impact the resources of teachers, such as classroom time, teacher incentives, and teacher capacity?
3. What are the effects of school meal interventions on the acquisition of reading skills in multilingual settings?

EFFECT OF SCHOOL FEEDING ON LITERACY

During COVID-19 induced nationwide school closures, no measures were taken to compensate for the disruption of school feeding, and HATUTAN school feeding activities were delayed or altered. Most students were out of school from March until July, and thus did not have exposure to any school feeding interventions—or to education outside of the *Eskola ba Uma* program—for a substantial portion of the school year. Delays in the approval of the national budget also delayed provision of SFP funds to schools; as a result, the SFP was not operational in most municipalities for most of the year. During school closures due to COVID-19, the HATUTAN program was able to provide students with take-home meal rations from the remaining commodities intended for SFPs.

Despite these challenges, during midline data collection, there was a significant improvement in the provision of school meals both during the day of data collection and the week before data collection in intervention schools. At midline, 89% of intervention schools reported providing meals to students the day of data collection and 84% the week before data collection. In contrast, only 33% of comparison schools reported providing meals to students the day of data collection and 26% the week before data collection. This difference is particularly stark when considered in contrast to baseline results, in which only 1% to 2% of intervention schools reported providing meals on either the day of or the week before data collection. Overall, while the school feeding intervention did not operate all days as intended due to COVID-19 in 2020, it still had clear successes at improving the number of meals for students.²⁵¹

Just as COVID-19 impacted the provision of school meals, it also affected student learning. As discussed at length in the “Literacy Results” section, grade 2 students’ literacy scores were substantially worse at midline than at baseline due primarily to lengthy school closures and ineffective at-home learning. However, our difference-in-differences regression analysis in this section suggested that the HATUTAN program did, indeed, have some positive effect at mitigating the impact of COVID-19 on learning—suggesting that, in response to Learning Agenda question 1 listed above, a combination of school meal interventions and educational interventions can help to improve education and literacy levels. This is evidenced by the fact that while average scores declined for students in both the intervention and comparison groups, scores for intervention students declined by less than would be expected given the results of comparison students.

Interestingly, we find that the program may have had a heterogeneous effect on students’ literacy scores for the worst-performing students—those who have no literacy ability and scored zero on the letter recognition subtask—compared to students with some literacy ability. At midline, overall, more students were “left behind” entirely and were unable to read a single letter or word. However, within intervention schools, there were fewer students “left behind”—i.e., with no ability to read letters—than

²⁵¹ HATUTAN provided schools meals for 10 out of the 12 planned weeks (January - March 2020). Students collected the remaining food stock for ten days of school meals as take home rations in May 2020.

would be expected given the results of students in comparison schools. In contrast, students with some ability to read letters appear to have benefitted less from the HATUTAN program, with little substantial effect on scores for these students in intervention schools compared to comparison schools.

In contrast, for invented word, familiar word, and passage reading, there were no significant changes in the percent of students unable to perform these tasks within the treatment group as compared to the control group, but the average non-zero score improved for treatment students relative to comparison students. This suggests that for word recognition, the HATUTAN program was more effective at improving skills among students with some prior ability, rather than for students with the lowest levels of literacy.

Overall, the implications of these findings for the learning agenda (and specifically Learning Agenda questions 1 and 3 listed above) are two-fold. First, the heterogenous effects on letter fluency for zero and non-zero scorers suggest that HATUTAN program activities and the school feeding program may be relatively more effective at reducing the number of students with no letter recognition ability, rather than improving the letter recognition skills of students who already have some ability in this area. In this case, it seems that the HATUTAN program had more of an impact on more disadvantaged students—those who were more likely to be “left behind” entirely—than on students who faced fewer challenges to learning. One possible mechanism for this effect is that school feeding increased enrollment of all students, including the most-disadvantaged students with the lowest base levels of literacy. As a result of changes to dropout rates, students with low base levels of literacy may have been able to gain some letter recognition skills—the most basic literacy task—but may have been unable to progress beyond this level of understanding given limited contact hours.

In contrast, the significant improvement in word recognition and passage fluency among non-zero scorers but lack of relative improvement in the percent of zero scorers may have been driven by the feeding program’s effects on attentiveness and memory. Among students with some word recognition ability, higher levels of attentiveness due to improved nutrition may have allowed for relatively improved learning outcomes in intervention schools for more difficult literacy skills. However, while students without any prior word recognition ability may still have benefitted from increased attentiveness, their low base skill level may not have allowed them to substantially improve their word recognition abilities. Overall, for both the most disadvantaged students and for students with more literacy ability, it seems that school feeding interventions were able to mitigate some of the negative impacts of COVID-19.

CHALLENGES TO IMPROVED LITERACY

In response to Learning Agenda questions 1 and 3 listed above, the midline results point to several specific challenges to improved literacy outside of the broader impact of the COVID-19 pandemic and issues facing the education system of Timor-Leste. At both baseline and midline, results suggested that students knew the names of letters relatively well—in general, when presented with a letter, far more students tended to name that letter correctly than incorrectly—but struggled with fluency, as they were not able to name very many letters within one minute. Most low overall scores were not due to students identifying letters incorrectly, but rather due to low reading speed, suggesting low levels of fluency.

There is also a clear gap between students’ abilities to recognize letters and their abilities to recognize words. This finding applies to both the cross-sectional (grade 2) and panel (grade 3 or 4) cohorts: Students have significantly more ability to recognize letters than to recognize words, especially invented words. The invented word task requires a strong understanding of phonemes in order to pronounce words that are unfamiliar and lack any meaning; this suggests that students may need more work in recognizing the sounds that specific letters and letter groups make. Similarly, with familiar words, average scores for grade 2 students in particular are extremely low, indicating that second grade students still lack adequate development of sound recognition, a fundamental skill needed for the recognition of both familiar and invented words.

The heterogeneous patterns seen in scores among the worst-performing students—those who have no literacy ability and scored zero on the letter recognition subtask—compared to students with some literacy ability also have implications for project strategy and activities beyond school feeding interventions. In particular, these results suggest that the HATUTAN programming may not be effectively improving teaching of phonemes and the relation between letter sounds and words to the weakest students, but may be improving these skills among students who already have some understanding of the concept. Students may benefit from future activities that seek to improve the transition from letter reading to word reading.

Notably, even among students who were able to read some invented and familiar words, 16% were unable to read a single word of the passage at both baseline and midline. For students who scored zero on passage fluency but greater than zero on both invented and familiar word fluency, scores on the invented and familiar word fluency subtasks were, on average, low, but not so low as to suggest that these students simply guessed words correctly but have no actual word recognition ability. Rather, these students seem to have some ability to read words in isolation, but struggle to read words in the context of a passage. These students could also struggle with attentiveness, making reading long passages more difficult. This pattern also reinforces the finding that students generally have low levels of fluency: Although they may be able to identify individual letters or words, they struggle to apply those basic skills to more difficult reading tasks.

Interestingly, at midline for the panel cohort, scores on the passage reading subtask exhibited both a ceiling and a floor effect, with a relatively high percentage of students scoring either 0% or 100%. This distribution of scores suggests that, at later grades, some students continue to get “left behind” entirely and have very little literacy ability. However, current teaching practices may be relatively successful at improving the reading skills of students who already have some literacy ability—thus resulting in a relatively high percentage of students who are able to read the entire passage.

Finally, unlike with other subtasks, HATUTAN programming had no significant effect on either the percent of students scoring zero on reading comprehension or on the reading comprehension of students scoring above 0%. It may be the case that HATUTAN programming does not sufficiently address reading comprehension skills. Alternatively, students may have too little general reading ability (in terms of letter and, especially, word recognition) for programming that targets reading comprehension to have a large impact.

EFFECT OF SCHOOL FEEDING ON TEACHERS

The McGovern-Dole learning agenda seeks to understand the impact of school meal interventions of the resources available to teachers, such as classroom time on task, teacher motivation, and teachers’ use of engaging teaching practices (Learning Agenda question 2 listed above). The program’s theory of change argues that classroom time on task may improve when students’ nutrition needs are met through school feeding; that teacher motivation may increase as teachers observe changes in students’ behavior; and that teachers may be more likely to see positive outcomes from engaging teaching practices when students are fed and able to pay attention, thus engendering a positive feedback loop whereby teachers increase their use of engaging practices.

The predictive analysis suggests that whether a student has eaten is, indeed, a significant predictor of student attentiveness. However, in general, we find few significant increases in student attentiveness or decreases in student hunger at midline in intervention areas compared to comparison areas, suggesting that the program has not yet had a substantial impact on attentiveness. It is worth noting, however, that the COVID-19 pandemic has substantially impacted household food security and the provision of school meals across the country; results may thus be affected by these dynamics.

As students become more attentive and engaged due to school feeding, teacher attendance may increase as teachers observe positive changes in students’ behavior and become more motivated. High levels of teacher attendance contribute to the quality of education by increasing the number of hours of instruction received by children. Furthermore, teachers who regularly attend classes may

have a better understanding of the needs of their students, thus allowing them to adjust lessons as necessary to improve learning outcomes.

At midline, average teacher attendance increased substantially among comparison schools; among treatment schools, day-of attendance increased but by less than comparison schools, while previous-day attendance decreased. The difference-in-differences regression analysis found that the relative improvement in attendance for comparison schools was large and significant; this result was not explained by teacher absences due to trainings. It is possible that the COVID-19 pandemic may have had effects on teacher motivation and thus attendance; if these effects were most salient in areas with the least resources—the intervention municipalities—then teacher attendance would have decreased at higher rates (or increased at lower rates) in these municipalities than elsewhere. Furthermore, the differential impact of the rainy season on intervention schools and the massive improvements in teacher attendance in several comparison municipalities that had very low attendance rates at baseline suggests that changes in teacher attendance were likely due to external factors, rather than the HATUTAN program. Overall, the unexpected direction of these findings and the large number of potential confounding variables make it difficult to draw a clear link between school feeding and teacher motivation/attendance.

While the use of engaging teaching practices did not decline at midline, there was no significant change in the total number of engaging teaching practices used. The difference-in-differences analysis suggested that there was also little significant change in the use of engaging teaching practices in intervention schools as compared to control schools, although there were some minor changes in the types of practices used. Among all engaging teaching practices, there was only a significant increase in teachers' use of games or exercises within treatment schools as compared to control schools at midline. While the use of some other teaching practices changed somewhat substantially in treatment schools as compared to control schools—such as asking open questions and use of the reading corner, which were observed 9 to 10 percentage points more frequently in treatment schools than expected given results in comparison schools—these results were not significant. However, there was a significant reduction in the use of traditional teaching practices in intervention schools.

Additionally, HATUTAN programming appears to have had little significant effect on teachers' use of questions and encouragement with boys and girls. There were no significant changes in gender-specific positive teaching practices among the treatment group at midline compared to the comparison group at midline. However, notably, encouragement of girls did increase rather substantially, although not significantly, among intervention schools.

Overall, these results do not show a clear link between school feeding and teacher capacity. However, particularly among longer-tenured teachers with more ingrained teaching habits, teaching practices may be expected to change rather slowly, and only as a result of continued training, feedback from school directors or other relevant professionals, and positive classroom results (as evidenced but higher levels of engagement from students). Endline findings may thus be more suggestive of any links between school feeding and the use of engaging teaching practices.

SCHOOL MEAL PROGRAM IMPLEMENTATION

The McGovern-Dole learning agenda aims to understand the community-level systems of governance and management that are required to successfully implement school meal programs. This relates to the HATUTAN program-specific learning agenda questions about the kinds of partnerships and exit strategies that are most effective at ensuring program sustainability. Specific Learning Agenda questions included in the Terms of Reference include the following:

1. What community-level systems of governance and management are required for the successful implementation and sustainability of school meal programs?
2. Which components of school meal programs, including food production, procurement, and preparation of meals, are the most sustainable in terms of operational efficiency and why? Does the cost-effectiveness of HATUTAN change over time?

3. What are the most effective methods for ensuring food safety within school meal programs taking into consideration the different systems of national, regional, local and community governance?
4. In what ways do school meal programs impact health equity in terms of poverty, gender, or geography?

COMMUNITY-LEVEL SYSTEMS

Of relevance to Learning Agenda question 1 above, most treatment schools in the baseline (69%) and the midline (89%) said that the director or coordinator was responsible for school feeding, whereas over a third of treatment schools (38%) said that PTAs had oversight of the feeding program at the baseline, which increased (55%) at midline. However, most treatment schools (82% baseline, 91% midline) said PTAs oversee the feeding program in some capacity, either exclusively or in collaboration with school staff.

Most of the respondents from treatment schools at the baseline (53%) and the midline (36%) reported not having a PTA meeting during the current school year. Another one-third of respondents at the baseline (34%) and one-fifth at the midline (22%) said that the last meeting was more than a month ago. Household surveys revealed most households (65% baseline, 63% midline) do not have a member who participates in the PTA, indicating limited participation and influence in activities involving school feeding. This may be a result of the lack of PTA meetings at the school and almost half (46%) of the PTAs having only one member, which implies limited functionality of the PTAs. However, most respondents said that their children's schools have a PTA (95% baseline, 99% midline) and that those PTAs are doing activities to improve school feeding (30% baseline, 72% midline).

Overall, participation of households in PTAs, as well as PTA membership and level of activity, remain low. This limits PTAs' ability to fully engage with parents and adequately provide oversight of the school feeding program.

FOOD PRODUCTION, PROCUREMENT, AND PREPARATION

Both the McGovern-Dole and HATUTAN program-specific learning agendas ask about the sustainability of meal program components, such as food production, local procurement, and food preparation (Learning Agenda questions 2 and 3 above). In terms of food preparation (Learning Agenda question 3), an improvement in the number of schools with kitchen space and access to clean water for food preparation increased. Kitchen space is available in most treatment schools (94% baseline, 97% midline) but not all treatment schools reported having access to clean water for preparing meals, which may raise a concern about sanitation. Only seven in ten intervention schools (69%) at the baseline reported having access to clean water for this purpose, but this increased to nine in ten intervention schools at midline (87%).

In relation to food production and local procurement (Learning Agenda question 2), more than half (55%) of the intervention schools that reported having a feeding program at the midline reported that their school does not buy local food from farmers for school feeding. The most common foods that intervention schools reportedly procured from local farmers for the school feeding at the midline were dark green vegetables (85%), vitamin A-rich foods such as pumpkins, carrots, or purple sweet potatoes (56%), and carbohydrate-rich foods such as potatoes, taro, or cassava (59%). At the midline, the primary reason given by most of the 47 intervention schools that do not purchase produce from local farmers was not having the budget to buy local produce.

STUDENT AND COMMUNITY HEALTH

The learning agenda aims to understand the impact of school feeding programs on student and community health (Learning Agenda question 4 above). McGovern-Dole specifically asks about health equity in terms of poverty, gender, and geography, while the HATUTAN program is interested in its potential to decrease health-related absences. For the HATUTAN program, the relevant Learning

Agenda question is as follows: “How effective are school based meals at preventing and/or decreasing health related absences?”

The study’s ability to measure the health of the community is limited by the data collected because it provides insight into knowledge and awareness of healthy practices but does not demonstrate healthy behaviors or health status. Proxies for community health include caregivers’ consumption of healthy food groups, ability to afford healthcare, and use of savings and loans for healthcare. Most of the caregivers in the four intervention municipalities did not mention consuming any food items belonging to the nine food groups contributing to the dietary diversity score, ranging from 64%-71% at the baseline and 62%-69% at the midline. Among those who mentioned consuming food from at least one of the food groups, most caregivers in treatment municipalities said that they consumed two (20% baseline, 35% midline) or three food groups (46% baseline, 35% midline) the previous day. A similar trend is observed between male and female caregivers at the baseline, where most of the caregivers consumed three food groups (41% of males, 46% of females), and at the midline, where most of them consumed two (32% of males, 35% of females) or three (32% of males, 36% of females) food groups. Male caregivers tend to eat more fruits and vegetables (22% baseline, 12% midline) than female caregivers (18% baseline, 6% midline). However, overall reductions in the consumption of fruits and vegetables were observed at the midline.

In terms of ability to afford going to the doctor or a clinic, most respondents at the baseline either said they cannot pay (37%) or they can sometimes pay (35%). However, at midline, most respondents said that there is no need to pay (55%), which was not a response option at the baseline. Given this limitation, the change in households’ ability to afford healthcare cannot be determined. Regarding the use of savings and loans for healthcare among treatment households, only a small number of respondents mentioned using their savings (15% baseline, 20% midline) and loans (15% baseline, 4% midline) for healthcare.

Additional relevant Learning Agenda questions relate to the impact of water and sanitation programs on learning and literacy outcomes: “How do WASH programs impact learning and literacy outcomes?” and WASH interventions are theoretically likely to affect learning and literacy by decreasing student illness and thus increasing attendance and attentiveness. Along these lines, regarding health-related absences, more than half of students (53%) were reported by caregivers in the treatment group to have missed at least one day of school in the last week due to sickness. Overall, at midline, students were less likely to miss school due to sickness: 78% of caregivers reported that their children did not miss school last week at the midline compared 48% of caregivers at the baseline. Female students were slightly more likely (24%) than male students (20%) to miss three or more days due to illness at the baseline, but this improved at the midline where households reported that their children missed school less and only 7% of boys and 7% of girls missed school in the last week. These findings show the likelihood of missing classes due to sickness has decreased at midline.

FOOD SAFETY AND WASH

One of the areas of focus of the McGovern-Dole learning agenda is on effective methods for food safety. Similarly, the HATUTAN program looks at WASH interventions. Effective school-level WASH interventions may contribute to ensuring the hygienic preparation and safety of food served to children. Relevant learning questions include the following:

1. How do WASH programs impact learning and literacy outcomes?
2. What are the best practices in sustaining WASH interventions delivered through a school meal project?

In relation to these questions, as reported earlier, kitchen space is available in most treatment schools (94% baseline, 97% midline) but only seven in ten schools (69%) at the baseline reported having access to clean water for food preparation. The number of schools that had clean water for food preparation increased at the midline (87%). At midline, 86% of intervention schools had at least one functional handwashing station in the school and 88% of comparison schools. Less than half of

intervention schools (41% baseline, 36% midline) reported having a handwashing station in their kitchen and only about one-third (39%) said they use detergent every day to clean the kitchen at the baseline, although this increased at midline (65%).

The baseline report noted that parents were concerned over food safety conditions in schools due to unhygienic food preparation methods that could cause sickness leading to children missing school. In the midline, however, most of the caregivers (85%) of children in treatment schools said that food served to children in school is prepared in a hygienic manner, which increased from the baseline (70%). The school and household survey data imply that there had been improvements in level of hygiene maintained in school during food preparation. However, it is not clear to what extent parents are aware of or certain about the hygienic preparation of food in schools.

In terms of maintaining clean storage spaces, about three-fourths of treatment schools at the baseline (73%) said that storage spaces were at least somewhat clean. However, the baseline report noted that results may have reflected a bias from the enumerators because photos of the locations suggest otherwise. At the midline, almost all the treatment schools (93%) said that storage spaces are clean. At the baseline, most treatment schools (73%) were reported to raise food off the ground using pallets, shelves, or another method; this increased at midline to 99%.

HEALTH, NUTRITION, AND AGRICULTURE

WASH INTERVENTIONS' IMPACT ON HEALTH KNOWLEDGE

In addition to the McGovern-Dole learning agenda questions, outside of the results described above, the midline study aims to further answer learning questions on how WASH projects impact learning (in addition to literacy) outcomes ("How do WASH programs impact learning and literacy outcomes?"). At the household level, the general trend observed is that there is an improvement in the percentage of caregivers that were able to identify healthy hygiene practices between baseline and midline in both the comparison and treatment groups. However, this is not a strong indicator of change in behavior but rather change in knowledge of healthy hygiene practices. Most caregivers in the treatment group were able to identify at least 17 out of 19 healthy hygiene practices (93%) at the midline, which was the standard set under the McGovern-Dole Custom Outcome #21. Most of the hygiene practices that respondents identified were related to handwashing, and predictive analysis indicates that having WASH facilities at home, such as a toilet and handwashing station, influences respondents' familiarity with healthy hygiene practices.

NUTRITION

The HATUTAN program learning agenda seeks to understand the extent to which school-based meals are effective in preventing and decreasing health-related absences. The key Learning Agenda questions are as follows:

1. How effective are school based meals at preventing and/or decreasing health related absences?
2. To what extent has Social Analysis and Action (SAA) been effective in addressing barriers to improved nutrition? Are there specific groups that benefitted more than others? Who are they and why did this happen?

In response to these questions, as noted in earlier sections, nationwide school closures without national measures taken to compensate for the loss of school feeding might dampen the effect on the outcome of interest. Students were mostly out of school from late March through July and there were delays in provision of SFP funds to schools throughout this period. However, at the time of midline data collection, schools in Timor-Leste had reopened; school feeding activities had also begun to resume, although SFPs generally did not fully resume until March or April except in intervention schools with HATUTAN programming. Additionally, we expect absences due to illness to be more

responsive in the short-term to school feeding than, for example, literacy scores. Time not spent in school will have a long-term effect on literacy scores, while the benefits of school feeding for school absences (if there are any) should become apparent much sooner.

Indeed, we can see a strong impact from the program on school feeding, suggesting that despite disruption from COVID the program was successful in implementing the school feeding program. There was almost no change in the percentage of comparison schools providing a meal to students between baseline and midline – about 30% provided meals at both rounds. However, there was a very large and significant increase in intervention schools providing meals, from 1% to 88%. The overall effect is somewhat biased due to lower than usual school feeding at baseline—data collection was carried out in different months for treatment and comparison—but correcting for this doesn't account for the increase seen. We can confidently conclude that the increase to nearly 90% providing meals reflects effective implementation of school feeding. Additionally, we can see a benefit in terms of student nutrition. In the cross-sectional analysis there was a positive difference in difference score for whether a student had eaten on the day of the survey that was not statistically significant, and in the panel analysis there was a greater difference in difference score that was statistically significant. Overall, we can conclude that the program was effective in providing school meals and that this led to improvements in student nutrition.

Most of the schools serving meals to students reported having a menu. Most school menus at the midline include carbohydrates, legumes and nuts, and dark green vegetables such as spinach, lettuce, and mustard greens. However, menus lack fruits. School menu dietary diversity is an important consideration in ensuring that nutritious and well-balanced meals are served to students. The school menu dietary diversity in comparison schools appear low at the midline, where only 30% of comparison schools mentioned one to three out of seven food groups. Among treatment schools at the midline, almost nine in ten (87%) mentioned food items that belong to one to three food groups. Dietary diversity among women of childbearing age and among children 6-23 months old is also low. Only 6% of children met the minimum acceptable dietary requirement of consuming food from at least four out of seven food groups, which is a slight improvement from 4% at the baseline. Fruit and vegetable consumption at the household level is also generally low, and men are more likely to consume more than women.

The next step is establishing whether there is a link between school feeding and nutrition indicators and health-related absences. The first thing to note is that disruption due to COVID-19 and other factors is likely to make analysis of the relationship more difficult, particularly when comparing treatment and comparison. Attendance fell for many grades, particularly in the treatment group. There were negative difference-in-difference scores for attendance rates, which was likely driven by treatment areas being harder-hit by natural disasters. Additionally, the analysis is complicated by a slight change in the question asking about health-related absences, with the recall period being changed from absences last month to absences last week.

To investigate a link, the days reported sick by the caregiver is regressed on whether the student is in a school that provided school meals on the day of the survey. Confining the analysis to the treatment group at midline, we find that being in a school serving meals was associated with 0.3 fewer days missed on average due to illness, although the result is not significant. Carrying out the same analysis but with whether a student had eaten as the predictor variable, we find that a student who had eaten on the day of the EGRA survey missed 0.4 days fewer on average the preceding week than a student who hadn't eaten, although again the result was not significant.

Overall, despite COVID-related disruptions in program implementation and data collection, there is strong evidence that the school feeding program was effective in increasing school meals served by schools, and that this had a positive effect on nutritional outcomes for students. Despite further complicating factors in analyzing health-related absences, we also find a link between school feeding and student nutrition as explanatory variables with fewer absences due to illness as the outcome variable, although the statistical significance was slightly weaker. We can therefore cautiously conclude that school meals seem to be effective in reducing absences due to illness.

AGRICULTURE

The HATUTAN learning agenda seeks related to agriculture seeks to understand the following:

1. How can a combination of local procurement during harvest time be supplemented with international food aid to promote locally and/or nationally sustainable school meals programs?
2. What kinds of partnerships with the private sector and/or host country governments are the most effective at ensuring project sustainability? Among successful partnerships, who are the key players and what are their roles? In what contexts do private sector and/or government partnerships work best and which contexts may be more challenging?

The extent to which these questions can be analyzed with quantitative data is limited, as we cannot assess the proportion of school meals produced with local produce or international aid. Additionally, quantitative indicators on sustainability are limited at present, although HATUTAN has developed a graduation and sustainability plan to measure progress. However, we can assess success in HATUTAN's school feeding program overall, and can also analyze the proportion of schools reporting that they procure local produce.

The success of the school feeding program has been reported in previous sections and will be briefly summarized here. While there was very little change in the percentage of comparison schools providing a meal to students between baseline and midline, there was a very large and significant increase in intervention schools providing meals, from 1% to 88%. Therefore, despite some bias due to baseline data for treatment and comparison being collected at different times of the year, we can confidently conclude that the program was effective in providing school meals. Additionally, there was a benefit in terms of student nutrition: Students in the treatment group were significantly more likely to report that they had eaten on the day of the survey. Therefore, the first main component of the learning agenda question can be answered positively: International food aid can be beneficial for school feeding programs.

However, at midline, a large percent of intervention schools—55%—reported that they do not purchase local produce, and only 11% of intervention schools reported that they purchase local produce “all the time.” There are likely multiple causes for this trend; intervention schools were chosen in part on the basis of greater deprivation and being more remote. If they were impacted to a greater extent by COVID restrictions, this might mean lesser availability of local produce and greater dependence on outside sources. However, given that the school feeding program used USDA commodities targeting the period of time the government school feeding budget was mostly not available, it is likely that this deterioration also reflects a weakening of demand for local produce. Therefore, the picture at present seems to be an effective school feeding program driven in the short term to a greater extent by international food aid, rather than food aid supplementing local procurement. This is also likely to have knock-on effects for sustainability, although the HATUTAN plan for program sustainability specifically includes a target of 100% of school meals provided without USDA support.

GENDER-BASED VIOLENCE

The HATUTAN program-specific learning agenda aims to understand the following: “To what extent does the synchronization of school meal programming with activities to address sexual and gender-based violence (SGBV) and school-based affect literacy outcomes and health/nutrition practices at household level?” It is worth noting that attitudes towards SGBV and gendered responsibilities/labor are rooted in social norms and influenced by socioeconomic status, both of which are unlikely to change substantially over a relatively short period of time. As such, at midline, we found few significant changes in SGBV or gender norms in intervention areas as compared to comparison areas.

At midline, there was a decline in the percent of respondents who believed that a husband of justified in beating his wife in any scenario. Intervention households saw a slight but insignificant improvement in attitudes towards spousal violence compared to comparison households for some scenarios.

Overall, these results suggested that fewer respondents believed that a husband was justified in beating his wife in most scenarios at midline, except if she argues with him. However, the similarity of results across comparison and intervention households suggests that this improvement in attitudes towards spousal violence is the result of a broader change in social norms or attitudes towards gender-based violence, rather than a result of any specific program interventions.

Analyzing violence within schools, at both baseline and midline, caregivers of male grade 2 children were significantly more likely to say that teachers use corporal punishment than caregivers of female children. Accordingly, at both baseline and midline and among both treatment and comparison groups, caregivers of both male and female children were more likely to say that corporal punishment was justified against boys than against girls. While intervention households were significantly more likely to state that teachers give a verbal warning at midline compared to the results expected given comparison households, there were no other significant changes in the discipline practices reported by intervention households compared to comparison households, including in corporal punishment. The use of corporal punishment and other negative discipline practices may affect student learning outcomes and decrease student attendance and motivation; future program activities should continue to address this issue, and should particularly focus on violence towards boys in schools, which appears to be more normalized than violence towards girls.

Notably, at midline, a higher percent of caregivers (15%) reported that they would not be able to report abuse of their child at school than at baseline (9%). However, only a very small percent of directors stated that abuse could not be reported. This suggests that there may remain barriers to accessing restorative justice for children abused at school, and that while directors may believe that there are effective avenues for reporting abuse and obtaining justice, caregivers do not necessarily agree or are not necessarily aware of these avenues for reporting.

In contrast to these results, at midline, intervention households saw a significant improvement in some indicators measuring students' access to and feelings of safety at school. Caregivers of male children in intervention households were significantly less likely to report that their child avoided school or was afraid to go to school than would be expected given the results of comparison household, and caregivers of female children were significantly more likely to report that it was safe for their child to walk to school than expected from the results of comparison households. While this improvement appears to have been slightly more salient for boys than for girls (with improvement in two indicators for boys, as opposed to just one for girls), both male and female students appear to have benefited. Furthermore, at midline, female students do not appear to face more risks to accessing education than male students. These results are particularly notable given the impact of COVID-19, which, in many countries, has made students more afraid of going to school due to fear of catching the disease. In this case, gender and violence-related interventions may have had an impact of learning outcomes by improving students' access to school.

Analyzing gender norms affecting children, we find that there were no significant changes in the gendered division of labor in treatment households at midline when compared to the changes seen in comparison households for either male or female students. There were also few significant differences in caregivers' perceptions of how much time male and female students spent on daily tasks or perceptions of caregivers on whether tasks made their male or female children late for school. However, participation in household tasks decreased fairly consistently across comparison and treatment households for both male and female students. The mechanism driving this consistent decrease in participation in household tasks is not entirely clear; however, given the insignificant difference-in-differences results, the HATUTAN program likely did not have a large effect on changing the gendered division of labor within households. Furthermore, there is still a strong gendered division of labor for children, with boys more likely to participate in agricultural activities and girls more likely to participate in household and caregiving activities. This division of labor may have implications for learning; for example, girls may be more likely to spend time on daily tasks that detract from their learning, but boys may be more likely to be pulled out of school entirely during peak agricultural seasons.

For adults, at midline, gender norms that dictate that women are primarily responsible for the care of children while men either take a supplementary role or are not involved in caregiving at all appear to have persisted in both intervention and comparison households. Furthermore, at midline, female caregivers were more likely to report making “minor” decisions—such as making a small household purchase—themselves, and to make “major” decisions—such as a large purchase—jointly with their spouse. In contrast, male caregivers were much more likely to report making “major” decisions by themselves than female caregivers. These results are suggestive of a gendered gap in labor and decision-making power over productive assets. Men often make the primary decisions that have major implications for household finances or food security, and women are often limited to making decisions that have smaller financial implications. This may, in part, be due to the implications of the gendered division of labor on earning power. Female caregivers who are responsible for child-rearing may not have substantial incomes and may rely on their husbands for money, thus reducing their decision-making power. This can have substantial implications for children’s learning outcomes and nutritional status, as studies have shown that, when given decision-making power over household spending, women tend to make purchases that benefit children more often than men.

SUSTAINABILITY

The Learning Agenda includes the following item on sustainability: “The following combination of four strategies will be assessed for their combined impact on sustainability: (i) strengthened capacity of school staff and authorities to manage school feeding projects; (ii) enhanced community participation in monitoring school feeding and education outcomes; (iii) private sector engagement; and (iv) building advocacy skills of stakeholder groups.” Sustainability is discussed in greater detail in the section on Program Implications below; however, we summarize several key findings here.

The HATUTAN program incorporates training within most activities to strengthen the capacity of school staff and authorities to manage SFPs. As discussed above, most treatment schools at midline stated that the school director or coordinator was responsible for school feeding, emphasizing the importance of these activities to ensure the sustainability of SFPs and SFP management. However, over half of treatment schools also reported that PTAs had oversight of SFPs at midline. This result, an increase from baseline, suggests that efforts to increase community participation in monitoring school feeding and education outcomes have had some impact to date, which may strengthen project sustainability.

Regarding private sector engagement, a majority of treatment schools at midline did not report buying produce from farmers for school feeding. This finding suggests that increased efforts to engage the private sector in SFPs may help improve sustainability.

Lastly, program activities aimed to build advocacy skills of key stakeholder groups. While the midline study did not collect data related to this item of the learning agenda, program activities involving advocacy have sought to improve SFP funding mechanisms at the national level. The program’s partnership with the government of Timor-Leste has also increased advocacy opportunities. However, there is likely potential to further strengthen the advocacy skills of stakeholders at the community and school level in order to further ensure that these activities are sustainable.

PROGRAM IMPLICATIONS

The midline study was designed to inform the continued development and implementation of the HATUTAN workplan. This section addresses evaluation questions from seven key areas: design/relevance, management and coordination, effectiveness, efficiency, sustainability, impact, and gender and power relations. Additionally, because of the major impact of COVID-19 on both program activities and related outcomes, we address implications specific to COVID-19.

DESIGN/RELEVANCE

The HATUTAN program design addresses a wide variety of factors known to affect literacy and health. The proposed strategies in the work plan considered resources, training, capacity building, infrastructure, advocacy, and partnerships. The design of the program was adapted at the onset of the COVID-19 pandemic to ensure that program activities were still relevant to the drastically changed context within Timor-Leste. Below, we analyze key activities from the work plan for design and relevance, analyzing if the interventions are relevant to the current priorities and needs of students, schools, and households in Timor-Leste.

TRAINING SCHOOL ADMINISTRATORS

The program argues that strengthening administrators' management skills and leadership is important both to improve school outcomes, as effective school administrators will be better able to train and guide teachers and manage school infrastructure and feeding, and to achieve program sustainability. At midline, HATUTAN and other partners had worked with the MEYS to provide guidance to school administrators to respond and adapt to the COVID-19 pandemic. Prior to the onset of the pandemic, HATUTAN field staff worked closely with school administrators to plan and schedule all school activities. Administrators also received trainings on developing student literacy, management and budgeting, and SGBV.

At midline, analysis of the correlations between school administrators' education levels, years of experience, and provision of coaching to teachers and overall literacy scores, teachers' use of engaging teaching practices, and provision of school meals did not find any significant results. However, midline data suggests that school directors are often responsible for school feeding programs—88% of respondents to the school survey stated that school directors were responsible for oversight of SFPs among all midline schools. Furthermore, in cases where a student is harassed at the toilets or a male teacher pays unwanted attention or harasses a girl, most respondents to the school survey reported that the school director was the first person to whom the student should report the abuse. As such, training of school administrators is particularly relevant to improve SFPs and to reduce the prevalence of SGBV and increase avenues for reporting abuse at schools. Future program activities may benefit from attempting to strengthen the relationships between school administrators and teachers in order to better enable administrators to improve teaching practices, and thus overall literacy scores.

ESTABLISHING AND STRENGTHENING PTAS

HATUTAN has identified active PTAs as an important sustainability strategy and schools with active PTAs have been observed to be better organized. Program activities have thus sought to reinvigorate the PTAs in intervention schools, facilitate community meetings on the quality of schools, and strengthen PTA capacity to monitor SFPs, teacher attendance, school budgets, and SGBV, among other activities. Midline data suggests that PTAs are involved in many activities related to schools, especially in intervention areas. More than half of PTAs reported being involved in improving school infrastructure, overseeing the SFP, monitoring safety and security, and monitoring student and teacher attendance. However, data also suggests that participation in PTAs remains relatively low in many areas. Activities that seek to strengthen PTAs may thus have particular relevance for improving school infrastructure, SFPs, and student and teacher attendance.

TRAINING TEACHERS

Better-trained teachers are more able to provide high-quality instruction to students, thus leading to improved literacy abilities. The HATUTAN program implemented a variety of activities to train and mentor teachers both to improve the quality of instruction, including increased use of engaging teaching practices and reduced use of ineffective or negative teaching practices, and to improve

nutrition and gender awareness and SGBV prevention. At baseline, qualitative data suggested that teachers understand the importance of using child-centered teaching strategies but had limited knowledge of how to operationalize them. Data at midline suggests that this may still be the case, with only a limited increase in the use of engaging teaching practices among intervention schools. Data on literacy scores and subtask scores also suggests that current teaching practices may not be effectively imparting knowledge on phonemes and the relations between letters and words to young students, and that overall reading fluency remains low. Given these results, teacher training remains of high relevance to the program. It may, however, be useful to analyze the design of teacher training activities to make sure that trainings target best practices for improving student fluency and helping young students read not just letters, but words.

PRODUCTION OF BOOKS AND SUPPLEMENTAL READING MATERIALS

The availability of reading materials both at school and at home can substantially help to strengthen students' overall reading abilities. HATUTAN activities in this area included the development and distribution of the *Lafaek* student and teacher magazines, sourcing literacy teaching and reading materials to assist at-home learning during COVID-19 and to be used in classroom reading corners, and supporting school administrators to manage their schools' reading materials.

The predictive analysis showed that at baseline, whether a student reads at home was a significant predictor of literacy results. While this result did not remain significant at midline, nor was there a significant relationship between the availability of reading materials in grade 2 classrooms and literacy scores, these results do not necessarily suggest that access to reading materials is not an important contributor to literacy, nor that schools and homes have sufficient reading materials. Provision of reading materials remains a relevant program activity, particularly given that at midline, 21% of treatment schools reported that the school did not lend story books for students to take home, and 24% of treatment schools reported that there were not enough books to lend to students. Furthermore, 31% of treatment households at midline were not observed to have any children's books or magazines in the house. The provision of reading materials may also improve teaching practices, as teachers are more able to incorporate activities that use the reading corner or have students read alone or in groups; this may then further improve learning scores.

PROVISION OF SCHOOL MEALS OR TAKE-HOME RATIONS

Provision of school meals to supplement the government-run SFP was a primary activity of the HATUTAN program. HATUTAN implemented a full-scale SFP with USDA-donated fortified rice, pinto beans, and refined vegetable oil at 435-445 schools in intervention municipalities. The program also provided take-home rations to students during COVID-19 induced school closures to ensure children had continued access to nutritious food that they would have received at school while they were at home. Take-home rations were provided for 416 out of 435 schools.

Results suggest that the provision of school meals may have helped mitigate the negative impact of COVID-19 on learning by increasing student attentiveness (see sections on "Literacy Results" and "Predictive Analysis – Student Attentiveness") and decreasing dropout rates (see "Student Attendance"). The provision of school meals continues to be highly relevant for students in Timor-Leste, helping to improve nutrition and literacy—both of which remain at relatively low levels throughout the country.

PARTNERING WITH FARMERS' GROUPS

The HATUTAN program sought to stimulate local rural agricultural markets and increase the production, consistency, and quality of nutritious foods as well as improve nutrition consumption in households. As a result of program activities, 117 keyhole gardens and 478 permagardens were established by midline. Additionally, the program sought to build linkages between schools and

farmers to increase demand for nutritious foods and to increase the quality of school meals, as well as to establish school gardens.

However, the demand for local produce among schools remains low due to feeding program budget limitations, which may affect the sustainability of the school feeding program. As a result, farmers will not produce for schools because there is no demand and schools tend to look to other sources, which may incur higher transportation costs if hiring a vehicle or increased labor if walking longer distances. Ensuring consistent and sufficient demand and supply of goods for the school feeding program will help in serving well-balanced, nutritious meals to students.

FORMING VSLAS

Through Community Development Agents, the HATUTAN program established and provided technical support to VSLAs in intervention communities and established 123 VSLAs with 2,101 members by midline. These groups serve as the foundation for which other trainings are layered, including trainings on nutrition, improved agriculture, and gender. Furthermore, VSLA members commonly report using their loans on education expenses (71%), food (63%), and to invest in businesses (30%), suggesting that VSLA membership may have positive effects on students' learning outcomes, household nutrition, and, potentially, gender empowerment if loans are used to invest in women's businesses.

Because VSLAs are the foundation for other trainings, their continued establishment and support is of strong relevance to program activities. However, it is worth noting that households that are not involved in VSLAs may also not be able to receive trainings if this is the primary mechanism by which training participants are recruited. As such, it may be worth exploring other mechanisms to recruit households for trainings outside of VSLA participation and, even more so, how to increase VSLA coverage among parents of schoolchildren.

TRAINING FOR GOOD HEALTH, HYGIENE, AND NUTRITION PRACTICES

Improved health, hygiene, and nutrition practices can help improve education outcomes by increasing student attendance and attentiveness in school. HATUTAN program activities in this area included the development of a social behavior change strategy; provision of training on optimal health, hygiene, nutrition, and gender-related practices; and provision of training on the safe preparation of nutritious meals for school cooks.

The training on health, hygiene, and nutrition is relevant to understanding the importance of related practices to ensure that schools meals are prepared in a hygienic manner to prevent absences due to sickness, that storage spaces are clean and secure from pests, and that well-balanced and nutritious meals are served to children. These practices are also relevant to the school feeding program, which contributes to improving student attendance, attentiveness, and performance. Moreover, hygienic practices are promoted to children in school with the availability of handwashing stations and toilets. However, given that knowledge of health, hygiene, and nutrition appears fairly high, it may be worth pivoting program activities to focus on behavior change, rather than knowledge.

EXTRACURRICULAR ACTIVITIES

The HATUTAN program intended to incorporate reading and health-focused extracurricular activities and games to increase contact time with students in a fun and cooperative environment and to develop students' self-confidence and voice, particularly for girls. Unfortunately, due primarily to restrictions due to the COVID-19 pandemic, HATUTAN had limited opportunities to work with teachers to initiate extracurricular activities. Given the low number of contact hours in most schools over the past year, extracurricular activities may be particularly useful to help students catch up after the challenges of the past year.

CAPACITY BUILDING AND ADVOCACY

With the onset of the COVID-19 pandemic, the HATUTAN program increased its focus on capacity building and advocacy at the national level for COVID-19 prevention and awareness. The program also included activities focused on advocating for policy changes related to the national SFP and education policies, including strengthening the government of Timor-Leste's capacity to deliver the SFP. After delays due to the COVID-19 pandemic, national budget issues further delayed provision of school feeding supplies to schools. This remains a highly relevant issue for school feeding across Timor-Leste; program activities should continue to advocate for flexible funding for the SFP and build capacity for SFP implementation.

MANAGEMENT AND COORDINATION

Management and coordination are an essential part of an effectively run program to ensure impact and sustainability. This includes both management and coordination of within-program activities to ensure that staff and other resources are distributed effectively and there are no inefficiencies in implementation. It also includes coordination with other relevant agencies and programs to ensure that program interventions are compatible with other interventions in target municipalities, such as the national school feeding program. If external coordination is weak, HATUTAN program activities may overlap with programs run by other agencies, generating inefficiencies. As a result, coordination processes should be continued in order to ensure maximum impact from the program.

EFFECTIVENESS

Overall, program effectiveness has been somewhat undercut by the COVID-19 pandemic, which has had substantial negative effects on education outcomes, nutrition, the provision of school feeding, and health, among other areas. However, given that the COVID-19 pandemic was an exogenous event that could not be controlled for or prevented by the HATUTAN program, and, further, given that program activities were relatively successful in pivoting to address issues relevant to the pandemic—such as providing take-home rations to students or increasing training on hygiene as related to COVID-19—in this section, we focus on issues other than COVID-19 that may have reduced program effectiveness and could potentially be addressed in future programming.

The effectiveness of program activities is dependent on a variety of factors unique to the context of each community and school. At baseline, three factors were identified as universally important and as having potentially large impact on effectiveness: school infrastructure, PTAs, and school administrators.

School infrastructure was identified as a potential constraint to effectiveness because an effective SFP requires infrastructure, such as a kitchen and a safe water source, to support it. At midline, the majority of intervention schools—97%—reported having a school kitchen, a slight increase from 94% of intervention schools at baseline. However, school kitchens often lacked sufficient supplies for safe food preparation, including a place to wash hands and soap, and 13% of schools reported that clean water was not available to prepare meals. Furthermore, only 46% of intervention schools had enough space to store food within or near the kitchen at midline and only 5% of schools had a canteen or a place for students to eat. While the HATUTAN program will aim to address these infrastructure issues by building the capacity of school administrators to apply for infrastructure-related grants or funding, this activity did not progress before midline, and infrastructure clearly remains a constraint to the establishment of effective and hygienic SFPs in all intervention schools.

PTAs were identified as playing a critical role in the success of SFPs and overall school management. PTA involvement increased at midline, a potentially positive sign for program effectiveness. The frequency of PTA meetings increased at midline: Among intervention communities, 24% of schools reported that the PTA had met last week at midline compared to 5% at baseline and 17% that the PTA

had met last month compared to 13% at baseline. Only 36% of schools reported that the PTA had not met at all at midline, compared to 70% of baseline schools in intervention areas. Furthermore, at midline, PTAs were significantly more involved in almost every factor affecting schools in intervention areas than in comparison areas (Table 146). Continuing to build on this progress in PTA involvement may further bolster program effectiveness, as well as improving sustainability. In particular, activities that seek to improve household participation in PTAs, which remains relatively low, may help improve related outcomes.

Table 146: PTA involvement at midline

	Intervention areas	Comparison areas	p
n	98	86	
School budget management	23.5%	30.2%	0.31
Learning quality	46.9%	30.2%	.02*
Improve school infrastructure	86.7%	64.0%	<0.001***
Oversee SFP	90.8%	57.0%	<0.001***
Monitor safety and security	71.4%	52.3%	0.008**
Monitor student attendance	70.4%	36.1%	<0.001***
Monitor teacher attendance	70.4%	32.6%	<0.001***
Monitor dropout	45.9%	33.7%	0.09
Does not do anything	1.0%	9.3%	.01*

* Significant at $p < 0.05$ ** Significant at $p < 0.01$ *** Significant at $p < 0.001$

Similar to the role of the PTAs, school administrators were also identified as crucial for program effectiveness due to many activities' reliance on the ability and motivation of administrators for implementation. Several program activities have thus focused on training school administrators in a variety of areas, including budgeting, procurement, human resources management, community mobilization, classroom management, and more. The effectiveness of program activities through the end of the program will rely on continued commitment from school administrators; as such, HATUTAN program staff should continue to pay close attention to administrators' involvement.

Outside of these three factors, midline data reveals several other constraints to program effectiveness. The first of these is the wide range of students' literacy abilities within classrooms. It is not uncommon for classrooms to include both students with no literacy abilities—i.e., no ability to read either letters or words—as well as students who are able to read a passage and answer at least some reading comprehension questions correctly. Literacy results on each subtask suggest that currently, teaching practices are not effective at improving the literacy skills on both of these groups of students at the same time. To improve program effectiveness, it may therefore be useful to provide teacher training on effective ways to engage all students in classes and teach to a wide range of skill levels and literacy abilities.

A second constraint relates to activities seeking to increase access to reading materials at school and at home. While the number of grade 2 classrooms with reading materials and, even more so, a reading corner increased at midline in intervention (and comparison) schools, it is not clear that this always translates into increased levels of access to reading materials for students. Twenty-one intervention schools did not report lending students storybooks to take home at midline; among these schools, 71% stated that books were not loaned to students because students were careless and might damage the books and 52% stated that students might lose the books. This suggests that mistrust of students or beliefs that books are “too nice” for student use might reduce the effectiveness of interventions seeking to increase access to reading materials.

Gender norms, and in particular the tendency of gender and social norms to change only very slowly, may also affect program effectiveness. As discussed in the “Gender and Power” section, there was

little significant change in various indicators of gendered responsibilities, decision-making ability, and SGBV at midline in intervention areas as compared to control areas. Traditional gender norms may limit program effectiveness by, for example, reducing the likelihood of families to invest in a girl's education or limiting the resources available to female caregivers to purchase nutritious foods. However, because gender norms are difficult to change, program activities must carefully consider the role of gender and power to maximize effectiveness. This is discussed further in the section on gender below.

EFFICIENCY

Efficiency refers to the extent to which program activities deliver, or are likely to deliver, results in an economic and timely way. If program activities are excessively costly but have only a small impact on outcomes of interest, the resources dedicated to the intervention may not be justified by its results, perhaps warranting changes to program design. Furthermore, even if the program's benefit-to-cost ratio is high, careful analysis of program efficiency can lead to improved use of limited resources. Efficiency is also important from an operational perspective in order to better understand the feasibility of achieving outcomes given limited resources and to ensure that programs are well conducted.

In the context of the HATUTAN program, it is important to analyze both the costs and benefits of program activities to ensure that program funding is spent effectively. Activities such as support for school feeding may be highly efficient if the cost to provide goods—in this case, rice, beans, and oil—is low, particularly because the benefits of school feeding on student attendance are evident. In contrast, activities that are highly costly in terms of time or resources but may not always have a clear impact on behaviors—such as, potentially, activities that seek to train parents or teachers—may be less efficient.

SUSTAINABILITY

In order to ensure that the benefits of HATUTAN program activities continues after conclusion of the program, sustainability must be at the forefront of all program-related decisions. The HATUTAN work plan focuses on sustainability through advocacy and capacity-building activities at the local, regional, and national levels. Additionally, the program intends to improve sustainability by training school administrators to better manage schools and apply for funding, training school cooks to provide more nutritious school meals, and training teachers to encourage the use of better teaching practices. The establishment of VSLAs and use of VSLAs as a platform to provide trainings to community members may also help improve sustainability. Additionally, strengthening PTAs will help increase the accountability of schools to the local community and thus enable greater sustainability. Municipal authorities have also been considered as potential actors to ensure project sustainability, but are highly constrained by a lack of resources.

HATUTAN has developed a graduation and sustainability assessment plan to measure progress towards graduation from the program for intervention schools and at the national level. This plan includes five targets: 1) improved national and municipal support for school feeding; 2) 100% of students provided with daily school meals without USDA support; 3) increased student literacy rates; 4) improved nutrition, health, hygiene, and gender equity practices; and 5) improved prevention and response to domestic, sexual, and gender-based violence. The scoring sheet will be refined for use in 2021. It is worth noting that progress on some of these indicators may inevitably be slower than others. Gender equity practices and student literacy, for example, tend to change slowly over time. As such, it may be worth exploring a staggered graduation plan whereby schools can graduate from specific areas of program support at different times.

TRAINING APPROACH

The HATUTAN program incorporates training within most of its activities: Ten of twelve activities rely on training as a critical element of the work. Strengthening the capacity of school staff and authorities to manage SFPs is likely to increase the sustainability of program activities, particularly in conjunction with the program's advocacy activities which seek to improve SFP funding mechanisms at the national level and with activities that seek to improve linkages between local farmers and schools. However, the program should ensure that administrator and teacher attrition do not undercut the effectiveness of trainings by continuing to develop training manuals and encouraging school administrators to document progress within schools.

PARTNERSHIPS

The HATUTAN program has partnered with the government of Timor-Leste, international development partners, farmers' groups, agricultural extension agents, community development agencies, and PTAs in implementation of various activities. During the pandemic, the program formed a close partnership with the Ministries of Health and Education and helped to inform the COVID-19 response and public information campaign as part of a national-level task force. HATUTAN also supported municipal-level COVID-19 task forces led by health department officials to improve public WASH facilities and public information campaigns. HATUTAN staff also participated in a consultative commission with nine ministry departments and five other development partners to coordinate the education response to COVID-19. The program's partnerships with government agencies and strong relationship with the government of Timor-Leste have opened advocacy opportunities, including activities related to the national budget and the funding of the school feeding program. Data from the HATUTAN program has also helped to shift MEYS planning addressing the loss of learning time due to COVID-19. These activities may help to improve program sustainability by addressing underlying issues affecting the provision of school feeding and limiting students' exposure to quality education. In addition to these partnerships, HATUTAN partnered with the Ministry of Agriculture and Fisheries as well as local AES agents to support nutrition-sensitive agriculture activities.

Outside of these national- and municipal-level partnerships, at the local level, the HATUTAN program seeks to enhance community participation in monitoring school feeding and education outcomes. The increased capacity of PTAs (described above in "Effectiveness" and "Design/Relevance") is an encouraging sign that program activities are successfully increasing community involvement. However, linkages between schools and farmers' groups appear to remain weak, with 54% of intervention schools stating that they did not buy local produce from farmers for school feeding at midline.

Moving forward, HATUTAN programming should attempt to establish better links between farmers' groups and schools in order to further improve sustainability. Furthermore, the program should continue to follow best practices around partnerships, including establishing clear expectations and shared accountability, investing in ongoing professional development for all partners, developing a plan for long-term sustainability, and implementing a communication plan to share progress and challenges.

POLICY IMPLEMENTATION

The involvement of community members, teachers, and parents is critical for the sustainability of program activities. Top-down policy implementation that only involves, for example, central school directors or government officials, is less likely to succeed as there may be less buy-in and more resistance from those responsible for carrying out the policy at the local level.

The HATUTAN program has taken a collaborative approach to the implementation of policies and activities, involving community members, teachers, and parents. The program has also worked to build the advocacy skills of various stakeholder groups by, for example, training school administrators in

proposal development. Moving forward, the program should continue to ensure that local community members are involved in the design and implementation of program activities in order to ensure that new practices continue to be carried out after the conclusion of the program.

IMPACT

Impact measures the extent to which the HATUTAN program has generated and is expected to generate significant positive or negative effects, whether intended or unintended. The majority of this midline report is dedicated to understanding the impact of HATUTAN program activities so far; below, we include a summary of findings.

LITERACY

At midline, due to the impact of COVID-19, overall literacy scores among grade 2 students worsened for both intervention and comparison groups. However, we find a significant and positive effect on literacy scores for the intervention group compared to the comparison group. In other words, while average scores for both groups declined at midline compared to baseline, average scores for treatment students exposed to the program declined significantly less than those for comparison students. The results for the panel cohort of students assessed at both baseline and midline are less suggestive of program impact, with scores improving by similar amounts in both intervention and comparison areas due to exposure to an additional year of education.

For grade 2 students' performance on specific literacy subtasks, analysis found that students in intervention schools performed significantly better than expected given results in comparison areas for the invented word fluency and passage reading subtasks. The program may also have had a positive impact on grade 2 students' scores for the remaining subtasks—letter recognition, familiar word fluency, and reading comprehension—but results are less conclusive for these subtasks. It is worth noting, however, that scores for grade 2 students on all subtasks remain very low, in part due to the COVID-19 pandemic and in part due to structural issues affecting literacy in Timor-Leste.

At both baseline and midline, female students performed better than male students. HATUTAN program activities do not appear to have had differential impacts by gender, with scores for both male and female students declining overall, but declining less within treatment areas than within comparison areas. Overall, there remains a large gap in scores between male and female students which HATUTAN programming seems not to have affected.

QUALITY OF INSTRUCTION

At midline, there was little change in the overall use of engaging teaching practices in either treatment or control schools. However, teachers in intervention schools were significantly more likely to use games or exercises in class than expected given results in comparison schools, and were substantially—though not significantly—more likely to ask open questions and use the reading corner. There was a substantial and significant decrease in the use of traditional teaching practices in treatment schools as compared to control schools; unfortunately, the HATUTAN program did not appear to have had a likewise positive effect at reducing the use of negative teaching practices. At midline, the use of corporal punishment increased by around 3 percentage points in both intervention and comparison schools; the prevalence of verbal and physical discipline is also likely considerably higher than observed in classroom observations due to social desirability bias. Indeed, caregiver perceptions of negative teaching practices remained high from baseline to midline, and there was no significant change in perceptions of negative teaching practices in treatment groups as compared to control groups. Analyzing the gender-specific prevalence of negative teaching behaviors, at midline, teachers used verbal and physical discipline more frequently with boys than with girls.

Unfortunately, the HATUTAN program also does not appear to have had a positive effect on teacher attendance at midline. Among comparison schools, at midline, teacher attendance taken the day of

the survey and recorded the previous day increased substantially; among treatment schools, teacher attendance the day of the survey increased, but by a smaller amount, and attendance the day before decreased substantially. This finding is likely due to external factors, such as the differential impact of the rainy season or interventions that increased attendance rates in some comparison municipalities; however, this effect warrants further examination in order to ensure that program activities are effectively addressing teacher attendance.

At midline, quantitative data suggests that poor families often have to use their savings to pay for student supplies. However, most households did not report that a lack of school supplies is major constraint to school attendance or learning to read. Within schools, access to literacy materials increased in both comparison and intervention areas at midline. Among intervention schools, there was a particularly notable increase in access to reading corners. However, the analysis did not find a significant increase in access to literacy materials in intervention schools as compared to comparison schools, implying that the increase may not be due to HATUTAN program activities.

While a higher percent of teachers reported having attended training on literacy education at midline than at baseline, there were no significant differences in the increase between intervention and comparison areas. Program activities do appear, however, to have had some impact on school administrators' provision of coaching to teachers: Administrators in treatment schools at midline were significantly more likely than expected to have provided training to teachers at least once given results in comparison schools.

STUDENT ATTENTIVENESS

Three main indicators for student attentiveness were used: (1) a self-reported measure of whether the student felt they could pay attention, (2) observed attentiveness, where an enumerator observed 10 students and reported how many were paying attention, and (3) working memory score as a proxy measure. In the cross-sectional and panel samples for self-reported attentiveness there were positive difference-in-differences scores, though these were not significant. As attentiveness is thought to depend on student hunger, this could point to a benefit of the program. However, it is important to point out that this indicator is flawed as most students are likely to reply positively due to desirability bias.

Observed student attentiveness showed a positive though not significant difference. This was the result of a decline in the number of students paying attention in the comparison group while the intervention group remained stable. Again, this could reflect improved attention as a result of the effective school feeding program. In the cross-sectional analysis (where all students included were in the second grade) working memory scores fell for both comparison and intervention, and for both the cross-sectional analysis and the panel analysis the difference-in-difference was negative. This suggests that working scores in the treatment group either fell further or rose slower than the comparison group. However, neither of these results were statistically significant. As an additional note, the decline in observed memory and working memory may be due to the impact both of COVID-19 and of natural disasters in the region, leading to increased school absences.

Factors likely to affect attentiveness were also studied, including student hunger and school feeding facilities. Whether a student had eaten anything on the day of the EGRA test had a positive difference-in-difference score for both the cross-sectional and panel analyses, and this was statistically significant for the panel. This is likely a benefit from the program: Student hunger decreased faster in the intervention group than the control. We also find that whether a student had eaten was a significant predictor for improved self-reported attentiveness, revealing a possible mechanism for the program to improve attentiveness. We find that student hunger decreased by a slightly greater amount for male students than female.

Finally, we also analyzed school-level food characteristics likely to influence student hunger (and, in turn, student attentiveness). The difference in difference score for whether the school served a meal on the day of the survey was very high (88 percentage points) and statistically significant, and was

driven mostly by a rapid increase in the intervention group from nearly no schools serving a meal to nearly 90% serving them. This finding provides strong evidence that the program was effective in implementing school feeding, a potential mechanism to reduce student hunger and thus increase attentiveness.

STUDENT ATTENDANCE

For five out of six grades we found negative difference-in-difference scores for student attendance; this result was only significant for grade 6. This suggests there was a deterioration in attendance in intervention schools compared to comparison schools. However, the most likely explanation for this is that students in intervention schools were worse-affected by natural disasters. When asked about the reason for school absences, 17% in the control group cited natural disasters as a reason compared to 35% in the treatment.

There was a negative difference-in-difference score for dropout rates for all grades, and this was statistically significant for all grades aside from grade 3. This suggests there was a benefit from the program and students were less likely to drop out, possibly due to parents being less likely to draw children out of school if there is an effective school feeding program or students being less likely to want to drop out. However, the response rate at baseline for the comparison group was quite low and this might bias the results. We are therefore unable to draw confident conclusions from this analysis.

We found positive difference-in-difference scores for number of days missed due to illness, meaning the trend over the study period was worse for the intervention group than the comparison group and they were likely to miss more days. The possible causal mechanisms behind a deterioration in this indicator are unclear. Of other factors studied, we found that students within a 30-minute walk and those who felt safe on their walk missed fewer days of school on average, while students who were afraid of school missed more days on average. These results were statistically significant. We also found that student who rely on transportation and had difficulty making friends missed more days of school, but these results were not significant.

SCHOOL FEEDING PROGRAM

We found that intervention schools served a higher-quality menu than comparison schools. However, we also found that relatively few treatment schools that were providing school meals purchased local produce from farmers, although sample size limitations mean that we cannot compare midline and baseline findings. This could be both a demand and supply issue as HATUTAN's community development agent (CDA) learning study shows that farmers do not see schools as a viable market because they are not creating a demand, and without the demand there will not be a supply. Finally, we also found a non-statistically significant positive difference in difference score for whether the PTA was responsible for school feeding, and a statistically significant difference-in-difference for the frequency of PTA meetings. This suggest the program was successful in improving the functioning of the PTA.

Procurement of produce from local farmers is hindered by budget limitations in purchasing supplies for the school feeding program; currently, budgets remain at only around half the amount that is required per student per day to purchase a nutritious, locally-sourced school meal. This may have implications on serving nutritious and well-balanced school meals. For example, the number of schools that procured fruits is very low, which may explain why there are few schools that included fruits in their school feeding menu.

NUTRITION, HEALTH, AND HYGIENE

Outcomes of the training may be seen through the higher number of schools at the midline that reported having access to clean water for food preparation, a clean storage space, a roof that does not leak, and a kitchen cleaned using detergent. Parents were reported to have been concerned over food safety conditions in schools at baseline due to unhygienic food preparation methods that could

cause sickness and, consequently, student absences. However, this perception of parents of children in treatment schools shifted at midline as most parents said that food served to children in school is prepared in a hygienic manner. Findings from the school and household survey indicate that there had been improvements in maintaining hygiene during food preparation. However, it is not clear how aware or certain parents are about the hygienic preparation of food in schools.

Dietary diversity of school meals appears to have improved, which may be linked to the training on nutrition practices. At the midline, intervention schools were far less likely to report serving meals with a dietary diversity score of 1 compared to comparison schools, and were more likely to serve meals that scored a 2 or 3 on the dietary diversity scale.

ECONOMIC EMPOWERMENT

In the economic empowerment section we analyzed savings, VSLA membership and loan use, and the effect of economic empowerment on nutritional and educational outcomes. While key indicators for savings had changed from baseline to midline—for example, a large decrease in the proportion of households with savings who used those savings for business investment—none of the difference in difference scores were significant. This suggests that the program had limited impact on savings, including whether a household had any savings, and whether they used those savings for food, education, debt, or investing in assets.

We also analyzed the impact of VSLAs, although it is important to note that most of the data collection for VSLA-related indicators was at midline and we were therefore unable to calculate difference-in-difference scores. Disaggregating savings use by VSLA participation showed little difference between the behavior of VSLA and non-VSLA participants. We found small differences in VSLA loan use by treatment and comparison: The intervention group spent statistically significantly more on agriculture, business investment, and debt. However, given we only have midline results and cannot assess trends over the program lifetime, we cannot draw strong conclusions as to program impact. Finally, we tested whether the program improved how VSLAs function and whether this had an effect on education and nutritional outcomes. We found no statistically significant results, again suggesting limited impact. However, it is important to point out that a richer analysis will be possible at later rounds of program evaluation.

AGRICULTURAL PRACTICES

This section analyzed agricultural practices with a focus on farmer training, keyhole gardening, and permagardens. In depth analysis on program impact for this section was limited, as data was only collected at midline for most indicators and the sample size was quite small. As for the economic empowerment section above, a more detailed analysis of impact will be possible at later evaluation rounds.

The main analysis for impact tested whether an interaction between being in the treatment group and any of the three main indicators—having received training on agriculture, owning a keyhole garden, and owning a permagarden—influenced nutritional and economic outcomes. Farmers in the treatment group who received training were significantly less likely to have had someone in their household go without eating in the past 30 days. Farmers in the treatment group who had a keyhole garden were more likely to spend savings on investment, while those with a permagarden were more likely to have had a household member go without food in the past 30 days. At this stage we must be cautious in attributing results for these indicators to program impact, however, and these instead may serve as possible trends for further analysis when more data is available.

GENDER AND POWER

Overall, there were few significant changes to gender and power dynamics among intervention areas at midline. There were no significant changes to the gendered division of labor for children within households; few differences in caregivers' perceptions of how much time male and female students

spent on daily tasks or whether tasks made students late for school; few differences in the control of productive assets; no differences in attitudes towards gender-based violence; no differences in caregivers' opinions of the skills and capacities of girls and boys; and few differences in the use of violence in schools. In contrast, there did appear to be a substantial positive change in children's safe access to schools at midline among treatment groups. This improvement appears to have been slightly more salient for boys than for girls, although both male and female students appear to have benefited. Impact on gender and power dynamics was likely low because, as they are rooted in social norms, these dynamics tend to change very slowly over time.

GENDER

As described at baseline, girls generally have higher engagement and achievement in school than boys, and many parents even consider girls to have more capacity for reading, writing, and math than boys. However, adult women have limited decision-making power in households, are often limited to working as caregivers, and are often less involved in community organizations such as farmer's group or receive fewer benefits from local services such as agricultural extension services. Clearly, despite the great potential evidenced by young female students, gender norms reduce the options available to girls as they grow older.

Girls appear to bear many obligations for both academic performance and household work. At both baseline and midline, girls were significantly more likely to be perform many household tasks than boys, particularly caregiving and housework. While boys were more involved in agricultural work than girls, around one-quarter of girls at midline still reported being involved in agriculture as well as other household tasks. At some point, girls may struggle to keep up with both their education and household responsibilities. Although not reported in the quantitative data, perhaps due to social desirability bias, parents may also be less likely to support girls through higher education due to gender norms that encourage girls to start families and stay at home to care for children.

Girls' confidence may also be challenged at a young age. While data from midline suggests that teachers encourage and ask questions to boys and girls at roughly similar rates—and, in fact, may use negative teaching practices slightly more frequently with boys—qualitative data from baseline suggests that girls are often viewed as “shy” by teachers and are teased by boys. Many girls are also conditioned to avoid asking questions in order to prevent embarrassment.

Overall, these findings suggest a need for interventions that sustain girls' successes at young ages through adulthood. Activities that provide an opportunity for dialogue with communities to challenge traditional gender roles may help achieve this goal. However, as noted above and in the “Gender and Power” section, gender norms tend to change slowly over time; at midline, there is thus little evidence to date of the impact of program interventions that seek to improve the status and opportunities of girls and women.

COVID-19 IMPACT

As discussed extensively throughout the report, the COVID-19 pandemic has had a massive impact both on program activities and on Timor-Leste more broadly. Learning outcomes have consistently declined on all subtasks and for overall literacy for all students. School closures have limited exposure to learning and delayed implementation of the SFP. Delays to school feeding were further exacerbated by the pandemic's negative effects on household food security and livelihoods, worsening nutrition outcomes for both children and adults. Many program activities were either delayed or altered to comply with COVID-19 restrictions.

While the development of multiple vaccines has brought into sight an end to the pandemic, Timor-Leste faces barriers to obtain sufficient vaccines and distribute to the entire population. Furthermore, the end of the pandemic will not undo the damage to learning, nutrition, health, livelihoods, and other outcomes caused this year. Next year, students moving on to later grades will be starting from a lower

level than students in previous years. The very low literacy abilities of this year's grade 2 students will carry forward the next year, and it will likely be difficult for teachers to both catch students up and teach them the expected curriculum to prepare them for continued progression through education.

Future program activities should carefully consider the implications of these dynamics on students, teachers, and schools. It may be useful, for example, to analyze whether the current trainings provided to teachers and administrators can effectively help schools catch students up after around half a year of missed education. It may also be useful for the HATUTAN program to engage in advocacy and policy dialogues at the national level around this issue. It will be important for programming not to consider the pandemic as a “thing of the past,” but rather to acknowledge the many long-term effects of the pandemic on outcomes of interest and adapt programming to address these effects.

Outside of the impact of the pandemic on schools and learning, the limited recall and even more limited implementation of COVID-19 prevention practices as well as the very low percent of households with handwashing stations with soap means that Timor-Leste may be very vulnerable to rapid spread of COVID-19 should an outbreak occur. Given the country's limited healthcare capacity, this represents a potential risk to the health and lives of Timorese citizens, and there is a need for immediate action to address such issues.

CONCLUSION

Conclusions drawn from the findings of the HATUTAN midline evaluation can inform program implementation moving forward to further improve health and literacy outcomes across the four intervention municipalities. In some cases, conclusions validate the approach to improving health and literacy for primary grade students; in others, they suggest gaps in impact that could be addressed with small changes to program activities or inputs. As discussed extensively throughout the report, midline findings have been massively affected by the COVID-19 pandemic and its related effects on Timorese schools and families. The downstream effects of the pandemic will continue for years after it ends, however; our conclusions here and throughout the report thus do not assume a “return to normalcy” between midline and endline, but rather acknowledge the new context within which the program is working.

LITERACY DEVELOPMENT AND QUALITY OF EDUCATION

The COVID-19 pandemic had a substantially negative impact on literacy scores, but the HATUTAN program seems to have mitigated some of this effect. While literacy scores worsened for all grade 2 students, they decreased by less in intervention municipalities than in comparison municipalities. The program appears to have had greater impact on literacy scores for grade 2 students than grade 3 or 4 students: There was no significant relative improvement in the literacy scores of intervention students in grades 3 or 4 that were re-contacted from the baseline, although there was an expected overall increase in scores due to these students' exposure to additional years of education. Despite these results, it is worth noting that literacy scores for grade 2 students remain extremely low.

There is a major gap in literacy ability between letter recognition and word recognition, and between word recognition and reading comprehension. As found at baseline, teachers appear to teach literacy skills by first focusing on letter recognition before moving on to words. As a result, students (particularly grade 2 students) have a very weak understanding of how letter sounds combine to form words, in addition to low levels of fluency overall.

Use of ineffective traditional teaching practices declined in intervention schools relative to comparison schools; however, there was little change in the use of engaging teaching practices. Unfortunately, the HATUTAN program did not appear to have had a positive effect at reducing the use of negative teaching practices or at increasing teacher attendance. Teachers appear to continue to struggle with implementing child-centered approaches in the classroom, despite some

improvements in student attentiveness, which the theory of change posits may lead to (and result from) improved quality of instruction.

Increase activities to improve quality of instruction. The above results suggest that teachers still face challenges to effectively teaching literacy skills, particularly to young students. As at baseline, teachers appear to teach literacy by focusing first on letter recognition before moving on to words. Effective literacy development, in contrast, occurs from more well-rounded instruction that includes concurrent focus on sounds, vocabulary development, and comprehension. It also necessitates engaging teaching practices that increase student interest in the content. Program activities that seek to strengthen both the use of engaging teaching practices and pedagogical strategies may help to improve student literacy.

School attendance did not improve in intervention areas relative to comparison areas. Data suggests that this was affected by external factors, particularly a greater incidence of natural disasters in intervention municipalities. However, we also found a positive difference-in-differences score for number of days missed due to illness, meaning that the number of days missed due to illness worsened more in intervention areas than in comparison areas at midline. Overall, these results suggest that many barriers remain to consistent student attendance.

Dropout rates in intervention municipalities decreased significantly compared to comparison municipalities. This suggests that program implementation may have made students less likely to drop out of school; a potential causal mechanism for this effect is that parents may be more likely to send children to school if there is an effective SFP, or students may be less likely to want to drop out if they are provided with school meals. This finding is notable given the disruption of the COVID-19 pandemic, which, in isolation, may have increased students' likelihood of dropping out due to the difficulty of returning to school after lengthy school closures.

Students in intervention municipalities appeared to have somewhat higher attentiveness than expected given changes in comparison municipalities. This was measured through positive difference-in-differences scores for both observed and self-reported attentiveness, although this finding was not confirmed by correspondingly positive changes in working memory scores. Confirming the theory of change, we find student hunger to be a significant predictor of attentiveness.

Further examine ways to improve student health and attendance. The HATUTAN program's support for school feeding appears to have had some positive effects on dropout rates and attentiveness, as posited in the theory of change. However, it has not corresponded to improved attendance rates. The increase in days missed due to illness is particularly striking, as, at midline, most caregivers had strong knowledge of healthy hygiene and nutrition practices, and most households reported that school meals were reported in a hygienic manner. However, food security and dietary diversity declined at midline, in part due to the COVID-19 pandemic. Given that relatively high levels of knowledge of healthy practices have not necessarily translated into an increase in health outcomes or behaviors, it may be worth examining the efficiency of program activities targeting health knowledge, and adjusting these activities to increase the focus on behavior change or food security.

SCHOOL FEEDING

Most schools have a PTA that provides oversight of school feeding, but PTA activity remains relatively low. At midline, there was an improvement in the frequency of PTA meetings in treatment schools relative to comparison schools, although 36% still reported that the PTA did not meet during the school year. There is also a low level of participation among households.

School meals have a relatively low level of dietary diversity, although intervention schools served a higher-quality menu than treatment schools. Most schools served carbohydrates, legumes and nuts, and dark green vegetables such as spinach, lettuce, and mustard greens at the midline, which is consistent with the baseline findings. However, food served to children in school are lacking in fruits. This low consumption of fruits may be linked to the low amount of fruit bought from local farmers.

Linkages between schools and farmers remain weak primarily due to limited budget for schools to buy produce from local farmers to supplement the SFP. As a result of limited and inconsistent SFP budgets, farmers do not view schools as reliable markets with which to market and sell their produce. The low demand for local produce thus also begets low supply.

Increase linkages between local farmers and schools to improve SFP sustainability and dietary diversity. Improving the consistency and quantity of SFP budgets will help increase the consistency of demand for local produce, giving farmers more incentives to grow and sell their produce to schools. HATUTAN programming should continue its capacity building and advocacy activities with the government of Timor-Leste in an effort to reduce delays in release of SFP funds, as well as explore other mechanisms to smooth SFP funding, including through increased involvement of PTAs.

HEALTH AND NUTRITION

Caregivers are highly knowledgeable about hygiene and nutrition practices. In general, more mothers were able to identify healthy practices at midline compared to baseline. However, it was not clear how their knowledge of hygiene and nutrition practices translated to practices as change in knowledge does not necessarily imply change in behavior. Investment in and implementation of a relevant, culturally appropriate social behavior change communication strategy will be crucial in understanding the link between intervention and behavior change.

Dietary diversity among women of childbearing age declined between baseline and midline in intervention municipalities, which may indicate poor nutrition practices that may have been affected by the ‘hungry season’ or limitations to financial access to purchase food during COVID-19. When looking at data on savings and loan usage, money is primarily used for food and education expenses, suggesting that household incomes may pose a major constraint on dietary diversity.

Caregivers reported consuming a predominantly carbohydrate-based diet with limited protein intake. Consumption of fruits and vegetables also declined, and men are likely to consume more than women. Meanwhile, dietary diversity of children 6-23 months old slightly improved, but the percentage of infants that meet the minimum acceptable diet remains low. On average, children in the treatment group consumed 2 food groups at the midline.

Most households had access to a toilet at home but less than half of treatment households reported having access to drinking water all year at midline. However, almost nine in ten reported collecting water from an improved water source. Urban households were slightly more likely than rural households to collect drinking water from an improved water source.

Medical expenses were not a predominant use of savings and loans. More than half of the households at the midline said that they do not have to pay to access health services. However, this is not an indication of accessibility of healthcare, but rather an indicator of affordability of health services. The data does not provide data to make conclusions in relation to healthcare-seeking behavior.

Strengthen activities that address gendered and economic barriers to health and nutrition.

Knowledge of good health and nutrition practices is high, but household incomes appear to remain a substantial barrier to the implementation of these practices. Activities that seek to improve household incomes or allow for consumption smoothing—such as farmer training and strengthening of VSLAs—may help households translate knowledge into practice. Additionally, women report low levels of decision-making power over major household decisions, including large household purchases and the sell or consumption of livestock. This dynamic may influence nutrition and health outcomes for children and other family members, as women have been found to spend a greater portion of household incomes on children, particularly through the purchase of more and healthier food and through spending on education. Strengthening activities that seek to improve the decision-making power of women may thus also help improve nutrition and health outcomes for families.

A sizable minority of respondents are unaware of COVID-19 preventive practices, and many respondents do not practice COVID-19 prevention behaviors. Twenty percent of respondents were not aware that handwashing could prevent COVID-19, 40% were not aware that wearing a mask could prevent COVID-19, and nearly 70% were not aware of the benefits of social distancing. In general, practices were reported less frequently than knowledge, suggesting that knowledge does not always translate into healthy behaviors.

Increase COVID-19 prevention behavior awareness and reinforce messaging on preventive practices. There is an urgent need to reinforce both knowledge and behaviors in order to prevent the spread of COVID-19 in Timor-Leste, particularly considering the escalating infection rates as of May 2021. In particular, there is a need to reinforce messaging on preventive practices with poor recall rates, such as social distancing, spitting in public, and avoiding gatherings.

ANNEX 1: RESEARCH QUESTIONS

The main objectives of the midline study were:

- To assess and highlight factors affecting effective, quality and efficient implementation of HATUTAN activities/interventions.
- To assess the progress (strengths and weaknesses) of the project (per each component) against stated outputs and outcomes to date; this will include an assessment of the relevance of the outputs and outcomes through a gender lens.
- To identify new barriers and trends associated to the ongoing COVID-19 crisis, providing recommendations on how to refine the intervention to mitigate its impact on project outcomes.
- To assess early evidence of changes in behavior and practices (both intended and unintended) and compare these with the changes that were expected to be promoted by project activities. Identify factors in the implementation or context that hold back or promote observed and intended changes.
- To assess the efficiency and effectiveness of project organization, management and coordination mechanisms, including quality and usefulness of partnership.
- To assess sustainability efforts to date and potential factors that may impede schools' ability to graduate and sustain activities post-project.
- To recommend the future orientation and emphasis of the project during its remaining time, including course corrections and adjustments to the Results Framework, project design, resource allocation, or implementation process as necessary.

Data collection was conducted for the following indicators, as outlined in the Terms of Reference:

- Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text
- Percentage of teachers adhering to improved learning practices in schools
- Percentage of schools with at least 80% of the teachers present during head counts
- Percentage of schools with reading corners in grade 2
- Number of teachers in target schools who demonstrate use of new and quality teaching techniques or tools as a result of USDA assistance
- Number of school administrators and officials in target schools who demonstrate use of new techniques or tools as a result of USDA assistance
- Percent of students with working memory scores equal or above 50%
- Percent of students who report that they did not consume any food during the school day
- Average student attendance rate in USDA supported classrooms/schools
- Percentage of parents (VSLA group members) using part of their savings or loans for education of their children
- Reduction in the number of days of absence from school due to illness
- Number of individuals who demonstrate use of new child health and nutrition practices as a result of USDA assistance
- Percentage of participants who are able to correctly identify keeping animals in the kitchen as a non-hygienic practice
- Number of individuals who demonstrate use of new safe food preparation and storage practices as a result of USDA assistance
- Percent of participants in program target groups (pregnant-lactating mothers, parents of school children, VSLA group members) who can identify at least three important nutrition/dietary recommendations
- Number of schools using an improved water source
- Number of schools using improved sanitation facilities

- Percent of targeted schools with improved food preparation and/or storage tools as a result of USDA assistance
- Number of policies, regulations, or administrative procedures in the following stages of development as a result of USDA assistance
- Value of new USG commitments, and new public and private sector investments leveraged by USDA to support food security and nutrition
- Percentage of schools purchasing food from local farmers for the school feeding program
- Percentage of VSLA groups integrating promotion and dissemination of health/hygiene/nutrition and specific gender messages in their regular activities

Evaluation questions include the following:

Design/Relevance

- How relevant/appropriate is the project concept and design to the local context and the education, health, and nutrition conditions?
- How well does the HATUTAN design match the needs, priorities and contexts of the target students, communities and other beneficiaries?
- Are there unforeseen assumptions and barriers? If so, what have been the extent and adequacy of adjustments?
- To what extent do the project activities and outputs contribute to intended effects and impact as declared in the Results Framework?

Management and Coordination

- How effective is the project's general implementation and management in terms of quality and efficiency of inputs and activities, including (but not limited to): adherence to work plans and budgets; management of food commodities; routine monitoring of project activities; and financial management?
- To what extent does the institutional set-up encourage involvement of key stakeholders (e.g. government agencies, local partners, and USDA)?
- What is the current level of cooperation (e.g. programmatic and organizational) among partners and what is the effectiveness of this partnership?

Effectiveness

- What is the achievement, to date, of HATUTAN activities, outputs and outcomes as detailed in the Results Framework?
- What are the major factors that have facilitated or impeded the progress of project implementation?
- For each component, which activities and implementation processes are acceptable to members of the target institutions, and communities and why?
- How well do implementation processes adhere to the underlying strategies, principles and established methodology and processes?
- Are there any specific sub-groups that benefitted more/less from project activities? If so, why?
- Which project strategies being implemented by the project have been more/less effective in contributing to positive results? Why?
- If targets are not being achieved, why not, and what have been the extent and adequacy of adjustments?
- To what extent has the project been successful in advocating for GOTL resources to fund infrastructure improvements (e.g. kitchens, storage rooms, latrines)? How many schools have received resources?
- Cost-effectiveness will be determined based on the triangulation of the size of the change observed in the outcomes of interest; the relative contribution of each intervention component to changes in outcomes; and the investment made in each of the intervention components.

Efficiency

- How efficiently have resources been allocated for project implementation?
- How could the efficiency of project support be improved?

Sustainability

- Are there specific concerns/opportunities in relation to the potential sustainability of the school feeding and literacy skills development strategies, considering the policy framework and mechanisms in place for its implementation?
- Level of adherence to intervention components (fidelity of implementation)
- Assessment of sustainability scores attained by schools, identifying sub-groups that are either highly successful (positive deviants) or lagging behind
- Is there evidence that certain benefits and/or activities are more likely to be sustained than others? If so, why?
- Are the five domains of sustainability identified by the project progressing at a similar pace or not? Why or why not?
- To what extent have local actors (community entities/associations, local government agencies, private sector, etc.) been engaged and committed (contributing time and resources) to ensuring sustainability? Which of them has demonstrated the greatest/least potential for contributing to sustainability? Why?
- What specific institutional capacity needs should the project address to foster greater host country ownership, engagement, and accountability? At the national level? At the municipal level? At the community level?
- What steps has the project taken to ensure the sustainability of adequate human, financial and other resources?
- What will help ensure greater likelihood of sustaining project achievements and activities post project in terms of sustained motivation, resources, capacity and linkages?
- The red flags or challenges that project implementers should be aware of and focus on during the second half of the project to further sustainability efforts.

Impact

- To what extent have project activities led to improved literacy among school-aged children?
- To what extent have project activities led to increased use of appropriate health, nutrition, hygiene/sanitation, and gender equality practices?
- What difference (early evidences of changes in knowledge, attitude, behavior, and practices) did project interventions make to the beneficiaries (considering the difference between what has happened with and what would have happened without the project)?
- Did any negative effects occur for beneficiaries? If so, what were they and why did they happen?
- To what extent has the project been successful in advocating for and securing timely distribution of resources and commodities to arrive at schools before the start of the school year (i.e. January)

Gender

- What differential outcomes and impacts (if any), are project activities having on women and girls? Men and boys?
- To what extent is the project actively addressing gender-specific barriers and constraints to education, health, nutrition and community engagement?
- To what extent has the project contributed to reducing incidences of violence, namely (a) domestic violence; (b) school-based violence; (c) sexual and gender-based violence?
- What specific actions and interventions has the project fostered to promote gender equality and how successful were they?

COVID-19 Impact

- How has the COVID-19 crisis impacted students, teachers and households? How does the negative impact of the crisis differ by gender, age and location?
- What adaptations are recommended to mitigate the persistent impact of the crisis on learning and student health?
- How can the SFP be adapted to continue to serve as a safety net for students in the context of a persistent pandemic, as well as in similar crises?

Learning and Recommendations

- What are the key/critical emerging learnings that HATUTAN should deepen to improve the quality of interventions?
- What should HATUTAN modify to improve implementation and achieve intended results; are any changes/adjustments in the results framework, in implementation strategies and approaches, in resource allocation and/or in management/coordination mechanisms necessary?

ANNEX 2: INDICATORS AND OUTCOMES

HATUTAN INDICATORS ASSESSED AT BASELINE AND MIDLINE

Indicator #	Indicator Description	Standard or Custom	Results		Targets				
			Baseline (Intervention Municipalities)	Midline (Intervention Municipalities)	Year 1	Year 2	Year 3	Year 4	Year 5/Life of Project
MGD SO1	Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text <i>Modified at baseline: Students were assessed 2-3 months after starting Grade 2</i>	Standard #1	12% can respond to 80% of questions correctly 9% male, 14% female	6% can respond to 80% of questions correctly 5% male, 8% female	N/A	N/A	15% (baseline +25%)	N/A	18% (baseline +50%)
MGD 1.1	Percentage of teacher adhering to improved learning practices in schools (using at least four engaging practices in class)	Custom	64% of teachers 61% male, 68% female 63% rural, 72% urban	69% of teachers 72% male, 67% female 68% rural, 78% urban	N/A	N/A	79% (baseline +20%)	N/A	83% (baseline +25%)
MGD 1.1.1	Number of schools with at least 80% of the teachers present during head counts	Custom	32% with 80% attendance recorded on the day of and day preceding the survey 39% with 80% attendance recorded on the day of the survey 55% with 80% attendance recorded on the day before the survey	44% with 80% attendance recorded on the day of and day preceding the survey 76% with 80% attendance recorded on the day of the survey 55% with 80% attendance recorded on the day before the survey 39% rural, 69% urban	N/A	29% (baseline +10%)	31% (baseline +20%)	34% (baseline +30%)	36% (baseline +40%)

			25% rural, 6% urban 18% Ainaro, 20% Ermera, 45% Liquica, 16% Manatuto	24% Ainaro, 44% Ermera, 55% Liquica, 73% Manatuto					
			37% of schools	56% of schools					
MGD 1.1.2	Percentage of schools with reading corners in grade 2	Custom	36% rural, 44% urban 36% Ainaro, 32% Ermera, 55% Liquica, 42% Manatuto	55% rural, 61% urban 70% Ainaro, 49% Ermera, 73% Liquica, 42% Manatuto	39% (baseline +5%)	41% (baseline +10%)	45% (baseline +20%)	48% (baseline +30%)	53% (baseline +40%)
			29% of students	28% of students					
MGD 1.2	Percent of students with working memory scores equal or above 50%	Custom	30% male, 28% female 29% rural, 30% urban 33% Ainaro, 26% Ermera, 35% Liquica, 27% Manatuto	26% male, 29% female 27% rural, 29% urban 24% Ainaro, 27% Ermera, 29% Liquica, 34% Manatuto	N/A	N/A	35% (baseline +20%)	N/A	41% (baseline +40%)
			14% of students	9% of students					
MGD 1.2.1	Percent of students who report that they did not consume any food during the school day	Custom	14% male, 13% female 12% age 5-7, 15% age 8- 10, 20% age 11-13 13% rural, 16% urban 14% Ainaro, 13% Ermera, 7% Liquica, 19% Manatuto	10% male, 9% female 9% age 5-7, 9% age 8-10, 10% age 11-13 9% rural, 11% urban 7% Ainaro, 10% Ermera, 7% Liquica, 13% Manatuto	N/A	N/A	9.0%	N/A	5.0%
			70% attendance rate	72% attendance rate					
MGD 1.3	Average student attendance rate in USDA supported classrooms/schools	Standard #2	69% male, 68% female 69% rural, 74% urban 63% Ainaro, 70% Ermera, 74% Liquica, 78% Manatuto	70% male, 75% female 71% rural, 75% urban 74% Ainaro, 69% Ermera, 66% Liquica, 79% Manatuto	70.0%	74% (baseline +5%)	77% (baseline +10%)	79% (baseline +13%)	81% (baseline +15%)

MGD 1.3.1	Percentage of parents (VSLA group members) using part of their savings or loans for education of their children	Custom	41% of parents (4.3% involved in savings groups) 51% Ainaro, 49% Ermera, 56% Liquica, 56% Manatuto	58% of parents (44% involved in savings groups) 82% of those trained on VSLAs 71% of those who took loans 55% Ainaro, 52% Ermera, 79% Liquica, 57% Manatuto	41.0%	50.0%	60.0%	70.0%	80.0%
MGD 1.3.2	Reduction in the number of days of absence from school due to illness	Custom	1.6 days missed due to illness 1.5 days male, 1.7 days female	0.52 days missed due to illness per week 0.51 days male, 0.53 days female	N/A	N/A	1.3 days (baseline -20%)	N/A	1.1 days (baseline -30%)
MGD 2.1	Percentage of participants who are able to correctly identify keeping animals in the kitchen as a non-hygienic practice	Custom	32.0%	77.9%	N/A	N/A	55% (baseline +20%)	N/A	80% (baseline +45%)
MGD 2.3	Percent of participants in program target groups (pregnant-lactating mothers, parents of school children, VSLA group members) who can identify at least three important nutrition/dietary recommendations (Mercy Corps)	Custom	46% 53% male, 45% female	65% 44% male, 65% female	N/A	N/A	60% (baseline +30%)	N/A	75% (baseline +63%)
MGD 2.4	Number of schools using an improved water source (WaterAid)	Standard #27	136	171	136	143	149	154	160
MGD 2.4	Number of schools using improved sanitation facilities (WaterAid)	Standard #28	156	175	156	159	163	165	165

MCGOVERN-DOLE STANDARD AND CUSTOM OUTCOMES

	Comparison Schools			Intervention Schools			Difference in Differences	
	BL	ML	Difference	BL	ML	Difference	DiD	p
Standard Outcome 1: Percentage of students who demonstrate that they can read and understand the meaning of grade level text (based on answering one comprehension question correctly)								
n	1,014	1,108		1,447	1,474			
Achieved	29.7%	14.7%	-15.0	23.5%	12.9%	-10.6	4.4	0.22
Standard Outcome 2: Percent of schools with an average student attendance rate of at least 80 percent								
n	25	56		37	70			
Achieved	24.0%	39.3%	15.3	29.7%	32.9%	3.2	-12.1	0.42
Standard Outcome 8: Number of schools with improved infrastructure (based on presence of four out of six components)								
n	28	71		98	95			
Achieved	21.4%	52.1%	30.7	62.2%	75.8%	13.6	-17.1	0.17
Standard Outcome 27: Number of schools with an improved water source								
n	88	88		98	98			
Achieved	54.6%	62.5%	7.9	61.2%	77.6%	16.4	8.4	0.40
Standard Outcome 28: Number of schools using improved sanitation facilities								
n	87	88		98	98			
Achieved	69.0%	70.5%	1.5	71.4%	79.6%	8.2	6.7	0.47
Custom 5: Percentage of teachers adhering to improved learning practices in schools (based on demonstrating four or more)								
n	45	87		98	98			
Achieved	71.1%	66.7%	-4.4	64.3%	69.4%	5.1	0.10	0.33
Custom 6: Percent of schools in which at least 80 percent of teachers were present on the day of data collection and the day prior								
n	67	61		66	70			
Achieved	31.3%	55.7%	24.4	31.8%	44.3%	12.5	-11.9	0.30
Custom 7: Percentage of schools with access to reading materials in classrooms								
n	90	88		99	98			
Achieved	47.8%	60.2%	12.4	56.6%	66.3%	9.7	-2.7	0.82
Custom 12: Percent of students who report they are attentive in class								
n	1,004	1,101		1,409	1,457			
Achieved	96.5%	94.6%	-1.9	95.5%	95.8%	0.3	2.4	0.10
Custom 13: Percentage of students who report that they did not consume any food during the school day								
n	1,012	1,107		1,442	1,470			
Achieved	11.7%	10.5%	-1.2	13.4%	10.5%	-2.9	-1.7	0.43
Custom 16: Percentage of days of absence from school due to illness								
n	374	615		478	723			
Achieved	48.0%	82.7%	34.7	47.4%	78.0%	30.6	-4.1	0.3
Custom 21: Percentage of participants who can identify important hygiene/sanitation practices								
n	362	625		163	982			
Achieved	88.7%	94.4%	5.7	80.4%	92.6%	12.2	6.5	0.08

Custom 23: Percentage of participants in program target groups (pregnant and lactating women, parents of school children, VSLA group members) who can identify at least three important nutrition/dietary recommendations								
n	366	596		482	982			
Achieved	53.8%	68.6%	14.8	45.8%	64.7%	18.9	4.1	0.36
Ministry of Agriculture Indicator: Number of schools procuring nutritious foods from local producers/farmers								
n	90	88		99	98			
No	11.1%	11.4%	0.3	4.0%	54.1%	50.0	49.8	<0.001***
Yes, sometimes	58.9%	68.2%	9.3	48.5%	36.7%	-11.8	-21.0	0.04**
Yes, all the time	30.0%	20.5%	-9.6	47.5%	9.2%	-38.3	-28.7	0.001***
Ministry of Agriculture Indicator: Percentage of daily school feeding foods procured from Timorese farmers								
n	80	78		95	45			
Dark green vegetables such as water spinach, lettuce, mustard, pumpkin leaves, cassava leaves	88.8%	73.1%	-15.7	90.5%	82.2%	-8.3	7.4	0.41
Pumpkin, carrot, purple sweet potato	68.8%	51.3%	-17.5	76.8%	53.3%	-23.5	-6.0	0.60
Potato, taro, yellow sweet potato, cassava	63.8%	50.0%	-13.8	75.8%	60.0%	-15.8	-2.0	0.86
Beans, peas, soybeans, peanuts	51.3%	26.9%	-24.3	66.3%	15.6%	-50.8	-26.4	0.01**
Beef, pork, sheep, goat, chicken, duck	48.8%	7.7%	-41.1	64.2%	13.3%	-50.9	-9.8	0.31
Eggs	41.3%	6.4%	-34.8	50.5%	2.2%	-48.3	-13.5	0.11
Rice, maize, bread	45.0%	44.9%	-0.1	46.3%	22.2%	-24.1	-24.0	0.04*
Cucumber, tomato, cabbage, eggplant	36.3%	6.4%	-29.8	46.3%	15.6%	-30.8	-0.9	0.92
Condiments	32.5%	2.6%	-29.9	25.3%	4.4%	-20.8	9.1	0.24
Fish (fresh or dried), shrimp, other seafood	23.8%	0.0%	-23.8	21.1%	4.4%	-16.6	7.1	0.31
Tofu, tempe	18.8%	1.3%	-17.5	17.9%	0.0%	-17.9	-0.4	0.94
Mango, papaya, honeydew melon, passionfruit, other yello fruits	10.0%	0.0%	-10.0	16.8%	8.9%	-8.0	2.0	0.76
Milk (not sweetened condensed milk)	5.0%	7.7%	2.7	10.5%	0.0%	-10.5	-13.2	0.009**
Coconut oil	5.0%	0.0%	-5.0	4.2%	2.2%	-2.0	3.0	0.44
Watermelon, tamarind, jackfruit	3.8%	0.0%	-3.8	4.2%	0.0%	-4.2	-0.5	0.88

* Significant at $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANNEX 3: DETAILED METHODOLOGICAL ANALYSIS

DEMOGRAPHIC COMPARISON OF BASELINE AND MIDLINE CROSS-SECTIONAL SAMPLES

We first analyze any differences in observable characteristics among students in the cross-sectional sample who took the EGRA. This sample includes all baseline students and the new cohort of grade 2 students who were contacted first at midline. Within the EGRA, data was collected on students' ages, genders, native languages, locations, and grades. While this data allows for a basic comparison of student traits across baseline and midline, there are a substantial number of unobserved characteristics (or characteristics observed only in the household survey, but not for all EGRA students) which could introduce bias to our results, such as student aptitude or economic status.

Within the cross-sectional sample, midline students were, on average, slightly older (7.7 years) than baseline students (7.6 years). There were also slight differences in gender composition: At baseline, 52% of students were male and 48% were female, while at midline, among the new cohort, 51% were male and 49% were female. While these differences in age and gender are slight, the tendency of older students to perform better in school and, within Timor-Leste, of girls to perform better than boys means that these demographic differences may introduce some bias to our results. Accordingly, we control for student age and gender when checking for the robustness of results found using the cross-sectional sample.

There were also some differences in native language among the cross-sectional sample at baseline and midline. At midline, the new cohort of students was somewhat more likely to speak Tetum-Prasa than at baseline (Table 147). This may be problematic because students who have Tetum-Prasa as their native tongue are likely to be at an advantage in school, particularly in early grades, because it is the language of instruction and examination. Other differences in native tongues may also be problematic if speakers of a particular language find it easier to understand Tetum, or if that language has more words/letter sounds in common with Tetum than another language. However, we note that the EGRA was designed to minimize the number of letters, words, and sounds included that would be unfamiliar to native speakers of a language other than Tetum-Prasa (see "Data Collection Tools – Early Grade Reading Assessment"); the most salient issue is thus that of Tetum native speakers. We therefore control for whether a student's native language is Tetum in our regression specification that checks for the robustness of the results.

Table 147: Differences in native language among cross-sectional sample

Mother tongue	Speakers at baseline	Speakers at midline, new cohort
n	2461	2582
Galolen	2.6%	2.4%
Kemak	19.6%	17.7%
Mambae	30.8%	26.1%
Bunak	4.2%	5.3%
Cairui	0.5%	0.5%
Mdiki	0.2%	0.0%
Makassae	0.1%	0.0%
Tetum-Prasa	64.2%	66.0%
Tetum-Terik	4.3%	6.6%
Tokodede	0.0%	7.1%
Other	14.6%	7.2%

Note: Does not sum to 100% due to ability to select multiple responses

Outside of these characteristics, we find little difference in the distribution of students by municipality, and all students in the cross-sectional cohort were in grade 2 at both baseline and midline. These five characteristics represent the extent to which we can detect observable differences across students; they do not, however, represent all potential differences across students. As described above, differences in characteristics such as ability and economic status may bias our results; lacking sufficient data for all students, we are not able to control for these differences.

Additionally—and notably—the COVID-19 pandemic has had a substantial effect on students at midline. Midline students have received significantly fewer days of instruction and total contact hours than comparable students at baseline, which is thus likely to worsen scores in a way unrelated to program impact. Because the impact of COVID-19 has been relatively homogenous across all students in Timor-Leste—school closures were instituted nationwide, as was the *Eskola ba Uma* program—this is not necessarily problematic for our analysis, as the impact of COVID-19 is essentially uniform across intervention and comparison groups. As a result, while we expect students' EGRA results to be worse at midline, in the absence of the HATUTAN program, we would expect results to be *consistently* worse across intervention and comparison groups. If we find that the intervention group performs relatively better than we would expect given the results of the comparison group—if, for example, the intervention group's scores worsen, but by fewer percentage points than the comparison group—we can conclude that the HATUTAN program may have had some effect at mitigating the negative impact of COVID-19 on learning.

This may be problematic only if the impact of COVID-19 was not homogenous across intervention and comparison groups; for instance, if some intervention municipalities are better-equipped to deal with COVID-19 through, for example, better access to radios to tune into at-home learning programs or smaller class sizes that do not necessitate division into shifts, then we may mistakenly attribute improvement in scores to program impact, rather than to the differentiated impact of COVID-19. We explore this further in the section “Differences Between Intervention and Comparison Groups.”

Outside of differences in students' characteristics recorded in the EGRA, we also analyze whether there are any observable differences in household survey respondents that may bias our results. We find some differences in head of household and caregiver education levels across the baseline and midline cohorts of the cross-sectional sample. At midline, heads of households had, on average, somewhat higher education levels, while caregivers had, on average, slightly lower education levels, although differences were generally minor. Differences in head of household and caregiver education levels could have implications for student literacy, nutrition, and hygiene, among other factors, as more highly-educated caregivers and heads of household may be more able to help their children learn or may have more knowledge of good hygiene and nutrition practices.

Considering head of household/caregiver occupation, the most substantive differences between the two samples are that heads of household were somewhat less likely to work as subsistence (own consumption) farmers, substantially more likely to work as farmers for both subsistence and sale, and somewhat more likely to be unemployed at midline than at baseline. Caregivers were also less likely to work as subsistence (own consumption) farmers and slightly more likely to work as farmers for both subsistence and sale at midline than at baseline, and were substantially more likely to be unemployed at midline. These differences could have implications for literacy and nutrition results as head of household/caregiver occupation is likely to be correlated with household income, and wealthier households may be more likely to be able to support their children's education or purchase more healthy foods. Additionally, unemployment or lack of economic stability may contribute to increased rates of gender-based violence, particularly in conjunction with the COVID-19 pandemic, which has seen increased rates of domestic violence in many countries. In this way, differences between employment status of respondents surveyed at baseline and midline may introduce bias to both EGRA results and results related to nutrition, hygiene, and gender-based violence.

Within the household survey, baseline and midline respondents reported speaking Tetum-Prasa at home at relatively similar rates: 71% at baseline and 70% at midline. Differences in language spoken at home were generally minor across baseline and midline groups.

In contrast, there were some differences in household sizes across baseline and midline. On average, the size of households at baseline was slightly larger (7.7 members) than at midline (7.4 members). Although average household size was fairly similar across baseline and midline, the distribution of household sizes varied somewhat, as did the composition of households. Baseline households had more female children on average (1.7 girls) than midline households (1.5 girls). Baseline households also had, on average, more children in school (2.9 children) than midline households (2.7 children). While it is difficult to understand the exact impact these differences in household size distribution may have, in general, household size is likely to have an impact on outcomes of interest: For example, larger households may have poorer nutrition outcomes because there are more mouths to feed, and households with more children in school may have poorer literacy outcomes because parents are less able to tutor each child—or better outcomes because children are able to help each other.

Midline households were more likely to report having savings (65%) than baseline households (51%). This difference could imply changes in income or in consumption-smoothing habits which would potentially have effects on outcomes such as nutrition, health, and schooling (as households have more money to buy nutritious foods, access healthcare services, or pay for school supplies). However, as the HATUTAN program did include a VSLA component, this variable is somewhat intertwined with program activities.

Household survey respondents were also asked to report on any disabilities faced by their second grade child, including physical and mental/emotional disabilities. Physical disabilities include trouble seeing, hearing, or walking, while mental/emotional disabilities include trouble with memory, self care, communication, and anxiety or depression. There were differences in reporting of disabilities across baseline and midline; at midline, 66% of households stated that their child had at least one mental disability compared to only 57% of baseline households, and 11% stated that their child had at least one physical disability compared to 8% of baseline households. Because students with disabilities may face more challenges to learn or to be attentive in class, differences in disability status may impact learning outcomes.

In addition to student- and household-level characteristics, we analyze differences across schools over baseline and midline as measured in the school survey, as differences in school characteristics may influence learning outcomes for students. On average, at midline, school directors had more years of experience (8.8 years) than at baseline (7.7 years). More-experienced directors may have better school management abilities or better ability to coach teachers, which would result in improved learning outcomes unrelated to program impact. However, at midline, school directors also had less education on average than at baseline. At midline, 20% of school directors had only a secondary degree, compared to 14% at baseline, and only 23% had a degree from a teacher training college at midline, compared to 28% of baseline directors. Less-educated school directors may be less able to provide training to teachers or may have weaker school management knowledge. This effect may thus work in the opposite direction as that of the greater experience of directors.

In general, there were only minor differences in the total number of teachers employed at schools at baseline and midline. However, there were some differences in the types of teachers employed for each cohort. At baseline, schools had on average more permanent teachers (5.0) than at midline (4.5) and fewer volunteer teachers (1.0) than at midline (1.3). As there may be differences in the education levels, training, or professionalism of volunteer teachers as opposed to permanent teachers, these differences in teacher type may have an impact on learning outcomes. The student-to-teacher ratio was also slightly higher at midline (19 students per teacher) than at baseline (17 students per teacher), which may further impact learning outcomes.

Finally, schools at midline were slightly less likely to have multigrade classes (48% of schools) than at baseline (52% of schools). Multigrade classes present a challenging learning and teaching environment, and thus schools with these classes may have worse learning outcomes. Schools at midline were also more likely to have morning shifts and less likely to have morning and afternoon shifts than at baseline; 80% of midline schools had morning shifts compared to 74% of baseline schools.

To better control for any bias that might arise from differences in school characteristics, we include school fixed effects (binary variables for each school) in a regression specification as a robustness check. This specification controls for possible structural differences in learning scores across schools due to characteristics inherent to those schools which do not change much over time or which change constantly over time, such as teacher experience or remoteness. The regression with both fixed effects and student-specific control variables is the most rigorous approach to deal with any potential confounding variables in the cross-sectional sample because it is able to control for any potential omitted variables that differ across schools but were not measured in the school survey.

DIFFERENCES BETWEEN INTERVENTION AND COMPARISON GROUPS

The above differences between baseline and midline groups in the cross-sectional sample do not necessarily pose a methodological challenge if characteristics vary uniformly across intervention and comparison groups. To illustrate, we again describe the impact of COVID-19. The pandemic had not occurred at baseline but had a significant impact at midline; it thus clearly had a heterogeneous impact on baseline and midline students. However, the impact of and response to COVID-19 was relatively uniform across all of Timor-Leste. As a result, while we expect midline students to perform worse than baseline students on learning assessments, we expect performance to be *uniformly* worse across intervention and comparison groups. The difference-in-differences model allows us to understand how learning outcomes differed across both baseline and midline and intervention and comparison groups; if, at midline, the learning outcomes of intervention groups declined by less than those of comparison groups, this may be suggestive of positive program impact.

This method of analysis only becomes problematic if characteristics change within just the intervention or the comparison group, rather than uniformly across both. For example, suppose the government of Timor-Leste implemented a program in two treatment municipalities that increased the number of trained teachers within treatment schools between baseline and midline. We may thus find that learning outcomes or the quality of instruction in treatment schools have improved relative to comparison schools, and mistakenly attribute this to the impact of the HATUTAN program, rather than to heterogeneous increases in the number of teachers due to a government program.

To attempt to control for potential differences between treatment and control groups, following selection of treatment schools, comparable sub-districts from comparison municipalities were matched with treatment sub-districts based on language, size, location, and typical livelihoods (Table 148). Comparison schools were selected from these sub-districts by identifying a set of schools with a similar average “remoteness” score as that of the treatment schools in the matching sub-district. Both treatment and comparison schools included only public schools that are not currently involved in interventions focused on reading and school feeding (other than nationwide programs that cover all municipalities). This selection process reduces some of the risk to validity described above; however, it was not possible to have perfect matches between treatment and comparison administrative posts. In some cases, the best matching comparison administrative post already had too many schools with similar interventions. In addition, there were two problems in matching languages between treatment and comparison groups. The Laclubar schools, for example, include students belonging to a very small linguistic group, which is not comparable to others, and while Manatuto Vila and Vemassee share many characteristics, the language makeup is different.

Table 148: Matched intervention and comparison administrative posts

Intervention municipality	Intervention administrative post	Comparison administrative post(s)
Ainaro	Ainaro	Liquidoe
	Hatubuilico	Turiscail/Liquidoe
	Hatu'udo	Zumalai
	Maubisse	Same
Ermera	Atsabe	Bobonaro
	Ermera Vila	Aileu
	Hatolia	Caelaco
	Letefoho	Caelaco/Bobonaro
Liquica	Railaco	Laulara
Manatuto	Maubara	Atabae
	Laclo	Remexio
	Laclubar	Alas/Fatuberliu
	Laleia	Vemassee
	Manatuto Vila	Vemassee
	Soibada	Alas/Fatuberliu

To better understand the potential impact of these issues, we analyze the characteristics described in the above section to see whether any of these characteristics also vary across intervention and comparison groups, and thus whether they may pose a methodological challenge. We note, as above, that our analysis is limited to only those variables observed in the EGRA or other surveys. Unobserved characteristics for which we are not able to control may also pose a threat to inferential validity.

We first analyze whether the COVID-19 pandemic may have had heterogeneous impacts on intervention and comparison groups. At midline, intervention schools had substantially higher numbers of enrolled students in grades 1 through 6 (156 students on average) and in grade 2 (31 students on average). Comparison schools, in contrast, had on average 108 total students enrolled and 22 students enrolled in grade 2. This latter finding is particularly notable as national COVID-19 restrictions require that classes with more than 25 students are divided into shifts; the larger class size of students in intervention schools may therefore have substantially impacted learning outcomes as COVID-19 restrictions necessitated a reduction of contact hours. Accordingly, in the school survey, 17% of intervention schools and only 12.5% of comparison schools reported operating in shorter shifts due to COVID-19 restrictions, and 2% of intervention schools (two schools) reported that grade 2 classes were not operating every day, compared to 0% of comparison schools.

However, in contrast to these results, both intervention and comparison schools reported that grade 2 students were attending class for an approximately equivalent number of hours per day—3.67 in comparison schools and 3.74 in intervention schools. While the reason for the reduction in contact hours in comparison schools is not entirely clear given the above data on classroom shifts, this suggests that the learning of students in intervention and comparison schools may have been similarly affected by COVID-19.

Looking now at demographic differences recorded in the EGRA, we do not find any significant differences in student ages across intervention and comparison groups at midline compared to baseline. At baseline, intervention students were, on average, around 0.2 years older than comparison students; at midline, intervention students remain around 0.2 years older than comparison students. Any impact on learning outcomes due to age will thus be controlled for in the difference-in-differences analysis. Similarly, we find only slight differences in gender composition: At baseline, there were

around 0.9 percentage points more male students in both comparison groups than intervention groups, while at midline, there were around 1 percentage point more male students in comparison groups.

The language spoken by students is somewhat more problematic. At baseline, students in treatment schools were around 6 percentage points more likely to speak Tetum-Prasa as their native language than students in comparison schools. At midline, in contrast, students in treatment schools were only 2 percentage points more likely to speak Tetum as their native language than students in comparison schools. In general, students who speak Tetum natively are more likely to perform well on learning assessments, as instruction and assessment both occur in Tetum. At midline, because relatively more students in comparison schools speak Tetum, we would thus expect midline comparison students to perform relatively better than baseline comparison students on the EGRA, all else held constant. As a result, we may underestimate the impact of the HATUTAN program on learning outcomes. To reduce the potential impact of this issue, as described above, we control for whether a student's native language is Tetum in some regression specifications.

Within the household survey, there may also be some problematic differences in head of household and caregiver education. At baseline, heads of household in treatment areas were 8 percentage points more likely and caregivers 5 percentage points more likely to have no education than in comparison areas. However, at midline, heads of household in treatment areas were only 1 percentage point more likely and caregivers 4 percentage points more likely to have no education than in comparison areas. In other words, at midline, heads of household and, to a lesser extent, caregivers in treatment areas had somewhat more education than at baseline. As described above, head of household/caregiver education has wide-ranging implications for outcomes in literacy, nutrition, and hygiene, among other areas; as a result, outcomes may improve in treatment areas at midline due to changes in the amount of education of household members, rather than due to program impact.

Looking at household members' occupations, the most salient differences for caregivers were in unemployment. At baseline, caregivers in comparison areas were 3.5 percentage points more likely to be unemployed than those in treatment areas; at midline, this gap had widened to 5 percentage points. In contrast, for heads of household, there were no relative changes in unemployment across treatment and comparison groups, but employment in farming changed substantially. At baseline, heads of household in treatment groups were 1 percentage point more likely to be farmers than in comparison groups; at midline, heads of household in treatment groups were instead 3 percentage points less likely to be farmers. This change becomes even more substantial when farming is split into subsistence and for sale: Baseline heads of household in treatment areas were 2 percentage points less likely to be subsistence farmers than in comparison areas but 2 percentage points more likely at midline, and were 3 percentage points more likely to farm for sale but 5 percentage points less likely at midline. This dynamic is complicated; however, overall, it points to changes in livelihoods and possibly economic status that may have differentiated implications for outcomes in treatment and comparison areas which could bias results.

As with students' native languages, there were some changes in language spoken at home across treatment and comparison groups. At baseline, comparison households were around 10 percentage points more likely to speak Tetum at home than treatment households; at midline, this gap had declined to only 4 percentage points. Students who speak Tetum at home may have better learning outcomes, particularly at lower grade levels; as such, the relative increase in Tetum-speaking treatment households at midline may bias learning assessment results upward for students in these households.

Outside of these household characteristics, we find no significant relative differences in household size or savings across intervention and comparison groups at baseline and midline. We also do not find a significant relative difference in households' reporting of physical or mental disabilities in their second grade child.

Continuing on to differences in schools, we find that at midline, directors in treatment schools had relatively more experience than at baseline. At baseline, directors in treatment schools had, on

average, 1.4 years less experience than those in comparison schools; at midline, directors in treatment schools had 0.1 years less experience than in comparison schools. Similarly, in treatment schools at midline, directors were relatively less likely to have only a secondary school degree and more likely to have a teacher training degree given results at baseline. This may have implications for results if, at midline, relatively more-experienced and better-educated school directors in treatment schools were better able to provide training to teachers or manage their schools, which may improve learning outcomes outside of HATUTAN program impact.

At midline, while the student-to-teacher ratio increased in both intervention and comparison schools, it increased relatively more in comparison schools. At baseline, treatment schools had an average student-to-teacher ratio of 19.0:1 and comparison schools 14.3:1; at midline, in contrast, the ratio was 19.6:1 and 18.1:1. Perhaps as a result of changing enrollment levels, at midline, comparison schools had relatively more multigrade classes compared to treatment schools, in which the average number of multigrade classes actually decreased.

Overall, the net impact of these differences between treatment and comparison areas is difficult to unpack. Some differences—for example, changes in students' native languages—may lead us to underestimate the impact of the HATUTAN program on learning outcomes. Others, such as the relative improvement in student-to-teacher ratios in treatment schools compared to comparison schools, may lead us to overestimate program impact. We bring up these issues not to imply that the methodology or results used in this report are invalid, but rather to systematically analyze potential pitfalls to inference, justify the use of control variables (such as those for student language or school fixed effects) where needed, and caution against overinterpretation of results.

As a result of our methodological analysis, we use a variety of regression specifications within the report of increasing methodological rigor, particularly focusing on EGRA results and the school survey. For the EGRA, we report results using a difference-in-differences model without controls. We then check for robustness using a difference-in-differences model which controls for student gender, age, and native language, and an additional model which controls for the aforementioned variables as well as school fixed effects. For outcomes related to schools, such as the quality of instruction, our robustness models include school fixed effects which control for the potential impact of variables such as director experience and teacher-to-student ratio, as well as, when relevant, classroom-level controls including teacher gender, education, and experience. We also control for the type of school—central or filial—as outcomes may vary across these school types given different levels of remoteness and access to resources.

ANALYSIS OF PANEL DATA AND RATES OF ATTRITION

The panel data—the cohort of students assessed at both baseline and midline—allows us to further check for the robustness of cross-sectional results. However, as described in the introduction of this section, a panel design is only inferentially valid if panel attrition is minimized. If students are not re-contacted and are not replaced by similar students, the statistical power of the study is weakened. Furthermore, if there are systematic differences between students who are re-contacted and students who fall out of the sample, results may be biased.

In this section, we analyze re-contact rates for the midline evaluation. We use the term re-contact to refer to a student who participated in the baseline evaluation and who was successfully located and interviewed in the midline evaluation. Attrition rates, in contrast, refer to students who participated in the baseline evaluation but who were not interviewed in the midline evaluation—the inverse of re-contact rates. At midline, 2,069 students were re-contacted out of the 2,461 total students evaluated at baseline—an 84% re-contact rate. Re-contact rates were roughly equal across treatment and comparison schools: 83.6% of comparison students and 84.4% of intervention students were re-contacted at midline. At midline, 100% of schools were re-contacted, and no new schools were added to the sample. Overall, these re-contact rates are relatively high.

However, beyond aggregate re-contact rates, an additional methodological concern is the extent to which attrition or re-contact is non-randomly distributed across the sample. A variety of factors can affect attrition rates. Attrition that is as if random—i.e., not correlated with the outcomes of interest, such as learning—does not produce bias in conclusions regarding change in program outcomes over time. The only methodological concern in that case is a reduction in sample size that may reduce the statistical power of the analysis. In contrast, non-random attrition can produce bias in estimates of program impact. As described above, if students who are predisposed to lower learning outcomes tend to fall out of the sample over time, we will overestimate the improvement in learning scores that has occurred over time. However, importantly, if attrition occurs equivalently in treatment and comparison schools, the results will not be biased, even in the case of non-random attrition. In other words, if low-performing students fall out of the sample at higher rates than high-performing students, but attrition of these students occurs at equal rates in the treatment and comparison schools, there will be no bias in our overall estimates of program impact.

To understand if attrition operates in a systematically different way between intervention and comparison communities, we analyze re-contact rates by several key demographic characteristics. Looking first at municipality, we find that re-contact rates are slightly higher, but similar, in comparison municipalities than in treatment municipalities—83% and 82% respectively. However, re-contact rates do vary by municipality (Table 149). Re-contact rates were particularly high in Aileu, Baucau, Manatuto, and Manufahi, and low in Ainaro and Bobonaro. If student, household, or school characteristics in some municipalities are substantially different than others, this heterogeneity in re-contact rates by municipality may affect inferential validity.

Table 149: Re-contact rates by municipality

Municipality	Group	Re-contact rate	Total re-contacted students
Aileu	Comparison	87.6%	205
Ainaro	Treatment	79.2%	285
Baucau	Comparison	91.7%	44
Bobonaro	Comparison	75.8%	270
Covalima	Comparison	80.8%	118
Ermera	Treatment	84.8%	535
Liquica	Treatment	83.2%	158
Manatuto	Treatment	91.4%	243
Manufahi	Comparison	91.7%	211

Looking at the demographic characteristics of students, we find that re-contact rates were higher for girls than for boys—86% of girls were re-contacted but only 82% of boys. However, re-contact rates for girls and boys were fairly similar across treatment and comparison schools: within comparison schools, 82.2% of boys and 85.2% of girls were re-contacted, while within treatment schools, 82.1% of boys and 86.8% of girls were re-contacted. Because girls tend to perform better in school than boys in Timor-Leste, the slightly higher re-contact rates for girls within treatment schools may have a minor effect on our estimates of program impact; as such, we control for student gender in some regression specifications.

We also find that successful re-contact is correlated with age: The likelihood that a student was re-contacted tends to decrease as the average age of the student increases (Table 150). In intervention schools, however, a greater percent of older students—those ages 8 or more—were re-contacted than in comparison schools. The potential impact of this dynamic is complicated; while older students perform better in school on average, it is important to consider the reasons why an older student may have dropped out of school and not been re-contacted. These students may be older because they

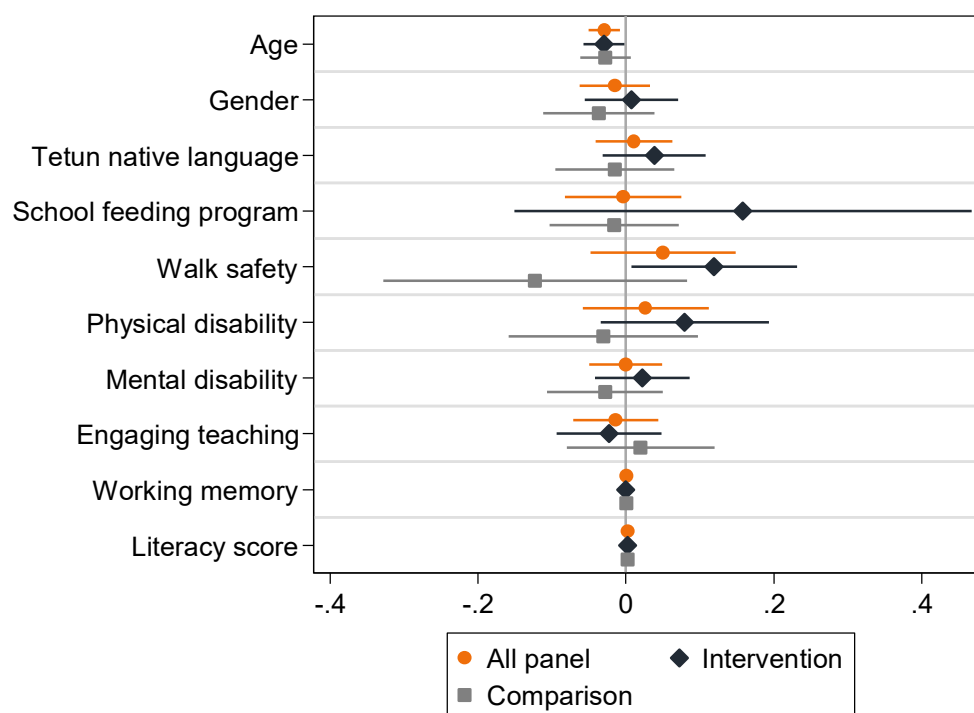
have had to repeat a grade, and as such, may actually have worse-than-average literacy abilities. If this is the case, then the differences in rates of re-contact across intervention and comparison groups may lead us to underestimate program impact.

Table 150: Re-contact rates by student age

	Intervention		Comparison	
Age	Re-contact rate	Total re-contacted students	Re-contact rate	Total re-contacted students
6 or less	92.9%	91	85.6%	77
7	85.8%	385	88.7%	331
8	85.0%	294	80.3%	191
9	85.9%	152	83.3%	80
10 or more	79.9%	111	70.6%	48

To analyze other potential predictors of re-contact, in the below figure, we report the results of a linear regression model predicting successful re-contact. The variables on the y-axis of the graph are included as predictors in the regression model; each model also controls for municipality (not shown in the figure). The results show the regression coefficient—the relationship between belonging to a sub-group, such as non-Tetum native speakers, and being successfully re-contacted—and the 95 percent confidence interval, which indicates our confidence in the result. A confidence interval that overlaps zero implies a low level of confidence that the regression coefficient is a predictor of re-contact rates.

Figure 24: Predictors of successful re-contact



As described above and shown in this figure, age is generally a predictor of successful re-contact: Older students are less likely to be re-contacted. Otherwise, we find few strong predictors of successful re-contact. The only significant and substantial effect we find is for students who report being able to

walk to school safely, who are significantly more likely to be re-contacted in intervention schools. If a student's walk is unsafe due to issues such as rain or poor roads, the student may miss more days of school or the enumerator may have more difficulty contacting the student due to remoteness and poor travel conditions. This may be a more pertinent issue in intervention municipalities if these municipalities are, for example, more vulnerable to natural disasters. There is, however, no significant difference in re-contact rates by walk safety across intervention and comparison groups; this dynamic thus does not pose a large risk to inferential validity.

Within this model, we also analyze whether working memory or literacy scores are significant predictors of re-contact. Working memory was not a significant predictor for any group; this suggests that innate student ability may not be affecting attrition in the evaluation, a positive finding for inferential validity. Furthermore, while the coefficient of literacy score on re-contact rates is significantly different than zero for the intervention group and all panel students, the effect size is very small: A one point increase in literacy score only corresponds to a 0.1 percentage point increase in likelihood of re-contact. As such, re-contact does not seem to be strongly correlated with literacy scores. Furthermore, this effect is relatively consistent across both intervention and comparison groups.

Overall, while these findings suggest that attrition is not entirely as-if random, it appears to operate similarly across both intervention and comparison schools. As a result, while we rely first and foremost on the cross-sectional sample for analysis, the panel sample provides strong data through which to check the robustness of our results.

ANNEX 4: DATA COLLECTION TOOLS

EARLY GRADE READING ASSESSMENT

- a. Ask for the consent of the Grade 2 teacher to conduct this activity with some of his/her students. Explain that the activity will take about 10 minutes. Explain to the teachers that the students will be randomly selected from the attendance list.
- b. In order to select 20 students from the Grade 2 class, enter the total number of students in the attendance list in the random number generator. You'll get a random number. Look for the number in the attendance list and ask the corresponding student if he/she would like to participate in the activity or not. If he/she doesn't want to join, generate another random number and call the corresponding student. If this student is absent or has dropped out of school, generate another number and so forth.
- c. Explain to the teacher that the students may feel shy or afraid if doing this activity in the presence of the teacher. It is recommended that the researcher conducts the activity alone with the child, although the teacher can watch from a distance and observe if the child is comfortable or not. If there is any issue, the activity can be stopped immediately.
- d. Ask for the child's consent to conduct this activity with him/her. If the child wants to stop the activity, he/she can do so at any time. Accept his/her decision and give thanks for the cooperation.

Questions for the child

- I. Have you eaten anything today?
☐ 1 Yes ☐ 0 No
- II. Do you feel that you can pay attention to the teacher?
☐ 1 Yes ☐ 0 No

Letter recognition

Please read the letters.

[Ask the child to read from the left to the right, from the top to the bottom. Mark in the tablet if the letter was read correctly or not]

m	i	a	L	T	s	u	N	e	R
1	2	3	4	5	6	7	8	9	10
B	o	k	t	d	v	E	F	U	N
11	12	13	14	15	16	17	18	19	20
h	t	j	x	p	Z	G	A	f	r
21	22	23	24	25	26	27	28	29	30
B	l	O	K	r	l	v	f	M	S
31	32	33	34	35	36	37	38	39	40
x	z	w	e	H	P	u	R	d	H
41	42	43	44	45	46	47	48	49	50
J	G	b	á	i	n	K	T	F	m
51	52	53	54	55	56	57	58	59	60
h	k	í	P	R	ñ	e	V	D	a
61	62	63	64	65	66	67	68	69	70
j	s	L	B	o	u	á	g	l	E
71	72	73	74	75	76	77	78	79	80
Z	k	P	o	L	n	M	J	A	b
81	82	83	84	85	86	87	88	89	90
Ú	é	v	H	g	í	z	ñ	W	r
91	92	93	94	95	96	97	98	99	100

Non-word reading

Now I'd like to ask you to read aloud the following words.

[Ask the child to read from the left to the right, from the top to the bottom.]

Give an example to the child to avoid confusing this task with the previous one: "Do you see the letters n and o here? This is the word "no".

Use the stopwatch to mark 60s, but do not tell the child that you are marking the time. Mark the number corresponding to the last word read by the child when reaching 60s.]

io	oa	due	biu	nai	beo	obu	alu	emi	eti
1	2	3	4	5	6	7	8	9	10
ulo	ime	dui	tuko	deda	naki	kusa	bato	numi	bima
11	12	13	14	15	16	17	18	19	20
sima	ruse	numi	niba	moke	telu	madi	ein	uan	bian
21	22	23	24	25	26	27	28	29	30
doin	ulus	mien	muon	edon	miik	anon	bais	elus	inis
31	32	33	34	35	36	37	38	39	40
moen	koar	daus	diur	boon	molun	dakal	kenas	milur	nemar
41	42	43	44	45	46	47	48	49	50
kadik	taden	biras	sodes	norau	salin	adiri	Ladis	amiluki	dokunar
51	52	53	54	55	56	57	58	59	60

Number of the last word read: ____

Mark the time spent (if the child spent one minute, mark 60s):

Frequent word reading

Please read the following words aloud.

[Ask the child to read from the left to the right, from the top to the bottom.]

Use the stopwatch to mark 60s, but do not tell the child that you are marking the time. Mark the number corresponding to the last word read by the child when reaching 60s.]

au	la	ita	ema	bei	tau	sai	nia	etu	rai
1	2	3	4	5	6	7	8	9	10
rua	iha	ida	busa	mana	soru	manu	same	tama	sira
11	12	13	14	15	16	17	18	19	20
taka	kuda	loke	toba	semu	moos	bani	hili	oan	ain
21	22	23	24	25	26	27	28	29	30
inan	ulun	aman	ibun	tein	maun	biin	inus	boot	udan
31	32	33	34	35	36	37	38	39	40
anin	isin	naan	tuur	Aileu	lalar	liman	talin	manas	tulun
41	42	43	44	45	46	47	48	49	50
talas	karau	besik	tudik	nanal	kabun	laran	belun	malirin	tuirmai
51	52	53	54	55	56	57	58	59	60

Number of the last word read: ____

Mark the time spent (if the child spent one minute, mark 60s):

Is the child able to recognize letters, to read words, or unable to read at all?

☐ 1 Cannot read any letters

☐ 2 Recognize letters only

☐ 3 Read words

[If the child cannot recognize letters or just recognize letters (cannot read any words), stop the activity here and mark “no response” in all subsequent questions. Tell a story to the child so the he/she feels ok about this activity. Please don’t simply send the child back to the class; he/she may feel humiliated because of his/her inability to read.]

Oral reading fluency - story

Please read the following story aloud for me.

[Ask the child to read from the left to the right, from the top to the bottom. Mark one minute in the stopwatch. Once finished, mark the number corresponding to the last word read by the child]

Anoi no Asu

Anoi¹ iha² asu³ ida⁴. Anoi⁵ nia⁶ asu⁷ metan⁸ no⁹ furak¹⁰.

Loron¹¹ ida¹², Anoi¹³ lori¹⁴ nia¹⁵ asu¹⁶ ba¹⁷ toos¹⁸.

Anoi¹⁹ sente²⁰ dukur²¹ no²² toba²³. Asu²⁴ halimar²⁵ no²⁶ halai²⁷ ba²⁸ dook²⁹.

Anoi³⁰ buka³¹ asu³² to'o³³ kole³⁴. Anoi³⁵ hetan³⁶ asu³⁷ iha³⁸ mota³⁹ sorin⁴⁰.

Asu⁴¹ tuur⁴² besik⁴³ Anoi⁴⁴ nia⁴⁵ Apá⁴⁶.

Apá⁴⁷ kail⁴⁸ ikan⁴⁹ barak⁵⁰ ona⁵¹. Apá⁵² dehan⁵³, "Mai⁵⁴ ita⁵⁵ ba⁵⁶ uma⁵⁷ hodi⁵⁸ han⁵⁹ ikan⁶⁰ tunu⁶¹!"

Reading comprehension – Level 1

Now I'm going to give you some time to read the story till the end.

[Allow the child to spend as much time as needed to read the story till the end]

Anoi and the dog

Anoi has a dog. Anoi's dog is black and pretty.

One day, Anoi brought her dog to the farm.

Anoi was tired and fell asleep. The dog was playing and run away.

Anoi looked for the dog until she got tired. Anoi found the dog on the margin of the creek.

The dog was sitting next to Anoi's Dad.

Dad had already caught a lot of fish. Dad said, "Let's go home to eat grilled fish!"

I'll ask a few questions. Could you please respond them?

[Ask the following questions to the child; allow enough time for him/her to respond each question]

Questions:

- What color is the dog?
[Black]
- Where did the dog run to?
[The dog ran to the creek (or to the margins of the creek)]
- Who was sitting with the dog on the creek side?
[Anoi's Dad]
- What was Dad doing at the creek?
[Dad was fishing]
- What will Anoi and her Dad eat?
[They will eat grilled fish.]

Reading comprehension - Level 2

Please read the following story:

Akai and his older sister live with their dad. Early in the morning, Akai was waiting for his sister to go to school. The sister had gone to fetch water with their dad. To help them, Akai cooked breakfast. When they got back, Akai served porridge to his sister. The sister was happy as she was feeling hungry already. Because they share the chores, Akai and his sister manage to get to school on time, even though they have to walk far.

The teacher met Akai's dad at the market. The teacher said, "Your children, Maria and Akai, are very hardworking." The dad responded, "My children help each other. That's why they are successful."

I'll ask a few questions. Could you please respond them?

[Ask the following questions to the child; allow enough time for him/her to respond each question]

-Who cooked the porridge in the morning?

[response: Akai cooked the porridge]

-How do Akai and his sister manage to arrive at school on time?

[Because they share the chores.]

-What is the name of Akai's sister?

[Maria.]

-What did the teacher say to the father?

[Your children are very hardworking]

-The father think that his children are successful because they do what?

[Because they help each other.]

Working Memory Assessment

I'm going to show you some drawings. Please try to remember them.

[Show each image separately to the student, saying out loud what is in each image (ball, cat, etc). After showing all images to the student, flip the cards so the student cannot see them anymore]

Instructions for Re-contacting Students

1. You have received a list of children who have participated in the baseline study. The list includes information on their current grades. You will locate each one of those children and conduct the assessment with them.
2. Go to the relevant grade/class and ask the teacher's permission to conduct the assessment with the child whose name is in the list. If the child is not in class on that day, you can conduct the assessment in his/her home instead (please ask the parents' permission first). If the child has moved to another location, is sick or has dropped out of school altogether, there is no need to conduct the assessment.
3. Ask the child if he/she agrees to participate, using the script provided. If the child refuses to participate, it's ok. If the child refuses but the teacher insists, decline politely and reassure the teacher and the child that this is not an issue. Please locate the next child and conduct the assessment with him/her instead.
4. DO NOT REPLACE THE CHILD WITH ANYONE ELSE FROM THE SAME CLASS. IF THE TEACHER DOESN'T WANT YOU TO CONDUCT THE ASSESSMENT WITH THIS CHILD AND SUGGESTS DOING IT WITH ANOTHER CHILD, POLITELY DECLINE AND INSIST IN CONDUCTING THE ASSESSMENT WITH THE ORIGINAL CHILD.
5. If the child agrees to participate, follow the same procedure as above. Remember that the child is free to stop the activity at any time.

CLASSROOM OBSERVATION TOOL

Description	Choice options
School location	1 Ainaro, 2 Ermera, 3 Liquica, 4 Manatuto
Enumerator code	1 Ainaro, 2 Hatu-udo, 3 Hatubuilico, 4 Maubisse, 5 Ermera, 6 Letefoho, 7 Atsabe, 8 Hatolia, 9 Railaco, 10 Liquica, 11 Maubara, 12 Bazartete, 13 Manatuto Vila, 14 Lacro, 15 Laclubar, 16 Laleia, 17 Barique-Natarbora, 18 Soibada
Enumerator name	
Today's date	
Municipality	
Subdistrict	
Village	
Hamlet	
EMIS	
School name	
is this a central or a filial school?	1 Central basic school, 2 Filial school
how many Grade 2 classes exist in this school?	
Choose a grade 2 class to carry out the observation. If there is only one class, there is no need to choose - just conduct the observation there. If there is more than one grade 2 class, enter the total number of classes in the "random number generator" and generate a random number. If grade 2a=1; grade 2b=2; 2c=3; 4d=4. Conduct the observation in the selected class. DON'T FORGET - YOU'LL ONLY OBSERVE TETUM CLASSES.	
[REQUEST THE DIRECTOR/COORDINATOR'S CONSENT BEFORE ASKING THE TEACHER'S CONSENT AS FOLLOWING] How are you doing? I am interviewing on behalf of the HATUTAN project, which aims to support education and children's nutrition in Timor-Leste. I would like to talk to you today and ask for your permission to sit in your class and observe the students. I will write down some notes to use in our research but we will not mention you by name or share your personal details with anybody outside of our team. The notes about your class will be mixed with the answers from other schools, and therefore the information for an individual school will not be identified. This observation will help us to learn more about the status of children's education in Timor-Leste. We will use this information to adjust our project's design. Do you accept to be part of this observation?	1 yes, 0 no
Teacher name	
Is the teacher a man or a woman?	1 male, 2 female
ASK THE TEACHER: How old are you?	
ASK THE TEACHER: What is your highest level of education?	1 secondary school, 2 teaching diploma (minimum qualification), 3 diploma issued by teacher training institute or faculty of education, 4 undergraduate course in another area (not education), 5 post-graduation, 99 don't know

ASK THE TEACHER: How many years have you been teaching in this school?

ASK THE TEACHER: How many years have you been teaching Grade 2 (at this school and in other schools)?

VERY IMPORTANT: OBSERVE ONLY THE TETUM CLASS.
PLEASE WAIT UNTIL THE TETUM CLASS HAS STARTED.

How many boys are attending the Grade 2 class?

How many girls are attending the Grade 2 class?

Write the name of the lesson the teacher is teaching now (ask the teacher or check the blackboard)

Start the observation. Select a part of the class to focus attention on.

Students spend most of the time copying from the board	1 observed, 99 did not observe
Students spend most of the time repeating after the teacher	1 observed, 99 did not observe
Students participate in reading activities with others (group reading, participating in reading games)	1 observed, 99 did not observe
Students read by themselves	1 observed, 99 did not observe
Teacher reads to the students	1 observed, 99 did not observe
Teacher uses the reading corner in literacy activities	1 observed, 99 did not observe
Teacher uses the Lafaek magazine in literacy activities	1 observed, 99 did not observe
Teacher uses games or exercises and students participate actively in those	1 observed, 99 did not observe
Teacher asks open questions (with more than one answer) - asks the students to reflect / think	1 observed, 99 did not observe
Teacher asks the students' opinion ("what do you think?" "what do you like?")	1 observed, 99 did not observe
When the student does not participate in class, the teacher calls on his/her and tries to engage him/her in activities	1 observed, 99 did not observe
Students work together in groups	1 observed, 99 did not observe
male and female students have equal access to desks, seats, learning materials (if some of them are sharing, male and female are sharing in equal proportions)	1 observed, 99 did not observe
Girls have less access to desks, seats and learning materials	1 observed, 99 did not observe
Teacher encourages male students ("good answer", "you are almost there")	1 observed, 99 did not observe
Teacher encourages female students ("good answer", "you are almost there")	1 observed, 99 did not observe
Teacher asks questions to male students	1 observed, 99 did not observe
Teacher asks questions to female students	1 observed, 99 did not observe
Teacher asks questions primarily to boys or girls (not to all students)	1 observed, 99 did not observe
Teacher uses angry voice tone or harsh language with students	1 observed, 99 did not observe
Teacher uses angry voice tone or harsh language with male students	1 observed, 99 did not observe
Teacher uses angry voice tone or harsh language with female students	1 observed, 99 did not observe
Teacher uses corporal punishment with girls	1 observed, 99 did not observe
Teacher uses corporal punishment with boys	1 observed, 99 did not observe

Select 10 students to observe for one minute. How many students are paying attention to the topic explained by the teacher?	
Wait until the class is over and then ask the following questions of the teacher.	
ASK THE TEACHER: Do you use formative assessments? Ask the teacher if he/she has records of formative assessments and their results. Mark "Observed" if there is documentation available; if not, thank the teacher and mark "not observed".	1 observed, 99 did not observe
ASK THE TEACHER: Do you use the Lafaek teachers' magazine?	1 yes, 0 no, 99 don't know
ASK THE TEACHER: How do you use the Lafaek teachers' magazine?	1 classroom management, 2 teaching mathematics, 3 teaching to read, 4 teaching Portuguese, 5 teaching Tetun, 6 teaching science, 7 health, 8 learn from other teachers' experiences, 9 information from the Ministry of Education, 10 learn about Timor-Leste, 11 peace education, 12 other
Other (explain)	
ASK THE TEACHER: How do you use the Lafaek Ki'ik magazine?	1 students practice reading in class, 2 prepare homework, 3 read stories to students, 4 games with students, 5 teach students to count, 6 teach students to read, 7 learn about Timor-Leste, 8 teach about science, 9 teach about health, 10 teach mathematics, 11 teach arts/ crafts, 12 teach Tetun, 13 other
Other (explain)	
ASK THE TEACHER: What are the main challenges to use the Lafaek teachers' magazine?	1 the Tetum used is too difficult, 2 the methodology is not appropriate, 3 activities are difficult to implement, 4 activities do not match the curriculum, 5 the topics are not appropriate for young children, 6 lack of materials, 7 director/coordinator does not authorize to use, 8 other
Other (explain)	
ASK THE TEACHER: What are the main challenges to use the Lafaek Ki'ik magazine?	1 the Tetum used is too difficult, 2 children don't speak Tetum, 3 stories are too long, 4 games are too complicated, 5 the math is too difficult for young children, 6 inappropriate drawings, 7 topics do not match the curriculum, 8 children don't want to use the magazine/ don't like it, 9 director/coordinator does not authorize to use, 10 other
Other (explain)	

SCHOOL SURVEY TOOL

Description	Choice options
GPS	
Enumerator code	
Enumerator name	
today's date	
municipality	1 Ainaro, 2 Ermera, 3 Liquica, 4 Manatuto
subdistrict	1 Ainaro, 2 Hato-Udo, 3 Hatobuilico, 4 Maubisse, 5 Ermera, 6 Letefoho, 7 Atsabe, 8 Railaco, 9 Liquica, 10 Maubara, 11 Bazartete, 12 Manatuto Vila, 13 Lacro, 14 Laclubar, 15 Laleia, 16 Barique-Natarbora, 17 Soibada
village	
hamlet	
School EMIS	
school name	
is this a central or a filial school?	1 Central basic school, 2 Filial school
Does the school have morning and afternoon shifts?	dadeer: morning, dadeer_lorokraik: morning and afternoon, lorokraik: afternoon
Does this school have a preschool?	1 yes, 0 no
CONSENT: How are you? I am interviewing on behalf of the HATUTAN project, which aims to support education and children's nutrition in Timor-Leste. I would like to talk to you today and ask for your permission to interview you. I'd like to ask some questions about this school. I will write your answers to use them in our research but we will not mention you by name or share your personal details with anybody outside of our team. Your answers will be mixed with the answers of other parents we have interviewed, making it impossible to identify individual answers. Your responses will help us to learn more about the situation of education in Timor-Leste. We will use this information to adjust our project's design. Do you consent in participating in this interview?	1 yes, 0 no
ENUMERATOR: Are you speaking with the school director/ coordinator or his/her representative?	1 director/coordinator, 2 representative
ENUMERATOR: What is the position of the respondent?	1 central basic school director, 2 deputy director, 3 technical advisor, 4 filial school coordinator, 5 teacher, 6 other
Name of the school director/ coordinator or his /her representative	
Contact number for the director/coordinator [DON'T RECORD THE NUMBER OF THE REPRESENTATIVE]	
For how many years has the director/coordinator been in his/her position?	
What is the highest grade in the school?	1 grade 4, 2 grade 5, 3 grade 6, 4 grade 7 (pre-secondary grade 1), 5 grade 8 (pre-secondary grade 2), 6 grade 9 (pre-secondary grade 3), 7 grade 1 secondary, 8 grade 2 secondary, 9 grade 3 secondary, 10 technical school, 11 preschool, 12 other

How many permanent teachers work in this school?	
How many contracted teachers work in this school?	
How many volunteer teachers work in this school?	
sum of permanent, contracted, volunteer teachers	
You said that there are \${Mestre_permanente_na_hira_iha_eskola_ne_e} permanent teachers, \${Mestre_kontratadu_na_hira_iha_eskola_ne_e} contracted teachers, and \${Mestre_voluntariu_na_hira_iha_eskola_ne_e} volunteer teachers. That means there are a total of \${total_teachers} at this school. Is that correct?	
How many teachers teach in Grade 2?	
How many of the grade 2 teachers are female?	
How many male teachers work in this school? [INCLUDING PERMANENT, CONTRACTED AND VOLUNTEERS]	
How many female teachers work in this school? [INCLUDING PERMANENT, CONTRACTED AND VOLUNTEERS]	
The number of male teachers and female teachers must add up to \${total_teachers}	
From the teachers who teach grades 1-3, how many have attended training on literacy teaching?	
What is the highest education level attained by the director/coordinator?	1 secondary school (including secondary level teacher training), 2 teacher training degree (lower level diploma-minimum qualification), 3 teacher training institute or Faculty of Education degree, 4 another undergraduate course (not on education), 5 post-graduation, 99 don't know, 6 other
How many male teachers have concluded bacharelato or teachers training college?	
How many female teachers have concluded bacharelato or teachers training college?	
How many teachers should be teaching in this shift?	
[COUNT HOW MANY TEACHERS ARE PRESENTLY TEACHING IN CLASS] How many teachers are teaching in class right now?	
[CHECK THE TEACHERS ATTENDANCE LIST] How many teachers were in school yesterday?	
How many grade 1 teachers normally work in this school?	
How many grade 1 teachers should be teaching in this shift?	
How many grade 1 teachers are teaching in class right now?	
How many grade 1 teachers are currently on leave?	
How many grade 1 teachers are currently attending training?	
How many grade 2 teachers normally work in this school?	
How many grade 2 teachers should be teaching in this shift?	
How many grade 2 teachers are teaching in class right now?	
How many grade 2 teachers are currently on leave?	
How many grade 2 teachers are currently attending training?	
Is the director /coordinator providing coaching to teachers on literacy teaching on a weekly, monthly, quarterly basis or not at all?	1 weekly, 2 monthly, 3 every trimester, 4 never provided coaching, 99 don't know

When was the last time the director/coordinator provided coaching to the Grade 2 teacher on literacy?	1 last week, 2 last month, 3 last year, 4 didn't provide, 99 don't know
if the director/coordinator is having challenges in providing coaching, who could provide help?	1 central school director, 2 deputy director, 3 technical advisor, 4 inspector, 5 municipal education officer, 6 NGO or project, 7 other, 99 don't know
Other (explain)	
Did grade 1-2 teachers participated in the teacher working group meetings in 2020?	1 yes, 0 no
Is there any project supporting literacy / reading development in this school?	1 yes, 0 no
Can you share the name of the project?	
Does the school have a PTA?	1 yes, 0 no
How many members take part in the PTA?	
When was the last time the PTA met at the school? (does not include one member visiting the school)	1 last week, 2 last month, 3 more than a month ago, 4 did not meet this year, 99 don't know
does the PTA do any activities in the following areas? (READ THE ANSWERS)	1 school budget management, 2 learning quality, 3 improve school infrastructure (buildings, toilets, kitchen, fence), 4 oversee the school feeding, 5 monitor safety and security, 6 monitor student attendance, 7 monitor teacher attendance, 8 monitor dropout, 0 does not do anything
does the school have a school improvement plan?	1 yes, 0 no
if a boy or a male teacher harasses or abuses a girl at the toilet, who is the first person the girl can report this to?	1 family/ relatives, 2 director/ coordinator, 3 deputy central school director or technical advisor, 4 teacher, 5 PTA, 6 municipal education officer, 7 police, 8 head of village/ head of the hamlet, 9 traditional leader, 10 clinic, 11 other, 12 cannot tell anyone, 99 don't know, 98 did not respond
if a male teacher pays unwanted attention to a girl against her wishes or harasses her, who is the first person she can report this to?	1 family/ relatives, 2 director/ coordinator, 3 deputy central school director or technical advisor, 4 teacher, 5 PTA, 6 municipal education officer, 7 police, 8 head of village/ head of the hamlet, 9 traditional leader, 10 clinic, 11 other, 12 cannot tell anyone, 99 don't know, 98 did not respond
how many classrooms does the school have?	
How many classes are multigrade classes?	
Are grade 2 classes operating in shorter shifts now due to COVID-19 restrictions?	1 yes, 0 no
Are grade 2 classes operating every day?	1 yes, 0 no
Grade 2 students are attending class for how many hours a day?	
how many toilets are available for the students in this school? (INCLUDE ONLY FUNCTIONAL TOILETS)	
How many toilets are available for female students (considering only functional toilets)	
How many handwashing stations are currently functional in this school (have water)? [OBSERVATION]	
How many of the handwashing stations have soap? [OBSERVATION]	
does the school have water?	1 yes, 0 no
Where does the school get water from?	1 well in school compound, 2 borehole in school compound, 3 piped water to the school, 4 piped water in another location (not at the school), 5 well or water pump in nearby location (not at the school), 6 no water available or water

	point is too far, 7 spring or water brought from springs through split bamboo "pipes", 8 rainwater harvesting
Does the school have electricity?	1 yes, 0 no
ASK TO VISIT GRADE 2] is there a reading corner in this classroom?	1 yes, 0 no
[OBSERVATION ONLY] Are there storybooks or magazines that can be used by Grade 2 students?	1 yes, 0 no
Do the children in this grade have the Lafaek magazine with them? [OBSERVATION]	1 yes, 0 no
OBSERVATION ONLY] is there sufficient light within grade 2 to allow students to read?	1 yes, enough light for the entire room, 2 enough light in some parts of the classroom, 0 no
Does the school lend story books for students to take home?	1 yes, 0 no
Why the school is not lending books to students?	1 may lose the books, 2 not enough books, 3 students are careless, 4 books were lost / stolen, 5 children don't know how to read, 6 other
How many students borrowed books last week?	
does the school have Lafaek archives? (copies of previous magazines for teachers to use)	1 yes, 0 no
does the school gives the Lafaek magazines to the students to take home?	1 yes, to all grades, 2 yes, to some grades only, 0 no
how do the teachers use the Lafaek teacher magazine?	1 classroom management, 2 teaching mathematics, 3 teaching to read, 4 teaching Portuguese, 5 teaching Tetun, 6 teaching science, 7 health, 8 learn from other teachers' experiences, 9 information from the Ministry of Education, 10 learn about Timor-Leste, 11 peace education, 12 other
Other (explain)	
how do the teachers use the children's Lafaek magazine?	1 students practice reading in class, 2 prepare homework, 3 read stories to students, 4 games with students, 5 teach students to count, 6 teach students to read, 7 learn about Timor-Leste, 8 teach about science, 9 teach about health, 10 teach mathematics, 11 teach arts/ crafts, 12 teach Tetun, 13 other
Other (explain)	
[ASK TO SEE THE ENROLMENT RECORDS] Male students enrolled in Grade 1	
Female students enrolled in Grade 1	
Male students enrolled in Grade 2	
Female students enrolled in Grade 2	
Male students enrolled in Grade 3	
Female students enrolled in Grade 3	
Male students enrolled in Grade 4	
Female students enrolled in Grade 4	
Male students enrolled in Grade 5	
Female students enrolled in Grade 5	
Male students enrolled in Grade 6	
Female students enrolled in Grade 6	
PLEASE GO TO EACH GRADE AND COUNT HOW MANY STUDENTS ARE IN CLASS RIGHT NOW. IF THE GRADE IS SPLIT ACROSS SEVERAL CLASSES, PLEASE COUNT THE TOTAL NUMBER OF STUDENTS IN THE GRADE (FOR EXAMPLE, IF GRADE 1 IS SPLIT ACROSS CLASSES A, B AND C, COUNT ALL	

THE MALE STUDENTS AND WRITE DOWN THE NUMBER;
THEN COUNT ALL THE FEMALE STUDENTS AND WRITE
DOWN THE NUMBER.

How many male students are in Grade 1 today?

How many female students are in Grade 1 today?

How many male students are in Grade 2 today?

How many female students are in Grade 2 today?

How many male students are in Grade 3 today?

How many female students are in Grade 3 today?

How many male students are in Grade 4 today?

How many female students are in Grade 4 today?

How many male students are in Grade 5 today?

How many female students are in Grade 5 today?

How many male students are in Grade 6 today?

How many female students are in Grade 6 today?

how many classes had attendance books?

CHECK THE ATTENDANCE RECORDS AGAIN AND LOOK FOR
INFORMATION ON THE NUMBER OF CHILDREN WHO HAVE
DROPPED OUT. BEFORE ENTERING THE DATA, PLEASE CHECK
WITH THE TEACHER IF THERE ARE OTHER DROPOUTS WHO
HAVE NOT BEEN MARKED AS SUCH BY THE SCHOOL.

How many male students dropped out from Grade 1 this
year?

How many female students dropped out from Grade 1 this
year?

How many male students dropped out from Grade 2 this
year?

How many female students dropped out from Grade 2 this
year?

How many male students dropped out from Grade 3 this
year?

How many female students dropped out from Grade 3 this
year?

How many male students dropped out from Grade 4 this
year?

How many female students dropped out from Grade 4 this
year?

How many male students dropped out from Grade 5 this
year?

How many female students dropped out from Grade 5 this
year?

How many male students dropped out from Grade 6 this
year?

How many female students dropped out from Grade 6 this
year?

Are teachers and students wearing masks? [OBSERVATION]

4 most teachers and students wear masks, 3 most teachers
wear masks but not students, 2 some teachers wear masks,
1 some students wear masks, 0 both teachers and students
do not wear masks

Is social distancing being practiced in the classes? [OBSERVATION]	2 most classes are observing social distance, 1 some classes are observing social distance, 0 none of the classes is observing social distance
Does the school buy local produce from farmers for the school feeding?	1 yes, sometimes, 2 yes, all the time, 0 no
What types of produce does the school buy from farmers?	1 rice, maize, bread, 2 pumpkin, carrot, purple sweet potato, 3 potato, taro, yellow sweet potato, cassava, 4 dark green vegetables such as water spinach, lettuce, mustard, pumpkin leaves, cassava leaves, 5 cucumber, tomato, cabbage, eggplant, 6 mango, papaya, honeydew melon, passionfruit or other yellow fruits, 7 watermelon, tamarind, jackfruit, 8 beef, pork, sheep, goat, chicken or duck, 9 fish (fresh or dried), shrimp or other seafood, 10 beans, peas, soybeans, peanuts, 11 milk (not sweetened condensed milk), 12 coconut oil, 13 condiments, 14 tofu or tempe, 15 eggs, 99 don't know, 98 did not respond
why isn't the school buying produce for school feeding from farmers or farmers' groups?	1 no budget to buy local produce, 2 farmers' produce is not sufficient, 3 farmers don't want to sell to the school, 4 poor quality of local produce, 5 production drops at certain times, availability of produce is uncertain, 6 local produce is not nutritious, 7 other
Other (explain)	
does the school have a menu for school feeding?	1 yes, 0 no, 99 don't know
who is responsible for the oversight of the school feeding in this school?	1 director or coordinator, 2 deputy director, 3 PTA, 4 teachers, 5 other
Other (explain)	
did the school provide meals to the students today?	1 yes, 0 no
what foods were included in today's meal?	1 rice, maize, bread or foods prepared with rice, maize or wheat, 2 pumpkin, carrot, purple sweet potato, 3 potato, taro, yellow sweet potato, cassava, sago, 4 dark green vegetables such as water spinach, spinach, lettuce, mustard, pumpkin leaves, cassava leaves, 5 Cucumber, tomato, cabbage, eggplant, 6 mango, papaya, honeydew melon, passionfruit or another yellow fruit, 7 watermelon, tamarind, jackfruit, 8 beef, pork, sheep or goat meat, chicken or duck, 9 fish (fresh or dry), shrimp or another seafood, 10 beans, peas, soybeans or peanuts, 11 fresh milk (not sweetened condensed milk), 12 sweetened condensed milk, 13 other, 14 eggs
is there any project supporting school meals in this school?	1 yes, 0 no
which project is supporting school meals in this school?	
does the school have a kitchen?	1 yes, 0 no
Which kind of stove is used to cook the meals?	gas: gas, wood: wood, electricity: electricity
is clean water available to prepare meals?	1 yes, 0 no
Is there a handwashing station at the kitchen?	1 yes, 0 no
Does the handwashing station at the kitchen have soap?	1 yes, 0 no
does the school have plates / cutlery for the students?	1 yes, 0 no
do you use detergent to clean the kitchen?	yes_every_day: yes, every day, yes_often: yes, often, yes_sometimes: yes, sometimes, no_never: no, never
Is there a scale in the kitchen?	1 yes, 0 no
[OBSERVATION ONLY] Are there animals in the kitchen (chicken, dog, cat or other)?	1 yes, 0 no, 99 did not observe
Is there a canteen/ space to eat at the school?	1 yes, 0 no

is there a place to store food at the kitchen or near the kitchen?	0 no, do not have, 2 some, 1 yes, enough
is the storage space within the school?	1 yes, 0 no
what material is the storage space floor made of?	cement: cement, mud: mud, gravel: gravel, wooden_bamboo: wood or bamboo, sand: sand, tile: tile
what material are the storage space walls made of?	bricks: bricks, wood_bamboo: wood or bamboo, sand: sand, tile: tile, mix: mix
what material is the storage space roof made of?	bamboo: bamboo, aluminium: aluminium sheet, tile: tile, branches_leaves: branches and leaves
Is the roof leaking? (can you see marks of water in the wall or floor?)	1 yes, 0 no
Is the storage space ventilated? (does it have windows, a fan or orifices for ventilation)	1 yes, 0 no
is the storage space clean or not?	yes: yes, storage space is clean (for example, there are no food remains or pests, and the food is correctly stored on the cement floor), mainly: yes, storage space is mostly clean (some kernels on the floor, most of the food is correctly stored, and there are no pests), somewhat: storage space is somewhat dirty (for example, some food remains and kernels around; materials are partially open and not correctly stored; non-food materials are stored in the same space), no: no, the storage space is dirty (dirty floor and walls, dusty bags, incorrectly stored and not closed, non-food items stored along with the food, presence of pests)
are there other items stored there?	1 yes, 0 no
are there pallets or another way of raising the food from the floor?	shelves: shelves, pallets: pallet, none: no shelves, food is placed on the floor, other: other
how do you secure the food?	lock: door with lock, watchman: watchman, camera: camera, none: no security, other: other

HOUSEHOLD SURVEY TOOL

Description	Choice options
Enumerator code	
Name of the enumerator	
Today's date	
Unique Student ID [INSERT FROM LIST OF INTERVIEWED STUDENTS]	
Student name [INSERT FROM LIST OF INTERVIEWED STUDENTS]	
Sex (student)	2 female, 1 male
School name (INSERT FROM THE LIST OF INTERVIEWED STUDENTS)	
School ID (INSERT FROM THE LIST OF INTERVIEWED STUDENTS)	
Was the student randomly selected from Grade 2 or was the student identified from the baseline interview list?	1 sampled from grade 2, 2 identified through the baseline list
How are you? I am interviewing on behalf of the HATUTAN project, which aims to support education and children's nutrition in Timor-Leste. I would like to talk to you today and ask for your permission to interview you. Feel free to consult other people in the household before you respond. I would like to ask you some questions about your household and the children who currently live here. Then, I may ask to speak with other members of the household if they agree to participate. I will write your answers to use them in our research but we will not mention you by name or share your personal details with anybody outside of our team. Your answers will be mixed with the answers of other parents we have interviewed, making it impossible to identify individual answers. Your responses will help us to learn more about the status of children's education, nutrition and health in Timor-Leste. We will use this information to adjust our project's design. Do you consent in participating in this interview?	1 yes, 0 no, 99 don't know
GPS coordinates	
Municipality	1 aileu, 2 ainaro, 3 baucau, 4 bobonaro, 5 covalima, 6 ermera, 7 liquica, 8 manatuto, 9 manufahi
Subdistrict	1 aileu-vila, 2 ainaro, 3 alas, 4 atabae, 5 atsabe, 6 baguia, 7 balibo, 8 barique, 9 baucau, 10 bazartete, 11 bobonaro, 12 cailaco, 13 ermera, 14 fatululic, 15 fatumean, 16 fohorem, 17 hato-udo, 18 hatolia, 19 hatu-builico, 20 laclo, 21 laclubar, 22 laleia, 23 laulara, 24 lequidoe, 25 letefoho, 26 liquica, 27 lolotoe, 28 maliana, 29 manatuto, 30 maubara, 31 maubisse, 32 maukatar, 33 quelicai, 34 railaco, 35 remexio, 36 same, 37 soibada, 38 suai, 39 tilomar, 40 turiscas, 41 vemasse, 42 venilale, 43 zumalai
What is your main occupation?	11 Civil servant, 12 Elected official (village chief, head of the hamlet), 13 Health worker (e.g. medical doctor, nurse, midwife), 14 Teacher, 15 Salesperson or service worker (e.g. retailer at a shop, market, or stall; restaurant worker), 16 Business owner, 17 Farmer (for own consumption), 18 Farmer (for sale), 19 Fisherman, 20 Artisan/ craftsman (weaver, ceramist, etc), 21

	Armed forces, 22 Police, 23 Student, 24 Traditional leader/ religious leader, 25 Retired, 26 Mason or plumber, 27 Temporary worker, 28 Does not have an occupation, 29 Other
Other (explain)	
What is your highest education level?	0 No education, 1 Literacy course, 2 Incomplete primary (did not reach Grade 6), 3 Completed primary (Grade 6), 4 Incomplete pre-secondary, 5 Completed pre-secondary, 6 Incomplete secondary or technical school, 7 Completed secondary or technical school, 8 Completed university, 9 Non-formal education (accelerated education)
Which language do you normally speak at home?	1 Mambae, 2 Galolen, 3 Tokodede, 4 Kemak, 5 Bunak, 6 Tetun Terik, 7 Tetun Prasa, 8 Idate, 9 Mdiki, 10 Makasae, 11 Other
Other (explain)	
How many people sleep and eat regularly in this household?	
How many of those are children less than three years of age?	
How many of the children age 5-15 are girls?	
How many of the children age 5-15 are attending school?	
How many of the children age 5-15 who are in school are girls?	
Among the women living in this household, are there any pregnant mothers?	1 yes, 0 no, 99 don't know
Among the women living in this household, are there any lactating mothers?	1 yes, 0 no, 99 don't know
Now I'd like to ask some questions about household finances.	
Does anyone in this household have savings?	1 yes, 0 no, 99 don't know
If yes, where are the savings kept?	1 VSLA (Savings Group), 2 Microfinance Group, 3 savings kept at home, 4 Bank, 5 Other, 99 Don't know
How many people in this household are participating in VSLA?	
Among those participating in VSLA, how many are female?	
What do you mainly use the savings for?	
Food	0 not mentioned, 1 mentioned
Health care	0 not mentioned, 1 mentioned
Education expenses	0 not mentioned, 1 mentioned
Agriculture (including livestock, seeds, plough, etc)	0 not mentioned, 1 mentioned
Invest in business	0 not mentioned, 1 mentioned
Funeral/ wedding/ traditional ceremony	0 not mentioned, 1 mentioned
Pay debt	0 not mentioned, 1 mentioned
investing in a household asset (e.g. phone, motorcycle, etc)	0 not mentioned, 1 mentioned
Have you, or the other people participating in VSLA in this household, borrowed money from the group during the past three months?	1 yes, 0 no, 99 don't know
What did you use the loan for? [DON'T READ THE ANSWERS]	1 Food, 2 health care, 3 Education expenses, 4 Agriculture (including livestock, seeds, plough, etc), 5 Invest in business, 6 Funeral/ wedding/ traditional ceremony, 7 Pay debt, 8 investing in a household asset (e.g. motorcycle, phone, tv, etc), 99 Don't know
Who made the final decision about how to use the loan?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my

	grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not applicable
How often did the VSLA group meet last month?	
What has been the greatest benefit of VSLA participation? [DON'T READ THE ANSWERS]	1 group solidarity, 2 obtain capital for business, 3 obtain money for family needs, 4 obtain information/ learning, 5 save money for future needs, 6 other
Other (explain)	
Who in the household has the final say on LARGE household purchases? (sale of large livestock, purchase of major household asset, etc)	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not applicable
Who in the household has the final say on SMALL household purchases? (food for daily consumption, small fees for transport, etc)	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not applicable
Where do you get information about doing small business from? [DON'T READ THE ANSWERS]	1 Lafaek magazine, 2 Lafaek facebook page, 3 Other, 0 Did not receive any information, 99 Don't know
What information did you get about doing small business? [DON'T READ THE ANSWERS]	1 cook bakso, 2 prepare a budget, 3 make tempeh, 4 plant vegetables in a banana trunk, 5 cook pizza, 6 make crafts (trash bin, painting, frame, bag), 7 make a broom, 8 make perkedel, 9 make biobriket from corn cobs, 10 plant vanilla, 11 make a leaf-shaped stool, 12 make corn kukus, 13 make coconut cookies, 14 soap making, 15 prepare terang bulan, 16 make avocado pudding, 17 make multi-color pudding, 18 make pickles, 19 make beef jerky, 20 make spicy popcorn, 21 market strategy, 22 savings, 23 fish farming, 24 other
Other (explain)	
Does the household receive the Lafaek ba Komunitade magazine?	1 yes, 0 no, 99 don't know
Do you or someone else in this household read the Lafaek ba Komunitade magazine?	1 yes, 0 no, 99 don't know
What did you learn from Lafaek ba Komunitade?	1 about health and hygiene practices (including nutrition/ cooking nutritious food), 2 about better family relationships, 3 about gender equality, 4 child care, teaching children, 5 crafts, 6 agriculture, 7 business/ finance, 8 environment, 9 about Timor-Leste, 10 stories, 11 child rights, 12 reading, 13 mathematics/ numeracy, 14 games, 15 other, 99 don't know
other (explain)	
Now I'd like to speak with \${naran_aluno}'s mother or primary caregiver. [IF YOU ARE ALREADY SPEAKING WITH THE STUDENT'S MOTHER OR CAREGIVER, THERE IS NO NEED TO ASK FOR CONSENT AGAIN – JUST CHOOSE YES]	
	1 yes, 0 no, 99 don't know
Hello, I am interviewing on behalf of the HATUTAN project, which aims to support education and children's nutrition in Timor-Leste. I would like to talk to you today and ask for your permission to interview you. Feel free to consult other people in the household	

before you respond.

I would like to ask you some questions about your household and the children who currently live here. Then, I may ask to speak with other members of the household if they agree to participate. I will write your answers to use them in our research but we will not mention you by name or share your personal details with anybody outside of our team. Your answers will be mixed with the answers of other parents we have interviewed, making it impossible to identify individual answers. Your responses will help us to learn more about the status of children's education, nutrition and health in Timor-Leste. We will use this information to adjust our project's design. Do you consent in participating in this interview?

Name of the mother/ caregiver

Gender - mother/ caregiver

2 female, 1 male

How old are you?

What is your relationship with \${naran_aluno}?

1 Mother, 2 Father, 3 Grandmother, 4 Grandfather, 5 Another relative, 6 Non-relative

[ASK OR RECORD] Are \${naran_aluno}'s parents alive?

1 Mother and father are alive, 2 Deceased mother, 3 Deceased father, 4 Double orphan, 99 Don't know

What is your main occupation?

11 Civil servant, 12 Elected official (village chief, head of the hamlet), 13 Health worker (e.g. medical doctor, nurse, midwife), 14 Teacher, 15 Salesperson or service worker (e.g. retailer at a shop, market, or stall; restaurant worker), 16 Business owner, 17 Farmer (for own consumption), 18 Farmer (for sale), 19 Fisherman, 20 Artisan/ craftsman (weaver, ceramist, etc), 21 Armed forces, 22 Police, 23 Student, 24 Traditional leader/ religious leader, 25 Retired, 26 Mason or plumber, 27 Temporary worker, 28 Does not have an occupation, 29 Other

Other (explain)

What is your highest education level?

0 No education, 1 Literacy course, 2 Incomplete primary (did not reach Grade 6), 3 Completed primary (Grade 6), 4 Incomplete pre-secondary, 5 Completed pre-secondary, 6 Incomplete secondary or technical school, 7 Completed secondary or technical school, 8 Completed university, 9 Non-formal education (accelerated education)

Which language do you normally speak at home?

1 Mambae, 2 Galolen, 3 Tokodede, 4 Kemak, 5 Bunak, 6 Tetun Terik, 7 Tetun Prasa, 8 Idate, 9 Mdiki, 10 Makasae, 11 Other

Other (explain)

How old is \${naran_aluno}?

Did \${naran_aluno} attend pre-school?

1 yes, 0 no, 99 don't know

What is the name of the school \${naran_aluno} attends?

Which grade is \${naran_aluno} in?

How long does it take for \${naran_aluno} to walk to school?

1 Less than 30 minutes, 2 Between 30 minutes and one hour, 3 Between one and two hours, 4 More than two hours, 99 Don't know

Does \${naran_aluno} take any form of transportation to go to school?

1 yes, 0 no, 99 don't know

How safe is it for \${naran_aluno} to walk to school every day? Is it very safe, fairly safe, fairly unsafe or very unsafe?

1 Very safe, 2 Fairly safe, 3 Fairly unsafe, 4 Very unsafe, 99 Don't know

What makes the journey to school unsafe? [DO NOT READ THE ANSWERS]	1 Long distance, 2 Traffic, 3 Poor roads, 4 Heat or rain, 5 River crossings, 6 Environmental disruptions (e.g. flood, landslides, fires), 7 Wild animals, 8 Risk of being verbally abused by other children or young people, 9 Risk of being physically abused by other children or young people, 10 Risk of being sexually abused by other children or young people, 11 Risk of being verbally abused by adults, 12 Risk of being physically abused by adults, 13 Risk of being sexually abused by adults, 14 Kidnappings, 15 Roadblocks, 16 Conflict, violence, open fighting, 17 Sorcery, traditional beliefs, 18 Other
How many days of school did \${naran_aluno} miss last week?	
What was the main reason why \${naran_aluno} did not go to school last week?	1 sick, 2 household chores or caring for other people within the household, 3 funeral, marriage, traditional ritual, 4 did not have school materials, 5 farm work, 6 helped parents with business, 7 was working for money, 8 did not want to go to school, 9 natural disaster, strong rains, 10 school was closed, 11 teacher did not attend, 12 parents or relatives don't want him/her to attend, 13 other
Other (explain)	
Does \${naran_aluno} sometimes don't want to go to school?	1 yes, 0 no, 99 don't know
Is \${naran_aluno} sometimes afraid of going to school?	1 yes, 0 no, 99 don't know
How many days of school did \${naran_aluno} miss due to sickness last month?	
Can \${naran_aluno} speak Tetun?	1 yes, 0 no, 99 don't know
Did \${naran_aluno} ever repeat a grade?	1 yes, 0 no, 99 don't know
Since \${naran_aluno} started school, did he/she ever drop out of school?	1 yes, 0 no, 99 don't know
Did \${naran_aluno} study at home when the school was closed last year due to COVID-19?	1 yes, 0 no, 99 don't know
How did \${naran_aluno} study at home while the school was closed due to COVID-19?	1 television (eskola ba uma program), 2 used books received from school, 3 used Lafaek magazine, 4 taught by parents or other relatives, 5 used mobile phone, 6 other
Other (explain)	
Does \${naran_aluno} have difficulty seeing?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
Does \${naran_aluno} wear glasses?	0 not mentioned, 1 mentioned
Does \${naran_aluno} have difficulty hearing or cannot hear at all?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
Does \${naran_aluno} have difficulty walking or climbing steps?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
Does \${naran_aluno} have difficulty remembering things or concentrating?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
Does \${naran_aluno} have difficulty with self care such as washing all over or dressing?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
Using your usual language, does \${naran_aluno} have difficulty communicating; for example understanding or being understood?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
In the last year, has \${naran_aluno} had any serious illnesses? [serious illness means an illness that put \${naran_aluno}'s life in danger]	0 not mentioned, 1 mentioned

Does \${naran_aluno} have difficulty making friends?	1 No, no difficulty, 2 Yes, some difficulty, 3 Yes, a lot of difficulty, 4 Cannot do at all, 99 Don't know
How often does \${naran_aluno} seem very anxious, nervous or worried?	1 Daily, 2 Weekly, 3 Monthly, 4 A few times a year, 5 Never
How often does \${naran_aluno} seem very sad or depressed?	1 Daily, 2 Weekly, 3 Monthly, 4 A few times a year, 5 Never
Does \${naran_aluno} spend time caring for younger or older family members?	1 yes, 0 no, 99 don't know
Does \${naran_aluno} spend time doing housework (e.g. cooking or cleaning)?	1 yes, 0 no, 99 don't know
Does \${naran_aluno} help with fetching water or firewood?	1 yes, 0 no, 99 don't know
Does \${naran_aluno} help with agricultural work (e.g. guarding livestock; planting, watering or harvesting crops)?	1 yes, 0 no, 99 don't know
Does \${naran_aluno} help with a family business or work outside the home (non-agricultural)?	1 yes, 0 no, 99 don't know
Usually how much time does \${naran_aluno} spend on those tasks on a day?	1 Whole day, 2 Half day, 3 Quarter day / a few hours, 4 A little time / an hour or less, 99 Don't know, 0 Does not do chores
Do those tasks ever cause \${naran_aluno} to arrive late at school?	1 Yes, many times, 2 Yes, sometimes, 3 No, 99 Don't know
Do those tasks reduce \${naran_aluno}'s time to study at home and do homework?	1 yes, 0 no, 99 don't know
Is there anyone at home who helps \${naran_aluno} with homework?	1 yes, 0 no, 99 don't know
Do you consider that the PTA is very active, somewhat active or inactive?	1 very active, 2 somewhat active, 3 inactive/ does not exist, 99 Don't know, 98 refused to respond
Do you or another person in this household participate in the school's PTA?	1 yes, 0 no, 99 don't know
Is \${naran_aluno}'s school PTA doing any activities to improve school hygiene?	1 yes, 0 no, 99 don't know
Is \${naran_aluno}'s school PTA doing any activities to improve school feeding?	1 yes, 0 no, 99 don't know
Is \${naran_aluno}'s school PTA doing any activities to improve student learning?	1 yes, 0 no, 99 don't know
Did this school provide meals to the students last week?	1 yes, 0 no, 99 don't know
I'll read some statements about the meals now being offered in this school. Please tell me if you agree completely, agree somehow, disagree somehow or disagree completely with each one of these statements.	
The quantity of the food is sufficient.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
The food is prepared in a hygienic manner.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
The food is available every day.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
The food is tasty.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
What does the teacher normally do when a child misbehaves in class? [DO NOT READ THE ANSWERS]	
Gives a verbal warning	0 not mentioned, 1 mentioned
Shouts at the child	0 not mentioned, 1 mentioned

Uses corporal punishment	0 not mentioned, 1 mentioned
Assigns chores to the child	0 not mentioned, 1 mentioned
Inform the parents	0 not mentioned, 1 mentioned
Have a conversation with the child	0 not mentioned, 1 mentioned
Are there any instances when it is justifiable for the teacher to use corporal punishment on boys?	1 yes, 0 no, 99 don't know
Are there any instances when it is justifiable for the teacher to use corporal punishment on girls?	1 yes, 0 no, 99 don't know
Who do you think is more skilled at reading and writing? [READ THE ANSWERS]	1 Boys and girls have the same capacity, 2 Girls have more capacity than boys, 3 Boys have more capacity than girls, 4 Neither boys nor girls have capacity, 99 Don't know
Who do you think is more skilled in mathematics? [READ THE ANSWERS]	1 Boys and girls have the same capacity, 2 Girls have more capacity than boys, 3 Boys have more capacity than girls, 4 Neither boys nor girls have capacity, 99 Don't know
Who does the teacher encourage to participate in classroom activities? [READ THE ANSWERS]	1 Encourage boys and girls equally, 2 Encourage girls more than boys, 3 Encourage boys more than girls, 4 Do not encourage girls or boys, 99 Don't know
Who is able to ask questions in class? [READ THE ANSWERS]	1 Boys and girls equally, 2 Girls more than boys, 3 Boys more than girls, 4 Neither boys nor girls, 99 Don't know
Who is able to ask the teacher for help? [READ THE ANSWERS]	1 Boys and girls equally, 2 Girls more than boys, 3 Boys more than girls, 4 Neither boys nor girls, 99 Don't know
Who is able to use the toilets at school? [READ THE ANSWERS]	1 Boys and girls equally, 2 Girls cannot use, 3 Boys cannot use, 4 Neither boys nor girls can use, 5 There are no toilets or their condition does not allow for use, 99 Don't know
I'm going to read some statements. Please let me know if you agree completely; agree in part; disagree in part; or disagree completely with each statement.	
\${naran_aluno} is learning well at school.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
\${naran_aluno} has enough books at school.	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
We parents are consulted when decisions are made in \${naran_aluno}'s school	1 Agree completely, 2 Agree partially, 3 Disagree partially, 4 Disagree completely, 99 Don't know
Do you think that \${naran_aluno} is having difficulties to learn to read?	1 yes, 0 no, 99 don't know
Which challenges does \${naran_aluno} face to learn to read?	0 Does not speak the language of instruction, 1 Teacher does not explain well, 2 Classes are too short, 3 Teacher is often absent, 4 Child struggles to pay attention, 5 No reading materials at school, 6 Child lacks stationery, 7 Child is a "slow learner", 8 Child has a disability, 9 Child is afraid of the teacher, 10 There is no teacher, 11 Child is often absent or late, 12 No reading materials for children at home, 13 Family is unable to help the child to learn, 14 Other
Other (explain)	
Does \${naran_aluno} receive the Lafaek magazine?	1 yes, 0 no, 99 don't know
Does the school allow \${naran_aluno} to bring the Lafaek magazine home?	1 yes, 0 no, 99 don't know

Do you have any children books or magazines at home? Can I see them?	1 observed, 0 did not observe
ENUMERATOR: Which reading materials (children's books or magazines) are available in this home?	1 Lafaek magazine only, 2 Children's books, 3 Both Lafaek magazine and children's books
Does \${naran_aluno} read at home?	1 yes, 0 no, 99 don't know
Is there anyone at home who helps \${naran_aluno} learn to read?	1 yes, 0 no, 99 don't know
Did \${naran_aluno} borrow any books from school during the past 3-4 months?	1 yes, 0 no, 99 don't know
Who reads the Lafaek magazine at home?	1 children, 2 youth, 3 parents, 4 grandparents, 5 other relatives, 0 No one
How do you use the Lafaek magazine at home?	1 Help children to learn, 2 Learn about health, hygiene, 3 learn about Timor-Leste, 4 Learn about business, finance, 5 Learn about agriculture, 6 Learn about gender, 7 Learn about careers, 8 Learn to read / count, 9 Read stories, 10 Play games, 11 Learn about childcare, good relationships at home, 12 Learn Portuguese, 13 Learn Tetum, 14 Learn about crafts, 15 Learn about the environment, 16 Other, 17 Does not use the magazine
Other (explain)	
How often do you or other household members read the stories in the Lafaek magazine for your children?	3 Very often, 2 Often, 1 Sometimes, 0 Never, 96 All household members are illiterate, 99 Don't know
How often do you or other household members play games with your children?	3 Very often, 2 Often, 1 Sometimes, 0 Never, 99 Don't know
Do you ever use the games in the Lafaek magazine with your children?	1 yes, 0 no, 99 don't know
Which games do you use in the Lafaek magazine? [DO NOT READ THE ANSWERS]	1 literacy games, 2 numeracy games, 3 health games, 4 crafts, 5 fine motor skills (connect the dots, maze), 6 Other, 99 Don't know
Do people in this household face any challenges to read the Lafaek magazine?	1 yes, 0 no, 99 don't know
What challenges do you face to read the Lafaek magazine?	1 Do not speak Tetun, 2 The Tetun used is too difficult, 3 None of the household members is literate, 4 The content is too difficult, 5 The content is irrelevant, 6 Design issues (small font, dark colors), 7 The content is inappropriate, 8 Visual disability, 9 Other
How important do you think it is for \${naran_aluno} to attend school?	4 very important, 3 somewhat important, 2 not very important, 1 not important at all, 99 don't know, 98 refused to respond
What is the maximum level of education you can support \${naran_aluno} to attend?	1 Primary (grade 6), 2 Pre-secondary (grade 9), 3 Secondary school (grade 12), 4 Technical school, 5 University, 99 Don't know
If anyone abuses or harasses \${naran_aluno} at school, who can you report it to?	1 Head teacher, 2 Police, 3 Social services, 4 Local authorities, 5 Cannot report, 6 Other (specify)
Enumerator: Please explain 'other' reporting mechanism	
Did \${naran_aluno} eat breakfast before leaving for school today?	1 yes, 0 no, 99 don't know
List as many examples as you can of important maternal child care practices.	
[DO NOT READ THE RESPONSES. MARK ALL THAT APPLY.]	
INITIATE BREASTFEED W/IN 1 HR OF DELIVERY	0 not mentioned, 1 mentioned
EXCLUSIVE BREASTFEED FOR SIX MONTHS	0 not mentioned, 1 mentioned

INTRODUCTION OF APPROPRIATE, SAFE, AND ADEQUATE COMPLEMENT FOODS AT 6 MONTHS UP TO 2 YEARS AND BEYOND	0 not mentioned, 1 mentioned
BREASTFEED FREQUENTLY ON DEMAND, BOTH DAY AND NIGHT	0 not mentioned, 1 mentioned
USE OF VARIETY OF NUTRITIOUS, LOCALLY AVAILABLE FOODS FOR INFANTS AND YOUNG CHILDREN	0 not mentioned, 1 mentioned
PREGNANT/LACTATING WOMEN RECEIVE APPROPRIATE CARE AND ENCOURAGED TO CONSUME ADEQUATE QUANTITIES OF NUTRITIOUS FOOD	0 not mentioned, 1 mentioned
WHEN INFANT UNABLE TO SUCKLE, EXPRESSED BREASTMILK FED BY CUP OR TUBE	0 not mentioned, 1 mentioned
MAINTAINS HEALTH CARD TO MONITOR GROWTH AND DEVELOPMENT OF CHILD	0 not mentioned, 1 mentioned
FEEDING FREQUENT MEALS AND SNACKS TO THE CHILD	0 not mentioned, 1 mentioned
TAKE THEIR CHILD/REN TO HEALTH PROMOTION SESSIONS OR HEALTH FACILITY	0 not mentioned, 1 mentioned
CONTINUE OR INCREASE BREASTFEEDING WHEN MOTHER OR CHILD IS SICK	0 not mentioned, 1 mentioned
FEEDING FOODS RICH IN IRON (red meat; poultry; eggs; tofu; beans; dark green leafy vegetables)	0 not mentioned, 1 mentioned
ENSURE TIMELY IMMUNIZATIONS	0 not mentioned, 1 mentioned
ENSURE CHILD SLEEPS UNDER TREATED MOSQUITO NET	0 not mentioned, 1 mentioned
CONTINUE BREASTFEEDING FOR 1 YEAR OR 2 YEARS	0 not mentioned, 1 mentioned
Now I would like to ask you about liquids or foods that you ate yesterday during the day or at night. I am interested in whether you had the item even if it was combined with other foods. For example, if you ate a porridge made with a mixed vegetable, you should reply yes to any food I ask about that was an ingredient in the porridge. Please do not include any food used in a small amount for seasoning or condiments (like chilies, spices, herbs), I will ask you about those foods separately. Yesterday during the day or night did you drink/eat any [ASK QUESTIONS BELOW)	1 Bread, cereals/porridge, noodles, rice, mash/residue or other foods made from grains such as maize or wheat., 2 Pumpkin, carrots, squash, orange flesh sweet potatoes or or any other dark yellow or orange fleshed roots, tubers and vegetables?, 3 White potatoes, white yams, white sweet potato, cassava, or any other foods made from roots?, 4 Any dark green leafy vegetables such as spinach, lettuce, chard, amaranth, pumpkin leaves, cassava leaves, bean leaves, sweet potato leaves, or cowpea leaves?, 5 Any other vegetables, like cucumbers, tomatoes, cauliflower, cabbage, broccoli, eggplant, etc.?, 6 Ripe mangoes, ripe papaya, melon, passionfruit or other fruits that are dark yellow or orange inside?, 7 Any indigenous (wild fruits), 8 Any other fruits like watermelon, tamarind, jackfruit etc., 9 Any liver, kidney, heart, blood or other organ meats from domesticated animals such as cow, pig, goat, chicken or duck?, 10 Any meat from domesticated animals, such as beef, pork, lamb, goat, chicken, or duck?, 11 Any organs from wild animals, such as game meat, bush rats, birds, wild pigeons, guinea fowl, deer, wild boar?, 12 Any flesh from wild animals, such as game meat, bush rats, wild birds, deer, wild boar, wild goat?, 13 Eggs, 14 Fresh or dried fish, shellfish or seafood?, 15 Any foods made from beans, peas, lentils, peanuts or other legumes such as cowpeas, pigeon peas?, 16 Any foods made from nuts and seeds such as pumpkin, sunflower seeds?, 17 Milk, cheese, yogurt or other milk products? (DO NOT count sweetened condensed milk here), 18 Any shea nut oils, other oils, fats, butter or foods made with any of these?, 19 Any sugary foods such as chocolates, sweets, candies, pastries, cakes or biscuits?, 20 Condiments for flavor,

	such as chilies, spices, herbs, or fish powder?, 21 Grubs, snails or insects?, 22 Foods made with red palm oil, red palm nut, or red palm nut pulp sauce?
Did you you take a vitamin supplement of iron and folic acid during your latest pregnancy?	1 yes, 0 no, 99 don't know
During the past 30 DAYS was there a time when you or others in your household went without eating for a whole day because of lack of money or other resources?	1 yes, 0 no, 99 don't know
During the past month, did your household face any of the following? [READ THE ANSWERS]	reduced_the_number_of_meals: reduced the number of meals, reduced_the_quantity_of_food_eaten_in_ea: reduced the quantity of food eaten, reduced_the_quality_of_the_food_eaten_in: reduced the quality of the food eaten
During the past three months, did this household face any natural disasters?	1 yes, 0 no, 99 don't know
During the past three months, did any member of this household experience a serious illness or died?	1 yes, 0 no, 99 don't know
During the past three months, did this household make any major contribution to traditional ceremonies?	1 yes, 0 no, 99 don't know
Who makes the decision on what children should eat in this household?	1 Myself alone, 2 Myself and my husband, 3 Myself and/or my husband in consultation with elders, 4 Elders/grandparents
If you or anyone in this household needs to go to the doctor/clinic, can you afford the costs associated with it? [READ THE ANSWERS]	1 all the time, 2 most of the time, 3 sometimes, 4 cannot sell
If \${naran_aluno} needs school materials, can you afford these all the time, most of the time, sometimes, or never?	1 all the time, 2 most of the time, 3 sometimes, 4 cannot sell
Who in the household has the final say on LARGE household purchases? (sale of large livestock, purchase of major household asset, etc)	1 Myself alone, 2 Myself and my husband, 3 Myself and/or my husband in consultation with elders, 4 Elders/grandparents
Who in the household has the final say on SMALL household purchases? (food for daily consumption, small fees for transport, etc)	1 Myself alone, 2 Myself and my husband, 3 Myself and/or my husband in consultation with elders, 4 Elders/grandparents
What is the main source of drinking water for your household?	1 Piped into dwelling, 2 Piped to the yard/plot, 3 Public tap, 4 Borehole, 5 Dug well (protected), 6 Dug well (unprotected), 7 River/lake, 8 Rainwater harvesting, 9 Trucked water, 10 Spring/ bamboo pipes
Is there a time of the year when drinking water becomes unavailable?	1 yes, 0 no
For how long is drinking water unavailable at that time of the year?	1 Some days, 2 A month, 3 More than a month, 99 Don't know
What type of toilet is used in this household?	1 No toilet, 2 Pit latrine, uncovered, 3 Pit latrine with a slab, 4 Improved pit latrine with ventilation, 5 Composting latrine, 6 Flush toilet
I'm going to show you some images. Please let me know if you should wash your hands before doing those things.	
before caring for children	1 yes, 0 no, 99 don't know
before feeding children	1 yes, 0 no, 99 don't know
before eating	1 yes, 0 no, 99 don't know
before working	1 yes, 0 no, 99 don't know
before cleaning	1 yes, 0 no, 99 don't know
before preparing food	1 yes, 0 no, 99 don't know

I'm going to show you some drawings. Please let me know if you should wash your hands after doing each of those activities.	
after caring for animals	1 yes, 0 no, 99 don't know
after reading books	1 yes, 0 no, 99 don't know
after cleaning children	1 yes, 0 no, 99 don't know
after picking the trash	1 yes, 0 no, 99 don't know
after using the phone	1 yes, 0 no, 99 don't know
after using the toilet	1 yes, 0 no, 99 don't know
Please point to the drawings showing good hygiene behaviors.	
drink boiled water	1 yes, 0 no
use the toilet to defecate/urinate	1 yes, 0 no
playing	1 yes, 0 no
clean a runny nose	1 yes, 0 no
keeping livestock in the kitchen	1 yes, 0 no
wash fruits and vegetables	1 yes, 0 no
bury the trash	1 yes, 0 no
cover the food after preparing it	1 yes, 0 no
write	1 yes, 0 no
use sandals/shoes	1 yes, 0 no
trim nails	1 yes, 0 no
wash clothes	1 yes, 0 no
throw trash outside	1 yes, 0 no
wash the dishes	1 yes, 0 no
brush teeth	1 yes, 0 no
play soccer	1 yes, 0 no
What behaviors help to prevent COVID-19? [DO NOT READ THE ANSWERS]	1 wearing a mask when leaving the home, 2 handwashing with soap, 3 staying at home, 4 maintaining social distance, 5 avoid gatherings, 6 staying away from sick people, 7 not spitting in public spaces, 8 if experiencing fever, coughing, sneezing or difficulty breathing, seek treatment immediately, 9 cover your sneeze or cough using a tissue or your elbow, 10 throw used tissues in the trash
What are you doing to prevent COVID-19? [DO NOT READ THE ANSWERS]	1 wearing a mask when leaving the home, 2 handwashing with soap, 3 staying at home, 4 maintaining social distance, 5 avoid gatherings, 6 staying away from sick people, 7 not spitting in public spaces, 8 if experiencing fever, coughing, sneezing or difficulty breathing, seek treatment immediately, 9 cover your sneeze or cough using a tissue or your elbow, 10 throw used tissues in the trash
where do you get information about health, hygiene and nutrition from?	1 SISCA/ community health volunteers, 2 health clinic, 3 training, 4 radio, television, 5 Lafaek ba Komunitade, 6 Lafaek's facebook page, 7 neighbors or family, 8 other
What information did you learn from Lafaek ba Komunitade about health, hygiene or nutrition?	1 make bakso, 2 make tempeh, 3 handwashing/ make a tippy-tap, 4 make perkedel with potatoes and meat, 5 make kukus from corn, 6 how to feed children under two years of age, 7 using onion as medicine, 8 healthy foods for mothers and children, 9 make beef jerky, 10 make pickles, 11 fish is good for children/ removing

	fish bones before feeding children, 12 myths about foods that should not be eaten by pregnant and lactating women, 13 other
other (explain)	
Who makes decisions on children's hygiene practices in your household? (ie: handwashing, bathing, etc)	1 Myself alone, 2 Myself and my husband, 3 Myself and/or my husband in consultation with elders, 4 Elders/grandparents
In your opinion, is a husband justified in hitting or beating his wife in the following situations:	
If she goes out without telling him	1 yes, 0 no, 99 don't know
If she neglects the children	1 yes, 0 no, 99 don't know
If she argues with him	1 yes, 0 no, 99 don't know
If she burns the food	1 yes, 0 no, 99 don't know
Is the respondent a farmer?	1 yes, 0 no, 99 don't know
Did you receive training on keyhole gardens and/or permagardens?	3 both on keyhole gardens and permagardens, 1 keyhole garden only, 2 permagarden only, 0 none, 99 don't know
Do you have a keyhole garden you are currently cultivating?	1 yes, 0 no
What are the primary crops you grow in the keyhole garden?	1 carrots, 2 onion, 3 garlic, 4 spinach, 5 mustard greens, 6 lettuce, 7 kangkung, 8 collard greens, 9 other
other (explain)	
Do you have a permagarden you are currently cultivating?	1 yes, 0 no
What are the primary crops you grow in the permagarden?	1 carrot, 2 peanut, 3 mustard greens, 4 Kangkung/ morning glory, 5 lettuce, 6 chilli, 7 cucumber, 8 parsley, 9 cabbage, 10 bok choy, 11 onion, 12 garlic, 13 tomato, 14 eggplant, 15 Other
Other (explain)	
Do you grow vegetables in your permagarden(s) for sale or for your own household's consumption?	1 for sale only, 2 for household consumption only, 3 both for sale and for household consumption, 99 don't know, 96 permagarden is not producing
We would like to understand what proportion of the vegetables grown in your permagarden(s) you sell and what proportion is used for your own household's consumption. Here we have 10 stones. Imagine that they are all the vegetables you are growing in the permagarden. I would like you to use them to show which proportion of vegetables grown in your permagarden are sold and the proportion is used for consumption. Please put the stones representing the vegetables sold here [point] and the stones representing the vegetables used for household consumption there [point]. ENUMERATOR: mark how many stones have been set aside as "sold".	
During the past six months, did you make any profit by selling produce from your permagarden?	1 yes, 0 no, 99 don't know
have you faced any challenges with the permagarden?	1 yes, 0 no, 99 don't know
What challenges did you face with the permagarden?	1 natural disaster, 2 limited production, 3 poor quality of seeds, 4 produce was stolen or damaged by others, 5 unable to sell the produce, 6 lack of tools, materials or seeds, 7 personal issues (illness, disability, lack of time, family issues, etc), 8 poor quality of produce, 9 limited amount of land, 10 lack of technical support, 11 lack of money to invest, 12 other
other (explain)	

Do you receive support from agriculture extension services/PPL?	1 yes, 0 no, 99 don't know
where do you get information about agriculture?	1 training, 2 agriculture extension services, 3 NGO worker, 4 revista Lafaek ba Komunidade
What information about agriculture did you learn from Lafaek ba Komunidade?	1 plant vegetables on a banana trunk, 2 planting vanilla, 3 make organic fertilizer, 4 other
Who has the final say in this household on decisions about what to plant in the keyhole garden?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
Who has the final say in this household on decisions about what to plant in the permagarden?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
Who has the final say in this household on whether or not to sell your produce?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
Who has the final say in this household on whether or not to sell or consume a chicken?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
Who has the final say in this household on whether or not to sell or consume large livestock?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
If you would like to start a small business, who would have the final say on it?	1 Myself, 2 My spouse, 3 Me and my spouse together, 4 Grandparent (parents of the respondent/spouse), 5 Myself and my grandparent, 6 Myself, my grandparent, my spouse, 7 My spouse and grandparents (parents of the respondent/ spouse), 8 Someone else, 9 Not
In this household, is there a child under two years of age? A child born after February 2019?	1 yes, 0 no, 99 don't know
[IF THE CHILD'S MOTHER IS A DIFFERENT PERSON, SEEK CONSENT; IF NOT JUST PRESS YES] Hello, I am interviewing on behalf of the HATUTAN project, which aims to support education and children's nutrition in Timor-Leste. I would like to talk to you today and ask for your permission to interview you. Feel free to consult other people in the household before you respond. I will write your answers to use them in our research but we will not mention you by name or share your personal details with anybody outside of our team. Your answers will be mixed with the answers of other parents we have interviewed, making it impossible to identify individual answers.	

Your responses will help us to learn more about the status of children's nutrition and health in Timor-Leste. We will use this information to adjust our project's design. Do you consent in participating in this interview?

What is the baby's name?

Now I would like to ask some questions about your child under 2 years of age \${bebe_naran}

In what year was your child born? (If the child has a health card can ask to see it to get the birthdate)

1 2019, 2 2020, 3 2021, 99 Other

In what month was your child born? (If the child has a health card can ask to see it to get the birthdate)

1 January, 2 February, 3 March, 4 April, 5 May, 6 June, 7 July, 8 August, 9 September, 10 October, 11 November, 12 December

Was the child born after February 2019?

1 yes, 0 no, 99 don't know

How old is your baby in months?

Has \${bebe_naran} ever been breastfed?

1 yes, 0 no, 99 don't know

Was the \${bebe_naran} breastfed during the day or night?

1 Day, 2 Night, 3 Both during the day and the night

Sometimes babies are breastfed by another woman or given breast milk from another woman by spoon, cup, bottle, or some other way. This can happen if a mother cannot breastfeed her own baby for various reasons, such as the mother is sick or away, mastitis, etc.

1 yes, 0 no, 99 don't know

Did \${bebe_naran} consume breast milk in any of these ways yesterday during the day or at night?

Now I would like to ask you about some medicines.

Was \${bebe_naran} given oral rehydration solution yesterday during the day or at night?

1 yes, 0 no, 99 don't know

Next I would like to ask you about some liquids that \${bebe_naran} may have had yesterday during the day or at night Did \${bebe_naran} have:

Did \${bebe_naran} drink plain water yesterday during the day or the night?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} drink infant formula such as (SGM 0-6, or S26) yesterday during the day or the night?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} have any milk such as tinned, powdered or fresh animal milk?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} have any juice or juice drinks?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} have any clear broth?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} have any thin porridge?

1 yes, 0 no, 99 don't know

Did \${bebe_naran} drink breastmilk only yesterday?

1 yes, 0 no, 99 don't know

Any other liquids

1 yes, 0 no, 99 don't know

How many times yesterday did \${bebe_naran} consume formula?

How many times yesterday did \${bebe_naran} consume any milk?

Other (explain)

Now I would like to ask you about (other) liquids or foods that \${bebe_naran} ate yesterday during the day or at night. I am interested in whether your child had the item even if it was combined with other foods. For example, if \${bebe_naran} ate a rice porridge made with a mixed vegetable, you should reply yes to any food I ask about that was an ingredient in the porridge.

1 Bread, cereals/porridge, noodles, rice, mash/residue or other foods made from grains such as maize or wheat., 2 Pumpkin, carrots, squash, orange flesh sweet potatoes or or any other dark yellow or orange fleshed roots, tubers and vegetables?, 3 White potatoes, white yams, white sweet potato, cassava, or any other foods made from roots?, 4 Any dark green leafy vegetables

Please do not include any food used in a small amount for seasoning or condiments (like chilies, spices, herbs, or fish powder), I will ask you about those foods separately. Yesterday, during the day or at night, did \${bebe_naran} eat any of the following?	such as spinach, lettuce, chard, Amaranth, pumpkin leaves, cassava leaves, bean leaves, sweet potato leaves, or cowpea leaves?, 5 Any other vegetables, like cucumbers, tomatoes, cauliflower, cabbage, broccoli, eggplant, etc.?, 6 Ripe mangoes, ripe papaya, melon, passionfruit or other fruits that are dark yellow or orange inside?, 7 Any indigenous (wild fruits), 8 Any other fruits like watermelon, tamarind, jackfruit etc., 9 Any liver, kidney, heart, blood or other organ meats from domesticated animals such as cow, pig, goat, chicken or duck?, 10 Any meat from domesticated animals, such as beef, pork, lamb, goat, chicken, or duck?, 11 Any organs from wild animals, such as game meat, bush rats, birds, wild pigeons, guinea fowl, deer, wild boar?, 12 Any flesh from wild animals, such as game meat, bush rats, wild birds, deer, wild boar, wild goat?, 13 Eggs, 14 Fresh or dried fish, shellfish or seafood?, 15 Any foods made from beans, peas, lentils, peanuts or other legumes such as cowpeas, pigeon peas?, 16 Any foods made from nuts and seeds such as pumpkin, sunflower seeds?, 17 Milk, cheese, yogurt or other milk products? (DO NOT count sweetened condensed milk here), 18 Any shea nut oils, other oils, fats, butter or foods made with any of these?, 19 Any sugary foods such as chocolates, sweets, candies, pastries, cakes or biscuits?, 20 Condiments for flavor, such as chilies, spices, herbs, or fish powder?, 21 Grubs, snails or insects?, 22 Foods made with red palm oil, red palm nut, or red palm nut pulp sauce?, 23 breast milk
ENUMERATOR: Did the child eat anything other than breastmilk?	1 yes, 0 no, 99 don't know
[IS THE CHILD UNDER SIX MONTHS OF AGE - BORN FROM SEPTEMBER 2020 ONWARDS]	1 yes, 0 no, 99 don't know
What are the main reasons for giving food to the baby?	
Lack of time for breastfeeding	0 not mentioned, 1 mentioned
Insufficient milk/ cannot breastfeed	0 not mentioned, 1 mentioned
I think that It's good for children	0 not mentioned, 1 mentioned
Recommendation of relatives	0 not mentioned, 1 mentioned
Traditional/religious belief	0 not mentioned, 1 mentioned
Other	0 not mentioned, 1 mentioned
Other (explain)	
Who makes the decision on what \${bebe_naran} should eat?	1 Myself alone, 2 Myself and my husband, 3 Myself and/or my husband in consultation with elders, 4 Elders/grandparents

ANNEX 5: TERMS OF REFERENCE

I. SUMMARY

Through this solicitation, CARE International in Timor-Leste (CITL) is seeking a consultant(s) (referred to in this document as “the contractor”) to conduct a gender-sensitive midline study to assess impact and identify factors that may affect the efficiency and effectiveness of the McGovern-Dole Food for Education Program HATUTAN in Timor-Leste. This study will be conducted 24 months after the project’s baseline.

For this study CITL requires a representative quantitative survey, inclusive of:

- Reading assessments (EGRA) conducted in the language of instruction (Tetum) with (i) two longitudinally tracked student cohort samples (treatment and comparison groups) and (ii) two cross-sectional samples of Grade 2 students (treatment and comparison). The reading assessments include a working memory assessment and questions to the student on food consumption at home during that day;
- Household surveys conducted with a subsample of parents of assessed students, assessing nutrition, health, hygiene and education practices; food security, savings and livelihoods; and attitudes and practices related to gender;
- School survey, including school management practices; head counts of teachers and students across all primary grades; provision of school feeding; and WASH conditions;
- Structured classroom observations, including teacher background information; use of child-centered and gender-equitable practices in class; and timed assessment of attentiveness.
- A survey with a sample of participants of farmers’ groups, assessing the extent to which their capacity for market engagement has been developed and emerging relationships with the school feeding program.

These tools were designed during the baseline study and will be adjusted as necessary in light of contextual changes, including but not limited to:

- The COVID-19 pandemic and its implications in terms of exposure time to education, and economic impact on households, including food and nutrition security impact;
- The impact of the delays in the approval of the 2020 and possibly 2021 national budget on the implementation of the school feeding program and other social safety nets;
- Natural disasters associated to shifting weather conditions and their impact on targeted project areas; etc.

In addition to these modifications, it is also expected that learning assessments will need to be modified to prevent pre-exposure bias, in particular for the longitudinally tracked cohort. The consultant is expected to use an equating procedure to assess potential differences in the difficulty levels of learning assessments.

The study will also include a qualitative component, obtaining data from focus group discussions, key informant interviews and participatory exercises that will add depth and context and serve to triangulate information from quantitative survey findings and analysis.

The contractor will provide an independent and rigorous midline study to enable the project to assess the effectiveness of the planned interventions in reaching its outcomes, and the extent to which sub-groups, including girls, students from minority language groups and students with disabilities, are benefitting from the project. The study will also assess the factors contributing to the expected outcomes and test the project’s Theory of Change. The report will also reflect on the lessons learned through implementation and provide recommendations for adaptive management based on findings.

The study will contribute directly to the McGovern-Dole Initiative's Learning Agenda. The HATUTAN learning questions are included in Annex 1 of this ToR. The evaluator is expected to consider the learning questions when designing the analysis framework and to include a section on the report responding to the learning questions, indicating which questions are part of the McGovern-Dole Initiative's Learning Agenda.

The contractor should strive to conduct the midline study by May 31, 2021, as agreed with USDA. While the midline study will be externally designed, led and reported on by the Contractor, CITL and CARE USA will provide input and be involved in its design, tool development, coordination with local stakeholders, and contribute to data validation and interpretation through the provision of contextual information on social and gender norms and cultural practices that may influence results. The evaluator will complete the analysis and interpretation of findings independently.

At the time of the preparation of this ToR, Timor-Leste has allowed schools to re-open and field activities are permitted. The situation remains fluid, however, with outbreaks currently taking place in neighboring countries and restrictions to international travelers' arrival. Given the current and potential future restrictions to travel resulting from the ongoing COVID-19 crisis, it is possible that the external evaluator may not be allowed to enter the country and/or conduct fieldwork with the data collection teams. If these restrictions are not lifted by the time the data collection is expected to take place (February-March 2021), or the external evaluator's health conditions do not allow for travel, CARE will put in place measures to allow for remote training of enumerators as well as for remote guidance to on-site project MEL supervisors to oversee data collection. Since all quantitative data will be collected electronically, the external evaluator will have immediate access to the datasets and will be able to provide real-time input to the field data collection teams and supervisors. In order to meet standards and principles of evaluation independence and credibility, the analysis and interpretation of midline findings will be finalized independently by the contractor.

HATUTAN beneficiaries, partners and key stakeholders will be central to the project evaluation management efforts, not only as recipients of targeted interventions, but also as active agents and informants who will support data validation, analyses and utilization. HATUTAN staff will also participate in the midline study as informants, providing the evaluation teams with all necessary project documentation and appropriate administrative/logistical support (e.g., providing contact details for key informants; providing information on security issues and local protocols; etc.), but will not be directly involved in data collection to avoid any influence on respondents.

II. PROJECT BACKGROUND

An effective and efficient School Feeding Program (SFP) nationwide has long been a high priority for the Government of Timor-Leste (GOTL). An effective Food for Education and Child Nutrition program responds strategically to several key interrelated development issues affecting the rural population targeted by HATUTAN. Over a five-year period (2018- 2023), HATUTAN works in partnership with the GOTL and key development stakeholders to strengthen literacy, overall education outcomes, nutrition, health, hygiene/sanitation and gender equality, with positive spillover effects to the rest of the nation. HATUTAN works in 440 pre-primary and primary schools across four of Timor-Leste's most disadvantaged municipalities, namely Ermera, Liquica, Ainaro, and Manatuto, reaching 90,000 school-aged children to enhance education, health, nutrition, hygiene/sanitation and gender equality. Complementary activities conducted in the areas targeted for school feeding are reaching a total of: 3,096 teachers; 502 school administrators and 460 food preparation vendors. Among the 440 targeted schools, 435 have provided daily meals prepared with USDA commodities; out of those, 220 schools in particularly disadvantaged circumstances received a broader intervention package. In these schools, the project is supporting 220 parent and teachers associations (PTAs); 48 Community Development Agents (CDAs); will mobilize a total of 320 village savings and loans groups during the life of the project; and is expected to reach a total of 9,174 pregnant and lactating women (PLW) and

parents with children under five years of age, through community-level health and nutrition interventions. HATUTAN's strategies are based on HATUTAN's Project-level Results Framework, which comprises two strategic objectives (MGD SO1: Improved literacy of School-Aged Children and MGD SO2: Increased Use of Health, Nutrition and Dietary Practices) with nine Intermediate Results (IRs) (Annex 2 - Project-Level Results Framework). Key project activities include: trainings for school administrators and teachers; establishing and strengthening PTAs; building capacity at local, regional and national level in all sectors; conducting advocacy at local regional, and national level for the timely disbursement of resources and commodities, for the development of a national School Health and Nutrition policy, and for resources to support infrastructure improvements; producing books and supplementary reading materials; supporting extra-curricular activities; providing school meals during the hungry season when the government SFP has delivery gaps; conducting trainings in food preparation, storage practices, and nutrition / health / hygiene / sanitation and gender equality; developing partnerships with farmer groups to supply food to schools; forming saving and lending groups; creating school garden laboratories; improving school feeding preparation and cooking areas; and promoting gender equality and the reduction of sexual, domestic, school-based and gender-based violence (GBV) with all program stakeholders. Achieving five foundational results will be critical to ensure sustainable improvements across the entire program, including: Increased Capacity of Government Institutions, Improved Policy and Regulatory Framework, Increased Government Support, Increased Engagement of Local Organizations and Community Groups and Improved Community Activism to Maintain Positive Changes in Social and Gender Norms.²⁵²

III. PURPOSE AND OBJECTIVES OF THE MIDLINE STUDY

The main objectives of the midline study will be to:

- To assess and highlight factors affecting effective, quality and efficient implementation of HATUTAN activities/interventions.
- To assess the progress (strengths and weaknesses) of the project (per each component) against stated outputs and outcomes to date; this will include an assessment of the relevance of the outputs and outcomes through a gender lens.
- To identify new barriers and trends associated to the ongoing COVID-19 crisis, providing recommendations on how to refine the intervention to mitigate its impact on project outcomes.
- To assess early evidence of changes in behavior and practices (both intended and unintended) and compare these with the changes that were expected to be promoted by project activities. Identify factors in the implementation or context that hold back or promote observed and intended changes.
- To assess the efficiency and effectiveness of project organization, management and coordination mechanisms, including quality and usefulness of partnership.
- To assess sustainability efforts to date and potential factors that may impede schools' ability to graduate and sustain activities post-project.
- To recommend the future orientation and emphasis of the project during its remaining time, including course corrections and adjustments to the Results Framework, project design, resource allocation, or implementation process as necessary.

²⁵² Other project specific information will be available to the Contractor after the agreement is signed

IV. INDICATORS FOR COLLECTION AND MIDLINE EVALUATION QUESTIONS

1) Indicators for collection -The Contractor will be responsible for collection data on all indicators listed below. Indicators in **bold font** correspond to standard USDA/McGovern-Dole indicators:

- **Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text**
- Percentage of teachers adhering to improved learning practices in schools
- Percentage of schools with at least 80% of the teachers present during head counts
- Percentage of schools with reading corners in grade 2
- **Number of teachers in target schools who demonstrate use of new and quality teaching techniques or tools as a result of USDA assistance**
- **Number of school administrators and officials in target schools who demonstrate use of new techniques or tools as a result of USDA assistance**
- Percent of students with working memory scores equal or above 50%
- Percent of students who report that they did not consume any food during the school day
- **Average student attendance rate in USDA supported classrooms/schools**
- Percentage of parents (VSLA group members) using part of their savings or loans for education of their children
- Reduction in the number of days of absence from school due to illness
- **Number of individuals who demonstrate use of new child health and nutrition practices as a result of USDA assistance***
- Percentage of participants who are able to correctly identify keeping animals in the kitchen as a non-hygienic practice
- **Number of individuals who demonstrate use of new safe food preparation and storage practices as a result of USDA assistance***
- Percent of participants in program target groups (pregnant-lactating mothers, parents of school children, VSLA group members) who can identify at least three important nutrition/dietary recommendations
- **Number of schools using an improved water source***
- **Number of schools using improved sanitation facilities***
- Percent of targeted schools with improved food preparation and/or storage tools as a result of USDA assistance
- **Number of policies, regulations, or administrative procedures in the following stages of development as a result of USDA assistance****
- **Value of new USG commitments, and new public and private sector investments leveraged by USDA to support food security and nutrition****
- Percentage of schools purchasing food from local farmers for the school feeding program
- Percentage of VSLA groups integrating promotion and dissemination of health/hygiene/nutrition and specific gender messages in their regular activities

**Indicator value should be determined based on an extrapolation of sample results*

*** Indicator will be determined using secondary data from the project's monitoring system and government records*

The contractor will ensure that rigorous practices are used to collect, tabulate, clean and analyze the indicator data.

2) Evaluation questions – The Contractor is expected to ensure that the study will provide valid and reliable data to respond to the evaluation questions listed below:

Design/Relevance

- How relevant/appropriate is the project concept and design to the local context and the education, health, and nutrition conditions?
- How well does the HATUTAN design match the needs, priorities and contexts of the target students, communities and other beneficiaries?
- Are there unforeseen assumptions and barriers? If so, what have been the extent and adequacy of adjustments?
- To what extent do the project activities and outputs contribute to intended effects and impact as declared in the Results Framework?

Management and Coordination

- How effective is the project's general implementation and management in terms of quality and efficiency of inputs and activities, including (but not limited to): adherence to work plans and budgets; management of food commodities; routine monitoring of project activities; and financial management?
- To what extent does the institutional set-up encourage involvement of key stakeholders (e.g. government agencies, local partners, and USDA)?
- What is the current level of cooperation (e.g. programmatic and organizational) among partners and what is the effectiveness of this partnership?

Effectiveness

- What is the achievement, to date, of HATUTAN activities, outputs and outcomes as detailed in the Results Framework?
- What are the major factors that have facilitated or impeded the progress of project implementation?
- For each component, which activities and implementation processes are acceptable to members of the target institutions, and communities and why?
- How well do implementation processes adhere to the underlying strategies, principles and established methodology and processes?
- Are there any specific sub-groups that benefitted more/less from project activities? If so, why?
- Which project strategies being implemented by the project have been more/less effective in contributing to positive results? Why?
- If targets are not being achieved, why not, and what have been the extent and adequacy of adjustments?
- To what extent has the project been successful in advocating for GOTL resources to fund infrastructure improvements (e.g. kitchens, storage rooms, latrines)? How many schools have received resources?
- Cost-effectiveness will be determined based on the triangulation of the size of the change observed in the outcomes of interest; the relative contribution of each intervention component to changes in outcomes; and the investment made in each of the intervention components.

Efficiency

- How efficiently have resources been allocated for project implementation?
- How could the efficiency of project support be improved?

Sustainability

- Are there specific concerns/opportunities in relation to the potential sustainability of the school feeding and literacy skills development strategies, considering the policy framework and mechanisms in place for its implementation?
- Level of adherence to intervention components (fidelity of implementation)

- Assessment of sustainability scores attained by schools, identifying sub-groups that are either highly successful (positive deviants) or lagging behind
- Is there evidence that certain benefits and/or activities are more likely to be sustained than others? If so, why?
- Are the five domains of sustainability identified by the project progressing at a similar pace or not? Why or why not?
- To what extent have local actors (community entities/associations, local government agencies, private sector, etc.) been engaged and committed (contributing time and resources) to ensuring sustainability? Which of them has demonstrated the greatest/least potential for contributing to sustainability? Why?
- What specific institutional capacity needs should the project address to foster greater host country ownership, engagement, and accountability? At the national level? At the municipal level? At the community level?
- What steps has the project taken to ensure the sustainability of adequate human, financial and other resources?
- What will help ensure greater likelihood of sustaining project achievements and activities post project in terms of sustained motivation, resources, capacity and linkages?
- The red flags or challenges that project implementers should be aware of and focus on during the second half of the project to further sustainability efforts.

Impact

- To what extent have project activities led to improved literacy among school-aged children?
- To what extent have project activities led to increased use of appropriate health, nutrition, hygiene/sanitation, and gender equality practices?
- What difference (early evidences of changes in knowledge, attitude, behavior, and practices) did project interventions make to the beneficiaries (considering the difference between what has happened with and what would have happened without the project)?
- Did any negative effects occur for beneficiaries? If so, what were they and why did they happen?
- To what extent has the project been successful in advocating for and securing timely distribution of resources and commodities to arrive at schools before the start of the school year (i.e. January)

Gender

- What differential outcomes and impacts (if any), are project activities having on women and girls? Men and boys?
- To what extent is the project actively addressing gender-specific barriers and constraints to education, health, nutrition and community engagement?
- To what extent has the project contributed to reducing incidences of violence, namely (a) domestic violence; (b) school-based violence; (c) sexual and gender-based violence?
- What specific actions and interventions has the project fostered to promote gender equality and how successful were they?

COVID-19 Impact

- How has the COVID-19 crisis impacted students, teachers and households? How does the negative impact of the crisis differ by gender, age and location?
- What adaptations are recommended to mitigate the persistent impact of the crisis on learning and student health?
- How can the SFP be adapted to continue to serve as a safety net for students in the context of a persistent pandemic, as well as in similar crises?

Learning and Recommendations

- What are the key/critical emerging learnings that HATUTAN should deepen to improve the quality of interventions?
- What should HATUTAN modify to improve implementation and achieve intended results; are any changes/adjustments in the results framework, in implementation strategies and approaches, in resource allocation and/or in management/coordination mechanisms necessary?

V. MIDLINE STUDY DESIGN AND METHODOLOGY

Following the evaluation methodology determined previously to the baseline, the contractor should use a mixed-methods quasi-experimental design, triangulating information from different sources and methods (e.g. quantitative and qualitative) to enhance the reliability and comprehensiveness of findings. In addition, the midline will use gender-sensitive and socially inclusive methods, ensuring that women, girls, men and boys from different groups of stakeholders are able to provide data in a safe and reliable manner and that their perspectives are adequately represented in data analysis. This may include conducting: sex-specific FGDs, so women feel comfortable and encouraged to participate, and women and men are able to share issues and concerns normally not discussed in mixed groups; appropriate approaches to engage child respondents; and the use of an analysis framework that allows the project to assess to what extent it is addressing gender-specific barriers and cultural constraints to nutrition, education, health, WASH, and community engagement, as well as barriers for children with disabilities and those from marginalized subgroups identified at the baseline.

While all tools are expected to replicate the baseline items in order to ensure comparability of results to assess progress, the project and the External Evaluator will work with the Ministry of Education, Youth and Sports (MEYS), Ministry of Agriculture and Fisheries (MAF), the Ministry of Health (MOH), KONSANTIL (National Council for Food Security) and the Prime Minister's Cabinet to review the tools and methodology for the midline. The midterm evaluation may incorporate additional areas of inquiry of interest of the aforementioned Ministries, for instance analysis of the impact of COVID-19 on learning patterns and household food security, thus contributing to sector learning and planning.

1) The quantitative component:

Learning survey - will assess changes in the literacy outcomes of beneficiary students. Progress in acquiring foundational reading skills will be measured assessing the differences in the average grade 2 scores from cross-sectional samples from intervention and comparison groups; and through a difference-in-differences analysis of the average scores for the longitudinally tracked intervention and comparison samples. It is expected that the intervention will result in, or exceed, an additional increase of 0.2 standard deviations over and above the average results of the comparison group. Student samples drawn from treatment and comparison schools should be refined using propensity score matching to ensure comparability of characteristics. Outliers should be removed.

Given the potential issues with exposure to the treatment in the upcoming months as a result of the COVID-19 crisis, the evaluator is expected to also conduct a treatment-of-the-treated (ToT) analysis, including only schools that have received the full intervention since project inception. It is expected that some schools may remain closed or be closed for extended periods of time in case of new outbreaks; therefore, a ToT analysis will allow the project to assess the impact of the model.

Sampling – The student sample drawn at the baseline will be longitudinally tracked; additionally, new samples²⁵³ of 1,014 girls and 1,056 boys²⁵⁴ will be randomly drawn from Grade 2 classes in the same

²⁵³ The sample size is calculated as $N=2*[(Z1-\alpha/2+Z1-\beta)/ES]$, where ES is the estimated effect size

²⁵⁴ The sample size is calculated as $N=2*[(Z1-\alpha/2+Z1-\beta)/ES]$, where ES is the estimated effect size.

99 treatment and 90 comparison clusters²⁵⁵ distributed across the four targeted and five comparison municipalities. The longitudinally tracked sample will include only students currently attending school, regardless of the grade they are attending. The household survey sample will be drawn solely from the parents of newly sampled Grade 2 students.

Data Collection Procedures - The quantitative data will be collected from students, teachers, school administrators and school records. At the student level, an adapted version of the Early Grade Reading Assessment (EGRA) will be used. The Tetum tool and the working memory test developed at the baseline will be revised to avoid pre-exposure bias. A subsample of the Grade 2 students will be asked to respond to both the previous and current EGRA and working memory tools to allow for the calculation of an equating factor.

Tools will be piloted with students in non-sampled schools before deployment and refined as necessary. Enumerators will use ONA for data collection, associated with paper-based copies of EGRA tasks for student use. Font size and type will be adapted to ensure readability by students, noting the limited access to corrective lenses in Timor-Leste. Electronic data collection will minimize errors and allow for real-time correction of data collection procedures.

Data collection procedures will be spot-checked by representatives of the MEYS at municipal and administrative post level, as well as by senior CARE MEL staff to ensure compliance with the required procedures and legal requirements in country. Neither MEYS representatives nor CARE staff will participate directly in surveys or interviews to avoid bias.

Structured Observations - At the teacher level, structured observations will be used in Grade 2 classrooms. Observations will be limited to language classes (Tetum) during which reading-focused activities are likely to take place. The structured classroom observation tool will be based on similar tools developed by CARE, which consider: (i) time spent on task; (ii) student engagement in class; (iii) student use of reading materials provided by the project; (iv) teacher's use of child-centered techniques; and (v) gendered behaviors in class.

School-level Data - A school survey will be conducted in each sampled school. It will include four modules: (a) an individual student module which will capture attendance and reading trimester scores for sampled students; (b) an attendance module, which will collect head counts in Grades 2 and attendance rates for each class; (c) a teacher module, which will capture information on the number and qualification of teachers, their background and attendance; and (d) a school feeding module, which will include data on school infrastructure (kitchens, sanitation), school feeding program implementation and procurement of locally produced items to complement the school meals.

Household Survey - A household survey will be conducted with caregivers of students sampled for EGRA and with a separate sample of PLW and/or mothers of children under five, randomly selected in the same sample points. To ensure representation in each municipality, the sampling frame of household survey will be stratified by municipality, and allocated proportionately, based on the overall distribution of households across all municipalities. The sample of 862^{256,257} household respondents will be randomly drawn from the cross-sectional Grade 2 student sample. The enumerators will then select eligible individuals in each household. The household survey will include household composition; socio-economic background of households; education, health, nutrition and WASH-related behaviors/practices; time allocation for children; gender-related attitudes and practices; and perspectives on SGBV. Different individuals will be eligible for interview, depending on the target group

²⁵⁵ Considering a minimum of 20 students per grade (considering class size in rural and remote areas) and potential contamination in 20%

²⁵⁶ The sample size is calculated as $N = [D * (Z_{1-\alpha} + Z_{1-\beta})^2 * [p_1(1-p_1) + p_2(1-p_2)]] / (p_2 - p_1)^2$

²⁵⁷ The Contractor should indicate by how much the number of households to be sampled will be pre-inflated to account for household non-response.

relevant for each chosen module. A “take all” approach that selected all eligible individuals per household for each module will be used. For example, the caregiver’s module will be administered to mothers of children under five and PLW, which means that all women in the household who meet the criteria will be selected and interviewed. Because all individuals who meet the eligibility criteria for each survey module will be selected, this level effectively has no further random sampling.

Farmer Group Survey: A subsample of household respondents currently participating in farmers’ groups will respond to an additional survey module (Farmer Group survey), assessing exposure to capacity building activities; perceived relevance; readiness for market engagement; emerging linkages with buyers, including schools; and barriers to market engagement.

Data Analysis - Average results for each EGRA task will be analyzed and reported separately for treatment and comparison groups (refined through propensity score matching). Data will be disaggregated by sex. Additional disaggregation levels - municipality, location (rural/urban), mother tongue of the student and disability- will be included to illustrate potential trends, as the sample is not sufficiently powered to determine statistically significant differences at municipal/mother tongue level. EGRA results will be triangulated with: (i) the findings of classroom observations; (ii) qualitative data collected from teachers; (iii) attendance data from head counts and school records; and (iv) data on household practices and socio-economic background. The analysis will seek to identify if the acquisition of reading skills is predicted (or not), by classroom practices; the availability and use of reading materials at home and school; attendance patterns; functionality of school PTA; childcare, health, nutrition and WASH practices at home; household participation in VSLA; and experience of SGBV.

The study will also include a comprehensive analysis of the efficiency and effectiveness of the implementation of the SFP; impact of participatory management on school governance, SFP implementation and infrastructure improvement; and WASH patterns. At household level, the study will assess the impact of the intervention on nutrition, health and hygiene practices as well as financial empowerment, market linkages (for farmers’ groups) and gender practices.

2) The qualitative component of the midline study will work toward two objectives. First, it will aim to collect and analyze data to complement the school and household surveys, and second, it will provide a robust understanding of the processes of change affecting education, food and nutrition security, gender (particularly SGBV), health, nutrition and hygiene/sanitation outcomes.

The qualitative study will cover all four targeted municipalities, all 18 administrative posts, and one village and one school in each administrative post. In making the selection of villages/schools the following specific criteria will be considered: geographic balance; ethnic balance; the food and nutrition security context; seasonal context (considering agricultural livelihoods); and variation in terms of maternal and child health and nutrition (MCHN) practices and logistics (e.g. safety, ease of access, impact of COVID²⁵⁸, etc.).

Participants of qualitative interviews will include PTAs; mothers (including pregnant women); fathers; teachers; and school principals.

Two types of interviews will be conducted, as follows:

- Focus group discussions (FGDs): with mothers, fathers, teachers and PTA members.
- Key informant interviews (KIs): local government officials and school administrators.

The contractor will be responsible to submit description of instruments that will be developed and the type of questions to be asked, e.g. key informant interview guide, focus group guides etc.

²⁵⁸ While the number of cases in Timor-Leste remains limited to date, the economic impact of the pandemic is likely to be disproportionately felt across target areas, depending on the main livelihoods.

At least two qualitative data collection teams (one per every two municipalities) should be formed, each comprising at least one native speaker of local languages. All focus group discussions and interviews should be recorded digitally, with the permission of the participants. All collected qualitative data will be transcribed and translated. Transcriptions and translations will be spot-checked for completion and correctness. The analysis will be conducted considering emerging themes in the qualitative data as well as pre-coded themes, thus allowing the project to capture unexpected impacts.

The external evaluator will be required to review and finalize the proposed sampling frameworks for both qualitative and quantitative samples. These should be of a sufficient size and representativeness allowing:

- reasonable levels of certainty that the findings are representative for the target population;
- reasonable ability to generalize the intervention's effectiveness to similar contexts; and
- reasonable ability to generalize the insights into what works and why for similar contexts.

3) Training of enumerators

The project will be responsible for hiring enumerators in country. The external evaluator is expected to design the training curriculum; provide remote training and guidance on field practicum; develop a post-test assessment to determine readiness for data collection; and provide guidance for team leaders. The training is expected to include a summary of the project; ethical protocols for research with children and vulnerable people; prevention of sexual abuse and other forms of violence and harassment during data collection; inclusive and gender-sensitive data collection; review of the tools; data quality assurance; and practical sessions. The training will include representatives from MEYS, MAF and MOH assigned by their respective entities to observe the sessions and provide input as appropriate. Given the ongoing COVID-19 crisis and restrictions on travel, the external evaluator is not expected to deliver the training in-person or to accompany staff to the field (except made in circumstances where the selected evaluation team is already present in country). CARE staff will provide on-the-ground support to ensure delivery of remote training and to enable remote spot-checks during practicums and data collection.

VI. MIDLINE STUDY DELIVERABLES AND REPORTING OUTLINE

a) Deliverables

The contractor is responsible for the following deliverables:

- Briefings – schedule of briefings will be determined jointly by contractor and project
- Inception report, inclusive of proposed midline approach, outline of revisions to tools, revisions to the sampling plan, data quality assurance protocol and analysis framework.
- Revised quantitative and qualitative instruments. The project will be responsible for translation and back-translation of edits to the instruments.
- Field procedure manual for enumerators and training curriculum.
- Midline workshop for training of enumerators and team leaders (the contractor will bear the cost of travel and attendance if in country).
- Datasets (both raw and clean) and data files (including audio recordings, transcriptions and translations) – submitted no later than three weeks after completing data collection. Depending on the contractor's proposal, the project may take the responsibility for transcription and translation of qualitative files in country.
- Draft midline study report and analysis/ syntax files – submitted no later than four weeks after completing survey data collection in the field
- Final midline study report – submitted no later than two weeks after receiving comments from the project on draft final report

- A 2-3 page stand-alone brief describing the evaluation design, key findings and other relevant considerations. It will serve to inform any interested stakeholders of the midterm evaluation, and should be written in language easy to understand by non-evaluators and with appropriate graphics and tables.
- Powerpoint summary of findings
- Revised version of the midline study report including feedback from USDA and the Ministry of Education, Youth and Sports to the final report.

b) Outline of midline study report

- Cover page, Table of contents, List of acronyms
- Executive summary – should be a clear and concise stand-alone document that states the most salient findings, conclusions and recommendations of the study and gives readers the essential contents of the midline report in the three to five pages.
- Introduction and contextual analysis
- Methodology and study design – should describe the methodology and design of the quantitative and qualitative components, constraints and limitations to the study, and issues in carrying out the study.
- Findings – should present results from the quantitative survey and qualitative study. Results should be presented in aggregated and disaggregated form, allowing for an in-depth understanding of impact for subgroups of interest. Results from the quantitative survey should be analyzed and discussed; and triangulated with other quantitative and qualitative data for interpretation and validation of results. Any bivariate and multivariate analysis undertaken should also be included. A comprehensive analysis of the factors affecting (i) the acquisition of reading skills and (ii) student nutrition and health should be included.
- Tabular summary of quantitative study results – should present findings of the quantitative survey in table form for all indicators and for the aggregate program area in Timor-Leste. Results should include all disaggregation levels listed in the project's Performance Monitoring Plan.
- Conclusions and recommendations – should provide high-level conclusions from the midline study and recommendations for the design and implementation of the future final evaluation. Recommendations must be relevant to the project and context and include concrete and realistic steps for implementing.
- Learning – should include a discussion of the project's learning questions (Annex 1), while also highlighting which questions are part of the McGovern-Dole Initiative's Learning Agenda.
- Annexes – all special documentation identified as necessary or useful (e.g. quantitative survey/qualitative study instruments, sampling plan for the quantitative survey, syntax files etc.).
- All final versions of international food assistance evaluation reports will be made publicly available. The evaluators shall provide a copy of the evaluation report that is free of personally identifiable information (PII) and proprietary information. Final versions of evaluation reports ready for publication should be accessible to persons with disabilities. For guidance on creating documents accessible to persons with disabilities, please see the following resources: <https://www.section508.gov/create/documents>.

VII. CONTRACTOR QUALIFICATIONS AND TEAM COMPOSITION

The selected firm/consultant(s) shall possess the following qualifications:

- Demonstrated experience and expertise in designing, organizing, and managing the implementation of the large-scale quantitative surveys in developing countries;
- Demonstrated experience and expertise in designing, organizing and conducting qualitative research, data collection and analysis in developing countries;
- Demonstrated experience and expertise in the statistical analysis of complex survey data and in analyzing data from mixed-method studies;
- Expertise in educational programming
- Expertise in maternal and child health and nutrition;
- Expertise in gender integration
- Ability to deliver high-quality written products

For planning purposes, the team for this study will consist of team leader and key personnel with defined technical expertise (in the areas of education, MCHN, gender) that will provide varying technical and subject matter expertise. The team should include local consultants with expertise, knowledge and experience in Timor-Leste or Southeast Asia and/or Pacific Regions.

VIII. MIDLINE STUDY MANAGEMENT

The midline study will be managed by the project's M&E Specialist, with technical oversight from CARE USA's Technical Advisor – M&E and Senior Research Advisor, Education. This individual will play a key role in soliciting evaluation bids, managing the proposal review and selection process, as well as directly overseeing all evaluation activities. The CARE USA advisors will be separated from the line management of the project in order to ensure independence and credibility in the decisions made during selection of bids, design and interpretation of findings.

CARE is responsible for arranging vehicle rental and drivers as needed for site visits and field work. CARE is also responsible for making hotel arrangements, providing workspace, internet access, printing and photocopying. The contractor will be required to bring its own computers and required software.

Ethical protocols

The contractor is required to set out its approach to ensuring complete compliance with international good practice with regards to research ethics and protocols particularly with regards to safeguarding children and vulnerable groups. The contractor will be required to sign CARE International in Timor-Leste's Child Protection Policy and the Policy on Protection from Sexual Harassment, Exploitation and Abuse. Compliance breaches with either policy will result in immediate termination of the contract, investigation of cases and legal action as applicable.

Adult individuals identified as eligible respondents for this study will be informed about the purposes and scope of the study; how the information provided will be used and confidentiality principles; and about their right to refuse to participate or to terminate their participation at any time. Enumerators will be required to obtain consent from parents/ teachers for children's participation in the study.

In line with CARE's principles on the protection of vulnerable individuals and children, the contractor is expected to ensure complete confidentiality of the information provided and safe storage of data, including during submission of data sets to CARE. Reports shall not include names, location markers or other information that allows for the identification of respondents. No visual documentation (photos or videos) of interviews or interviewed individuals will be allowed.

IX. MIDLINE SCHEDULE/TIMELINE

The contractor must submit a timeline of activities as part of its proposal, which should follow the timeline set forth in Section VI of this Scope of Work and in the table below.

Timeframe of key midline activities:

Activity	Responsible	Period
Prepare the final midline TOR and share with key stakeholders (including USDA) for review/comments and approval	CARE International in Timor-Leste (CITL)	August 14, 2020
Recruitment and procurement of evaluator (“the contractor”)	Deputy Chief of Party, HATUTAN	October 30, 2020
Inception report	The Contractor	November 30, 2020
Midline workshop	The Contractor	January 14-19, 2021
Primary field data collection activities completed	The Contractor	April 02, 2021
Submit draft midline study report	The Contractor	April 30, 2021
Review the draft midline report and provide feedback	The project/ MEYS ²⁵⁹	May 15, 2021
Submit the final midline study report	The Contractor	May 31, 2021
Review final midline report & submit to USDA for approval	The program	June 18, 2021
Revise the final midline report to incorporate feedback from USDA (and MEYS if not incorporated pre-submission)	The Contractor	July 31, 2021 (estimated)
Follow-up & dissemination with stakeholders	The program	June 2021
Develop an action plan to apply findings; review and adjust (if necessary) the Results Framework and Performance Indicators, implementation strategies, targets, etc	The program	June 30, 2021

X. BUDGET

The consultant must, in addition to a technical proposal, submit a budget showing the projected level of effort (LOE) for each proposed full-time and/or short-time member of team. Other costs that should be included are international travel and per diem, in-country costs for data collections and interviewing, communications, report preparations and other costs as appropriate. A six-day work week is authorized when working in-country.

²⁵⁹ In case the timeline does not allow for an immediate response from MEYS, the consultant will revise the report to incorporate MEYS feedback after submission to USDA.

XI. EVALUATION CRITERIA FOR PROPOSALS

Submitted proposals will be evaluated on the merit of the proposed approach, including the following criteria:

- 30% - Technical approach as illustrated in the description of proposed methodology;
- 25% - Timeline reflecting proposed activities, which emphasizes the ability to meet propose deadlines;
- 25% - Key personnel and composition of the technical team
- 20% - Past performance. The offeror should also include in the submission a list of references, preferably of USDA and USDA-funded programs, related to the completion of an evaluation study for a McGovern-Dole Food for Education Program.

XII. INTELLECTUAL PROPERTY

CARE International in Timor-Leste shall solely and exclusively own all rights in and to any work created in connection with this midline study, including all data, documents, information, copyrights, patents, trademarks, trade secrets or other proprietary rights in and to the work. In accordance with 2 CRF 200.315 Intangible Property, USDA has the right to

- (1) Obtain, reproduce, publish, or otherwise use the data produced during the evaluation; and
- (2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes.

XIII. SUBMISSION

The deadline for submission is on 26 October 2020 by 10 am Dili, Timor-Leste time.

Submissions should be emailed to: Angelina.Santos@careint.org

And questions about TOR should be emailed to Angelina.Santos@careint.org by 15 October 2020

Proposals should include the following:

- Technical proposal – bidders approach to the midline
- Financial proposal
- Qualifications – how does the bidder meet the qualifications, including CVs of all key involved persons
- References

TOR ANNEX 1: LEARNING AGENDA QUESTIONS

The midterm evaluation is expected to contribute to the following areas of inquiry in the USDA/McGovern-Dole learning agenda:

School meal program implementation

- What community-level systems of governance and management are required for the successful implementation and sustainability of school meal projects?
- Which components of school meal programs, including food production, procurement, and preparation of meals, are the most sustainable in terms of operational efficiency and why? Does the cost-effectiveness of HATUTANs change over time?
- What are the most effective methods for ensuring food safety within school meal projects taking into consideration the different systems of national, regional, local and community governance?

- In what ways do school meal projects impact health equity in terms of poverty, gender, or geography?

Education

- In what ways do the combination of school meal interventions and educational interventions improve education and literacy levels? How can these combinations improve cost-effectiveness?
- In what ways do school meal interventions impact the resources of teachers, such as classroom time, teacher incentives, and teacher capacity?
- In addition to the McGovern-Dole Learning Agenda questions, the midterm evidence will also contribute to the following learning questions:

Literacy

- What are the effects of school meal interventions on the acquisition of reading skills in multilingual settings?

Health

- How do water, sanitation and hygiene (WASH) projects impact learning and literacy outcomes?
- What are the best practices in sustaining WASH interventions delivered through a school meal project?

Nutrition

- How effective are school-based meals at preventing and or decreasing health related absences?
- To what extent has Social Analysis and Action (SAA) been effective in addressing barriers to improved nutrition? Are there specific groups that benefitted more than others? Who are they and why did this happen?

Agriculture

- How can a combination of local procurement during harvest time be supplemented with international food aid to promote locally and/or nationally sustainable school meals projects?
- What kinds of partnerships with the private sector and/or host country governments are the most effective at ensuring project sustainability? Among successful partnerships, who are the key players and what are their roles? In what contexts do private sector and/or government partnerships work best and which contexts may be more challenging?

Gender-based violence

- To what extent does the synchronization of school meal programming with activities to address sexual and gender-based violence (SGBV) and school-based affect literacy outcomes and health/nutrition practices at household level?

Sustainability

- The following combination of four strategies will be assessed for their combined impact on sustainability: (i.) strengthened capacity of school staff and authorities to manage school feeding projects; (ii.) enhanced community participation in monitoring school feeding and education outcomes; (iii.); private sector engagement; and (iv.) building advocacy skills of stakeholder groups.

TOR ANNEX 2: RESULTS FRAMEWORK

Project-level Results Framework
FY18 McGovern-Dole Food for Education
Program in Timor-Leste (CARE-Mercy
Corps-WaterAid)

