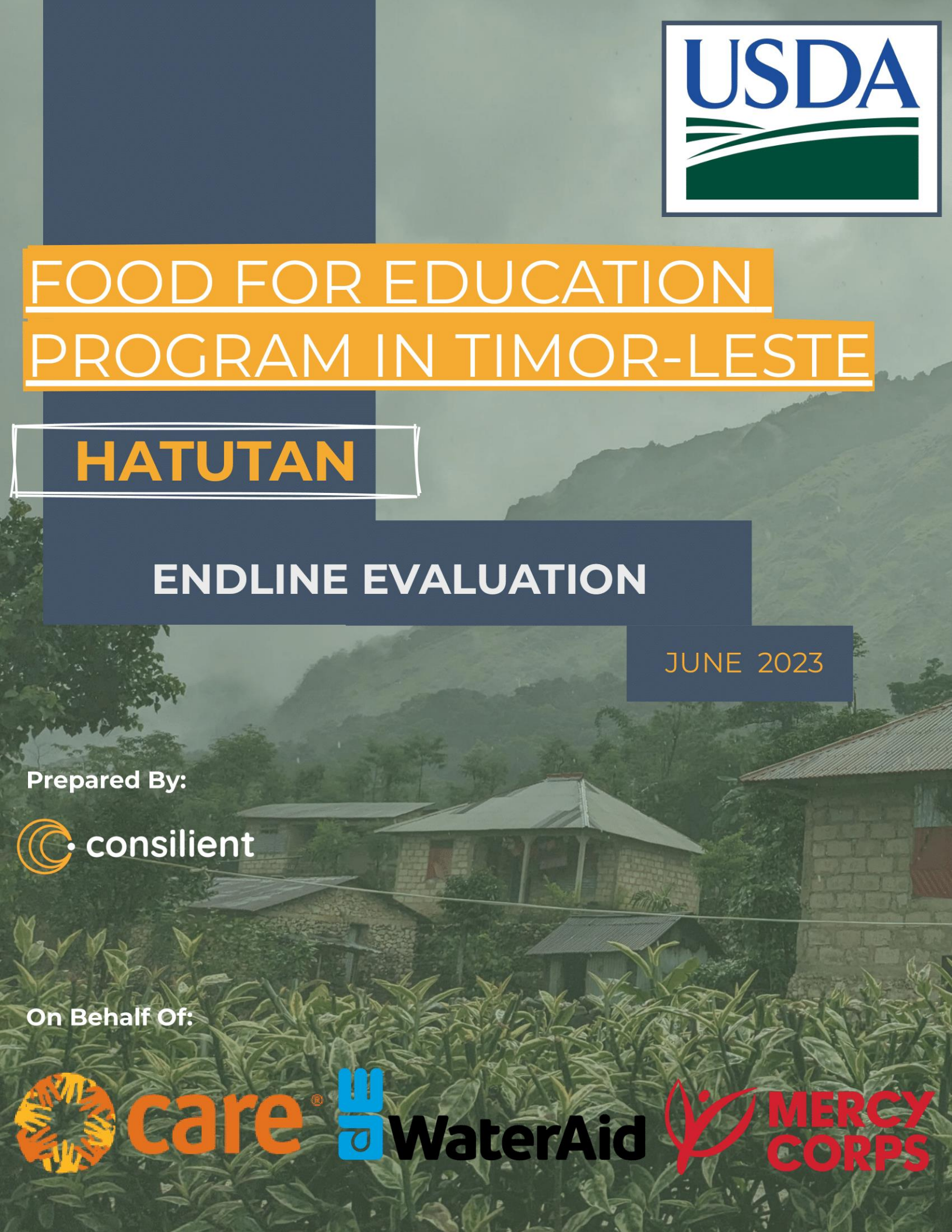


# FOOD FOR EDUCATION PROGRAM IN TIMOR-LESTE

## HATUTAN

## ENDLINE EVALUATION

JUNE 2023


Prepared By:

consilient


On Behalf Of:

care

WaterAid

MERCY CORPS

# TABLE OF CONTENTS

**List of Tables** ... **iii**
**List of Figures** ... **v**
**Acronyms** ... **vii**
**Executive Summary** ... **viii**
**Introduction** ... **1**
Timor-Leste Overview ... 1
HATUTAN Program Overview ... 7
**Methodology** ... **9**
Evaluation Objectives ... 9
Evaluation Design ... 9
Data Collection Tools ... 10
Fieldwork ... 14
Proposed and Achieved Samples ... 15
Challenges and Limitations ... 16
**Literacy Outcomes** ... **19**
Overall Literacy Scores ... 19
Literacy Subtask Scores ... 26
Predictors of Literacy ... 39
**Quality of Instruction** ... **43**
Teaching Practices ... 43
Teacher Skills and Knowledge ... 50
Teacher Attendance ... 52
Literacy Instruction Materials ... 55
Predictors of Engaging Teaching Practices ... 57
**Student Attentiveness** ... **58**
Attentiveness ... 59
Working Memory ... 60
Student Hunger ... 63
Predictors of Student Attentiveness ... 65
**Student Attendance** ... **68**
Student Attendance ... 68
Dropout Rates ... 72
Health-related Absences ... 75
Other Factors Affecting Attendance ... 76
Predictors of Student Attendance ... 77
**School Management** ... **79**
Quality of School Management ... 79
Parent-Teacher Associations ... 80
**School Feeding Program** ... **83**
Provision of School Meals ... 83
Management and Quality of the SFP ... 85
Food Preparation and Storage ... 88
Purchase of Local Foods ... 93
Community Contribution to School Feeding Program ... 97
**Health and Nutrition** ... **99**

Nutrition Practices ..... 100
Body Mass Index ..... 107
Knowledge of Nutrition ..... 109
Health Practices ..... 110
Knowledge of Health and Hygiene ..... 113
Access to Clean Water and Sanitation ..... 116
Access to Healthcare ..... 121
Predictors of Health and Nutrition ..... 123
**Agricultural Practices ..... 127**
Permagarden Use ..... 127
Challenges Faced by Farmers ..... 129
Association with Nutritional and Economic Outcomes ..... 130
Farm Sales ..... 131
**Economic Empowerment ..... 132**
Savings ..... 132
VSLA Loans ..... 134
VSLA Membership, Nutrition, and Investment ..... 135
**Gender and Power ..... 135**
Division of Labor ..... 136
Household Decision-making ..... 140
Control of Productive Assets ..... 141
Claiming Rights and Meaningful Participation in Public Decision-making ..... 143
Control Over One's Body ..... 144
Violence and Restorative Justice ..... 146
Aspirations and Strategic Interests ..... 150
**Learning Agenda ..... 153**
Education and Literacy ..... 153
School Feeding Program Implementation ..... 156
Health and Nutrition ..... 157
Agriculture ..... 159
**Program-specific Outcomes ..... 160**
Design/Relevance ..... 160
Management and Coordination ..... 162
Effectiveness ..... 163
Efficiency ..... 165
Sustainability ..... 166
Impact ..... 168
Gender and Social Equity ..... 169
**Conclusions and Recommendations ..... 170**
Literacy and Education ..... 170
School Feeding ..... 172
Health and Nutrition ..... 172
**Annex 1: Theory of Change ..... 155**
**Annex 2: Indicators and Outcomes ..... 156**
HATUTAN Indicators Assessed at Baseline, Midline, and Endline ..... 156
McGovern-Dole Standard and Custom Outcomes ..... 160
**Annex 3: Methodological Analysis ..... 162**
Demographics of Achieved Sample and Differences In Intervention and Comparison Samples ..... 162
Differences in Samples Across Rounds ..... 166

**Annex 4: Food Group Categories** ..... **172**
**Annex 5: Sample Characteristics** ..... **175**

# LIST OF TABLES

Table 1: Comparison of partial and full support schools ..... 8
Table 2: EGRA sections and scoring ..... 11
Table 3: School survey and classroom observation sample ..... 15
Table 4: Learning assessment and household survey sample ..... 16
Table 5: Analysis of changes in overall literacy scores from baseline to endline ..... 20
Table 6: Analysis of changes in overall literacy scores from midline to endline ..... 21
Table 7: Change in overall literacy scores by gender, grade 2 students ..... 25
Table 8: Analysis of changes in letter recognition scores ..... 27
Table 9: Analysis of changes in invented word reading scores ..... 30
Table 10: Analysis of changes in familiar word reading scores ..... 32
Table 11: Analysis of changes in passage reading scores ..... 35
Table 12: Analysis of changes in reading comprehension scores ..... 37
Table 13: Use of engaging teaching practices (% of classrooms) ..... 44
Table 14: Change in use of traditional teaching practices ..... 46
Table 15: Change in use of negative teaching practices ..... 47
Table 16: Change in caregiver perceptions of negative teaching practices ..... 47
Table 17: Change in negative teaching practices towards girls and boys ..... 49
Table 18: Treatment of male and female students by gender of teacher at endline ..... 50
Table 19: Changes in teacher training and education ..... 52
Table 20: Change in teacher attendance on the day of survey ..... 54
Table 21: Changes in access to reading corner and reading materials ..... 56
Table 22: Predictors of engaging teaching practices ..... 58
Table 23: Difference-in-differences analysis of working memory ..... 62
Table 24: Percent of students who ate the day of the EGRA, by intervention municipality ..... 64
Table 25: Change in average attendance rates for grade 2 ..... 68
Table 26: Average attendance by gender ..... 69
Table 27: Attendance rates by municipality ..... 70
Table 28: Change in reasons for missing school ..... 70
Table 29: Dropout rates disaggregated by municipality ..... 73
Table 30: Dropout rates disaggregated by gender ..... 74
Table 31: Days missed because of illness ..... 76
Table 32: Average days missed weekly because of illness, disaggregated by gender ..... 76
Table 33: Change in access to public spaces and services ..... 77
Table 34: Individual-level predictors of student absence ..... 78
Table 35: School-level predictors of student absence ..... 78
Table 36: Change in administrator provision of coaching ..... 80
Table 37: Frequency of PTA meetings ..... 81
Table 38: PTA involvement ..... 81
Table 39: Meals provided to students, by study group ..... 84
Table 40: School feeding program responsibility, by study group ..... 85
Table 41: School feeding program menu available, by study group ..... 85
Table 42: Menu items by study group among schools with meals served ..... 86

Table 43: School menu dietary diversity score, by study group ..... 87
Table 44: Frequency of detergent use and availability of food storage, by treatment group ..... 90
Table 45: Availability of school facilities, by study group ..... 91
Table 46: Storage sanitation, by study group ..... 91
Table 47: Schools buying local produce from farmers, by study group ..... 94
Table 48: Local produce schools bought from farmers, by study group ..... 95
Table 49: Schools buying local produce from farmers, by municipality and study group ..... 96
Table 50: PTA involvement in school feeding, by study group ..... 98
Table 51: Change in consumption by food group, female caregivers of child-bearing age ..... 101
Table 52: Change in consumption by food group, children under age 2 ..... 102
Table 53: Foods consumed by students at endline ..... 107
Table 54: Change in achievement of McGovern-Dole Custom Outcome #23 ..... 110
Table 55: Change in handwashing frequency ..... 111
Table 56: Change in achievement of McGovern-Dole Custom Outcome #21 ..... 116
Table 57: Change in access to improved water source ..... 117
Table 58: Change in school toilet availability ..... 120
Table 59: Change in handwashing infrastructure at schools ..... 120
Table 60: Change in use of savings and loans for healthcare ..... 123
Table 61: Change in permagarden use ..... 128
Table 62: Change in permagarden use by municipality ..... 129
Table 63: Challenges faced by farmers at endline ..... 129
Table 64: Difference in nutritional outcomes by permagarden use ..... 130
Table 65: Difference in economic outcomes by permagarden use ..... 131
Table 66: Change in households selling crops ..... 131
Table 67: Change in percentage making a profit from crop sales, by municipality ..... 132
Table 68: Change in prevalence and location of savings ..... 133
Table 69: Change in use of savings ..... 133
Table 70: Change in households taking loans from VSLAs ..... 134
Table 71: Difference in difference regression for VSLA loan use ..... 134
Table 72: Difference in nutritional and economic outcomes by VSLA membership ..... 135
Table 73: Percentages of second-grade children performing various daily household tasks ..... 136
Table 74: Change in the gendered division of children's labor within households ..... 137
Table 75: Time spent by second-grade children on various daily household tasks ..... 138
Table 76: Change in time spent on household tasks and impact on schooling ..... 138
Table 77: Caregivers' reports on the impact household tasks on their child's schooling ..... 139
Table 78: Change in household decision making ..... 141
Table 79: Change in control of productive assets ..... 142
Table 80: Change over time in whether respondents receive support from AES or training on keyhole gardens, by gender ..... 144
Table 81: Change in attitudes towards gender-based violence ..... 145
Table 82: Change in teachers' discipline practices ..... 147
Table 83: Teacher's use of the different discipline practices, by student's gender ..... 148
Table 84: Reporting of abuse at school ..... 149
Table 85: Change in perceptions of gendered literacy and numeracy skills, as well as perceptions of gendered experiences at school ..... 151
Table 86: Key infrastructure within intervention schools ..... 163
Table 87: Summary of HATUTAN sustainability ..... 167
Table 88: Summary of HATUTAN impact ..... 168
Table 89: Differences in demographic characteristics of students by treatment group ..... 163

Table 90: Students with disabilities by treatment group ..... 163
Table 91: Differences in head of household and household characteristics by treatment group ..... 164
Table 92: Differences in caregiver characteristics by treatment group ..... 165
Table 93: Differences in school characteristics by treatment group ..... 166
Table 94: Differences in demographic characteristics of students by round ..... 167
Table 95: Students with disabilities by round ..... 168
Table 96: Differences in head of household and household characteristics by round ..... 169
Table 97: Differences in caregiver characteristics by round ..... 170
Table 98: Differences in school characteristics by round ..... 171
Table 99: Food group categorization, caregivers and children under 2 ..... 172
Table 100: Food group categorization, students ..... 173

# LIST OF FIGURES

Figure 1: Overall literacy scores from all rounds, grade 2 students ..... 19
Figure 2: Percent of students scoring 0% overall on EGRA ..... 22
Figure 3: Overall literacy scores from all rounds, grade 2 students scoring greater than 0% ..... 23
Figure 4: Overall literacy scores from all rounds by gender, grade 2 students ..... 25
Figure 5: Letter recognition scores from all rounds, grade 2 students ..... 27
Figure 6: Letter recognition scores from all rounds, grade 2 students scoring greater than 0% ..... 28
Figure 7: Invented word reading scores from all rounds, grade 2 students ..... 29
Figure 8: Invented word scores from all rounds, grade 2 students reading more than zero WPM ..... 31
Figure 9: Familiar word reading scores from all rounds, grade 2 students ..... 32
Figure 10: Familiar word scores from all rounds, grade 2 students reading more than zero WPM ..... 33
Figure 11: Passage reading scores from all rounds, grade 2 students ..... 34
Figure 12: Passage reading scores from all rounds, grade 2 students scoring greater than 0% ..... 35
Figure 13: Reading comprehension scores from all rounds, grade 2 students ..... 36
Figure 14: Reading comprehension scores from all rounds, grade 2 students scoring greater than 0% ..... 38
Figure 15: Predictors of literacy, individual level ..... 40
Figure 16: Predictors of literacy, school level ..... 42
Figure 17: Change in McGovern-Dole Custom Outcome #5 for teaching practices ..... 45
Figure 18: Change in percent of teachers attending literacy training ..... 51
Figure 19: Change in teacher attendance ..... 53
Figure 20: Change in students self-reporting that they can pay attention in class ..... 59
Figure 21: Change in students observed paying attention in class ..... 60
Figure 22: Change in working memory score ..... 61
Figure 23: Change in percent of students who ate the day of the EGRA ..... 63
Figure 24: Change in households going without food during 30-day period ..... 65
Figure 25: Predictors of working memory ..... 66
Figure 26: Predictors of observed student attentiveness ..... 67
Figure 27: Change in dropout rates, grade 2 students ..... 72
Figure 28: Change in caregiver-reported student dropout and re-enrollment rates ..... 75
Figure 29: Change in school meal provision ..... 84
Figure 30: Change in access to clean water to prepare school meals ..... 89
Figure 31: Consumption by food group, female caregivers of child-bearing age ..... 100
Figure 32: Change in dietary diversity score, women of childbearing age ..... 104
Figure 33: Change in dietary diversity score, children ages 6 months to 23 months ..... 104

Figure 34: Change in attainment of minimum acceptable diet, children ages 6 months to 23 months .... 105
Figure 35: Change in use of exclusive breastfeeding, children under 6 months of age ..... 106
Figure 36: Actual and theoretical BMI distributions, intervention students ..... 108
Figure 37: Change in nutrition knowledge ..... 109
Figure 38: Situations when caregivers washed hands ..... 112
Figure 39: Child vaccination prevalence ..... 113
Figure 40: Change in handwashing and hygiene scores ..... 114
Figure 41: Handwashing and hygiene knowledge, intervention caregivers at endline ..... 115
Figure 42: Change in household access to a toilet ..... 118
Figure 43: Change in household access to handwashing station ..... 119
Figure 44: Change in affordability of healthcare ..... 122
Figure 45: Predictors of nutrition knowledge ..... 124
Figure 46: Predictors of caregiver dietary diversity ..... 125
Figure 47: Predictors of student dietary diversity ..... 126
Figure 48: Predictors of health knowledge ..... 127
Figure 49: Change in permagarden use ..... 128
Figure 50: Change in attitudes towards gender-based violence ..... 146

# ACRONYMS

| | |
|---|---|
| AES | Agricultural Extension Services |
| BCG | Bacille Calmette-Guérin (tuberculosis vaccine) |
| BL | baseline |
| BMI | Body Mass Index |
| CBA | curriculum-based assessment |
| DDS | dietary diversity score |
| DiD | difference in differences |
| EGRA | Early Grade Reading Assessment |
| EL | endline |
| FGD | focus group discussion |
| (S)GBV | (sexual and) gender-based violence |
| GoTL | Government of Timor-Leste |
| HATUTAN | *Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun* (Food to Support Nutrition and Education) |
| HoH | head of households |
| KII | key informant interview |
| LRP | locally and regionally procured |
| MEYS | Ministry of Education, Youth, and Sports |
| ML | midline |
| OPV | Oral Poliovirus Vaccine |
| PTA | Parent-Teacher Association |
| SFP | School Feeding Program |
| TVET | technical and vocational training and education |
| U.S. | United States |
| USDA | United States Department of Agriculture |
| VSLA | Village Savings and Loan Association |
| WASH | water, sanitation, and hygiene |

# EXECUTIVE SUMMARY

In this report, we present findings from the endline evaluation of the HATUTAN (*Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun* or Food to Support Nutrition and Education) program. Running between 2018 and 2023, the program worked 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 program operated in 443 schools and communities within four of Timor-Leste's most deprived municipalities, Ainaro, Ermera, Liquica, and Manatuto, supporting an estimated 105,967 target beneficiaries including school-aged children, teacher, school administrators, and community members. Key activities included support for the government-run School Feeding Program (SFP) and trainings targeting teachers, school administrators, and community members.

## METHODOLOGY

The endline evaluation was designed to understand the impact of the HATUTAN program on its target beneficiaries. The evaluation assesses the impact of HATUTAN and provides useful evidence to inform the McGovern-Dole Program Learning Agenda and Government of Timor-Leste (GoTL) initiatives related to school feeding, the education sector, and school health and hygiene. Following the evaluation design used for the HATUTAN baseline and midline, this endline evaluation primarily uses a difference-in-differences approach, comparing the progress observed in primary schools supported by the HATUTAN program—"intervention schools"—to progress observed in a comparison group of schools. By comparing progress in intervention and comparison schools, we are better able to understand whether improvements are due to HATUTAN program activities or are rather due to external factors affecting all schools in Timor-Leste.

Data collection occurred between February 6, 2023 and March 31, 2023. In total, data was collected from 98 intervention communities and 84 comparison communities. This included school surveys and classroom observations at target schools in all communities, 1,468 literacy assessments with grade 2 students in intervention schools, 1,041 literacy assessments with grade 2 students in comparison schools, 741 household surveys in intervention communities, and 602 household surveys in comparison areas.

One key challenge facing the endline evaluation was the timing of data collection. Students were assessed at the beginning of the school year; however, the standard definition for assessment requires assessments to be conducted with students at the end of grade 2. As such, results in this study may be biased downwards compared to other McGovern-Dole funded initiatives assessing impact at the end of grade 2.

## FINDINGS

### Learning and Education

**The HATUTAN program had a significant positive impact on grade 2 students' literacy levels.** Following a decline in literacy scores at midline due to the COVID-19 pandemic and resultant school closures, at endline, literacy scores had rebounded to levels above baseline for intervention students. We find that the HATUTAN program may have both mitigated the negative impacts of COVID-19 on learning, as shown at midline, and allowed students to learn more quickly after returning to school post-COVID, although the program's impact on literacy post-COVID appears to have been more modest.

However, literacy abilities remain very weak, and many students remain unable to read words. The average overall score on the EGRA was only 12.0% for intervention students, and only 20.4% of intervention students demonstrated the ability to read and understand the meaning of a grade-level passage. Furthermore, **HATUTAN's impact appears to have been greater for students who were at least able**

**to recognize letters**; in contrast, the program appears to have had little impact on reducing the percent of students scoring 0% on literacy assessments.

**HATUTAN had a significant impact on lessening the use of traditional teaching practices in intervention schools.** However, we find no significant impact on the overall use of engaging or negative teaching practices, although we find a possible reduction in the use of corporal punishment towards girls. Additionally, we find that teacher attendance may have declined among intervention schools relative to comparison schools at endline. The data suggests that the reduction in attendance may be due to teacher training programs; while these programs may be helpful to strengthen the quality of instruction, their negative impact on attendance should be addressed.

In addition to quality of instruction, students' literacy abilities are also affected by their attentiveness and attendance at school. HATUTAN intended to address these areas by supporting the SFP to reduce student hunger, among other activities.

**We find a significant positive impact of HATUTAN on student attentiveness** as measured through classroom observations. We also find a significant relative reduction in intervention student hunger at endline. These positive findings suggest that HATUTAN activities effectively targeted attentiveness.

In contrast, **we find a decline in student attendance among intervention schools, while attendance in comparison schools improved.** In other words, we do not find positive impact of HATUTAN on student attendance. While the exact reason for this is unclear, we note that caregivers frequently reported that natural disasters had caused their children to miss school, a factor outside the control of the HATUTAN program. However, sickness was also frequently cited as a reason for absence, suggesting a need to further bolster health, hygiene, and nutrition practices.

In a more positive finding, **we find a decline in dropout rates from baseline to endline**. This decline, however, was relatively consistent across intervention and comparison schools, suggesting that the reduction may be due to broader conditions across Timor-Leste rather than due to HATUTAN.

## School Feeding Program

One of the primary aims of the HATUTAN program was to assist the Government of Timor-Leste in the successful implementation of the school feeding program in all basic education and preschools throughout the school year. **We find a significant positive impact on school meal provision at endline compared to baseline.** However, school meal provision substantially declined since midline in both intervention and comparison schools due to delays in government provision of SFP funding, suggesting a need for continued advocacy with the Government of Timor-Leste to improve the consistency and reliability of funding.

Furthermore, **school meals have a relatively low level of dietary diversity.** Purchasing patterns and dietary diversity of school meals continue to be constrained by the reality that the average cost of a nutritional diet using nutritious, locally available food items is too high. Continuing to support linkages between schools and local farmers will remain an important component of HATUTAN II.

**The majority of schools have a PTA involved in overseeing the SFP.** However, participation in PTAs and their levels of activity remain limited; in intervention areas, almost half of PTAs had during the current school year. Given that PTAs address many areas of relevance to HATUTAN, continued support for these bodies may be an important component of HATUTAN II.

## Health, Nutrition, and Agriculture

**HATUTAN appears to have had little impact on nutrition practices,** including the quality of diets consumed by women of child-bearing age and children under the age of 2 years. Concerningly, we find evidence suggesting a widespread decrease in dietary quality over the past five years, with very low consumption of protein-rich foods.

Of additional concern, **almost half of the grade 2 students in intervention schools and over half of those in comparison schools have underweight BMIs**. This finding emphasizes that weak dietary diversity and nutritional practices have a tangible impact on students' health outcomes, and reiterate the importance of the SFP and other activities seeking to improve food consumption and dietary quality.

**Nutrition knowledge was relatively low across all rounds, with very limited change at endline.** While caregivers do have some knowledge of healthy nutrition practices, more work is needed to expand knowledge.

**Handwashing practices deteriorated from midline to endline** for both intervention and comparison groups. We also find a decline in the prevalence of handwashing stations within households and schools. Given that the midline evaluation occurred during the COVID-19 pandemic, these results are likely a reflection of higher-than-average levels of handwashing due to COVID-19.

However, in a more positive finding, across all rounds, **knowledge of handwashing and hygiene practices was high for both intervention and comparison groups.** We also find possible slight improvement in knowledge of hygienic practices due to the HATUTAN program.

Overall, these findings suggest that knowledge of nutrition, health, and hygiene may not be effectively translating into good practices. As such, simply improving knowledge may not translate into improved behaviors. It may be more effective to focus on other barriers to improved nutrition, such as cultural norms, economic circumstances, gender and power dynamics, and preferences for processed and sugary foods.

Regarding agricultural practices and participation in Village Savings and Loan Associations, we find no significant impact of HATUTAN on cultivation of permagardens or on farm sales. However, **we find a significant positive impact on whether households had savings at endline.**

## Gender and Power

In line with broader patterns in Timor-Leste, we find that grade 2 girls generally have higher school achievements than boys, including higher literacy levels. However, adult women have limited decision-making power in households, and more so for younger women than older ones. Furthermore, women are often less involved in community organizations such as farmer's groups and receive fewer benefits from local services such as agricultural extension services. Thus, despite the great potential evidenced by young female students, gender norms seem to reduce the options available to girls as they transition into adulthood.

At endline, **we find little evidence on change in gender norms as a result of HATUTAN programming.** This points towards a need for further interventions that sustain girls' successes at young ages through adulthood. Activities that challenge gender norms, in combination with interventions that help girls stay in school as they transition from childhood to adulthood, may help achieve this goal. However, it is important to note that gender norms tend to change slowly over time.

## HATUTAN Program Efficacy

Overall, this evaluation shows evidence of moderate program impact. We additionally find that **the HATUTAN program was highly relevant to the needs and priorities of schools and communities in Timor-Leste, and was managed efficiently.** However, the sustainability of progress made to date appears somewhat limited. As such, the successful implementation of HATUTAN II over the next five years will be vital to reinforce progress made during HATUTAN.

# INTRODUCTION

The HATUTAN program (*Hahán ne'ebé Atu fó Tulun ho Nutrisaun no Edukasaun* or Food to Support Nutrition and Education) was a five-year initiative (2018 – 2023) to build a partnership between schools and communities in order to improve literacy, learning, health, and nutrition for children and adults in Timor-Leste. The program worked 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.

To achieve these objectives, the program supported, 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. An effective and efficient SFP is a high priority for the Government of Timor-Leste and responds strategically to key interrelated development issues affecting rural populations in the country. Key program activities included strengthening and supplementing the government-sponsored SFP, building school capacity through trainings for teachers and administrators and provision of resource materials, and supporting 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 also sought to conduct trainings related to nutrition, health, and other topics, and to promote gender equality and the reduction of gender-based violence.

The HATUTAN program was 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 was implemented by a consortium led by CARE International with Mercy Corps and WaterAid. The lead Timorese government partner was the Ministry of Education, Youth, and Sports (MEYS), in collaboration with the Ministry of Health, Ministry of State Administration, and Ministry of Agriculture and Fisheries.

This report presents the endline evaluation of the HATUTAN program. This evaluation focuses on change in key outcomes over the past five years of program implementation. We utilize a quasi-experimental difference-in-differences approach, analyzing the relative change in program intervention schools and similar comparison schools in order to better understand the direct impact of HATUTAN on schools, students, households, and farmers. This approach is discussed more in the *Methodology* section; first, we provide an introduction to the Timor-Leste context and the HATUTAN program.

## TIMOR-LESTE OVERVIEW

Timor-Leste gained independence in 2002 following more than three centuries of colonial rule by Portugal and decades of conflict with Indonesia. 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,[1] and destroyed approximately 70% of the country's infrastructure.[2] 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

[1] "Timor-Leste: Background and U.S. Relations," *Congressional Research Service,* June 27, 2019, https://fas.org/sgp/crs/row/IF10320.pdf.
[2] Nicole Stout, "Infrastructure in Timor-Leste Growing According to Strategic Plan," *The Borgen Project*, February 23, 2018, https://borgenproject.org/infrastructure-in-timor-leste.

population of 1.3 million live below the national poverty line,[3] 75% are chronically or mildly food insecure,[4] and around 30% of the adult population (ages 15 and above) are illiterate.[5] Furthermore, while governance has improved substantially since independence and the country has held four free and fair elections, many government institutions still have low capacity, especially at the subnational level.

## Education

The education system in Timor-Leste consists of four levels: preschool, basic education, secondary education, and higher education (university and polytechnic). "Basic education" includes primary and pre-secondary schooling, 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 central basic schools, which usually provide education for all three cycles and manage clusters of filial schools, which usually provide first and second cycle classes. A minority of filial schools provide only first cycle classes. Filial schools are generally located in remote and rural areas.

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, development of strategic education plans, and for encouraging the establishment of Parent-Teacher Associations (PTAs) in all schools. The academic council is responsible for strengthening curriculum delivery, providing pedagogical support and training, and improving teacher performance and professionalization.[6]

Education is a high priority for the Government of Timor-Leste; indeed, 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."[7] Current priorities include achievement of universal completion of basic education by 2030, elimination of illiteracy (particularly among youth ages 15-24), and achievement of gender parity in school teaching and administration. 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.[8]

Along these lines, the government has pushed for universal enrollment in basic education and committed around 10% of the annual national budget to expenditure related to education.[9] Government expenditure on education has included investment in infrastructure, which increased the number of preschools, primary schools, and secondary schools from 943 in 2002 to 1,715 in 2017;[10] 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.[11]

[3] 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.
[4] "Timor Leste: Chronic Food Insecurity Situation 2018-2023," *Integrated Food Security Phase Classification*, January 17, 2019, https://www.ipcinfo.org/ipc-country-analysis/details-map/fi/c/1151924/?iso3=TLS. See also "Timor Leste: IPC Acute Food Insecurity Analysis November 2022 – September 2023," *Integrated Food Security Phase Classification*, February 14, 2023.
[5] "Literacy rate, adlt total (% of people ages 15 and above) – Timor-Leste," *World Bank*, October 24, 2022, https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=TL.
[6] Timor-Leste Ministry of Education, *National Education Strategic Plan 2011-2030* (Dili, Timor-Leste: Ministry of Education, 2011).
[7] World Bank, *Timor-Leste Basic Education Strengthening and Transformation* (Washington, D.C.: World Bank, 2020).
[8] Timor-Leste Ministry of Education, *National Education Strategic Plan*, 2011.
[9] World Bank, *Timor-Leste Basic Education*, 2020.
[10] "Número de escolas e de alunos em Timor-Leste quase duplicou nos últimos 15 anos – PM," *Diario de Noticias*, 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.
[11] "Timor-Leste: Education and Literacy," *UNESCO Institute for Statistics*, accessed February 23, 2020, http://uis.unesco.org/en/country/tl.

However, despite this focus on education and recent related improvements in 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), is low, an issue discussed more in the section *Literacy Outcomes*. Similarly, a curriculum-based assessment conducted in 2017 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.[12] In accordance with these poor educational outcomes, repetition and dropout rates are relatively high for students in primary school, especially at early grade levels.

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.[13] Teacher certifications range from full teacher training qualifications to emergency waivers qualifying an individual to serve as a teacher.[14] 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.

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-Prasa—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.[15]

In addition to instruction challenges, many students face low levels of access to learning resources and insufficient infrastructure. While initiatives by the 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 of teachers' capacity to use 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, and many students lack time to access libraries, which may instead be predominantly used by teachers.[16] In secondary school, average class sizes are also generally high.[17]

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 curriculum-based assessments for both math and language learning, and have lower dropout and repetition rates than boys.[18]

## School Feeding Program

Starting in 2005, 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.

[12] World Bank, *Timor-Leste Basic Education*, 2020.

[13] Within the data collected during classroom observations for the endline evaluation, 46.5% of teachers had only a secondary education, 38.3% had a teaching diploma, and the remaining 15.8% had a diploma from a teacher training institute or higher. These statistics, however, likely overestimate the prevalence of teachers with just a secondary education.

[14] 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).

[15] World Bank, *Timor-Leste Basic Education*, 2020.

[16] Ibid.

[17] Timor-Leste Ministry of Education, "Statistical Data: Average Class Size," accessed February 23, 2020, http://www.moe.gov.tl/pt/emis/dados-estatistico.

[18] Fasih et al., *Using EGRA for an Early Evaluation*, 2019.

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.[19] The government has fully managed the program since 2011. In addition to government support, local programs rely on PTAs for regular monitoring and selection of cooks.

The SFP provided equipment as well as 25 cents per child per day ($0.42 from 2023 onwards) in budgetary support for schools to buy local produce to supplement meals. The GoTL also provided rice to primary schools for meals until 2017. However, purchase of local produce is limited due to constraints including limited and seasonal local production, a lack of linkages with farmers, insufficient budget to purchase more expensive local products, and the greater convenience of buying pre-made food from nearby vendors – especially for schools without a kitchen. As a result, the food composition of school meals is frequently poor, with only 35% of basic education schools meeting a recommended composition of two vitamins-rich foods, one protein-rich item, and one carbohydrate as of 2019.[20]

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; in 2018 delivery plummeted further, and 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. Indeed, as of February 2023, schools had yet to receive SFP funding from the government (although some schools in Manatuto used some of the remaining funds from the 2022 fiscal year to initiate school feeding for a short period of time – around 10 days at most). 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.[21]

## 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.[22] 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,[23] 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.[24] 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. Rural and poor households receive, on average, poorer quality healthcare than urban or wealthier households.[25] Additionally, Timor-Leste has one of the highest tuberculosis incidence rates in the world, and the incidence

[19] Stephen Lister, Jane Keylock, and Trish Silkin, *Timor Leste: An evaluation of WFP's portfolio (2008-2012)* (Rome: World Food Program, 2013).
[20] CARE and Julie Imron, *School Feeding Program Study Report: Timor-Leste* (Dili: CARE, 2019).
[21] Ibid.
[22] Sophie Cousins, "Health in Timor-Leste: 20 years of change," *The Lancet World Report* 394 (2019): 2217-8.
[23] 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).
[24] Cousins, "Health in Timor-Leste," 2019.
[25] World Bank, *Timor-Leste COVID-19 Emergency Support Project: Project Information Document* (Washington, D.C.: World Bank, 2020).

of non-communicable diseases has risen; these diseases now account for 62% of all deaths in the country.[26]

Malnutrition also remains a severe problem in the country. The 2020 Timor-Leste Food and Nutrition Survey found that 47% of children under 5 were stunted, or too short for their age, an indication of chronic undernutrition; 8.6% of children under age 5 were wasted, or too thin for their height, an indication of acute malnutrition; and 32% of children under age 5 were underweight. The stunting rate has decreased slightly since 2016 (from 50% to 47%), and a similar pattern was observed in wasting (from 11% to 8.6%), while the proportion of underweight children went from 37.7% to 32%. While there is a positive trend, malnutrition remains 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 2020, the Food and Nutrition Survey found that 19% of women were underweight. As of 2016, 23% of women aged 15-49 were anemic.[27] 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.[28]

Low levels of access to improved sanitation and poor hygiene practices exacerbate health and nutrition challenges. Access to improved water sources and sanitation facilities remains limited, especially in rural areas. Handwashing practices are also generally weak; while these practices improved during COVID-19, they have since declined.[29] 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.

## 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, the payment of bride prices, and customary rules regarding property rights, inheritance, and succession to traditional offices. Those practices vary between patrilineal and matrilineal groups, however. Although its prevalence has declined over time, early marriage is also a persistent issue; a relatively high proportion of women are married by age 20, while the average age of marriage for men is much higher.[30] As of 2016, 8.4% of the girls ages 15-19 had ever been married or in a union, and 5.2% have ever given birth.[31]

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.[32]

[26] Cousins, "Health in Timor-Leste," 2019.
[27] General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016.*
[28] Integrated Food Security Phase Classification (IPC), *Timor-Leste: Chronic Food Insecurity Situation 2018-2023* (Rome: IPC, 2018).
[29] General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016.*
[30] Asian Development Bank (ADB), Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment* (Mandaluyong City, Philippines: ADB, 2014).
[31] General Directorate of Statistics, Ministry of Planning and Finance and Ministry of Health, *Timor-Leste Demographic and Health Survey 2016.*
[32] Athena Nguyen, Alison Darcy, and Louise Kelly, "CARE Rapid Gender Analysis: COVID-19 Timor-Leste," *CARE*, April 27, 2020.

Timor-Leste has successfully increased girls' enrollment in primary and secondary schools, with girls' enrollment rates now exceeding boys' at lower primary school levels. Girls' education outcomes far surpass those of boys, with higher transition rates into secondary school (67% compared to 63% of boys) and gross enrolment at secondary level (86% compared to 72% of boys).[33] However, girls may face gender-related barriers to education, such as sexual harassment, early pregnancies, and lack of adequate sanitation facilities.[34] Both girls and boys may also face violence at school and in the household. Women are also less likely to attend and complete tertiary studies and technical and vocational education and training (TVET) than men. Additionally, only around 37% of teachers are female[35] and relatively few women work in the Ministry of Education, particularly in decision-making positions, which poses a significant challenge to improving challenging gender dynamics within the education system.[36]

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. 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.[37]

Furthermore, 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 33% of women ages 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).[38] 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, self-blame, distance, and cost.[39]

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.[40] A study on causes of school dropouts found that 35% of girls at risk of dropout in grades 4-6 feel unsafe traveling to and from school, and 26% do not feel safe at school.[41] 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.[42]

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,[43] a child and family welfare system to protect children, and a National Commission on the Rights of the Child.[44] However, in many cases, community leaders and elders are responsible for dispensing justice rather than police or

[33] Ministry of Education, Youth, and Sports, 2019 EMIS
[34] Ibid.
[35] UNESCO Institute for Statistics. "Secondary education, teachers (% female) - Timor-Leste." Accessed October 24, 2022. See https://data.worldbank.org/indicator/SE.SEC.TCHR.FE.ZS?locations=TL.
[36] ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*, 2014.
[37] Nguyen, Darcy, and Kelly, "CARE Rapid Gender Analysis," 2020.
[38] Ibid.
[39] ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment,* 2014.
[40] "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.
[41] USAID/Creative Associates International, *School Dropout Prevention Pilot Program – Situational Analysis/ Timor-Leste,* pg.24
[42] ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*, 2014.
[43] Ibid.
[44] Nguyen, Darcy, and Kelly, "CARE Rapid Gender Analysis," 2020.

the judicial system. This system is problematic in cases when customary justice does not provide sufficient safeguards for women's and children's rights.[45] 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.

### Impact of COVID-19

The first case of COVID-19 in Timor-Leste was reported in March 21, 2020, following which the government declared a state of emergency and enacted public health measures, including restrictions on international travel, school closures, restrictions on gatherings and businesses, and hygiene measures, to reduce the spread of the virus. These measures helped reduce the health burden of COVID-19 on Timorese, with only 138 total deaths recorded.[46]

However, public health measures, voluntary changes in behavior, and global economic shocks post-COVID had substantial impact on economic activity and food security. As of December 2022, Timor-Leste's gross domestic product (GDP) per capita (not including oil revenues) had not yet returned to pre-pandemic levels. The price of fuel and some imported commodities rose sharply in 2022, inflation reached 7.9 percent in August 2022, and food inflation rose by 8.3 percent.[47] The increase in fuel, food, and other commodity prices has negatively affected livelihoods, aggravated malnutrition, and had a direct impact on rural markets given the dependency on staples brought from the capital Dili (such as rice).

## HATUTAN PROGRAM OVERVIEW

The HATUTAN program operated within 443 schools and communities in the municipalities Ainaro, Ermera, Liquica, and Manatuto to improve education, nutrition, health, hygiene/sanitation, economic empowerment, and gender equality. The program focused 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 for pre-primary and primary school children.
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, included in Annex 1, argued that by providing schools meals, teacher training, and related support, school enrollment and student literacy skills 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.

[45] ADB, Government of Timor-Leste, and UN Women, *Timor-Leste Country Gender Assessment*, 2014.

[46] World Health Organization, "Timor-Leste COVID-19 Homepage," accessed May 10, 2023, https://covid19.who.int/region/searo/country/tl. Note from the program: The actual morbidity may be underestimated due to families underreporting deaths occurring at home in rural areas.

[47] World Bank, *Improving the Quality of Public Spending is Critical to Accelerating and Sustaining Economic Development in Timor-Leste,* Dec. 2, 2022, https://www.worldbank.org/en/news/press-release/2022/12/02/world-bank-improve-the-quality-of-public-spending-to-accelerate-development-in-timor-leste.

The program's four target municipalities were selected due to having poor education and health indicators in the country. Within these municipalities, as of October 2022, the program had reached an estimated 105,967 target beneficiaries, including 90,000 school-aged children, 1,419 teachers, 503 school administrators, 805 food preparation vendors, 219 parent-teacher associations, 48 Community Development Agents, 3,799 community members through Village Savings and Loans Associations (VSLAs), and 9,174 pregnant and lactating women and parents with children under the age of five. In total, HATUTAN operated in 443 schools, which includes 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 had a national-level advocacy component to address barriers to SFP implementation and improved education outcomes.

HATUTAN provided two packages of support in target areas: "partial package" and "full package." The partial package included provision of commodities for school feeding (oil, rice, and beans) between January and March, copies of supplementary literacy materials (including magazines and books for early grade readers), and training for school administrators on SFP management, and encompassed all the pre-school and basic education schools in the four municipalities. The full package was implemented in 219 schools and their surrounding communities, representing about half of the public preschools and primary schools in the target municipalities. The support included provision of commodities for school feeding (oil, rice, and beans) between January and March; provision of literacy materials (storybooks and educational magazines); coaching of school directors/coordinators[48] and teachers;[49] 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 included 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.

**Table 1: Comparison of partial and full support schools**

| Type of Package | Partial | Full |
|---|---|---|
| Interventions included | 1. Provision of commodities for school feeding<br>2. Provision of literacy materials<br>3. Training for school administrators on SFP management | 1. Provision of commodities for school feeding<br>2. Provision of literacy materials<br>3. Training for school administrators on SFP management and other topics<br>4. Coaching of teachers<br>5. Mobilization and trainings of PTAs<br>6. Implementation of school dialogue and improvement plan<br>7. Support for extracurricular activities<br>8. Training of parents on VSLAs, agriculture, health, WASH, and gender |
| Number of schools | 232 | 219 |
| Addressed in endline evaluation? | No | Yes |

[48] On school management, improving school learning environments, providing support to teachers on literacy skills development, SFP delivery, improved health, hygiene, and gender practices.
[49] On classroom management and literacy skills development.

# METHODOLOGY

In this section, we discuss the research design and methodology for the HATUTAN endline evaluation. We begin with a discussion of the evaluation objectives, followed by the evaluation design, data collection tools, sampling methodology and achieved sample, and limitations and assumptions. More detailed methodological analysis is included in Annex 3.

## EVALUATION OBJECTIVES

The endline evaluation was designed to understand the impact of the HATUTAN program on schools, students, teachers, families, farmers, VSLA members, and other stakeholders. The evaluation assesses the impact of HATUTAN and provides useful evidence to inform the McGovern-Dole Program Learning Agenda and Government of Timor-Leste (GoTL) initiatives related to school feeding, the education sector, and school health and hygiene. Along these lines, the endline evaluation was designed to address the following objectives:

1. Assess the impact of HATUTAN's intervention against program indicators;
2. Identify key school- and household-level factors contributing to HATUTAN's impact, and map those against program interventions and the program's Theory of Change;
3. Conduct a gender and power analysis, identifying if and how gender norms and power relationships have affected HATUTAN results;
4. Assess the efficiency of HATUTAN's interventions and its cost effectiveness; and
5. Generate evidence to contribute to the program's learning agenda, the McGovern-Dole Program's global Learning Agenda, the GoTL's implementation of the national SFP, education sector planning, and public interventions on school health and hygiene.

Overall, the endline evaluation seeks to determine impact and provide learning on key program activities, including activities addressing literacy, the quality of instruction, student attentiveness and attendance, school feeding programs, nutrition and health knowledge and practices, economic empowerment, agricultural practices, and gender and power dynamics. The evaluation additionally assesses program implications for design relevance, management and coordination, effectiveness, efficiency, sustainability, impact, and gender sensitivity. The evaluation further provides key recommendations for CARE, implementing partners, and Timorese government officials to learn from and adapt the HATUTAN II program.

## EVALUATION DESIGN

Following the evaluation design used for the HATUTAN baseline and midline, this endline evaluation uses a mixed-methods quasi-experimental design, triangulating information from different data collection tools and using quantitative and qualitative methods to enhance the reliability and comprehensiveness of findings. The evaluation primarily utilizes a difference-in-differences approach, comparing the progress observed in primary schools supported by the HATUTAN program—"intervention schools"—to progress observed in a comparison group of schools selected in similar, neighboring municipalities but unaffected by HATUTAN programming. By comparing progress in intervention and comparison schools, we are better able to understand whether improvements are due to HATUTAN program activities or are rather due to external factors affecting all schools in Timor-Leste, such as the COVID-19 pandemic or government initiatives. For example—as demonstrated in the midline evaluation—in the context of the COVID-19 pandemic, if we evaluated only intervention schools, we might find that literacy results in HATUTAN schools declined from 2020 to 2021, suggesting that the program may have reduced student performance. In

contrast, a difference-in-differences approach would likely show that the program had helped mitigate negative external effects, and that as a result, intervention students' literacy results had declined by *less than* those of comparison students.

To enhance the validity of this evaluation approach, comparison schools and communities were intentionally selected to match the socio-economic characteristics of intervention schools, particularly considering linguistic backgrounds, livelihoods, and geographies. Where possible, we avoided selecting comparison schools and communities that been impacted by project interventions similar to HATUTAN, as "contamination" with similar interventions would bias results. Overall, the selection of comparable schools and communities with, where possible, no similar interventions allows for more confident attribution of any findings to the impact of the HATUTAN program, rather than external factors.

Although this evaluation generally follows the design used for the baseline and midline, unlike at midline, this study did not recontact students who were assessed in prior evaluation rounds. Instead, we focused on revisiting schools assessed in prior rounds in order to establish a cross-sectional sample of grade 2 students—i.e., a new group of students who had not been assessed at baseline or midline. This methodology allows us to understand the impact of HATUTAN on one of its main target groups; however, use of a cross-sectional sample also faces several limitations, discussed at length below and assessed in Annex 3.

All methods are gender sensitive and socially inclusive. The fieldwork was conducted to ensure that women, men, girls, boys, and individuals from minority groups (such as persons with disabilities) were able to provide data in a safe and open environment; specific steps taken during fieldwork include conducting gender-specific focus group discussions with mothers and fathers and using appropriate approaches for the engagement and protection of child respondents. Our approach to the analysis similarly ensures that perspectives from all gender and minority groups are adequately represented; we use an analysis framework that allows for the assessment of differential impacts based on gender and additionally analyze the extent to which the program addresses gender-, disability-, and other subgroup-specific barriers and cultural constraints.

## DATA COLLECTION TOOLS

Tools were designed to replicate the baseline and midline surveys in order to ensure comparability of results. Minor additions or removals of questions were made at endline to ensure relevancy of the tools (by, for example, removing questions related to social distancing in schools) and in response to requests made by government partners. Additionally, minor modifications were made to the learning assessment, discussed more below.

The evaluation utilized the following tools:

- Quantitative tools
    - Learning assessment: Early Grade Reading Assessment (EGRA) and working memory test
    - Household survey
    - Farmers' group/VSLA survey
    - School survey
    - Classroom observation
- Qualitative tools
    - Focus group discussions (FGDs) with parents, teachers, and PTA members
    - Key informant interviews (KIIs) with government officials, school administrators, and health clinic representatives

Tools were translated into Tetum by Tetum-speaking CARE staff; additionally, the EGRA was originally developed in Tetum and back-translated into English for quality control purposes. All translations were verified by an independent translator.

Full tools are included in Annex 6.

## Learning Assessment

Learning assessments were used to measure grade 2 students' progress in acquiring foundational reading skills and to measure working memory. The learning assessment tool also included several questions related to food consumption, student attentiveness, and student height/weight.

The endline evaluation used an adapted version of the EGRA conducted at midline and baseline. The EGRA included five subtasks: (1) letter recognition, (2) invented word reading (i.e., recognition of letter groups and corresponding sounds), (3) familiar word reading, (4) short story reading, and (5) reading comprehension (basic and advanced). Tasks on letter recognition, invented word reading, and familiar word reading generally had a progressive increase in difficulty of letters/words; subsequent sections also generally had a progressive increase in level of difficulty. As such, students who were unable to read letters were not asked to perform subsequent subtasks, and students who were unable to read words (invented and familiar) were not asked to read a short story or answer comprehension questions. The below table shows the number of items and administration methods for each subtask.

**Table 2: EGRA sections and scoring**

| Subtask | Total Possible Score | Time Given | Notes |
|---|---|---|---|
| Letter recognition | 100 | 60 seconds | Students who could not read letters did not proceed |
| Invented word reading | 60 | 60 seconds | |
| Familiar word reading | 60 | 60 seconds | Students who could not read any words (invented or familiar) did not proceed |
| Passage reading | 61 | 60 seconds | |
| Reading comprehension | 10 | Untimed | Two groupings of five questions, related to two passages with increasing levels of complexity |

Given the multilingual context of Timor-Leste, learning assessments were conducted in the language of instruction (Tetum-Prasa) but with instructions given in the student's local language where applicable. At baseline, Tetum speakers were consulted to identify a list of words common to Timorese children regardless of their location of residence. This list of words was refined to exclude words with unfamiliar or difficult sounds or ambiguous meaning in other local languages, as identified by speakers of those languages. Enumerators were also instructed to accept any correct response to letter and word identification regardless of pronunciation, accent, or use of Portuguese/Tetum letter names, and were trained to recognize differences in pronunciation or recognition of letter names in order to ensure these instructions were applied during data collection. Lastly, to additionally enhance inclusivity, font size and type of the EGRA was selected to improve readability by students, considering the limited access to corrective lenses in Timor-Leste.

In addition to these five EGRA tasks, students were also administered a pictorial working memory test during the assessment to measure student attentiveness. For this test, students were presented with a set

of 19 images of common objects/animals, shown and read the name of each image individually by the enumerator, and asked to remember the image for later. The child was then asked to recall as many images as possible without looking at the images.

As some students in the sample may have been repeating a grade, they may have been exposed to the EGRA and working memory test administered at midline. As such, to minimize the risk of pre-exposure bias, the learning assessment and working memory tests were modified by, for example, rearranging the position of letters (e.g., b-c-a instead of a-b-c) or words. To ensure that differences in performance on the learning assessments are not due to differential difficulty of the assessments, following fieldwork, a subset of grade 2 students were administered both the midline and endline assessments, from which the similarity of the assessments could be calculated.

## Household Survey

The household survey was conducted with families of grade 2 students assessed with the EGRA. The household survey largely replicated the midline survey and included the following topical areas:

- Socioeconomic background of household heads
- Participation in VSLAs and economic empowerment
- Support for education
- Perceptions of school conditions, management, and safety, including school feeding program
- Quality of schoolteachers and education
- Student attendance
- Student disability
- Housework responsibilities of children
- Nutrition, health, and sanitation knowledge and practices
- Dietary diversity
- Gender-related attitudes and practices, including participation in decision-making, perceptions of child capacity, and perspectives on education for girls and boys
- Perspectives on gender-based violence (only administered to female caregivers if no other individuals were present)
- Use of agricultural practices
- Knowledge of and attitudes towards breastfeeding and child nutrition/health (for households with children under the age of 2 and babies under 6 months of age)

Within households, the head of household, grade 2 student's caregiver, and a mother of a child under the age of 2 were interviewed separately, as applicable.

## Farmers' Group/VSLA Survey

A component of HATUTAN programming included developing partnerships with farmers' groups to supply food to schools. As such, a subsample of household respondents participating in farmers' groups and VSLAs responded to an additional survey module assessing exposure to capacity building activities, use of savings and loans, income, perceived relevance of VSLA and farmers' group activities, farming practices, types of commodities produced, readiness for market engagement, and emerging linkages with buyers (including schools). Because relatively few respondents to the household survey were also engaged in farmers' groups or VSLAs, a booster sample of farmers outside of targeted households was also surveyed.

## School Survey

School surveys were conducted with school directors/coordinators in each sampled school. These surveys included school background, school feeding, teacher, and student attendance modules. The school school

background module collected information on school characteristics, school administrator practices and exposure to training, the availability and role of the PTA, book loaning, and safeguarding practices. The school feeding module included data on school infrastructure (kitchens, sanitation), implementation of the school feeding program, and procurement of locally produced items to complement school meals. The teacher module included data on the number and qualifications of teachers and teacher attendance. Lastly, the attendance module included data on enrolment, head counts of Grade 2 attendance, and recorded attendance rates.

## Classroom Observation

Classroom observations were conducted in grade 2 Tetum language classes. Teachers were randomly selected from among all teachers teaching grade 2. The classroom observation tool considered the following:

- Time spent on task
- Student engagement in class
- Student use of reading materials
- Teacher use of child-centered techniques and reading practices
- Gendered behaviors in class
- Safeguarding and corporal punishment
- Use of formative assessments

Researchers conducting classroom observations were provided with a list of teaching practices related to the above thematic areas. The use (or absence) of teaching practices was then measured as either “observed” or “not observed.”

## Focus Group Discussions

Focus group discussions were conducted with students’ parents (men and women separately) and teachers (mixed gender). FGDs with parents focused on understanding linkages between household characteristics and learning outcomes (including attendance and enrolment) and the activities conducted by the PTA. These FGDs help us understand changes (or lack thereof) over the program cycle in key areas including nutrition, food security, WASH, and education.

FGDs with teachers included content similar to that covered in parent FGDs, but emphasized school governance and other school-based factors affecting learning outcomes. FGDs with teachers also provided insight into program implementation, helping to identify successes and challenges that may inform future government or HATUTAN II programming.

Within each administrative post, FGDs were randomly assigned to schools so that multiple FGDs were not clustered in a single school. This design was intended to maximize representation of schools with varying characteristics and localities. Randomly assigning FGDs (rather than clustering all FGDs in a small number of schools) also controls for potential eligibility biases; for example, questions on PTA activities in the parents’ FGD are contingent on a school having a functioning PTA, which is itself indicative of school governance quality and a degree of parental support and engagement in education. If eligibility to be selected for those questions in qualitative interviews was limited to schools with functioning PTAs, all qualitative interviews could have been biased to represent the highest functioning school systems participating in the HATUTAN program.

### Key Informant Interviews

Key informant interviews included local government officials, school administrators, and health clinic representatives.[50] These interviews included questions on perceived management responsibilities, previous training and perceptions about training, PTA engagement and perceptions of its value, student and teacher attendance, school governance issues, the relationship between schools and the implementing organization(s), and gender dynamics within school systems.

## FIELDWORK

Fieldwork was managed by CARE with support for training and quality assurance provided by Consilient Research. In this section, we describe processes for enumerator selection, training, fieldwork, and data quality control.

### Enumerator Selection and Training

Enumerators were pre-selected by CARE from a pool of applicants based on previous data collection experience and skills in relevant local languages; around 30% of selected enumerators had participated in previous HATUTAN evaluations. The pre-selected enumerators were assessed with a test, and those who passed the test were then interviewed, following which enumerators were selected to participate in training.

Enumerators were trained for seven days by a joint CARE and Consilient team. Training topics included an introduction to the program and the evaluation, child protection, prevention of sexual harassment and abuse, research ethics and informed consent, confidentiality and data security, use of electronic data collection forms, working with children, review of quantitative and qualitative tools, and data quality control practices. Training also included mock practice sessions for all quantitative tools and a field pilot of the learning assessment, school survey, and household survey tools.

Enumerators were regularly assessed during training, and only those enumerators who achieved minimum standards on assessments and demonstrated proficiency in tool administration during the field pilot were selected. The highest-performing enumerators, as measured through assessments and performance in the field pilot, were selected as team leaders and provided with brief additional training on qualitative tools and team management.

### Fieldwork Overview and Quality Control

Data collection began on February 6, 2023 and ended on March 31, 2023. Teams were distributed to municipalities based on language skills and spent an average of two days at each school collecting data with students, teachers, school directors/coordinators, and households. To guide the fieldwork, teams were provided with tracking tools, including individual and school tracking sheets containing the identifier and demographic information for target respondents.

During the first week of data collection, teams were accompanied by Consilient and CARE staff who monitored the progress of data collection and adherence to fieldwork procedures, including proper tool administration. This also allowed team leaders to quickly clarify any procedural questions. We note that while CARE staff visited the data collection teams to ensure that they were present in the correct schools and following procedures appropriately, as well as to answer any questions from the teams, CARE's program monitoring officers did not directly observe data collection with respondents to avoid introducing bias to answers.

[50] Health clinic representatives are not linked to a specific school area, but rather to a broader administrative post due to the limited number of facilities in rural areas.

At the end of each day of data collection, completed surveys were uploaded to the secure online data management platform Ona (where connectivity allowed). The submitted data was then downloaded regularly by the Consilient and CARE teams for quality control and data cleaning. A quality control tool was developed in Stata and used regularly to track the number of submitted surveys, calculate results by school and enumerator, and flag potential issues within the quantitative data. These potential issues included implausible EGRA results, possible cases of EGRA misadministration, outlying values, contradictory attendance and enrolment records, and illogical or incoherent text- or numeric-based responses, among others. All inconsistencies and mistakes were discussed with teams in the field and, if necessary, corrected in the data. To further ensure data quality, several quality checks were scripted into the tools to reduce data entry-related errors; these included measures such as age restrictions, constraints for numeric values, and calculations for learning assessment scores.

## PROPOSED AND ACHIEVED SAMPLES

In this section, we describe the sample achieved for the endline evaluation for each tool described above. In preparation for the endline evaluation, CARE developed a sampling strategy across HATUTAN intervention municipalities (Ainaro, Ermera, Liquica, and Manatuto) and comparison municipalities (Aileu, Baucau,[51] Bobonaro, Covalima, and Manufahi). A sample of 98 intervention schools was drawn from a list of school communities receiving a "full package" of HATUTAN interventions; 88 comparison schools were also selected in comparable communities. School surveys and classroom observations were intended to be conducted in all 186 schools. From these 186 schools, the endline evaluation sought to sample 1,014 girls and 1,056 boys from grade 2 classes (around 20 students from each school).[52] The household survey sample was randomly drawn from among parents of sampled grade 2 students, with a target of 8 households per school. Lastly, the study also sought to interview a booster sample of farmers' group representatives.

The table below shows the breakdown of samples for the school survey and classroom observation by intervention status and municipality. In total, 98 intervention schools and 87 comparison schools were sampled; among comparison municipalities, four fewer schools are visited in the final sample than proposed.[53] Classroom observations were conducted in all comparison schools and 94 intervention schools; classes could not be observed in some intervention schools due to teacher absences.

**Table 3: School survey and classroom observation sample**

| | School survey | Classroom observation | Proposed sample |
|---|---|---|---|
| **Municipality (intervention)** | | | |
| Ainaro | 27 | 25 | 27 |
| Ermera | 41 | 39 | 41 |
| Liquica | 11 | 11 | 19 |
| Manatuto | 19 | 19 | 11 |
| Total (intervention) | 98 | 94 | 98 |

[51] Schools in Baucau were dropped from the sample for analysis due to low levels of comparability with intervention schools, as these schools had much higher learning assessment performance than other schools (intervention and comparison) in the sample. As such, they are not included below.
[52] This sample size was calculated considering a 95% confidence level, 80% power, an effect size of 0.2, a design effect of 2, a 15% attrition rate and disaggregation by gender. The sample was calculated as $n = 2*[(Z(1- \alpha/2)+Z(1-\beta))/ES]^2$.
[53] Three comparison schools in Baucau were dropped as they were deemed to be poorly matched with intervention schools.

| | | | |
|---|---|---|---|
| **Municipality (comparison)** | | | |
| Aileu | 22 | 22 | - |
| Bobonaro | 31 | 31 | - |
| Covalima | 10 | 10 | - |
| Manufahi | 21 | 21 | - |
| Total (comparison) | 84 | 84 | 88 |
| **Grand total** | **182** | **178** | **186** |

Table 4 shows the sample for learning assessments and household surveys. The sample of learning assessments conducted for the endline exceeded the target number. More boys than girls were sampled within intervention municipalities.

**Table 4: Learning assessment and household survey sample**

| | Learning assessment – girls | Learning assessment – boys | Household survey |
|---|---|---|---|
| **Municipality (intervention)** | | | |
| Ainaro | 174 | 206 | 203 |
| Ermera | 290 | 329 | 316 |
| Liquica | 87 | 103 | 86 |
| Manatuto | 139 | 140 | 136 |
| Total (intervention) | 690 | 778 | 741 |
| **Municipality (comparison)** | | | |
| Aileu | 120 | 106 | 151 |
| Bobonaro | 213 | 191 | 230 |
| Covalima | 64 | 79 | 74 |
| Manufahi | 126 | 142 | 147 |
| Total (comparison) | 523 | 518 | 602 |
| **Grand total** | **1,213** | **1,296** | **1,343** |

In addition, 286 farmers' group surveys were conducted in intervention municipalities: 98 in Ainaro, 66 in Ermera, 76 in Liquica, and 46 in Manatuto. This booster sample supplemented sample size for questions regarding VSLA participation and agricultural practices also asked in the household survey.

Further information about the demographics of the achieved sample is contained in Annex 3.

## CHALLENGES AND LIMITATIONS

The evaluation methodology is associated with a number of limitations described in this section. These limitations do not imply that the methodology of the evaluation is systematically flawed or invalid; discussion is rather intended to note potential threats to inference.

**Use of cross-sectional sample:** This endline evaluation uses a cross-sectional sample of grade 2 students assessed at endline, but not in previous rounds. Use of this sample has some key advantages: For example, because grade 2 students were also assessed at baseline and midline, we expect students' assessment results and experience schooling to be broadly similar at endline since they are of similar ages and have had similar exposure to education. However, because the cross-sectional sample is comprised of an entirely new group of students, 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 endline in intervention schools was, coincidentally, more motivated or had greater aptitude for reading than students observed at baseline, the endline students would perform better than expected on the EGRA. 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 could result in positive or negative estimates of impact that are driven by the unusual nature of the sample, rather than actual impact.

Analysis of the cross-sectional sample may also be affected if the HATUTAN program had 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 enrolment 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.

These dynamics are explored more in Annex 3.

**Difference-in-differences assumptions and limitations:** Inferences drawn via difference-in-differences analysis rely on two key assumptions. The first assumption is that in the absence of an intervention (i.e., under the counterfactual condition) the change in outcomes over time in treatment and comparison schools will be similar. This assumption is often referred to as the "parallel trends assumption" because it assumes that, in the absence of intervention, trends in outcomes in the intervention group would evolve in parallel to those in the comparison group. This assumption is generally untestable under a difference-in-differences framework. In the absence of randomization, the parallel trends assumption may or may not hold, though the selection process leaves no reason to expect divergent trends.

The second assumption is that comparison group respondents are not exposed to the intervention or to similar interventions affecting learning, nutrition, and other outcomes of interest. Informally, this assumption is often stated as the "no spillover" assumption. The second assumption is testable following data collection by directly measuring whether comparison group respondents received any of the program's interventions.

If both assumptions are satisfied, inferences drawn using difference-in-differences are expected to be unbiased. However, we note that, in the context of this endline evaluation, there is expected to be some contamination among comparison schools from other program (i.e., not HATUTAN) interventions. This may cause us to underestimate the impact of the HATUTAN program.

**Differences between intervention and comparison schools:** While comparison schools were carefully selected to come from similar areas and levels of "remoteness" and to have a similar linguistic background as intervention schools, intervention and comparison schools are not perfectly balanced, a factor explored more in Annex 3. Given that we observe some demographic differences across comparison and intervention schools/groups, it is likely that these samples are also imbalanced in terms of other potentially important, but unobserved, factors that may bias analysis.

The main implication of this limitation is that we cannot be sure that observed differences across intervention and comparison groups are not partially a product of unobserved, systematic differences across these 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 cannot be certain that we have accounted for all potential confounders, and thus cannot claim that our estimates are completely unbiased.

**Timing of data collection:** Due to the school calendar in Timor-Leste and the program's implementation timeline, students were assessed at the beginning of the school year. As such, assessed grade 2 students have skills comparable with students starting that grade; however, the standard definition for assessment requires assessments to be conducted with students at the end of grade 2. As such, results in this study may be biased downwards compared to the results of other learning assessments conducted using the "standard" process (although this limitation has affected all rounds and intervention and comparison students equally, thus not affecting the validity of the difference-in-differences approach). We note that grade 3 students were not assessed as there was a high risk of biased results due to high dropout rates at the end of grade 2, generally among the lowest-performing students.

**Inaccuracy of school record-keeping:** Collecting student and teacher attendance data from school records is generally challenging, as records are often low quality and may be partially or entirely incomplete. This evaluation triangulates attendance across multiple sources, including through headcounts of students and teachers, to provide a more holistic picture of attendance rates, rather than relying on school records.

**Accessing schools and respondents:** Families in Timor-Leste frequently travel for extended periods of time in order to attend traditional ceremonies, which makes it difficult to contact some respondents for household surveys or to administer the learning assessment. Additionally, fieldwork 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 tools 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 students' ages were unusually high or low. This reduces the sample size for some variables.

**Floor and ceiling effects:** Within the EGRA and 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 difficult for most students, scores will be clustered around 0% with little variation. We observe floor effects in several EGRA subtasks, discussed more in the section *Literacy Outcomes*. These effects dampen our ability to differentiate between intervention and comparison school, thus reducing our ability to draw conclusions about the potential impact of the program in intervention schools.

**Generalizability of results:** Sample sizes for several municipalities and languages are small. While we still include analysis of results disaggregated by these (and other) subgroups, results should be taken as indicative, not definitive, due to sample size limitations. Results cannot be generalized at national level, as they are representative only of program areas and comparison areas, primarily rural and remote.

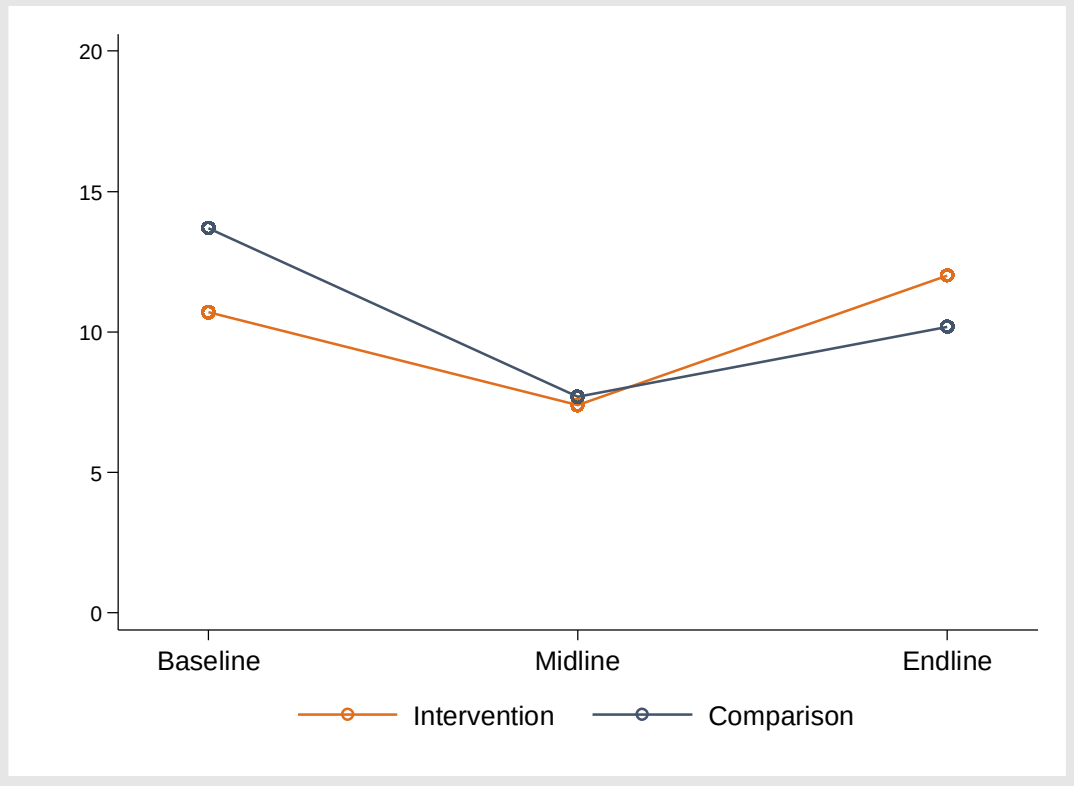
# LITERACY OUTCOMES

In this section, we discuss literacy outcomes for grade 2 students assessed using EGRA. We first discuss overall literacy scores, then analyze results by each of the five subtasks—letter recognition, invented word reading, familiar word reading, passage reading, and reading comprehension. We focus primarily on changes in outcomes from baseline to endline, using midline data to supplement this analysis and provide a broader picture of trends over time and the impact of COVID-19 on learning. We also conduct a predictive analysis to better understand factors associated with improved literacy outcomes.

## OVERALL LITERACY SCORES

We first analyze overall literacy scores, calculated as the simple average of scores on each of the five literacy subtasks. The below figure shows literacy scores for grade 2 students across all rounds. This figure first shows that literacy scores declined from baseline to midline due to the impact of COVID-19 on learning. At endline, overall literacy scores have increased for both intervention and comparison students, but have increased more substantially for intervention students: Over the past two years since midline, intervention students' average literacy scores increased by 4.6 percentage points, while comparison students' average scores increased by only 2.5 percentage points.

**Figure 1: Overall literacy scores from all rounds, grade 2 students**



This pattern means that intervention students' average literacy scores have increased slightly from baseline to endline, by 1.3 percentage points, while comparison students' scores have fallen since baseline by 3.5 percentage points. It also means that intervention students' literacy scores have now surpassed comparison students' scores, despite falling substantially behind comparison students at baseline and slightly behind at midline. In other words, these patterns suggest that **the HATUTAN program may have both mitigated the negative impacts of COVID-19 on learning, as shown at midline, and allowed students to learn more quickly after returning to school post-COVID.**

To further explore these dynamics, we now run a difference-in-differences analysis to measure the significance of relative improvements in intervention students. The below table shows results from baseline to endline. We find a significant improvement in overall literacy scores for intervention students compared to comparison students within all regression models, including the most robust model controlling for a number of differences across intervention and comparison students and schools.[54] Across all models, we find that intervention students' overall literacy scores improved by around 5 percentage points more from baseline to endline than expected given the results of comparison students. In other words, **students supported by the HATUTAN program have had significantly greater improvement in literacy scores since baseline than students who did not receive program interventions.**

**Table 5: Analysis of changes in overall literacy scores from baseline to endline**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| Regression | BL | EL | BL | EL | DiD | P-value |
| No controls | 10.7 | 12.0 | 13.7 | 10.2 | 4.8 | 0.005* |
| Student-level controls[55] | - | - | - | - | 5.2 | 0.003* |
| Student- and school-level controls[56] | - | - | - | - | 4.4 | 0.006* |
| Student-, school-, and household-level controls[57] | - | - | - | - | 4.9 | 0.03* |

Looking now at results from midline to endline, the below table shows that, as discussed above, intervention students' overall literacy scores increased by around 2.1 percentage points more from midline to endline than expected given the results of comparison students. This relative improvement, however, is significant only for one of the difference-in-differences models, that controlling for student-level differences in age, gender, and native language.

[54] We note that student-, school-, and household-level controls are the same across all tables throughout this section, and are thus defined only within the footnotes for Table 5.
[55] Including student gender, age, and whether the student speaks Tetum-Prasa natively.
[56] Including student gender, age, whether the student speaks Tetum-Prasa natively, and school fixed effects.
[57] Including student gender, age, whether the student speaks Tetum-Prasa natively, school fixed effects, whether the caregiver reported that the student has difficulty with memory, whether the caregiver speaks Tetum-Prasa natively, whether the HoH speaks Tetum-Prasa natively, and the size of the household.

**Table 6: Analysis of changes in overall literacy scores from midline to endline**

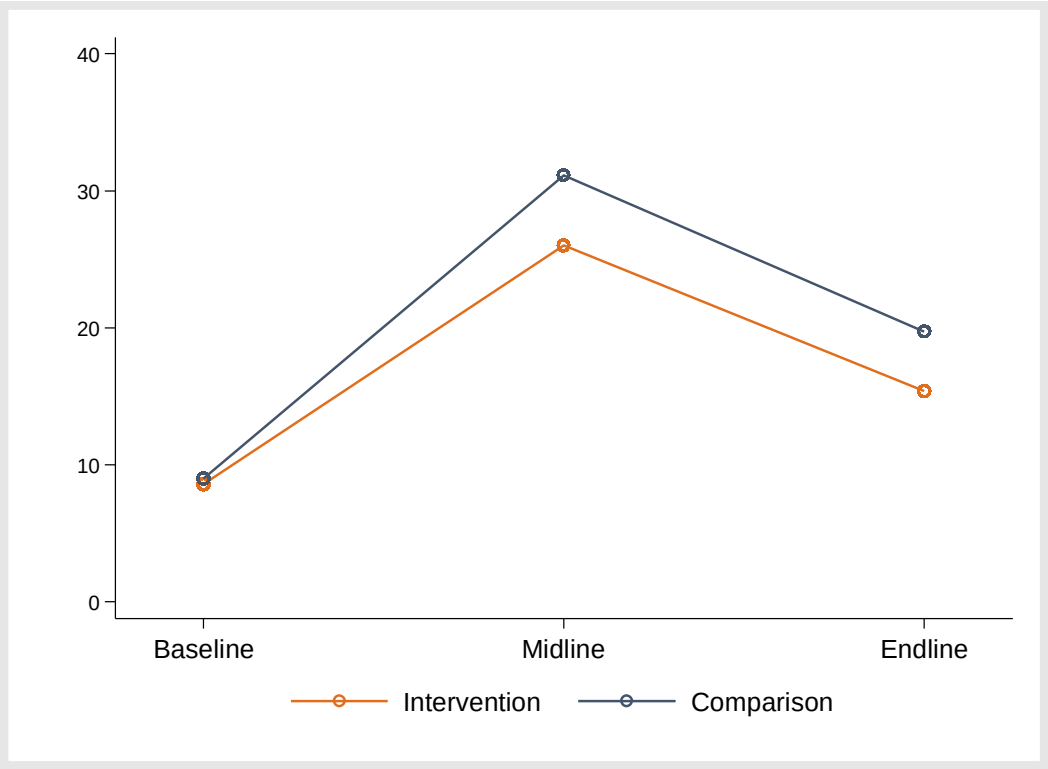
| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| Regression | ML | EL | ML | EL | DiD | P-value |
| No controls | 7.4 | 12.0 | 7.7 | 10.2 | 2.1 | 0.06 |
| Student-level controls | - | - | - | - | 2.3 | 0.04* |
| Student- and school-level controls | - | - | - | - | 2.1 | 0.05 |
| Student-, school-, and household-level controls | - | - | - | - | 1.9 | 0.19 |

Despite the lack of significance in most regression models in midline-to-endline comparisons, the above results remain indicative of potential program impact. Rather than reducing the validity of above findings, these results rather suggest that school feeding program functionality may have played an important role in strengthening literacy. At midline, the SFP was fully functional; at endline, in contrast, only a fraction of schools were providing meals to students, as will be discussed further in the *School Feeding Program* section. As such, results from midline to endline reflect the program's impact on literacy despite a reduction in school feeding (which tends to reduce student attentiveness). Results from baseline to endline, in contrast, show HATUTAN's impact on literacy with little change in school meal provision. In other words, these results suggest that **school meal provision has an important impact on literacy results, but HATUTAN may also have supported improvements in student literacy outside of its impact on the SFP.** Additionally, these results may also suggest that the HATUTAN program may have been more effective at mitigating the negative impacts of COVID-19 on grade 2 students than in supporting improvements in literacy upon return to school – although we again note that these results still suggest a potential improvement in intervention students' literacy abilities relative to comparison students post-COVID.

Overall, these findings have important implications for HATUTAN II. They suggest, first and foremost, that HATUTAN has indeed had a positive impact on literacy learning among grade 2 students. However, the program appears to have had slightly more impact on supporting grade 2 students during COVID than after their return to school, potentially due to the decrease in SFP functionality at endline. As such, lessons learned from program implementation during COVID, such as the need to pivot activities to respond to environmental conditions, should be carefully considered when designing HATUTAN II activities.

In addition to analyzing results for all students assessed with the EGRA, we now analyze changes in the percent of students scoring zero overall, as well as changes in scores for students who scored greater than 0% overall. The below figure shows that a substantial portion of both comparison and intervention students scored 0% in all rounds of data collection, with the number of zero scorers drastically increasing at midline due to COVID and falling, though not to baseline levels, at endline. The percent of zero scorers increased at a lower rate for intervention schools than comparison schools from baseline to endline, and decreased at a higher rate from midline to endline; however, these differences were not significant. In other words, at endline, **the HATUTAN program does not appear to have had a statistically significant effect on reducing the percent of students with no literacy abilities.**

**Figure 2: Percent of students scoring 0% overall on EGRA**

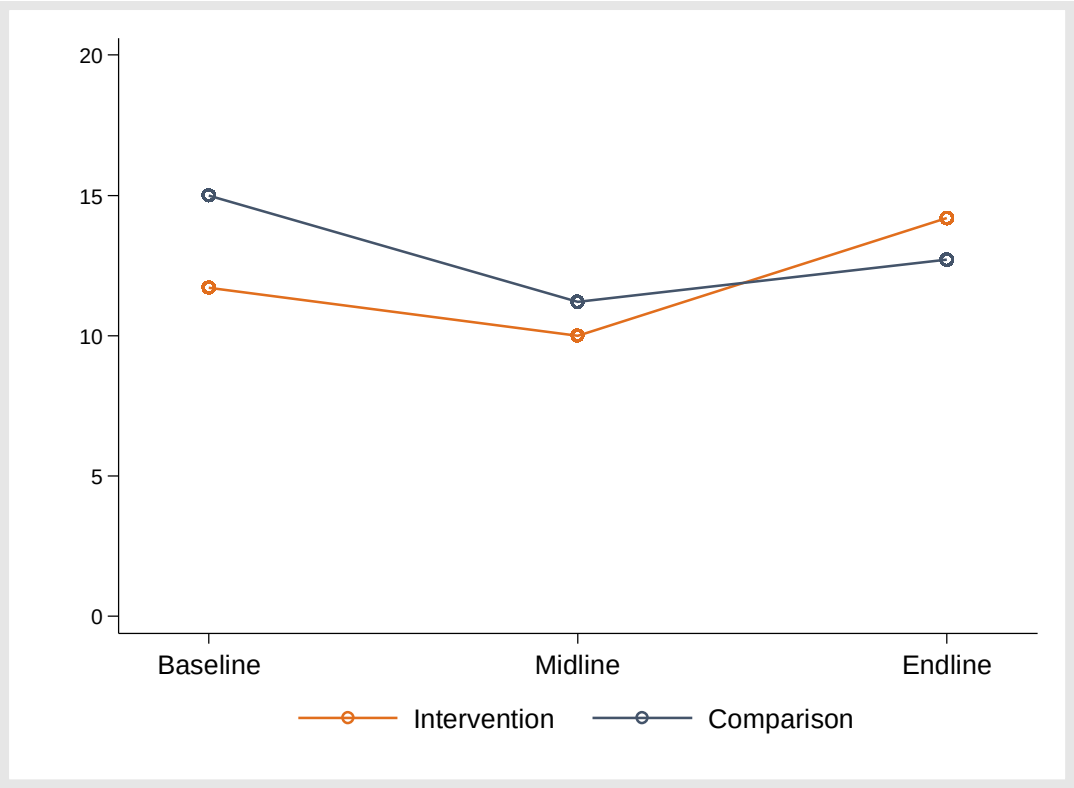



We now examine results for students who scored greater than zero percent. For this group, we find that intervention students' scores were consistently better than expected given the results of comparison students: Literacy scores decreased by less for intervention students than for comparison students between baseline and midline, and increased by more for intervention students between midline and endline. As a result, at endline, intervention non-zero scorers' average scores had surpassed those of comparison students.

Indeed, at endline, intervention non-zero scorers' average scores had improved by 2.5 percentage points, while comparison non-zero scorers' scores had declined by 2.4 percentage points. This relative improvement of 4.9 percentage points among intervention students is significant,[58] and suggests that while HATUTAN may not have had a significant effect on students with no literacy abilities, **students in HATUTAN program schools with at least a basic ability to recognize letters have seen a significant improvement in their overall literacy abilities relative to students unaffected by the program.** This finding applies to both male and female students.

[58] The result remains significant in all specifications including control variables.

**Figure 3: Overall literacy scores from all rounds, grade 2 students scoring greater than 0%**




As with above results, we note that this relative improvement was not significant from midline to endline, again suggesting the important role that the SFP may have played in strengthening student literacy by improving attentiveness. In addition, we note a similar result from midline, whereby the percent of zero scorers increased from baseline to midline, but the mean score of students who scored above 0% improved from baseline to midline in intervention schools. Although the percent of zero scorers declined at endline, it declined fairly uniformly for both comparison and intervention students. This finding, in combination with midline findings, suggests a dichotomous effect of the HATUTAN program on learning: The program appears to be more effective in strengthening learning for students who were at least able to recognize letters, but has less impact on students who have not yet obtained these skills.

**Box 1: Love to Read Groups**

Between midline and endline, HATUTAN launched a new initiative for community-based reading groups led by youth volunteers and parents called *Hadomi Lee* (Love to Read). These groups aimed to provide support to teachers, help children who are struggling with reading via educational after school activities, and encourage parents to read with their children. The groups are also intended to provide support to rural children in the event of future school closures.

At endline, 86.7% of intervention schools serving over 15,000 children reported that they had a Love to Read group. Of these schools, 84.7%, or 72 schools in total, reported that the Love to Read group was active, and 54.1%, or 46 schools in total, reported that the group had met within the last month.

Qualitative data suggests that Love to Read groups have successfully supported literacy learning within communities. A school administrator from Ermera, for example, stated the following:

> *In my opinion, the Love to Read group has so many advantages. Children learn to read at school, and they also continue their learning at home based on their schedule organized with the Love to Read group. This means that they are very unlikely to forget what they have learned at school. They also become aware that whatever they learn at school can also be learned at home or at community level.*
>
> - KII with administrator, Ermera, Int. 4

A teacher in Liquica also spoke positively about the Love to Read group, stating that it was a useful addition to the community given that they did not have a local kindergarten:

> *The Love to Read Group may help children learn at home, and when they go to school they will already grasp a little bit. We don't have a kindergarten here, thus we accepted children immediately to learn. The Love to Read Group assists a little bit to develop children who know how to read a little bit.*
>
> - FGD with teachers, Liquica, Int. 42

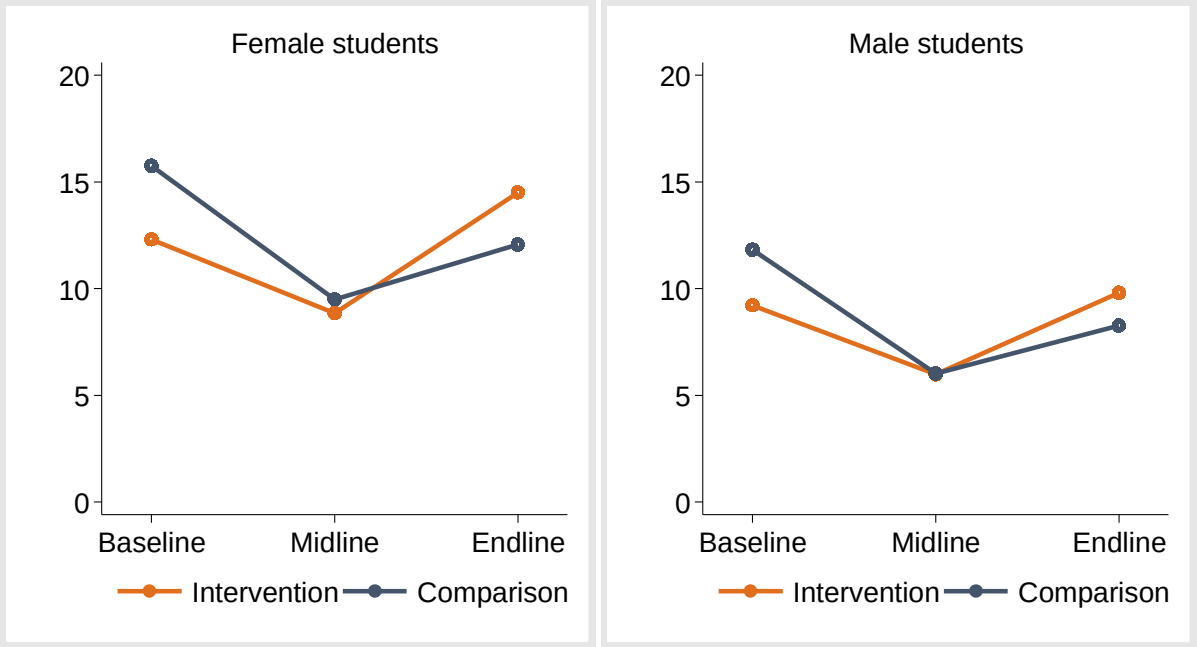
However, one teacher in Ainaro noted that the groups had limited reach because children would not participate if their parents were not also participating. This teacher further described the group's positive impact but reliance on parental involvement as follows:

> *The Love to Read group contributed to the school and assisted the teachers with the aid of parents. If the children are hesitant or fearful of the teacher, the parents assist them in teaching and they learn easily. This school established the group, and they do their activities one or two days per week, but the parents recently stopped doing the work, and we have no idea why.*
>
> - FGD with teachers, Ainaro, Int. 37

Overall, these findings suggest that Love to Read groups have been largely successful in achieving their goals, and should be continued in HATUTAN II or supported through other interventions.

We now disaggregate overall literacy results by student gender to uncover any potentially heterogeneous program impacts on male and female students. The below figures show changes in female and male grade 2 students' overall literacy scores across all evaluation rounds. We first find that male students' scores have consistently fallen behind female students' scores, in line with broader learning patterns among primary-aged students in Timor-Leste. Furthermore, among intervention schools, female students' average literacy scores increased by 2.2 percentage points from baseline to endline and 5.7 percentage points from midline to endline. Male students' average scores, in contrast, increased by only 0.6 percentage points from baseline to endline and 3.8 percentage points from midline to endline.

**Figure 4: Overall literacy scores from all rounds by gender, grade 2 students**




The below table shows difference-in-differences results for female and male students.[59] We find that, from baseline to endline, both female and male students' average literacy scores increased at a greater rate among intervention schools than for comparison schools. However, female students' average scores demonstrated a greater relative improvement than male students'. Furthermore, intervention female students' scores had a relatively greater improvement from midline to endline than male students. Overall, these findings suggest that **HATUTAN had a greater impact at improving literacy abilities among female students than male students.**

**Table 7: Change in overall literacy scores by gender, grade 2 students**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| **Baseline-Endline** | **BL** | **EL** | **BL** | **EL** | **DiD** | **P-value** |
| Female students | 12.3 | 14.5 | 15.8 | 12.1 | 5.9 | 0.007* |
| Male students | 9.2 | 9.8 | 11.8 | 8.3 | 4.1 | 0.02* |
| **Midline-Endline** | **ML** | **EL** | **ML** | **EL** | **DiD** | **P-value** |
| Female students | 8.9 | 14.5 | 9.5 | 12.1 | 3.1 | 0.05 |
| Male students | 6.0 | 9.8 | 6.0 | 8.3 | 1.6 | 0.18 |

With these results in mind, it is important to again note the gender context in Timorese education, whereby male students tend to perform worse than female students at low grades, then gradually overtake female students as they age. In this context, greater program impact on female students actually represents a widening of the gender learning gap among primary-aged students. As such, it may be advisable for HATUTAN II to explore ways to better engage male, as well as female, primary school students.

[59] For the sake of brevity, regression specifications with control variables are not included below.

Lastly, it is important to note that while we have found promising signs of HATUTAN program impact, **overall literacy scores remain very low for both girls and boys across all municipalities.** The average score for intervention students was just 12% at endline, out of a possible 100% score. In absolute terms, this represents very weak literacy abilities. This challenge was also emphasized in the qualitative data; teachers, parents, and school administrators from all municipalities stated that students had difficultly learning, particularly students who were shy,[60] who did not speak Tetum-Prasa natively,[61] or who did not have parental support to learn at home.[62] These dynamics are explored in greater detail in the predictive analysis below.

To further explore this, we now analyze results by EGRA subtask.

# LITERACY SUBTASK SCORES

## Letter Recognition

For the first EGRA subtask, students were given a paper with 100 letters and were asked to read those letters aloud in order. Students were given one point for each letter that was read correctly within 60 seconds.

At both baseline and midline, analysis suggested that students knew the names of letters fairly well, and thus rarely identified letters incorrectly. Low average scores were rather due to low levels of fluency, as students were not able to name very many letters within one minute. Endline results generally conform to these findings; we find that on average, endline intervention students named only 13.7% of letters incorrectly, a decrease since midline. Furthermore, 30% of intervention students at endline did not name any letters incorrectly. However, on average, intervention students only attempted to read around 29 letters before the maximum time had elapsed. While this suggests a need to further strengthen reading fluency, we do note that at midline, intervention students only attempted to read 21 letters on average before time elapsed; as such, intervention students have seen increases in letter reading speed since midline.

The below figure shows changes in letter recognition scores across all evaluation rounds and treatment groups. As with overall literacy, we find that from baseline to midline, scores for intervention grade 2 students fell less substantially than those for comparison students, and from midline to endline, average scores for intervention students rose by more than those for comparison students. As a result, intervention students' average letter recognition scores have now surpassed those of comparison students by 3.1 percentage points (with an average score of 24.1% for intervention students and 21.0% for comparison students).

[60] E.g., FGD with fathers, Ainaro, Int. 23; FGD with fathers, Ainaro, Int. 14

[61] E.g., FGD with fathers, Ermera, Int. 15; FGD with fathers, Liquica, Int. 18; FGD with mothers, Ainaro, Int. 25; FGD with mothers, Emera, Int. 27

[62] E.g., KII with administrator, Manatuto, Int. 7; FGD with fathers, Manatuto, Int. 20

**Figure 5: Letter recognition scores from all rounds, grade 2 students**

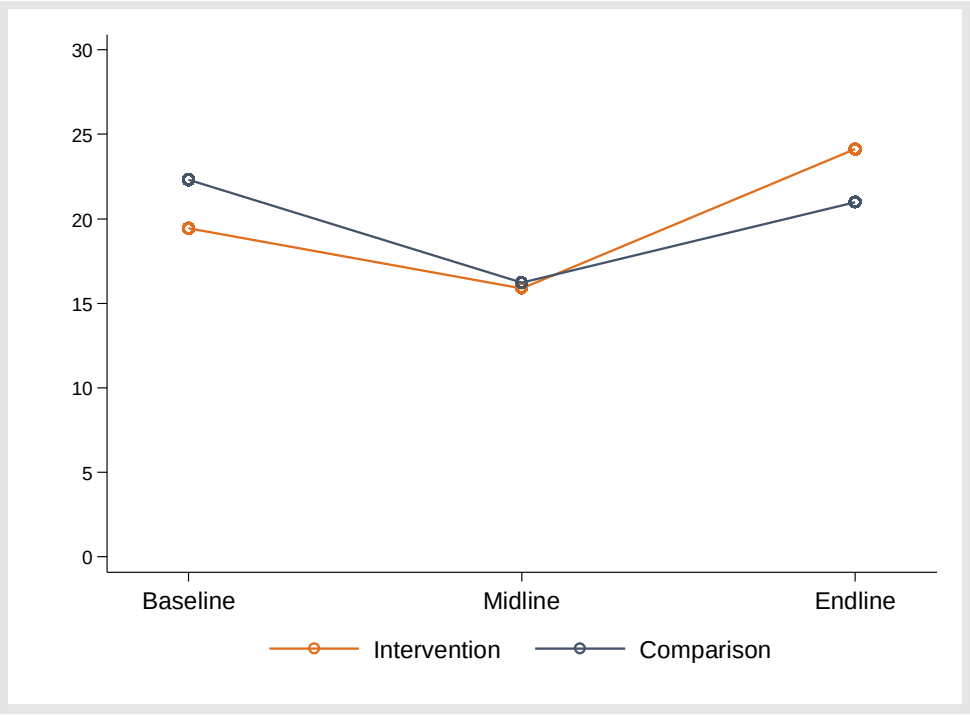



Table 8 further presents results of the difference-in-differences analysis for all students from baseline to endline and from midline to endline. As shown in the above figure, we find substantial relative improvements for intervention students compared to comparison students at endline. From baseline to endline, intervention students' average letter recognition scores improved by 6.2 percentage points more than would be expected given the results of comparison students, from 19.5% to 24.1% (while comparison students' average scores declined slightly). This significant relative improvement remained significant when accounting for potential differences across intervention and comparison students and schools.

From midline to endline, intervention students still saw a significant relative improvement in letter recognition scores; however, this improvement was less substantial than the baseline-endline improvement. Overall, these results suggest first and foremost that **the HATUTAN program had a positive effect on letter recognition abilities among intervention students.** However, as with overall scores, this effect appears to have been slightly stronger when considering longer-term change since baseline, rather than change since schools have reopened post-COVID.

**Table 8: Analysis of changes in letter recognition scores**

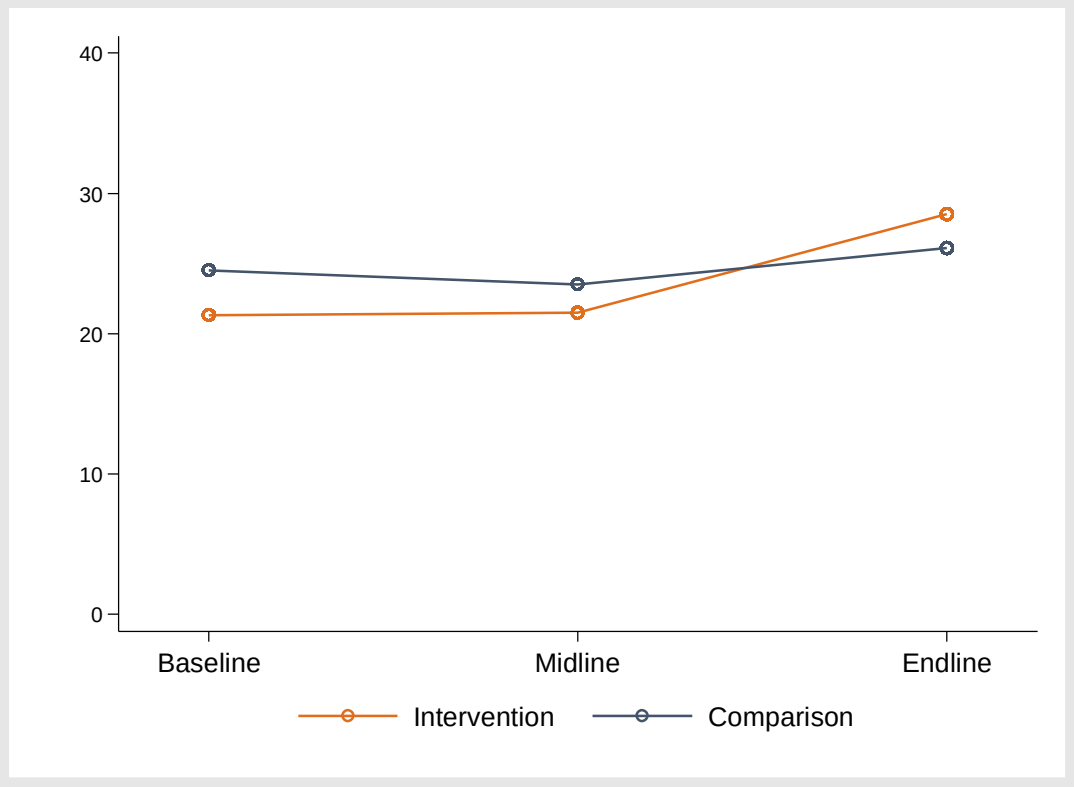
| | Baseline-Endline | | Midline-Endline | |
|---|---|---|---|---|
| | DiD | P-value | DiD | P-value |
| No controls | 6.0 | 0.004* | 3.5 | 0.02* |
| Student-level controls | 6.3 | 0.002* | 3.6 | 0.02* |
| Student- and school-level controls | 5.4 | 0.006* | 3.4 | 0.03* |
| Student-, school-, and household-level controls | 6.3 | 0.01* | 3.9 | 0.04* |

This progress is particularly important because letter recognition provides a foundation for more advanced literacy skills. In the qualitative data, for example, one mother stated that many students had difficulty reading because they had "not really pick[ed] up the alphabet/ABCs."[63] However, other respondents also emphasized that while letter recognition provides a foundation for literacy, it does not necessarily translate into word recognition; for example, a mother in Ermera stated that "children can learn alphabets easily, but they often have difficulties forming sentences."[64] This statement will be further evidenced through the results for subsequent subtasks.

Lastly, we analyze letter recognition results for non-zero scorers.[65] Figure 6 shows the change in scores for this group of students across round and treatment group. We find relatively little change in overall letter recognition scores among non-zero scorers from baseline to midline, suggesting that post-COVID shocks—such as economic and food security shocks, as well as increased teacher absences—may have had a substantial impact in preventing students from learning any literacy skills at all (thus resulting in a higher percent of zero scorers), but had relatively less impact on students who were already able to recognize letters.

At endline, then, we find that the average score of non-zero scorers increased in both comparison and intervention schools, but increased more for intervention students. For these students, average letter recognition scores increased by 7.2 percentage points, from 21.3% to 28.5%. For comparison students, in contrast, scores increased by only 1.6 percentage points, from 24.5% to 26.1%. This relative improvement in intervention students from midline to endline was significant.

**Figure 6: Letter recognition scores from all rounds, grade 2 students scoring greater than 0%**




[63] FGD with mothers, Manatuto, Int. 32

[64] FGD with mothers, Ermera, Int. 28

[65] Results for zero scorers on letter recognition are the same as results for zero scorers on the EGRA overall, as students who could not read letters were not asked to complete subsequent subtasks and thus scored 0 overall.

Overall, these results suggest, as above, that **the HATUTAN program had a greater impact among students who were able to recognize at least some letters.** For these students, at endline, letter recognition skills improved at a greater rate than for students unaffected by the program. In contrast, as seen in Figure 2, the HATUTAN program appears to have had little effect on enabling students to learn to recognize at least some letters if they were not already able to read letters.

## Invented Word Reading

For the second EGRA subtask, students were asked to read invented words to assess their abilities to recognize and read the sounds made by letters and letter groupings (i.e., phonics). Students were given a paper with 60 invented words and were asked to read the words aloud in order; they were marked as reading a word correctly if their pronunciation of the word demonstrated recognition of the word, even if it was not pronounced "correctly," to account for differing accents and pronunciations across native languages. Students were not asked to attempt this task if they could not read any letters; as such, only 82% of students (820 comparison students and 1,233 intervention students) attempted this task at endline, and the remaining 18% were given a score of 0.

The below figure shows first that average invented word reading scores were very low across all rounds, never exceeding 6 words per minute (WPM). The HATUTAN program, however, seems to have had some success in both reducing the learning loss in invented word reading due to COVID, and in boosting invented word reading scores for intervention students after students returned to school. As a result, while intervention students' average invented word reading scores fell 2.5 words per minute below comparison students' scores at baseline, their scores now exceed those of comparison students by 1.8 words per minute. This further means that intervention students' average invented word reading scores have increased since baseline, while comparison students' scores have decreased.

**Figure 7: Invented word reading scores from all rounds, grade 2 students**

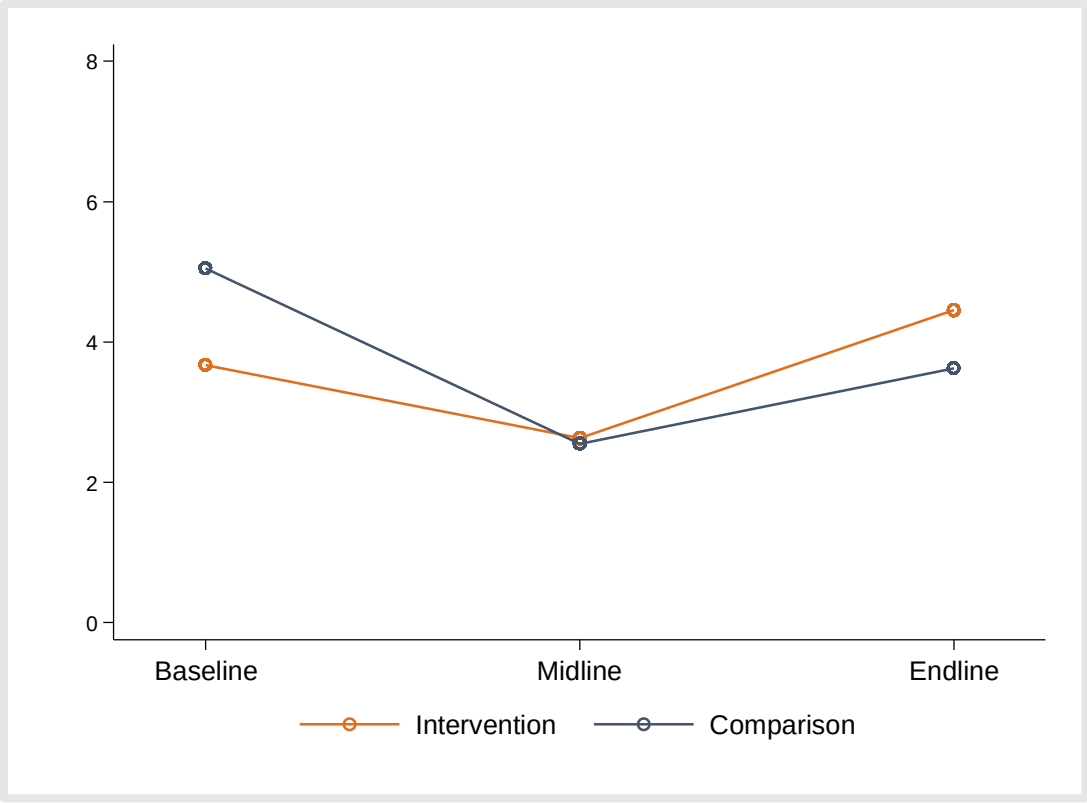


Table 9 shows the results of the difference-in-differences analysis on these results. We find a significant improvement in scores for intervention students relative to comparison students between baseline and endline, which is robust to the inclusion of control variables. In contrast, as seen in previous sections, we find a slight but insignificant relative improvement from midline to endline. This again suggests that **the HATUTAN program had a positive effect on invented word reading skills**, but that this effect was less substantial in the post-COVID period.

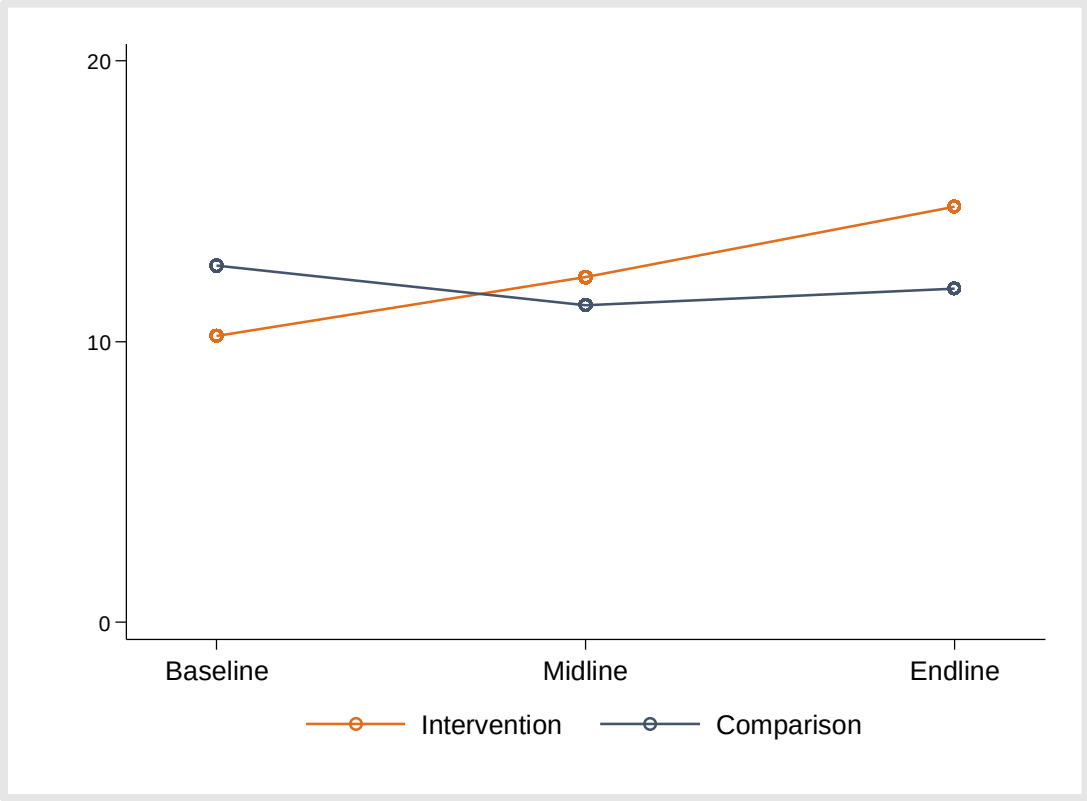
**Table 9: Analysis of changes in invented word reading scores**

| | Baseline-Endline | | Midline-Endline | |
|---|---|---|---|---|
| | DiD | P-value | DiD | P-value |
| No controls | 2.2 | 0.006* | 0.7 | 0.20 |
| Student-level controls | 2.4 | 0.005* | 0.8 | 0.17 |
| Student- and school-level controls | 2.0 | 0.009* | 0.7 | 0.20 |
| Student-, school-, and household-level controls | 2.3 | 0.03* | 0.5 | 0.48 |

Lastly, we analyze the percent of zero scorers as well as results for students reading zero WPM. We first find that a very high percentage of students read zero WPM on this task across all rounds, with relatively few differences across intervention and comparison groups. Indeed, at baseline, 60.3% of comparison students and 64% of intervention students read zero WPM; at midline, 77.6% of comparison students and 78.7% of intervention students read zero WPM; and at endline, 69.6% of comparison students and 70.0% of intervention students read zero WPM. While this suggests that the gap between intervention and comparison students has lessened slightly over the HATUTAN implementation period, this finding is not significant.

Figure 8, however, shows that invented word scores among students reading more than zero words per minute improved significantly more for intervention students than for comparison students: From baseline to endline, invented word scores improved by 4.6 words per minute for intervention students while declining by 0.8 words per minute among comparison students.

**Figure 8: Invented word scores from all rounds, grade 2 students reading more than zero WPM**



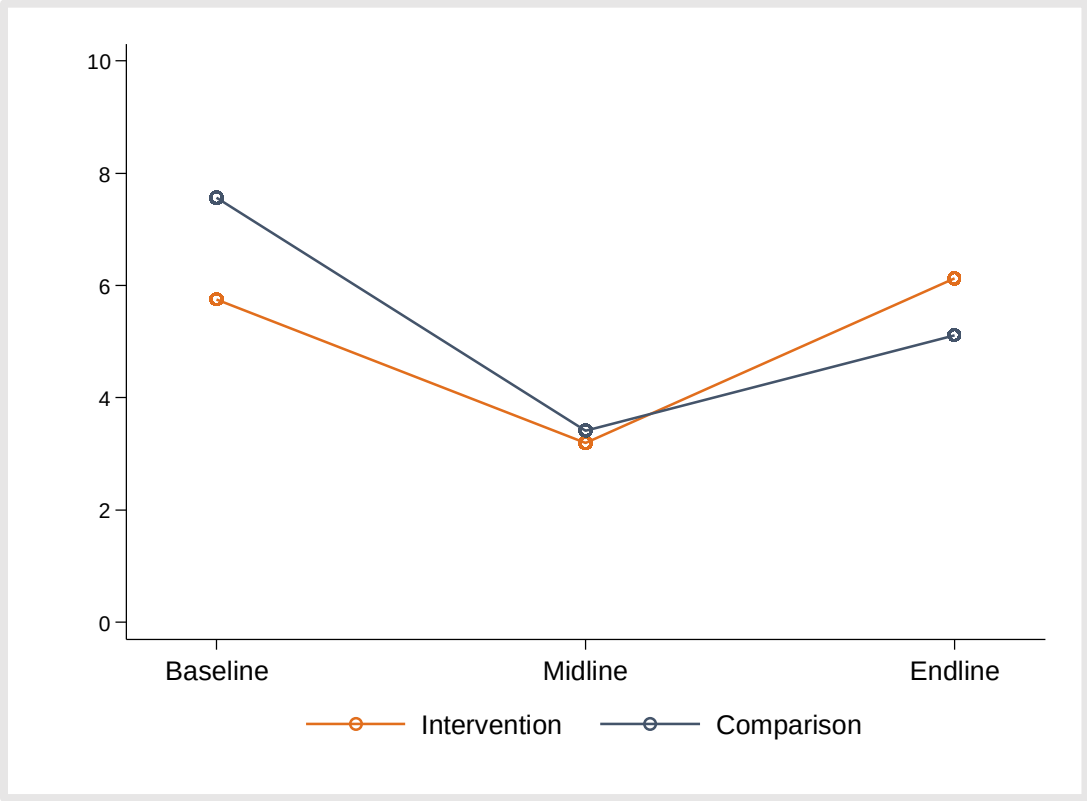

These results again suggest that the HATUTAN program may have had a greater impact among students who were already able to recognize some letter groupings than in reducing the percentage of students who cannot read at all (or, in this case, who are only able to read letters).

## Familiar Word Reading

In the third subtask, students were asked to read familiar words (e.g., if the subtask were in English, words such as “cat” or “dog”, starting with shorter, structurally simpler words and progressing towards longer words with more complex structures). They were presented with a paper with 60 words and asked to read the words aloud in order until 60 seconds had elapsed. Their familiar word score is then calculated as the percent of words read correctly (accounting for pronunciation differences) within the 60 second time limit.

Figure 9 shows very similar patterns to those seen for overall scores and subtasks analyzed so far. While intervention students had substantially lower average scores than comparison students at baseline, the combination of a slower rate of decline in scores between baseline and midline and a faster rate of improvement in scores between midline and endline means that at endline, intervention students have now surpassed comparison students. Indeed, intervention students’ average familiar word reading scores were 6.1 WPM at endline, while comparison students’ scores were 5.1 WPM.

**Figure 9: Familiar word reading scores from all rounds, grade 2 students**




The below table shows the difference-in-differences analysis for baseline-endline and midline-endline changes. As with most results so far, we find a significant relative improvement among intervention students' scores from baseline to endline, but a positive but insignificant improvement in relative scores from midline to endline. In other words, **the HATUTAN program had a positive effect on familiar word reading skills**, but this effect was not significant between midline and endline.

**Table 10: Analysis of changes in familiar word reading scores**

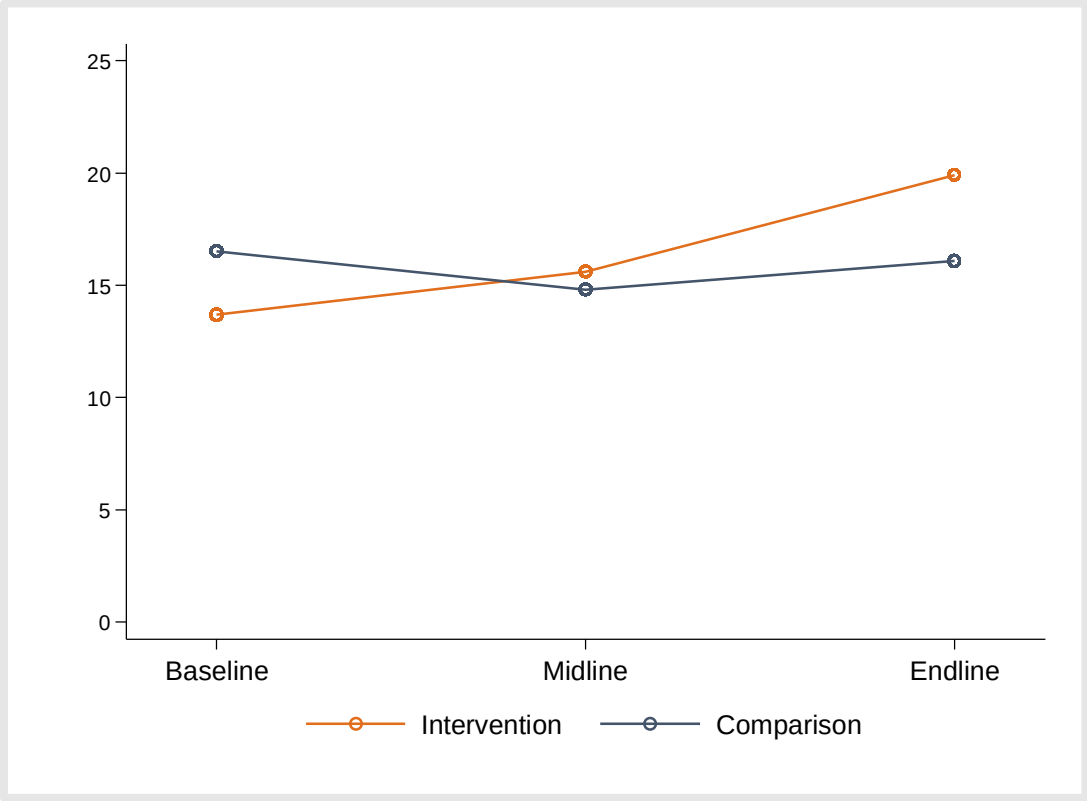
| | Baseline-Endline | | Midline-Endline | |
|---|---|---|---|---|
| | DiD | P-value | DiD | P-value |
| No controls | 2.8 | 0.009* | 1.2 | 0.11 |
| Student-level controls | 3.1 | 0.006* | 1.3 | 0.08 |
| Student- and school-level controls | 2.7 | 0.009* | 1.2 | 0.11 |
| Student-, school-, and household-level controls | 3.0 | 0.04* | 0.8 | 0.36 |

Looking at the percent of students reading zero WPM on this subtask, we note that while rates of zero scorers are similar across comparison and intervention groups, rates remained substantially higher at endline than at baseline. At baseline, 58.2% of intervention read zero WPM on this subtask; this percent increased to 79.5% at midline as a result of COVID, and then decreased by around 10 percentage points to 69.1% at endline. While this is a substantial decrease since midline, the gap between baseline and endline results suggests that grade 2 students may still be substantially behind in more-advanced literacy

skills, perhaps due to lingering effects from the COVID-19 pandemic and subsequent shocks. We note also that we find no significant impact of HATUTAN on the percent of zero scorers for this subtask.

For non-zero scorers, once again, results are more positive. Figure 10 shows that average scores for intervention non-zero students consistently improved from baseline to endline, and improved at a greater rate between midline and endline than between baseline and midline. In contrast, comparison non-zero scorers' average scores declined from baseline to midline, and improved only marginally between midline and endline; these students' average scores were lower at endline than at baseline.

**Figure 10: Familiar word scores from all rounds, grade 2 students reading more than zero WPM**




As such, we find a large and significant improvement in familiar word scores among intervention non-zero scorers compared to comparison non-zero scorers. From baseline to endline, intervention students' average familiar word reading scores improved by 6.5 words per minute more than expected given the results of comparison students. From midline to endline, relative improvement was somewhat less, at 3.0 words per minute, but remained significant. As with letter recognition and overall scores, this shows the HATUTAN program's relative impact among students with at least some ability to recognize words.

## Passage Reading

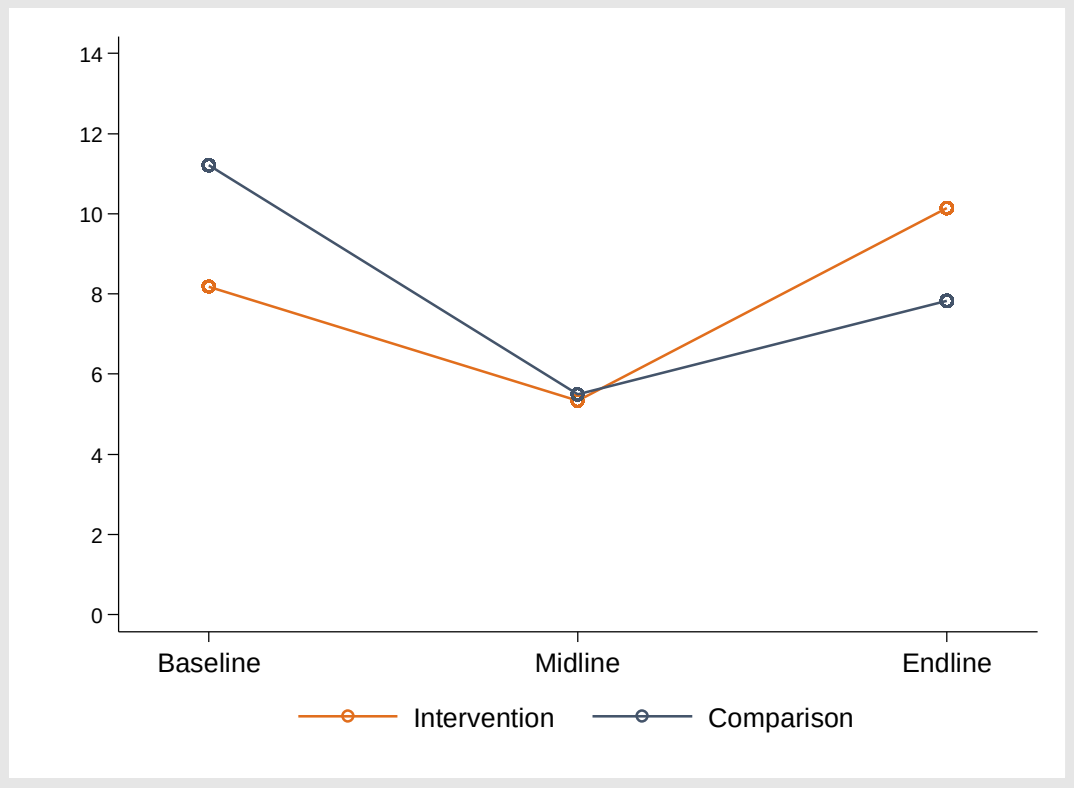
The fourth subtask asked students to read a 61-word passage aloud; students were marked based on the number of words read correctly within 60 seconds. Only students who were able to read either invented or familiar words were asked to attempt this subtask. As such, only 30.5% of students at endline (438 intervention students and 317 comparison students) attempted this subtask.

First, we note that among students who attempted this subtask, only 2% scored 0. In other words, most students who were able to read at least one invented or familiar word were also able to read words within the passage. At midline, in contrast, 16% of students who read invented or familiar words were unable to read any passage words; it was hypothesized that this was due to low levels of fluency or due to students

feeling overwhelmed by the task of reading a passage. It is a promising sign that this dynamic has improved at endline.

The below figure first shows that average passage reading scores were very low, in part due to the large percentage of students who did not attempt this task (as they either could not read words or letters). However, as with other results, **intervention students have seen relatively greater improvement in passage reading abilities compared to comparison students at endline**. For intervention students, a relatively less severe decline in scores between baseline and midline and a greater increase in scores between midline and endline means that these students' average passage reading scores now fall 2.3 percentage points above comparison students'.

**Figure 11: Passage reading scores from all rounds, grade 2 students**



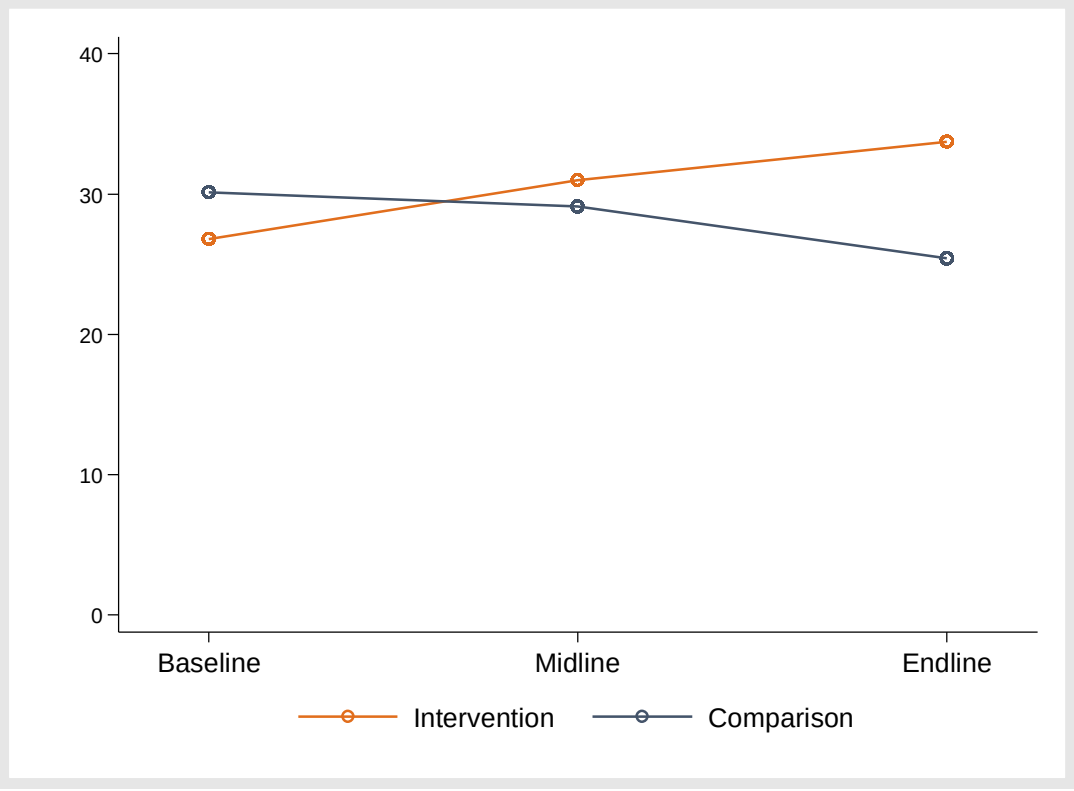

The below table shows results of the difference-in-differences analysis for passage reading scores. We again find a positive and significant impact of the HATUTAN program for intervention students' passage reading scores from baseline to endline when compared to comparison students. Results from midline to endline, in contrast, are significant only in one regression specification, and have a lower effect size of only around 2.5 percentage points.

**Table 11: Analysis of changes in passage reading scores**

| | Baseline-Endline | | Midline-Endline | |
|---|---|---|---|---|
| | DiD | P-value | DiD | P-value |
| No controls | 5.3 | 0.004* | 2.5 | 0.06 |
| Student-level controls | 5.9 | 0.003* | 2.8 | 0.04* |
| Student- and school-level controls | 5.0 | 0.005* | 2.4 | 0.06 |
| Student-, school-, and household-level controls | 5.9 | 0.02* | 1.8 | 0.25 |

We now analyze results for non-zero scorers, shown in Figure 12 below.[66] This figure shows a pattern quite similar to that found in Figure 10 on familiar word reading, suggesting that students who were able to read familiar words had similar levels of success, on average, reading a passage. As with familiar words, we find consistent improvement in non-zero scorers' passage reading abilities from baseline to endline, while scores for comparison students have deteriorated.

**Figure 12: Passage reading scores from all rounds, grade 2 students scoring greater than 0%**




As such, **we find a large and significant impact of the HATUTAN program on passage reading scores for students who were able to read at least some words in the passage.** This effect is particularly substantial from baseline to endline; we find that at endline, intervention students performed 11.6 percentage points higher than expected given the results of comparison students between these two

[66] We do not analyze results for zero scorers as very few students scored 0% on only one of the invented word and familiar word reading subtasks. As such, results for zero scorers for passage reading are highly similar to those analyzed in the previous two subsections.

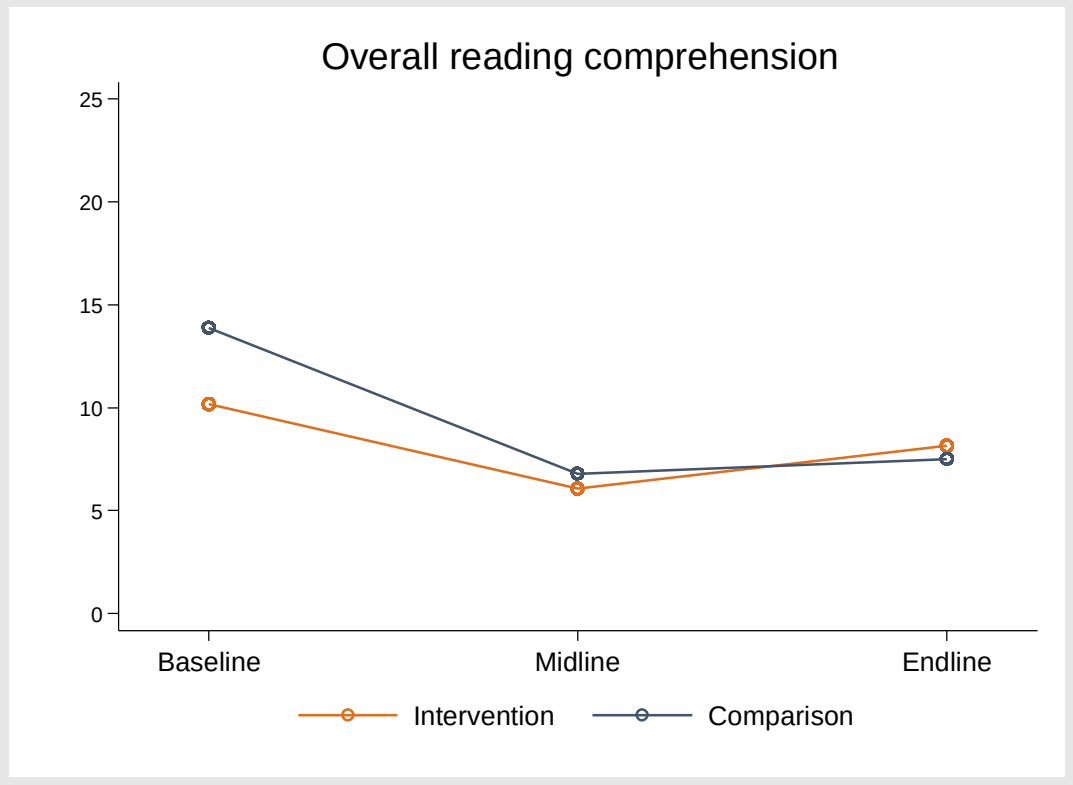
rounds. While the effect size is less from midline to endline, it is still fairly large and remains significant: Between midline and endline, intervention students performed 6.5 percentage points better than expected given the results of comparison students.

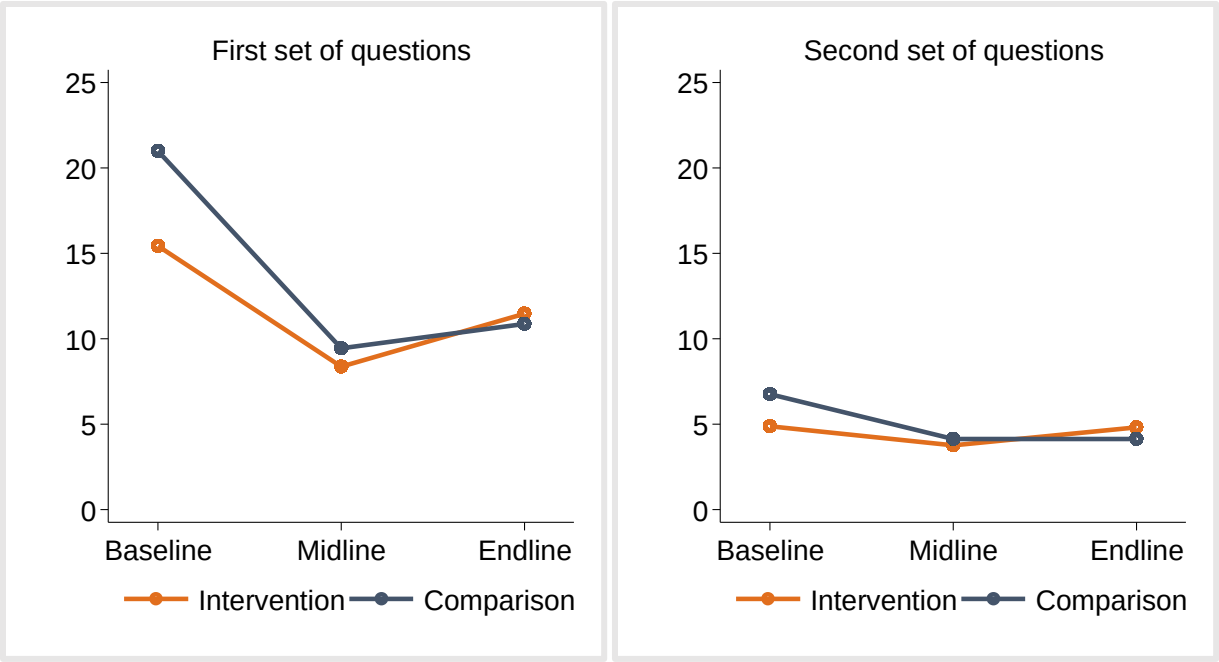
## Reading Comprehension

In the fifth EGRA subtask, students were asked to answer reading comprehension questions. The first set of five questions required literal understanding of the passage read in subtask 4; students were allowed to re-read this passage with no time limit before answering the questions. The second set of five questions required students to read a longer, more complex passage and to answer somewhat more complex, inferential questions. Neither of these tasks was timed.

The below figures show overall reading comprehension scores and scores on the first five questions, which involved reading and interpreting a relatively simple, short text, and the second five questions, which required students to read and interpret a longer and more complex passage. First, we note that average scores on the first set of questions were substantially higher than those on the second set of questions, reflecting the increased difficulty of the second set of questions. Second, we note that average scores for reading comprehension have declined for both intervention and comparison students since baseline, and have only increased marginally since midline. Scores have declined more severely for comparison students than for intervention students, but neither group saw a substantial improvement in scores from midline to endline, in contrast to results discussed above.

**Figure 13: Reading comprehension scores from all rounds, grade 2 students**






To further examine this pattern, the table below shows difference-in-differences results for overall reading comprehension (as patterns in scores for the two subcomponents of this task mirror overall results). **For the period from baseline to endline, we find that intervention students' scores are significantly higher than expected given the results of comparison students**: Intervention students' average scores declined by 2 percentage points from baseline to endline, while comparison students' scores declined by 6.4 percentage points. However, we find no significant effect of the HATUTAN program on intervention students' reading comprehension scores from midline to endline.

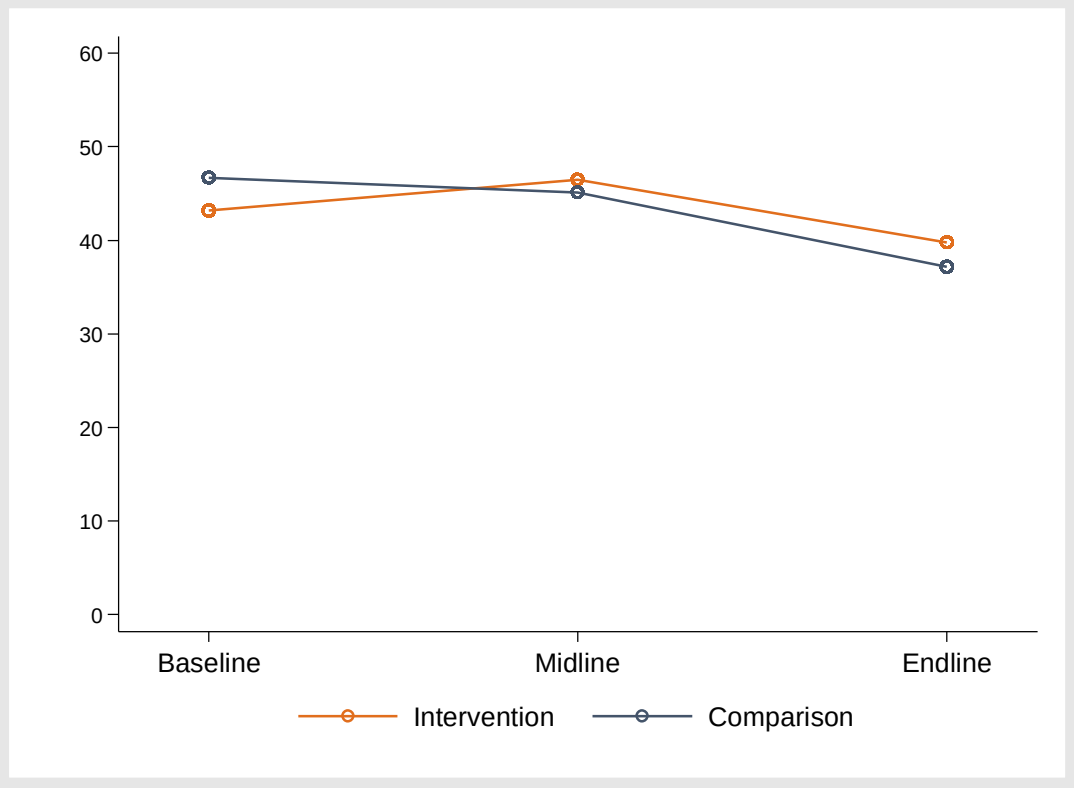
Furthermore, while these results suggest that the HATUTAN program did have a mitigating effect on reading comprehension losses, it is still important to note that scores decreased for intervention students at endline, in contrast to improved scores for all other subtasks. This may suggest that **the HATUTAN program has been more effective at strengthening skills such as letter and word recognition than in addressing higher-level skills needed to interpret the meaning of written passages.** This may be the case if HATUTAN programming does not sufficiently address reading comprehension skills or, alternatively, if students' literacy skills are too weak to substantially benefit from programming targeting comprehension.

**Table 12: Analysis of changes in reading comprehension scores**

| | Baseline-Endline | | Midline-Endline | |
|---|---|---|---|---|
| | DiD | P-value | DiD | P-value |
| No controls | 4.4 | 0.03* | 1.4 | 0.32 |
| Student-level controls | 4.9 | 0.02* | 1.6 | 0.25 |
| Student- and school-level controls | 4.1 | 0.04* | 1.7 | 0.23 |
| Student-, school-, and household-level controls | 3.4 | 0.23 | 1.4 | 0.44 |

We now look at scores among students who scored greater than 0%. Interestingly, we find that intervention non-zero scorers had the highest reading comprehension scores at midline, in sharp contrast to results for other subtasks. Scores for both intervention students and comparison students declined at endline. Considering the period from baseline to endline, intervention non-zero scorers' average scores declined significantly less than expected given the results of comparison students. However, there was little difference in the decline in scores between treatment groups from midline to endline.

**Figure 14: Reading comprehension scores from all rounds, grade 2 students scoring greater than 0%**




Lastly, we note that within the HATUTAN results framework, 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. Globally, this is generally measured as the percent of students who can respond to at least 80% of reading comprehension questions correctly. However, because many 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 intervention students at endline, 20.4% of students demonstrated that they could read and understand the meaning of a grade-level passage (McGovern-Dole Standard Outcome #1).** This represents an increase from the midline value of 13.0% but a decrease from the baseline value of 23.5%. Only 1.7% of students were able to answer 80% of reading comprehension questions correctly, a marginal increase from midline when 1.5% of intervention students were able to answer 80% of questions correctly.

## PREDICTORS OF LITERACY

In this section, we test the relationship between student- and school-level characteristics and overall literacy scores. We first look at individual variables, such as attendance and age, and then analyze variables measured at the school level, such as student-teacher ratio.[67]

At midline, the predictive model included ten student-level variables: age, gender, working memory, caregiver education level, preschool attendance, school absences, whether the student reads at home, whether a toilet is available at home (a proxy for student health and hygiene), the caregiver's nutrition level (a proxy for household nutrition status), and whether the student studied at home during COVID. We include the first nine of these variables in our analysis, as well as controls for whether the student speaks Tetum-Prasa as a first language or has a physical or cognitive[68] disability. In our first model, we also control for factors which vary at the school level, but only report values at the individual level.[69]

Figure 15 shows results for predictors of literacy at the individual level. Within this figure, the markers represent the effect size between the variable and outcome (e.g., all else held constant, at endline, as student age increases by 1 year, overall literacy scores increase by 1.1 percentage points). The bars represent the confidence interval. Bars which do not overlap with zero represent a significant effect.

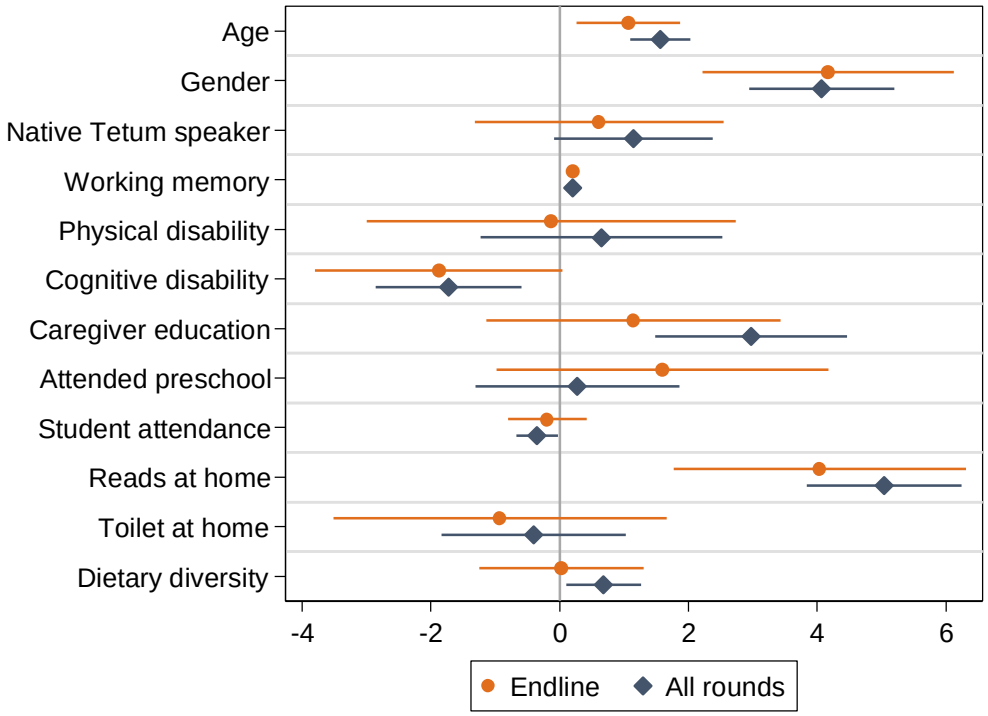
We run two regressions, one for only grade 2 students evaluated at endline, and one for all grade 2 students evaluated at baseline, midline, or endline. For both regressions, we find a significant and positive relationship between overall literacy scores and student age (all else held constant, older students have significantly higher average literacy scores), gender (female students have significantly higher average literacy scores), working memory (students with higher working memory scores have significantly higher average literacy scores), and whether the student reads at home (students who read at home have significantly higher average literacy scores). For the regression which includes all grade 2 students evaluated at baseline, midline, and endline, we find that students with more-educated caregivers and caregivers with higher levels of dietary diversity also have significantly higher average literacy scores. We also find that students who have missed more days of school (i.e., students with poorer attendance) and who have cognitive disabilities have significantly lower overall literacy scores.

[67] 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. 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.

[68] Cognitive disabilities include difficulties with memory or concentration, communication, and self care.

[69] We use school fixed effects in the first model. 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.

**Figure 15: Predictors of literacy, individual level**




These results are, in general, unsurprising. First, we expect older students to have stronger literacy abilities due to maturation effects – i.e., because older students have had longer to learn. Second, in Timor-Leste, female students tend to outperform male students at early grades. Higher working memory scores imply greater student attentiveness and (tautologically) memory, aspects correlated with learning. Qualitative interviews reiterated both of these dynamics, with some respondents stating that young students have more difficulty learning[70] and that girls tended to perform better in school.[71]

Third, reading at home helps strengthen students' literacy abilities and, vice versa, students are only likely to read at home if they already have some literacy abilities; the two dynamics are thus closely related. Fourth, caregivers with higher education levels are more likely to have literacy abilities themselves, and thus be able to help their child learn to read. One father reiterated this dynamic in an FGD:

> *[The teachers] called us to school [and asked], "Why can't your children read?" We responded to them, "Our children cannot read, that's why we send them here to learn from you. As parents we cannot read either, so we cannot help them."*
>
> - FGD with fathers, Manatuto, Int. 20

Fifth, greater dietary diversity is associated with household nutrition status; students with better nutrition may also have improved health outcomes and attentiveness, allowing them to attend school more often and learn more effectively while in school. Sixth, students who frequently miss school have less exposure to education and literacy learning, and therefore may have more difficulty reading. Lastly, students with

[70] E.g., KII with administrator, Ainaro, Int. 1; KII with administrator, Manatuto, Int. 7
[71] E.g., KII with administrator, Manatuto, Int. 8; FGD with fathers, Manatuto, Int. 17; FGD with mothers, Ainaro, Int. 26

cognitive disabilities, including difficulties with memory and communication, may struggle more in school if not given proper accommodation.

What is more surprising is the variables which are not correlated with overall literacy scores. We find no significant relationship between overall literacy scores and native language, whether the student has a physical disability, whether the student attended preschool, and whether the student's household has a toilet.

For the last dynamic, it is possible that the benefits of household toilet ownership are captured in the variable for caregiver dietary diversity; we expect that caregivers with higher dietary diversity are more likely to be economically well-off, and thus also more likely to have a toilet. Both variables may also serve as cross-correlated proxies for student health. Indeed, we find a positive and significant relationship between households' toilet access and caregiver dietary diversity, suggesting that this cross-correlation may indeed be occurring.

For student attendance in preschool, the lack of significance may imply that preschools currently do not effectively provide a basis for student literacy learning. This finding may warrant further investigation and, if indeed the case, interventions at the preschool level to ensure that students have a solid start. However, it may also be the case that preschool attendance has a stronger impact on letter recognition skills than other literacy skills; given that letter recognition skills were relatively high for all students, this may mask the effect of preschool attendance within the regression.

The result for physical disability, meanwhile, is encouraging, as it suggests that students' physical disabilities—in eyesight, hearing, or walking—do not serve as a significant barrier to literacy learning in grade 2. We note, however, that the prevalence of physical disabilities was relatively small within our samples, at only 6.9% of all grade 2 students and 6.3% of endline students. We also note that despite this promising finding, it remains highly important to ensure that students with disabilities are given proper accommodations within classes and are properly included in education. Indeed, these results only reflect students with disabilities who are able to attend school, not those who are kept at home due to their disability.

Lastly, regarding student native language, the reason for a lack of significance is unclear, as we would expect students who do not speak the language of instruction to face more difficulties in learning. It is possible that, given low overall student literacy levels, there is insufficient variation in scores at the grade 2 level to capture the difficulties faced by non-native Tetum-Prasa speakers—although we note that we also found no significant relationship between native language and score for non-zero scorers. It is also possible that the impact of student language on learning is captured within the control variables for individual schools, as many schools tend to include students from the same language group (whether Tetum or a minority language). Indeed, we find a significant relationship between school and native language as well as between municipality and native language, suggesting that this may be the case.

Furthermore, qualitative interviews emphasized that student language has a substantial influence over literacy skills and learning. In Liquica, for example, a father stated the following:

> *Language or dialect are the challenges. For example, the Tokodede dialect right now dominates in this school, but teachers have the obligation to introduce Tetum… Well, this is the obstacle… In my opinion, in school, teachers should stay firm and forbid them speaking Tokodede in the school. Otherwise, they will speak Tokodede all the time and Tokodede will dominate at school and at home. At the end, they cannot speak Tetum fluently.*
>
> - FGD with fathers, Liquica, Int. 18

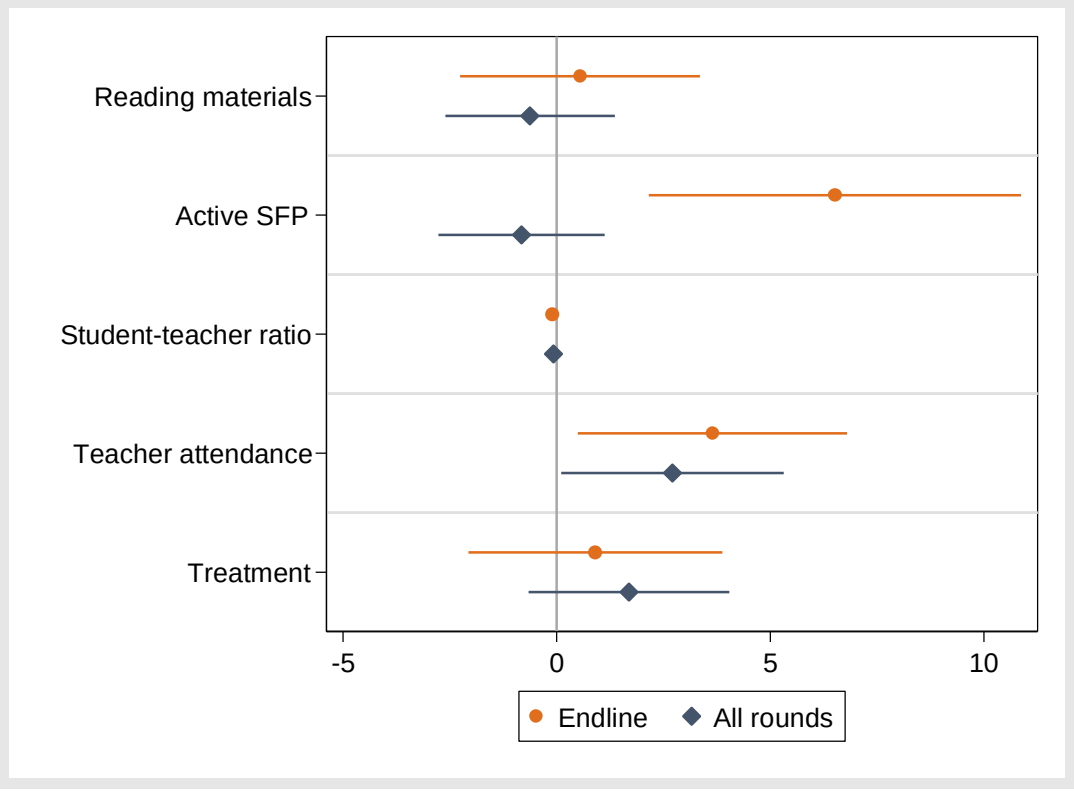
Respondents from other native language groups, such as Bunak speakers in Ainaro and Kemak speakers in Ermera, echoed this sentiment.

> ***Note from the program:*** The proportion of self-reported Tetum speakers is abnormally high, considering the areas where the program is being implemented and the distribution of other languages in country. It is possible that self-reporting has been influenced by desirability bias, affecting the results. The proportion of caregivers self-reporting being Tetum speakers is much higher than the proportion of students (73% compared to 65%) further reinforcing this hypothesis.

We now turn to predictors at the school level. We include two school-level variables also included at midline: availability of reading materials and grade 2 student-teacher ratio.[72] We also include a variable for whether the school feeding program is active and for grade 2 teacher attendance on the day of data collection.

Figure 16 shows the regression results. We find that for endline schools, whether the SFP was active and higher levels of grade 2 teacher attendance were significantly correlated with higher average literacy scores. We also find a very small but significant negative correlation between grade 2 student-teacher ratio and overall literacy scores; in other words, we find, surprisingly, that *larger* class sizes are associated with slightly higher overall literacy scores. While this result is unexpected, it may be correlated with the remoteness of schools, as very remote schools tend to have fewer students and, often, worse learning outcomes.

**Figure 16: Predictors of literacy, school level**




Looking at data from all rounds, we similarly find a positive and significant relationship between grade 2 teacher attendance and literacy outcomes, as well as a small negative relationship between student-teacher

[72] We do not include use of engaging teaching practices as this was measured during classroom observations, and may thus not be applicable to all students within the school.

ratio and literacy. We do not, however, find a significant relationship between SFP activity and overall literacy. In the below section analyzing SFP outcomes, we note that at endline, most SFPs were inactive due to government delays. Active SFPs were only recorded in Ainaro, Bobonaro, and Manatuto; as such, at endline, this variable may serve more as a proxy for learning outcomes in those municipalities than as an indicator of the actual relationship between SFPs and learning.

Overall, this predictive analysis suggests first that student-level demographic characteristics have an influence on literacy outcomes. This finding is unsurprising, but is also not addressable by the HATUTAN program (or any program). Of more relevance for the program—though less conclusive—is the finding that some household characteristics that can change over time may have an influence on learning outcomes. The most relevant of these characteristics is caregiver dietary diversity, an aspect targeted by HATUTAN and which may have a positive effect on literacy levels. Given that working memory scores were also significantly associated with higher literacy outcomes, student attentiveness appears to have an important effect on learning, and again may be influenced through HATUTAN interventions addressing household nutrition as well as school feeding – areas discussed further in subsequent report sections.

Lastly, we find the most significant school-level relationship between teacher attendance and learning. This finding suggests that future interventions targeting teacher attendance may have an outsized effect on literacy levels. These dynamics are further discussed in the next section.

# QUALITY OF INSTRUCTION

To enhance students' literacy outcomes, it is critical to consider the quality of instruction. This includes various elements such as regular teacher attendance, availability of school materials, access to high-quality and effective literacy instruction materials, and the presence of knowledgeable school administrators and teachers who employ engaging and effective teaching practices. To assess the quality of instruction, we conducted classroom observations of grade 2 Tetum language classes during both baseline and endline surveys. Our data collectors recorded whether teachers used engaging or ineffective teaching practices and whether there was any evidence of gender bias in teaching practices. In addition, we collected data on teacher attendance, the availability of school supplies, and the backgrounds of teachers and school administrators.

## TEACHING PRACTICES

In order to understand the differences in teaching practices, it is important to distinguish between engaging, traditional, and negative teaching practices. Engaging teaching practices have been shown to increase student participation and learning outcomes. These practices include asking open-ended 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, using the Lafaek magazine in literacy activities, encouraging students, and asking questions to students. In contrast, traditional teaching practices that have been excessively employed in schools throughout Timor-Leste involve having students spend the majority of their time copying from the board and requiring students to repeat back what the teacher has said for most of the time. Negative teaching practices, including using a harsh tone with students or employing corporal punishment, can lead to negative emotional and psychological outcomes for students. It is crucial for teachers to understand the impact of their teaching practices on student learning and to incorporate engaging practices while avoiding ineffective and negative ones.

## Engaging Teaching Practices

At endline, the comparison schools demonstrated an average increase of twice the number of engaging teaching practices,[73] while intervention schools showed only a marginal increase. At baseline, teachers in comparison schools used an average of 2.3 engaging teaching practices, while those in intervention schools used on average 4.9 engaging teaching practices. By the endline, teachers in comparison schools used an average of 4.8 engaging teaching practices, while those in intervention schools used an average of 5.1 engaging teaching practices. The findings reported in Table 13 suggest that, **although the mean number of engaging teaching practices in the intervention schools increased only slightly, there was a favorable change in the distribution of the number of teaching practices.** Specifically, the percentage of classrooms employing a low number of engaging teaching practices decreased by approximately 11 percentage points from baseline to endline, while the percentage of classrooms using a high number of engaging teaching practices increased by 7 percentage points.

**Table 13: Use of engaging teaching practices (% of classrooms)**

| | Comparison Schools | | Intervention Schools | |
|---|---|---|---|---|
| | BL | EL | BL | EL |
| n | 90 | 84 | 98 | 98 |
| Average # practices | 2.3 | 4.8 | 4.9 | 5.1 |
| Low (0-3) | 66.6% | 22.6% | 35.7% | 24.5% |
| Moderate (4-6) | 20.1% | 53.6% | 39.8% | 43.9% |
| High (7-9) | 13.3% | 23.8% | 24.5% | 31.6% |

According to the difference-in-differences regression analysis, **there was no significant difference in the number of engaging teaching practices used at endline in intervention schools compared to comparison schools**. The analysis employs two specifications. The first specification is a DiD regression on all surveyed schools, whether at baseline or endline, which controls for teacher education and experience, teacher gender, classroom size, and school type (central or filial), all of which may influence the use of engaging teaching practices. The results suggest that teachers in intervention schools only used slightly more engaging teaching practices on average than those in comparison schools, which was not statistically significant. The second specification is a restricted DiD regression on a sample of schools where classroom observations were conducted both at baseline and endline. This analysis controls for potential unobserved differences between schools that may affect the use of engaging teaching practices and bias our results. The restricted sample includes all 98 intervention schools and 84 comparison schools assessed at both baseline and endline. The findings confirm that there were no significant changes in the number of engaging teaching practices used in intervention schools relative to comparison schools at endline.

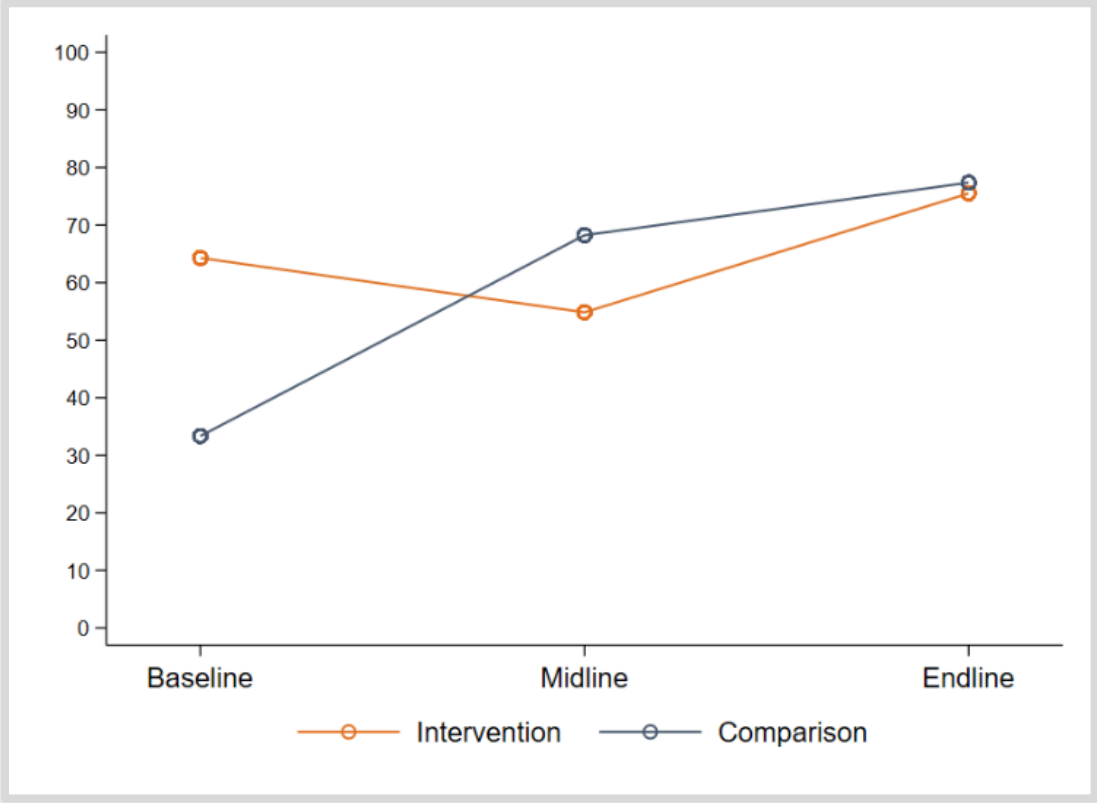
The findings also reveal some changes in the types of engaging teaching practices used. At baseline, teachers were most likely to ask open questions (observed in 87% of all classes), read to the classroom (67%), and engage inactive students (62%), while they were least likely to use the reading corner (26%) or have students participate in group work (39%). However, at endline, teachers were relatively less likely to ask open questions (observed in 74% of all classes), but more likely to read to the classroom (93%) and engage inactive students (69%). Despite these changes, the difference-in-differences analysis indicates that **there was no significant change in the use of different types of engaging teaching practices**

[73] 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.

**between intervention and comparison schools because of HATUTAN programming. Therefore, the impact of the program on the types of engaging teaching practices used in intervention schools remains inconclusive.**

McGovern-Dole Custom Outcome #5 measures the percent of teachers who adhere to improved learning practices in schools, as defined by the demonstration of a minimum of four engaging teaching practices during classroom observation. Despite Figure 17 showing an improvement in this indicator for intervention schools at endline, a difference-in-difference regression reveals that this result is not significant when compared to the comparison schools. Interestingly, the comparison schools show a higher improvement than the intervention schools, resulting in a negative coefficient for the difference-in-difference regression, albeit insignificant.

**Figure 17: Change in McGovern-Dole Custom Outcome #5 for teaching practices**




Teachers were also observed using the Lafaek magazine in literacy activities, another positive teaching practice that can help students improve their literacy skills. At midline,[74] 7% of teachers in intervention schools were observed using the Lafaek magazine for literacy activities, and this decreased to 6% at endline. On the other hand, in comparison schools, 16% of teachers were observed using the Lafaek magazine in literacy activities at midline, and this decreased to 7% at endline, indicating a decline of around 9 percentage points. While this relatively smaller 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, there may be inadequate encouragement or training for teachers to effectively use the Lafaek magazine in literacy activities.** This trend could also be attributed to the fact that teachers are now being evaluated based on the use of other resources, which were previously absent. Despite the magazine's alignment with the curriculum, it may have served as a temporary solution to fill a gap.

[74] Teachers were not observed using Lafaek magazine in literacy activities at baseline.

## Traditional Teaching Practices

Table 14 displays the outcomes of data collectors' observations of traditional teaching practices in classrooms. **At endline, there was a substantial and statistically significant decrease in the use of traditional teaching practices in intervention schools compared to comparison schools.** The average number of traditional teaching practices used (ranging from 0 to 2) decreased by 0.4 at endline in intervention schools, while comparison schools showed an increase of 0.7. The difference-in-difference regression indicates statistical significance. Additionally, the prevalence of students copying from the board decreased substantially in intervention schools: At endline, students were observed copying from the board around 21 percentage points less frequently than at baseline, while in comparison schools, students were observed copying from the board only around 6 percentage points less than at baseline. However, unlike the change in the practice of students repeating the teacher, which was significant, this change is not statistically significant when running a difference-in-difference regression. These results remain robust when controlling for teacher, classroom, and school characteristics and limiting the sample to schools assessed at both baseline and endline**. Overall, it seems that the HATUTAN program had limited success in promoting engaging teaching practices, but it was effective in decreasing the use of traditional teaching practices.**

**Table 14: Change in use of traditional teaching practices**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | DiD | P-value |
| Average # of practices | 1.6 | 1.2 | 0.6 | 1.3 | -1.1 | <0.001* |
| Students copy from the board | 81.6% | 60.6% | 60.5% | 54.8% | -15.3 | 0.18 |
| Students repeat the teacher | 74.5% | 63.8% | 65.1% | 79.7% | -25.3 | 0.01* |

## Negative Teaching Practices

In contrast to the previous results, the effect of the HATUTAN program on negative teaching practices appears to be limited. Table 15 presents the results of the analysis of negative teaching practices, showing that there was almost no change in the use of negative teaching practices from baseline to endline in intervention schools, while there was a modest increase of 0.3 in comparison schools. The difference-in-difference regression suggests that the decrease in negative teaching practices was significant, although the result is not robust when controlling for teacher, classroom, and school characteristics.

Examining specific negative teaching practices, the use of corporal punishment did not change significantly in the intervention schools but increased by approximately 11 percentage points in comparison schools, a substantial but not significant change. The practice of employing an angry or harsh tone with students increased marginally in intervention schools, but this result did not attain statistical significance either. Overall, **the findings suggest that the HATUTAN program had limited success in reducing negative teaching practices, and additional program interventions are necessary to effectively discourage the prevalence of such practices among teachers.**

**Table 15: Change in use of negative teaching practices**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | DiD | P-value |
| Average # of practices | 0.5 | 0.4 | 0.2 | 0.5 | -0.3 | 0.04* |
| Teacher uses angry voice | 34.7% | 35.1% | 35.7% | 34.5% | 1.6 | 0.87 |
| Teacher uses corporal punishment | 11.2% | 11.2% | 1.1% | 11.9% | -10.8 | 0.24 |

Moreover, as noted in the initial report, the observed incidence of verbal and physical disciplinary measures in classrooms is likely to be lower than their actual prevalence due to social desirability bias. In fact, at endline, the household survey revealed that 33% of respondents reported that teachers use corporal punishment, while only 12% of classrooms showed evidence of it. Additionally, 10% of respondents reported that teachers use chores as a form of discipline, and 12% of caregivers stated that their children sometimes feel scared to attend school.

Table 16 indicates that, in intervention areas, caregiver perceptions of teacher use of corporal punishment decreased slightly from baseline, although the difference was not statistically significant**.** On the other hand, perceptions of teachers using chores and an angry tone increased in intervention areas compared to comparison areas, but again, not significantly. In addition, caregiver perceptions of children feeling frightened to go to school[75] decreased by 8 percentage points in intervention areas compared to comparison areas, which was statistically significant. These findings reinforce the notion that more interventions are necessary to reduce negative teaching and disciplinary practices in schools, and that the HATUTAN program has yet to achieve significant reductions in these practices.

**Table 16: Change in caregiver perceptions of negative teaching practices**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | DiD | P-value |
| n | 378 | 741 | 482[76] | 601 | | |
| Teacher uses angry voice | 40.2% | 35.9% | 39.4% | 33.9% | 1.2 | 0.89 |
| Teacher uses corporal punishment | 36.1% | 34.0% | 30.4% | 30.9% | -2.6 | 0.69 |
| Teacher assigns chores | 26.3% | 10.4% | 28.8% | 10.1% | 2.8 | 0.73 |
| Student is afraid to attend school | 16.5% | 11.6% | 10.1% | 13.3% | -8.1 | 0.04* |

According to qualitative interviews, there is a broad range of perspectives on the effectiveness of teaching practices and the prevalence of engaging, traditional, and negative practices. Many interviewees, including a father,[77] teachers,[78] and school administrators,[79] believe that positive reinforcement and patience are

[75] The question did not further investigate the underlying reasons why the child exhibits fear of attending school.
[76] n = 479 for the "student afraid to attend school" indicator.
[77] FGD with fathers, Ainaro, Int. 13
[78] FGD with teachers, Ainaro, Int. 37; FGD with teachers, Manatuto, Int. 44; FGD with teachers, Oe-cusse, Int. 45
[79] KII with school coordinator, Ainaro, Int. 1; KII with school coordinator, Liquica, Int. 6

more effective than punishment or violence for improving children's academic performance, while negative practices were found to scare children and discourage participation. One teacher in Manatuto[80] municipality stated that when children misbehave in class, he asks them to sing or play a game using illustrated books to regain their focus. Another teacher from Ainaro municipality mentioned that she asks children to sing together to capture their attention.[81] With regards to traditional teaching practices, a teacher in Oe-cusse municipality noted that teachers had moved away from the traditional practices where students copied what the teacher wrote or dictated.[82] However, a teacher and a school administrator from Ainaro mentioned that repeating letters after the teacher is still a commonly used method,[83] especially when children are first learning the alphabet.

However, several respondents reported using or desiring to use corporal punishment and other negative teaching practices. For instance, a school teacher mentioned that while it is now illegal to beat students, teaching without the use of corporal punishment is more challenging.[84] Meanwhile, an administrator from Ermera acknowledged that it can be challenging for teachers to abandon corporal punishment since some still rely on old Portuguese and Indonesian discipline techniques.[85] Some teachers acknowledge that regulations prohibit corporal punishment, but they still view it as an educational tool rather than a form of punishment.[86] The intention, according to these teachers, is to educate the children and help them make better choices. Qualitative data shows that some teachers even seek to justify abuse by mentioning that they ask the children whether they believe that being beaten makes them smarter or not and whether it is an act of love or punishment. According to those teachers, the children typically respond that being beaten is a way to show love and help them understand.

In addition to discussing engaging, traditional, and negative teaching practices, educators also shared their methods for teaching literacy. Most educators stated that they begin by teaching students the alphabet, followed by syllables formed by letter combinations in some cases, and only then move on to teaching words after students have learned the alphabet.[87] These findings help to clarify the gap between students' abilities to recognize letters and their abilities to read words. However, many teachers have also mentioned using innovative techniques, such as showing children pictures to help explain words.[88]

Teachers, parents, and school administrators also raised concerns about large classroom sizes as a major challenge to implementing engaging teaching practices and effectively managing classrooms. For example, one teacher stated, "We have 50 pupils in grade 2, and as a result, they chase each other at the back of the classroom and do not pay attention to the teacher. We are unable to deal and are compelled to be patient."[89] Many teachers also noted that large class sizes can present a challenge due to the wide range of skill levels among students in the classroom, making it difficult to ensure that all students are engaged and effectively learning during lessons.

## Gender-Specific Differences in Teaching Practices

We now examine whether there are gender differences in teaching practices and whether teachers treat male and female students differently. Our analysis shows that, at endline, on average, female teachers

[80] FGD with teachers, Manatuto, Int. 43
[81] FGD with teachers, Ainaro, Int. 37
[82] FGD with teachers, Oe-cusse, Int. 45
[83] FGD with teachers, Ainaro, Int. 38; KII with administrator, Ainaro, Int. 2
[84] FGD with teachers, Ainaro, Int. 37
[85] KII with administrator, Ermera, Int. 3
[86] FGD with teachers, Liquica, Int. 42
[87] FGD with teachers, Ainaro, Int. 38; FGD with teachers, Manatuto, Int. 44
[88] FGD with teachers, Ainaro, Int. 37; FGD with teachers, Liquica, Int. 41
[89] FGD with teachers, Liquica, Int. 42

used slightly more engaging teaching practices than male teachers, approximately the same number of traditional practices, and slightly more negative practices. However, when we compare the use of teaching practices across intervention and comparison groups at baseline and endline, we find that **the HATUTAN program generally did not have a significant effect on female or male teachers' practices compared to those who were not exposed to the program.** The only exception is that there was a significant decrease in the use of traditional teaching practices among male teachers in intervention schools at endline compared to comparison schools: In intervention schools, the average number of traditional teaching practices used decreased among male teachers by 0.4, while it stayed the same in comparison schools.[90]

In addition to classroom observations, data collectors were tasked with observing potentially gender-biased positive and negative teacher behaviors such as differential encouragement or questioning of male and female students, and use of corporal punishment or angry voice with male and female students. At endline, teachers were observed to encourage girls (55% of classrooms) slightly more than boys (53% of classrooms). Moreover, teachers asked questions to boys (65% of classrooms) slightly more than girls (64% of classrooms). Additionally, at endline, in 19% of classrooms, teachers were observed asking questions primarily to either boys or girls, rather than to students of both genders.[91] **The HATUTAN program, however, did not appear to have a significant impact on teachers' differential use of encouragement and questioning with boys and girls.**

In order to understand the prevalence of negative teaching behaviors across genders, we examined teacher behavior towards boys and girls separately at endline. We found that teachers used angry voices or harsh tones more often with boys (31% of classrooms) than with girls (25% of classrooms) and were slightly more likely to use corporal punishment with boys (10% of classrooms) than with girls (7% of classrooms). Despite the lack of impact on positive teaching behaviors, **the HATUTAN program may have had a positive effect on reducing the use of corporal punishment towards girls.** Table 17 indicates that there was a 2-percentage point decrease in the use of corporal punishment on girls in intervention schools, while comparison schools saw an approximately 10 percentage point increase. Overall, the difference-in-differences regression analysis suggests a decrease of around 11 percentage points in the use of corporal punishment on girls in intervention schools compared to comparison schools. This result, however, is not significant.

**Table 17: Change in negative teaching practices towards girls and boys**

| | Intervention | | | Comparison | | | DiD (BL-EL) | P-value |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | | |
| n | 98 | 98 | 94 | 43 | 84 | 84 | | |
| Angry voice with boys | 26.5% | 29.6% | 30.9% | 25.6% | 21.4% | 29.8% | 0.2 | 0.99 |
| Angry voice with girls | 23.5% | 16.3% | 22.3% | 23.3% | 15.5% | 28.6% | -6.5 | 0.52 |
| Corporal punishment on boys | 8.2% | 7.1% | 11.7% | 2.3% | 6.0% | 8.3% | -2.5 | 0.75 |
| Corporal punishment on girls | 6.1% | 8.2% | 4.3% | 0.0% | 3.6% | 9.5% | -11.3 | 0.05 |

[90] 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.
[91] 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 6 classrooms, teachers primarily asked questions to female students, and in 8 classrooms, teachers primarily asked questions to male students.

Finally, we examine whether there is a difference in how male and female teachers treat their male and female students. At baseline, male teachers exhibited a more positive attitude towards male students, while female teachers treated their male and female students similarly. Nonetheless, the endline data indicated a notable rise in the proportion of female teachers in intervention schools who asked questions to male students and a decrease of 9 percentage points in their use of physical punishment on female students compared to their counterparts in comparison schools. Male teachers in intervention schools exhibited a tendency to offer more encouragement to female students, ask them more questions, and use less angry tone and corporal punishment towards them (see Table 18). **The findings suggest that the program may have successfully motivated male teachers to enhance their treatment of female students. This could be attributed to the fact that head teachers and PTAs received gender-based violence training as part of the program intervention.** However, as an unintended outcome, it seems to have caused a slight decrease in equity in teaching practices towards male students. Further investigation is necessary to determine the root causes of these divergences.

**Table 18: Treatment of male and female students by gender of teacher at endline**

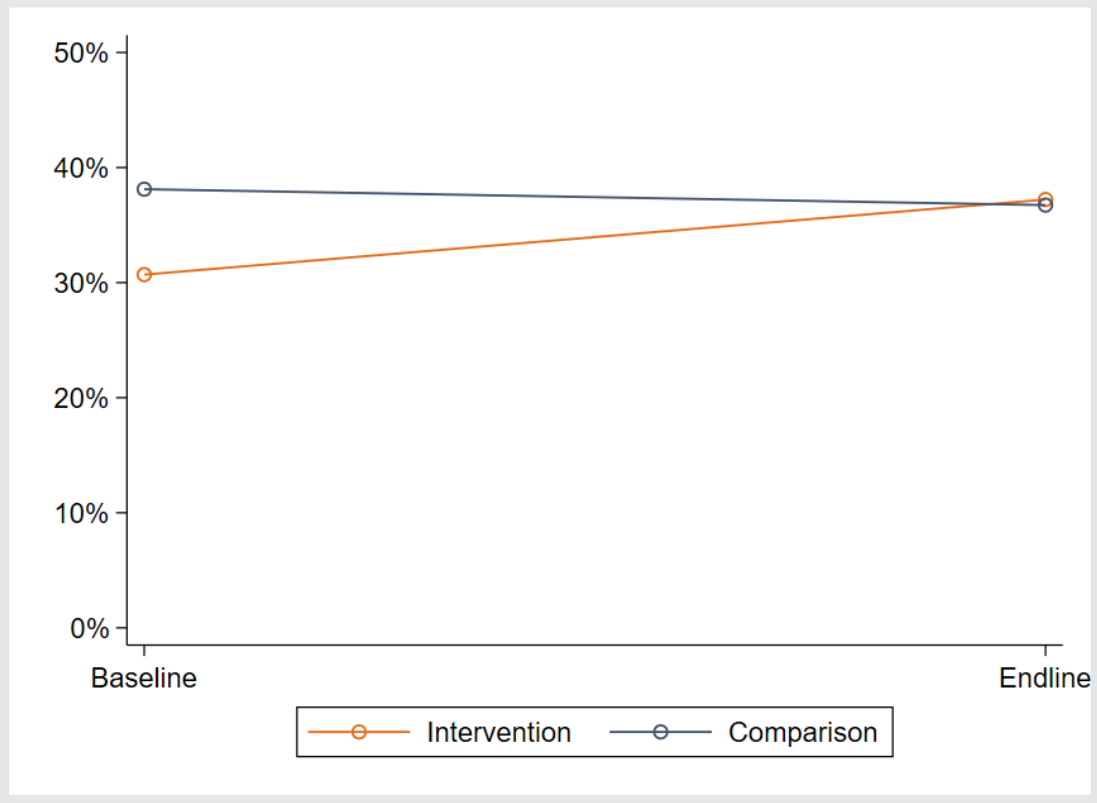
| | Comparison Schools | | Intervention Schools | |
|---|---|---|---|---|
| | Female teachers | Male teachers | Female teachers | Male teachers |
| n | 48 | 36 | 53 | 41 |
| Encourages female students | 54.2% | 50.0% | 50.9% | 65.9% |
| Encourages male students | 54.2% | 52.8% | 45.3% | 63.4% |
| Asks questions to female students | 66.7% | 69.4% | 60.4% | 61.0% |
| Asks questions to male students | 56.3% | 72.2% | 73.6% | 58.5% |
| Uses angry voice with girls | 35.4% | 19.4% | 28.3% | 14.6% |
| Uses angry voice with boys | 35.4% | 22.2% | 41.5% | 17.1% |
| Uses corporal punishment on girls | 14.6% | 2.8% | 5.7% | 2.4% |
| Uses corporal punishment on boys | 10.4% | 5.6% | 15.1% | 7.3% |

## TEACHER SKILLS AND KNOWLEDGE

In difficult learning environments such as those commonly encountered in Timor-Leste, the quality of instruction may be heavily influenced by the educational and skill levels of teachers. The findings at baseline suggested that teachers were undergoing a transitional period, during which they were assimilating new methods of classroom management that eschew violence, learning to work with young pupils, adopting participatory pedagogical techniques in the classroom, becoming more cognizant of the need to engage disengaged students, and experimenting with novel methods to captivate their pupils. Participating in educational courses and literacy training programs may accelerate this transition process and enhance the competencies and knowledge of teachers, thereby advancing the learning outcomes of students.

At endline, the school survey data revealed a higher average percentage of teachers who reported attending literacy education training compared to baseline.[92] Specifically, the percentage of intervention teachers attending literacy training increased by 7 percentage points from 30.7% to 37.2%, while comparison schools witnessed a 1.4 percentage point decrease from 38.1% to 36.7% (Figure 18). **Although the difference was not statistically significant, the findings indicate a positive trend in the proportion of teachers who attend teacher training.** There are various teacher trainings available which are implemented in all Government Basic Education primary schools in some of the intervention municipalities. It is worth noting that the survey did not inquire about the specific training received by the teachers, although it is probable that the question pertained to the INFORDEPE in-service training, which often takes place in urban centers and requires teachers from remote schools to be away from their schools for several days.

**Figure 18: Change in percent of teachers attending literacy training**




Turning to teacher education, the endline data revealed a substantial increase in the percentage of both male and female teachers who had completed a bacharelato or teacher training college program[93]. Male teachers exhibited a 29 percentage point increase, while female teachers displayed a 26 percentage point increase. The difference-in-differences regression analysis indicated that **intervention schools experienced greater improvements in teacher education of male teachers compared to comparison schools, but that these results were not statistically significant** (Table 19). It is worth highlighting that

[92] 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 132 baseline schools and 130 endline 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.

[93] It is important to highlight that internationally, equivalency courses are generally considered to be of lower quality compared to a university degree in teaching.

this type of in-service training programs may primarily benefit adult teachers who participate in them by helping them secure better employment and increased pay, but it may harm the students in the short term due to teacher absences, and it may not necessarily lead to significant improvements in teacher skills. It should also be noted that inconsistencies in data collection and indicator design, as well as variations in the number of schools with zero male or female teachers, resulted in a significant variation in the sample size of schools for this analysis. Therefore, caution is advised when interpreting the results.

**Table 19: Changes in teacher training and education**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | DiD | P-value |
| Attended literacy training | 30.7% (n=75) | 37.2% (n=71) | 38.1% (n=56) | 36.7% (n=59) | 7.9 | 0.48 |
| Completed bacharelato or teacher training college (male) | 41.0% (n=98) | 69.6% (n=98) | 54.8% (n=86) | 77.0% (n=84) | 6.4 | 0.51 |
| Completed bacharelato or teacher training college (female) | 33.3% (n=88) | 59.3% (n=93) | 40.2% (n=76) | 67.6% (n=77) | -1.4 | 0.90 |

In the qualitative data, trainings were mentioned as a common reason for teacher absences, but they were also recognized as valuable opportunities for teachers to acquire knowledge and experience, especially for those without formal education degrees.[94] Furthermore, school administrators emphasized the significance of teacher education, particularly for those teaching first-grade students:

> *I believe that all teachers have similar abilities, but those who teach in first grade require higher abilities because they work with young students who have recently transitioned from pre-school. Therefore, first-grade teachers need to possess exceptional abilities.*
>
> - KII with administrator, Ainaro, Int. 1

## TEACHER ATTENDANCE

In order to enhance the quality of education, it is crucial that teachers attend classes consistently and frequently, as this ensures that children receive a greater number of hours of instruction. Additionally, teachers who attend classes regularly may possess a better grasp of the requirements and potential of their students, and as a result, can adapt their teaching approaches to foster better learning outcomes. The school survey included teacher attendance as a variable of interest, and the data collected by enumerators included the count of permanent, contract, and volunteer teachers at each school, as well as the number of teachers present during the visit and on the previous day.[95]

On average, at endline, 80% of teachers from all grades were present on the day of data collection, while 61% had attended on the previous day based on school records. These results were somewhat different from those of the midline and baseline surveys. During the midline survey, 90% of teachers attended on the day of data collection and 76% attended the day before. During the baseline survey, 62% of teachers attended on the day of collection while 74% attended the day before. At baseline, the overreporting of

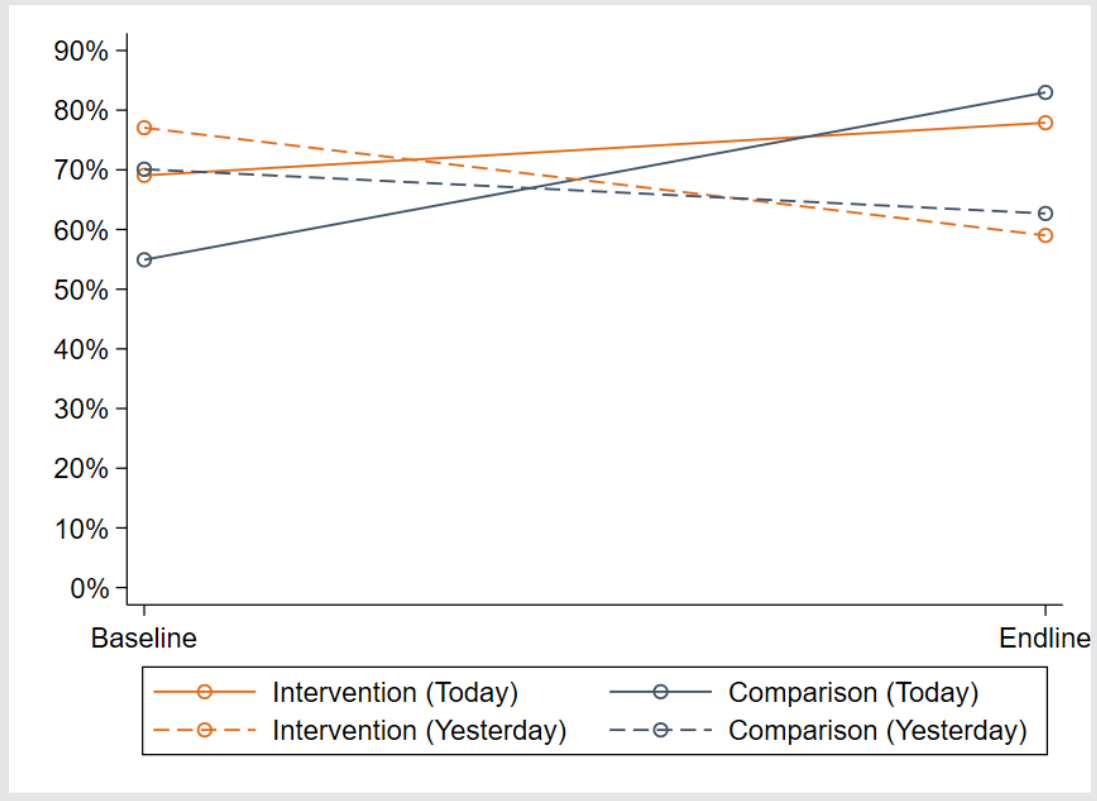
[94] KII with administrator, Ermera, Int. 4; FGDs with teachers, Ermera, Int. 39; FGDs with teachers, Manatuto, Int. 44
[95] At baseline, due to a data collection error, teacher attendance data was unreliable and thus removed for 46 schools.

teacher attendance rates by school records compared to headcounts may have accounted for the disparity between previous-day and day-of attendance rates. Conversely, at both midline and endline, the higher percentage of teacher attendance on the day of data collection may have occurred because school directors encouraged teachers to attend, as they were aware of the scheduled visits by the data collection teams. Therefore, these results only suggest a possible improvement in teacher attendance rates and should not be viewed as conclusive.

As part of the data collection process, data collectors were also instructed to record headcounts for grade 2 teachers. The findings at endline showed that on average, 82.5% of grade 2 teachers were present on the day of data collection, which was similar to the attendance rates recorded at baseline and midline. This consistent attendance pattern may have been due to the fact that teachers were aware of the presence of data collectors and the observation of grade 2 classes on the day of data collection. This possible explanation should be taken into account when interpreting the results.

When comparing intervention and comparison schools, there are notable differences in teacher attendance rates at endline. In comparison schools, teacher attendance taken on the day of the survey increased by 28 percentage points while previous-day attendance decreased by approximately 7 percentage points compared to baseline. On the other hand, in intervention schools, teacher attendance on the day of the survey increased by approximately 9 percentage points while previous-day attendance decreased by around 18 percentage points compared to baseline (see Figure 19).

**Figure 19: Change in teacher attendance**




Using a difference-in-differences regression analysis, **we find a significant relative difference in attendance between comparison and intervention schools taken on the day of the survey, with comparison schools[96] outperforming intervention schools** (see Table 20). These findings remain

[96] Visits to comparison schools were also planned ahead of time.

significant even after controlling for differences in school type, which may be a proxy for remoteness and thus impact teacher attendance, and whether the school has a PTA, which may play a role in enforcing teacher attendance.

**Table 20: Change in teacher attendance on the day of survey**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 66 | 98 | 67 | 84 | | |
| Teacher attendance: day of survey | 69.1% | 77.9% | 54.9% | 83.0% | -19.3 | 0.01* |

After conducting further analysis, it appears that the disparity in teacher attendance rates between intervention and comparison schools can be linked to teacher training programs, as teachers who attend such trainings are unable to be present in school and conduct their regular teaching duties on the same day. Additionally, teachers may also be absent from school to prepare for graduation ceremonies[97] or other school-related events. More intervention schools (nine schools in total) reported grade 1 or grade 2 teachers attending trainings at endline, in contrast to comparison schools where only one school reported teacher trainings. However, the usefulness of data on teacher absences collected through the household survey was limited, as the majority (65%) of respondents did not know how many times their child's teacher had been absent that week. Parents may not be aware of what is happening in their child's classroom or what they are learning. The child may not feel comfortable or confident enough to share their experiences at school with their parents. This disconnect could mean that the child is not receiving the necessary support from their parents to succeed academically.

Examining the results by municipality reveals that the increase in teacher attendance rates varied across different municipalities. For instance, in Covalima, a comparison municipality, day-of attendance increased by 51 percentage points from low baseline level of 32%. In contrast, in Liquica, an intervention municipality where day-of attendance was already high at baseline (75%), attendance increased by only 10 percentage points at endline. These findings suggest that factors at the municipality level may have also influenced attendance rates in some comparison municipalities, rather than the impact of the HATUTAN program. Another possible explanation for the lower teacher attendance rates in intervention areas is the effect of weather and natural disasters. As intervention schools are typically located in remote areas with poor infrastructure, roads in these areas are more likely to be washed out during the rainy season, making it difficult for teachers to attend school.

Finally, results for the McGovern-Dole Custom Outcome #6 for teacher attendance indicate that around 18% of schools achieved the outcome of having at least 80% of teachers in attendance on both the day of data collection and the prior day at endline, which is a decrease from the 32% of schools that achieved the outcome at baseline.[98] The difference-in-difference analysis suggests that intervention schools were about 9 percentage points less likely to achieve this outcome at endline than would be expected based on the comparison schools, but this result was not statistically significant.

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

[97] Related to the conclusion of the trainings attended by teachers.
[98] We do not include schools that did not have teacher attendance records in the analysis.

> *This issue is affecting our children's education because we have very few teachers. When a teacher goes for training, which can last for a week or even a month, our children do not receive the education they need. This is why we sometimes feel like our children are second-class learners compared to those in the city. We would like to send them to school in the city, but our economic situation does not allow for it. As farmers, we can only afford to keep them in our village schools. However, due to the shortage of teachers, when our government sends them for training, our children miss out on valuable lessons.*
>
> - FGD with fathers, Ermera, Int. 16

Parents emphasized the need for teachers and schools to better plan for trainings, ensuring that another teacher covers the lessons for the teacher who is absent due to training, so that students are not left without a teacher. In Ainaro municipality, a mother gave an example of this happening: if teachers were attending a training, they made sure that teachers who teach the same subjects covered for each other.[99] However, most respondents who mentioned training stated that it had a negative impact on teacher attendance, which ultimately affected students' learning. In recognition of the negative impact of teacher training on students, one school coordinator stated that he had discussed with his superior the possibility of rescheduling all training for holidays so that teachers would not have to abandon their classrooms.[100]

In addition to training, several respondents mentioned that bad weather or road conditions had an impact on teacher attendance, and that understaffing of schools made the situation worse. One father in Liquica municipality mentioned that teachers occasionally choose not to go to work due to excessive rain that can last for up to two or three days. This affects the education of the students as they miss out on valuable lessons.[101]

## LITERACY INSTRUCTION MATERIALS

Literacy instruction materials are crucial to the quality of education as they provide teachers with access to resources that can enhance their lesson plans and classroom activities, and improve students' access to supplementary materials, which can aid learning. In the school survey, data collectors observed whether grade 2 classrooms had a reading corner and reading materials suitable for children in the 2nd grade.

At the endline, significant improvements were observed in the access to literacy instruction materials at the intervention schools. The intervention schools experienced a remarkable increase in reading corners, with almost 28 percentage points more schools having a reading corner at endline than at baseline. Conversely, comparison schools showed a decline in both the reading corner and reading materials. The difference-in-differences analysis indicates that intervention schools exhibited a substantial increase in access to reading corners or both reading corners and reading materials at endline, beyond what would be anticipated based on the outcomes in comparison schools (Table 21). These results are significant and remain robust even after controlling for other variables such as school type and the presence of PTA in the school. The significant magnitude of the findings suggests that **the HATUTAN program had a meaningful impact on enhancing access to reading corners and reading materials in grade 2 classrooms.**

[99] FGDs with mothers, Ainaro, Int. 26
[100] KII with administrator, Liquica, Int. 5
[101] FGDs with fathers, Liquica, Int. 19

**Table 21: Changes in access to reading corner and reading materials**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 98 | 98 | 87[102] | 84 | | |
| Reading corner | 37.8% | 65.3% | 29.9% | 26.2% | 31.2 | <0.001* |
| Reading materials | 57.1% | 66.3% | 46.0% | 41.7% | 13.5 | 0.44 |
| Reading corner and materials | 34.7% | 58.2% | 27.8% | 22.6% | 28.7 | 0.01* |

Overall, across intervention and comparison schools, at endline, 47% of grade 2 classrooms had a reading corner while 55% had reading materials. This represents a modest increase from baseline, where only 34% of classrooms had a reading corner and 52% had reading materials.

In accordance with McGovern-Dole Custom Outcome #7, the percentage of schools with age-appropriate reading materials in classrooms was measured. However, at baseline, the survey did not assess the age-appropriateness of reading materials[103], and thus only the percentage of schools with reading materials in classrooms was measured. Table 21 shows that at endline, 66% of intervention schools and 42% of comparison schools achieved this outcome. The outcome improved for intervention schools but decreased by 4 percentage points for comparison schools from endline to baseline. These results reiterate that **the HATUTAN program had a positive impact on the availability of age-appropriate reading materials in the intervention schools, as compared to the comparison schools.**

In the household survey, caregivers were also asked about their perceptions of their children's access to reading materials at school. However, the results from this survey differ from the positive results obtained from the school survey. At baseline, 95% of caregivers reported that their children had enough books at school, while at endline, only 80% agreed with this statement, indicating a significant decrease. This decrease in reported book availability may suggest a growing awareness among parents about their child's learning environment or an increase in their engagement with the school. Nevertheless, only 1% of caregivers identified "no reading materials at school" as a challenge affecting their child's ability to learn to read. It is probable that many caregivers are not fully aware of the number of literacy materials available at their children's schools, and therefore, these results should not be considered as definitive evidence of a decline (or lack of improvement) in access to reading materials at endline, particularly when compared to the positive and more reliable findings from the school survey.

Finally, the study also collected data on schools' practice of lending story books to students to take home. According to the school survey, at endline, 84% of schools in the intervention group were lending books to students, while the corresponding figure for schools in the comparison group was only 54%. However, it was observed that only 50% of intervention schools and 77% of comparison schools provided books for students to borrow in the past week. The most common reasons cited by schools for not lending books were concerns that students may lose or damage them, and insufficient availability of books. Similarly, the household survey found that 84% of caregivers in the intervention group reported that their child's school was lending books to students at endline, while the corresponding figure for caregivers in the comparison

[102] n = 90 for the “reading corner and materials” indicator.
[103] 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.

group was 55%. Again, the most common reasons cited by caregivers for not lending books were concerns about students being careless or losing the books, and insufficient availability of books.

Qualitative data revealed several challenges regarding the use of literacy materials in and out of schools. Many respondents stated that they did not have sufficient books or magazines to lend to students, and the available books did not align with the curriculum:

> *The availability of books is limited. While there are some extra books available for students in grades one through three to borrow, the lent books are not aligned with the curriculum. Due to the shortage of books, only two to five students are allowed to borrow a book each day, and once they finish reading it, they must return it so that other students can borrow it.*
>
> - KII with administrator, Ermera, Int. 3

Many parents have also mentioned that schools do not lend books to students, but only lend the Lafaek magazine.[104] This could be the reason for the difference between the increase in access to literacy materials reported in the school survey and the significant decrease in reported lending of books among caregivers.

***Note from the program:*** The Lafaek magazines are actually provided to students to take home as their individual property. However, schools may lend magazine copies provided to their libraries.

Several other challenges to the use of literacy materials were mentioned in the qualitative data. Some respondents noted that children lose or damage books,[105] while others mentioned that students are too afraid or shy to take books home.[106] Additionally, a school administrator noted that many students forget to return books on time, which prevents others from borrowing them.[107] These findings indicate a need for improved access to age-appropriate literacy materials and related infrastructure, as well as a shift in attitudes toward these materials. Schools should not refuse to lend books or magazines to students, and efforts should be made to encourage responsible handling and timely return of borrowed materials.

## PREDICTORS OF ENGAGING TEACHING PRACTICES

We now employ a linear regression model to identify the factors that predict the use of engaging teaching practices. The model utilizes the number of engaging teaching practices observed during classroom observations, with a maximum score of nine, as the outcome variable. Predictor variables included teacher gender, education level, experience, class size, availability of reading materials, presence of electricity, presence of a PTA, and whether the director provides coaching to teachers. Table 22 presents the results of the regression analysis for baseline, midline, endline, and all data combined.

At baseline, larger class sizes were found to be negatively correlated with the use of engaging teaching practices, while access to reading materials was positively correlated with their use. At midline, the presence of Parent-Teacher Associations (PTAs) was a significant predictor of engaging teaching practices. However, at endline, the availability of reading materials was once again found to be positively correlated with engaging teaching practices, while the presence of a PTA had a significant and adverse effect on the use of such practices. Furthermore, the study found that the availability of electricity was a significant predictor of engaging teaching practices when all data were combined, with a positive coefficient.

[104] FGDs with mothers, Ermera, Int. 28; FGDs with mothers, Ainaro, Int. 25
[105] FGDS with teachers, Ermera, Int. 40; KII with administrator, Liquica, Int. 5
[106] KII with administrator, Ainaro, Int. 2
[107] KII with administrator, Manatuto, Int. 7

These findings highlight the **importance of resources, such as reading materials and electricity, in promoting effective teaching practices**.

**Table 22: Predictors of engaging teaching practices**

| | Baseline | | Midline | | Endline | | All data | |
|---|---|---|---|---|---|---|---|---|
| | Coef. | P-value | Coef. | P-value | Coef. | P-value | Coef. | P-value |
| n | 136 | | 177 | | 178 | | 491 | |
| Gender | 0.55 | 0.22 | 0.37 | 0.28 | 0.20 | 0.18 | 0.35 | 0.09 |
| Education | 0.82 | 0.06 | 0.20 | 0.46 | 0.07 | 0.65 | 0.36 | 0.10 |
| Experience | -0.02 | 0.67 | 0.02 | 0.55 | 0.02 | 0.73 | 0.01 | 0.68 |
| Class size | -0.03 | 0.03* | -0.01 | 0.50 | 0.00 | 0.61 | -0.01 | 0.13 |
| Electricity | 0.44 | 0.10 | 0.31 | 0.36 | 0.53 | 0.15 | 0.49 | 0.04* |
| Reading materials | 1.10 | 0.01* | -0.28 | 0.47 | 0.70 | 0.03* | 0.41 | 0.07 |
| PTA | -0.50 | 0.55 | 1.17 | 0.01* | -1.23 | 0.01* | -0.42 | 0.54 |
| Coaching | 0.26 | 0.28 | -0.43 | 0.12 | 0.02 | 0.95 | 0.10 | 0.65 |

To further examine the relationship between engaging teaching practices and class size, we conducted a simple regression analysis for each of the nine engaging teaching practices. Our results suggest that there is a **significant negative correlation between class size and teachers engaging students who are not participating in class**. In larger classes, it may be more challenging for teachers to effectively engage students who are quiet or not paying attention due to difficulties in managing large numbers of students and identifying which students are participating.

We also explored the relationship between the availability of reading materials and specific engaging teaching practices. Our findings suggest that access to reading materials is significantly and positively correlated with the use of a reading corner in class, highlighting the importance of resources in promoting engaging teaching practices.

Lastly, we examine the specific engaging teaching practices associated with availability of electricity in schools. Our results indicate that teachers in schools with access to electricity are more likely to implement group work, use games, and ask open-ended questions to students. It is likely that the association between engaging teaching practices and the presence of electricity in schools is due to the fact that such schools are situated in less remote areas and have better teacher quality and supervision.

# STUDENT ATTENTIVENESS

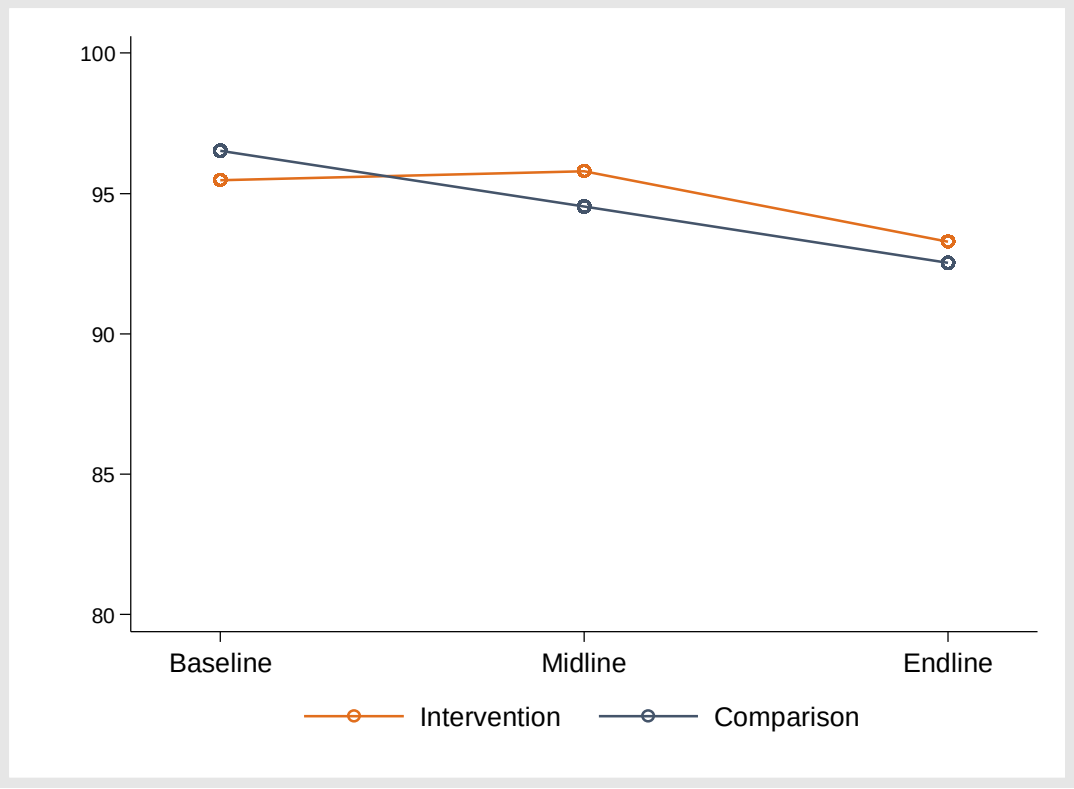
This section analyzes student attentiveness, highlighted in the log frame as a factor that may influence student literacy. We begin by analyzing changes over time in self-reported and observed student attentiveness. We then analyze working memory as a proxy measure for attentiveness, followed by a brief discussion of student hunger and its impact on attentiveness. Lastly, we conduct a predictive analysis to better understand factors which influence student attentiveness.

# ATTENTIVENESS

We first utilize two indicators to measure attentiveness: a self-reported measure, where students were asked if they felt they were able to pay attention in class, and observed student attentiveness, where enumerators observed ten students during classroom observations and reported how many are paying attention. The former of these measures is likely to overreport attentiveness, as it is subject to social desirability bias. In contrast, while the second measure is not prone to this type of bias, it represents only a snapshot of student attentiveness in one class and at one point in time. It may therefore also be reflective of engaging teaching practices or other dynamics, rather than solely a measure of attentiveness. Regardless, the two measures, combined with working memory (discussed below), provide validating information to triangulate findings.

Figure 20 shows the change in self-reported student attention across round and treatment group. We first note that self-reported attentiveness was high in all rounds. However, compared to baseline, we find a decline in reported attentiveness for both intervention and comparison students; for intervention students, this decline comes after a slight increase from baseline to midline. The decline in attentiveness was relatively more severe for comparison students; for these students, self-reported attention declined from 96.5% percent of students at baseline to 92.5% at endline, a 4.0 percentage point decrease. In contrast, for intervention students, attention declined by only 2.2 percentage points, from 95.5% at baseline to 93.3% at endline. This relative difference, however, was not significant.

**Figure 20: Change in students self-reporting that they can pay attention in class**

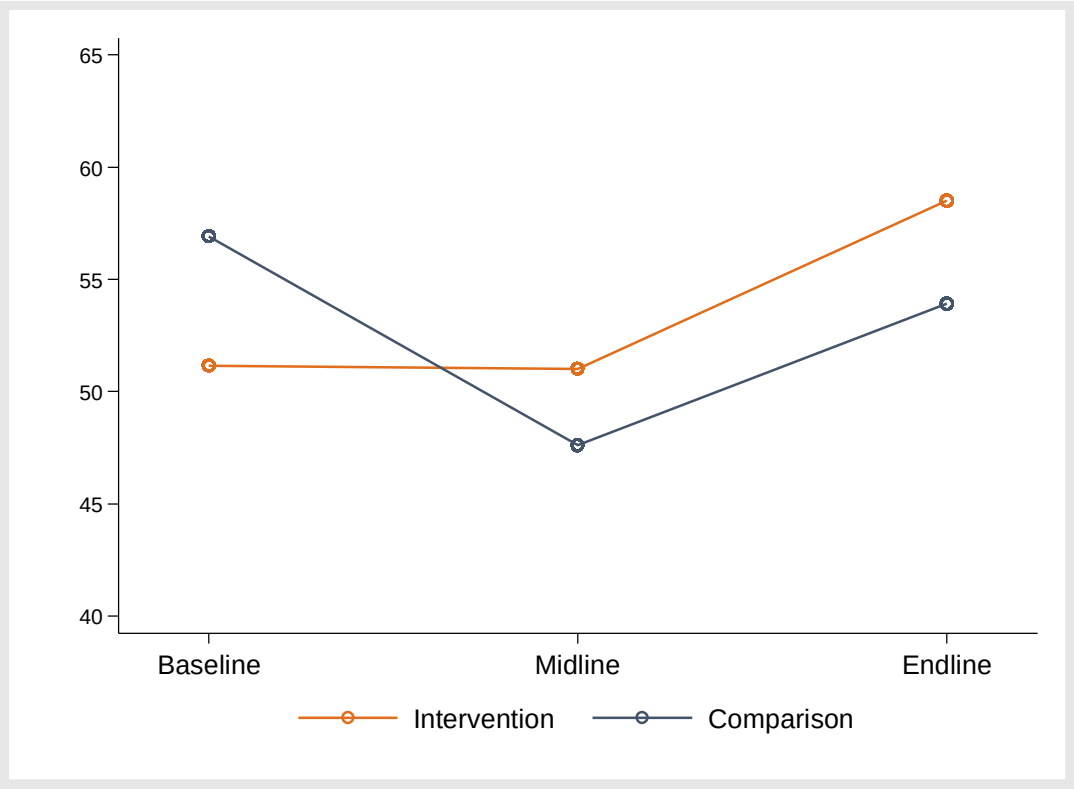



Further examining this measurement, we find that there were no significant differences in self-reported attention by student gender. Differences across municipalities were significant but not highly substantial at endline, ranging from a low of 86.6% in the comparison municipality Covalima to a high of 97.0% in the comparison municipality Manufahi.

Overall, while this measurement suggests that there may have been a slight decline in student attentiveness at endline, the impact of social desirability bias is clear as over 90% of students reported that they were able to pay attention. As such, we now analyze observed attentiveness in the below figure.

This figure shows that between baseline and endline, observed student attentiveness increased in intervention schools while slightly decreasing in comparison schools. In intervention schools, an average of 51.1% of students were observed paying attention at baseline; this increased by 7.4 percentage points at endline, to 58.5%. In comparison schools, in contrast, observed student attention declined from 56.9% at baseline to 53.9% at endline (a 3 percentage point decrease). This relative difference, with 10.4 percentage points more intervention students paying attention than expected given the results of comparison students, was significant. In other words, **HATUTAN may have had a significant positive impact on student attentiveness** as measured through classroom observations.

**Figure 21: Change in students observed paying attention in class**



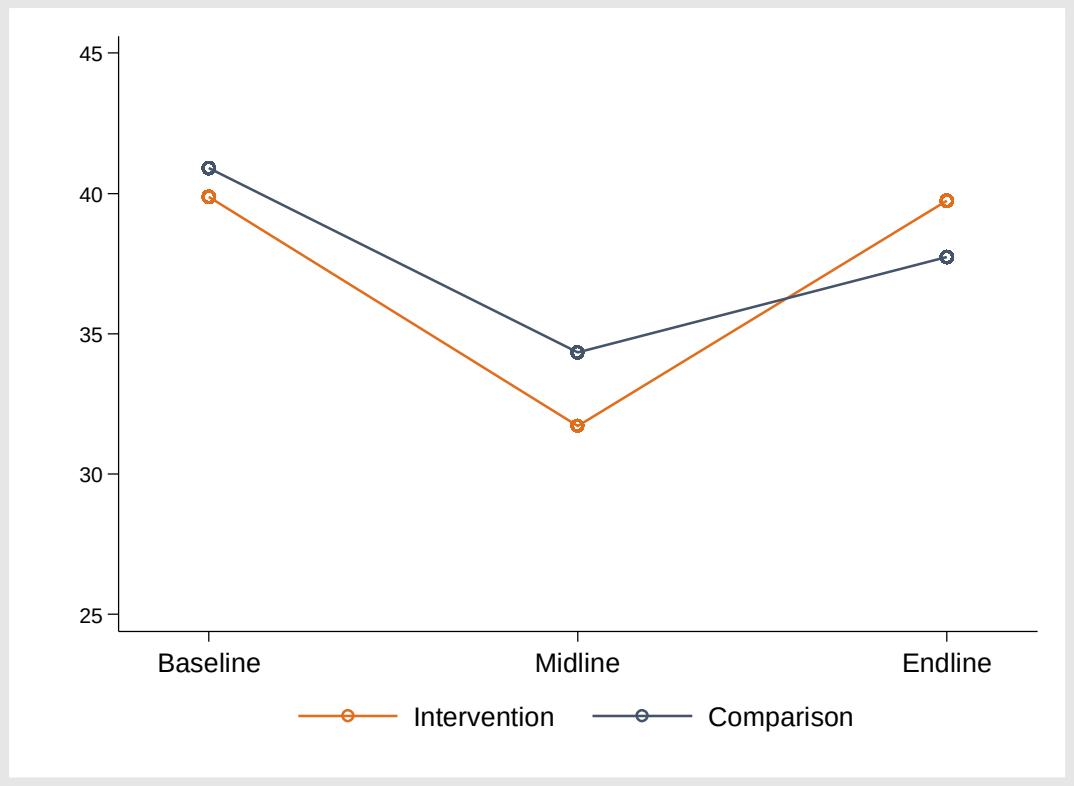

In addition to this positive finding, we note that observed student attentiveness fell substantially below self-reported attentiveness, as expected due to social desirability. In order to further validate these measurements, we now discuss working memory.

## WORKING MEMORY

Working memory was assessed using a pictorial test; students were shown and read the names of 19 images, and shortly thereafter were asked to name as many images as they could remember without looking at the pictures. Scores were then calculated as a percentage out of 100. Working memory score is used as a proxy measure as it is thought to depend on student attentiveness and is a more objective and direct measure than observed or self-reported attentiveness. However, we note that working memory may also be influenced by non-attentiveness-related factors, such as short-term memory. As such, this measure should be considered to provide further validating information to the analysis above.

Figure 22 shows the change in working memory score across rounds and treatment groups. For both intervention and comparison students, we find a decline in working memory from baseline to midline, followed by an increase at endline. Both the decline and increase were relatively more substantial for intervention students.

**Figure 22: Change in working memory score**




As working memory score may be influenced by other factors discussed above, we include several control variables in our difference-in-difference analysis, including students' overall EGRA scores,[108] gender, age, and whether the caregiver reported that the student has difficulty with memory.[109] The below table shows the results of the analysis with and without controls.

We find that while working memory scores improved more for intervention students than for comparison students from baseline to endline, this difference was not significant. Table 23 shows that this difference remains insignificant when controlling for various confounding factors; indeed, the effect size decreases and p-values increase when controlling for overall EGRA score and whether the student has trouble with memory. This may suggest that in addition to serving as a proxy for attentiveness, working memory scores are also capturing other dimensions which can affect student performance, such whether the student has a learning or physical disability.

[108] We note the relationship between working memory and EGRA scores may operate in both directions. In other words, low EGRA scores may be related, in part, to intelligence, which may also affect working memory. However, students' ability to remember—and thus perform well on the working memory test—may also affect their performance in school. Regardless of this directional issue, including EGRA scores in some regression specifications allows us to control, at least in part, for inherent student capacity factors unrelated to attentiveness that may still affect working memory.

[109] We note that this may, at least in part, also be a measure of student attentiveness. However, it may also capture student disabilities which affect memory and concentration, such as Attention Deficit Disorder or other related disabilities, and is thus useful to include in some regression specifications.

**Table 23: Difference-in-differences analysis of working memory**

| Regression | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| No controls | 39.9% | 31.7% | 39.7% | 40.9% | 34.3% | 37.7% | 3.0 | 0.19 |
| Controlling for gender and age | - | - | - | - | - | - | 3.0 | 0.19 |
| Controlling for EGRA score, gender, age | - | - | - | - | - | - | 1.4 | 0.52 |
| Controlling for EGRA score, gender, age, memory disability | - | - | - | - | - | - | 0.2 | 0.94 |

Overall, this analysis suggests that **the impact of HATUTAN on student attentiveness is inconclusive, but may have been positive** given the results for observed student attentiveness discussed above.

Despite these potential positive findings, in qualitative interviews, several respondents emphasized that attentiveness was a challenge. While few respondents provided a reason for this challenge, in Ermera, a father stated that large class sizes were detrimental to attentiveness, especially for students who had difficulty concentrating:

> *Sometimes in the classroom there are so many students, not in accordance with the [recommended] standard… Therefore, disorganization occurs. So, their attention to the teacher isn't quite there… [students] who are listless sometimes can't cope.*
>
> - FGD with fathers, Ermera, Int. 16

Teachers described a variety of practices used to improve student attentiveness during class, as discussed more in the above section *Quality of Instruction.* Positive practices included entertaining or amusing students, encouraging students to follow the example of their attentive peers, playing games, and moving around the classroom.[110] However, teachers also described the use of negative and abusive behaviors, including hitting children or yelling at them. In some cases, teachers also did not seem to understand why these behaviors were negative. A teacher in Ainaro, for example, stated that prohibition of corporal punishment was detrimental: "Because hitting, pinching ears, and punishing are prohibited, children can do whatever they want. We make them pay attention by shouting without touching them."[111] As discussed above, this emphasizes that teachers may believe that corporal punishment and other negative behaviors have a positive effect on attentiveness, which may drive continued use of these behaviors.

In addition to this analysis, the percent of students with a working memory score above 50% was identified as a key indicator to track for program performance. At endline, we find that 30.1% of intervention students achieved a working memory score above 50%. This is a substantial increase from midline, when only 14.4% of grade 2 intervention students achieved this standard, and a slight increase from the baseline value of 29.0%. However, this value falls slightly below the target of 35% of students.

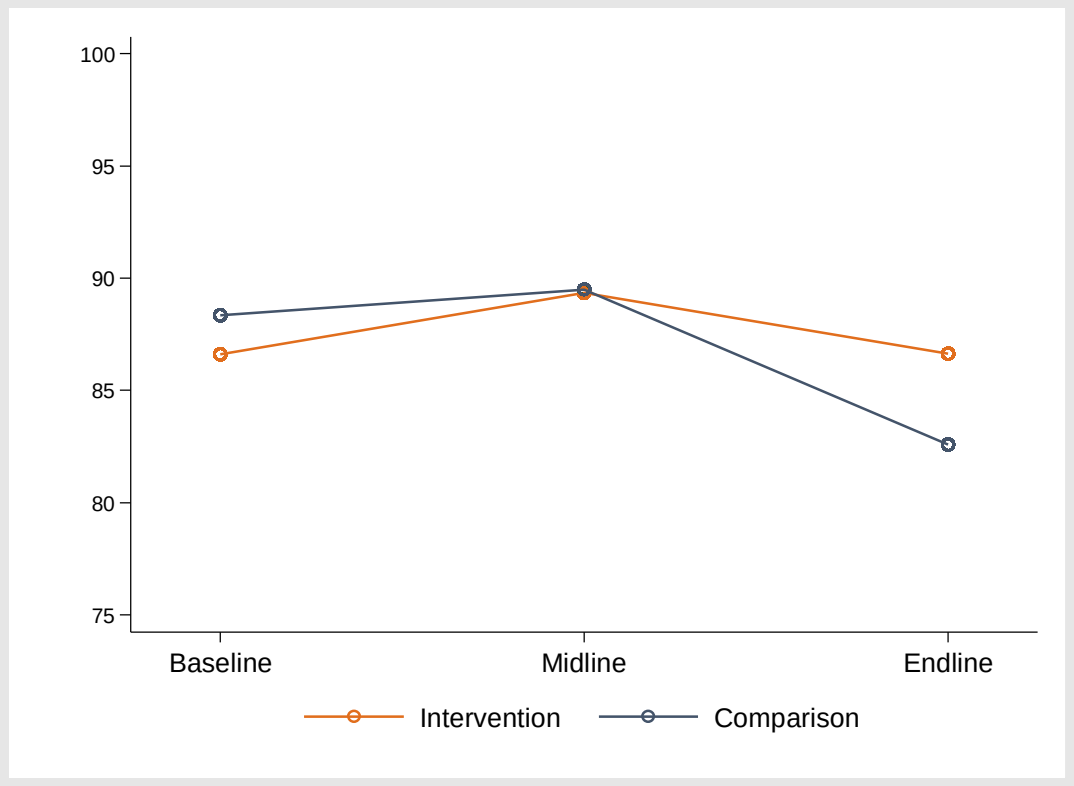
[110] E.g., FGD with teachers, Ainaro, Int. 37; FGD with teachers, Ermera, Int. 39
[111] FGD with teachers, Ainaro, Int. 37

## STUDENT HUNGER

We now examine student hunger, a factor expected to affect student attentiveness and of direct relevance to HATUTAN program activities. Indeed, in an interview, a school administrator stated that during the previous year he saw that "school meals can improve the children's ability, intellect, and attention to the lessons they are given."[112] We note that detailed analysis of foods eaten by students is included in the section *Health and Nutrition*. In this section, we only analyze the percent of students who reported eating on the day of the EGRA and the percent of households that report food scarcity.

Figure 23 shows the percent of students who said they had eaten the day they were assessed with the EGRA. We find a high percentage of students—around 88% on average—stated that they had eaten something that day (McGovern-Dole Custom Outcome #13). For both intervention and comparison students, we find that the frequency with which students reported eating increased slightly at midline, before falling at endline. This may in part reflect the delay in implementation of school feeding programs at endline, discussed more below and in the *School Feeding Program* section.

**Figure 23: Change in percent of students who ate the day of the EGRA**




Looking at the relative change in intervention and comparison areas, we find a more substantial decline in students eating among comparison students than intervention students. Among comparison students, the percent of students who had eaten declined by 5.8 percentage points from baseline to endline, from 88.3% to 82.5%. In contrast, only 0.1 percentage points fewer intervention students reported having eaten at endline compared to baseline. This relative difference, whereby intervention students reported having eaten at a significantly higher rate than expected given comparison students, was significant. In other words, **the HATUTAN program may have had a significant impact on reducing student hunger at endline.**

[112] KII with administrator, Ermera, Int. 3

To further understand the potential impact of HATUTAN on student hunger, we test whether there was a link between whether a student ate on the day of the EGRA and whether the school was providing meals. At midline, we found a statistically significant, positive relationship between school meal provision and whether students had eaten that day. At endline, we still find a positive relationship—students in schools that served meals were 4.3 percentage points more likely to report having eaten—but this relationship is no longer significant. This may be due to the lower prevalence of school feeding at endline than at midline; at endline, only 10.4% of schools had provided meals to student on the day of data collection, compared to 60.1% of schools at midline. This dynamic is discussed more in the section *School Feeding Program*.

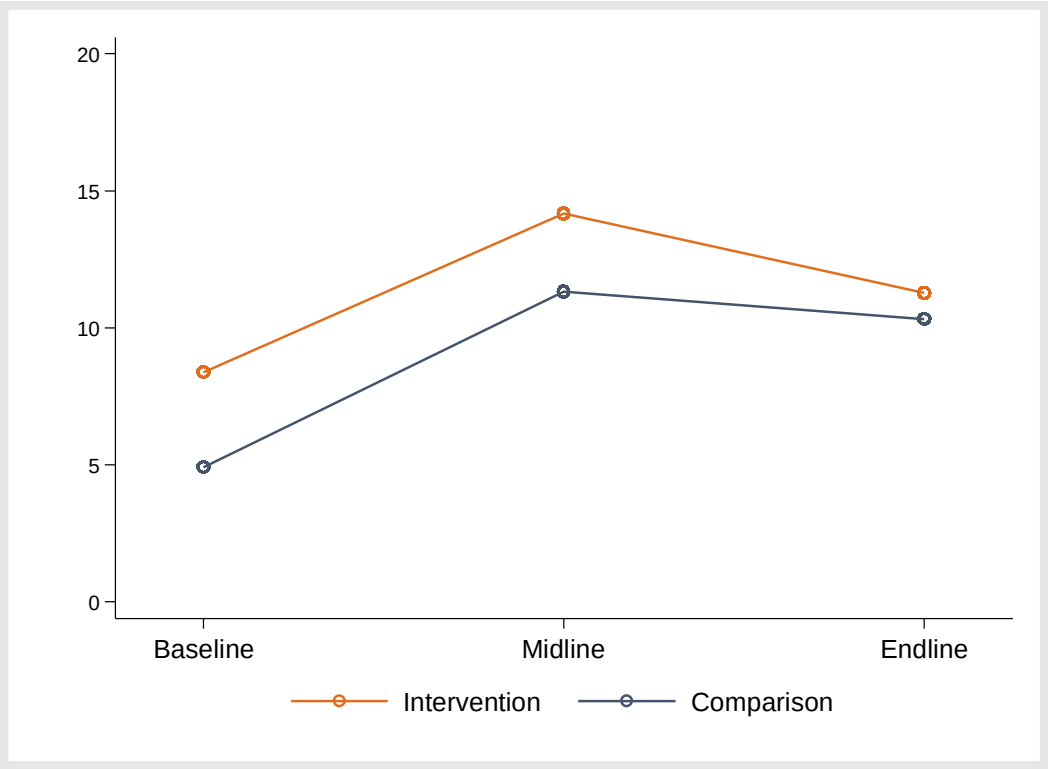
Additionally, we find no significant or substantial differences by gender in the frequency with which students reported eating; in other words, male and female students reported having eaten on the day of the EGRA at very similar rates. We do, however, find differences in the rate at which students ate across municipalities, although with no substantial differences in baseline-to-endline change by municipality. The below table shows municipality-specific results for intervention students; we find that students in Liquica were significantly more likely to report having eaten than students in Manatuto, although results in Manatuto had begun to catch up to those in other municipalities—albeit slowly—at endline.

**Table 24: Percent of students who ate the day of the EGRA, by intervention municipality**

| Municipality | BL | ML | EL | Difference (BL to EL) |
|---|---|---|---|---|
| Ainaro | 86.0% | 90.9% | 86.8% | 0.8 |
| Ermera | 87.5% | 88.3% | 87.1% | -0.4 |
| Liquica | 93.2% | 92.6% | 91.6% | -1.6 |
| Manatuto | 80.7% | 87.4% | 82.0% | 1.3 |

We now examine household-level indicators of student hunger. In particular, within the household survey, respondents were asked whether there was a time during the past 30 days when they or others in their household had gone without food. Figure 24 shows the change over time in this indicator. We find that household food insecurity increased at midline, in concurrence with the COVID-19 pandemic which increased food prices in Timor-Leste. Food insecurity has since declined, but remains higher at endline than at baseline.

**Figure 24: Change in households going without food during 30-day period**




Food insecurity increased at a similar rate for intervention and comparison households from baseline to midline, but declined at a greater rate for intervention households between midline and endline than for comparison households. As a result, although 3.5 percentage points more intervention than comparison households stated that they had gone without food within the past 30 days at baseline, at endline, rates were similar across intervention and comparison groups, at 11.3% and 10.3% respectively. This relative improvement among intervention households, however, was not significant in a difference-in-differences analysis.

Overall, these results suggest that student hunger may have an impact on attentiveness, but that the majority of students eat on school days and do not come from acutely food insecure households. We note, however, that this analysis does not include the quality of the diet eaten by students; even if students eat during school days, if the food eaten is very low quality (or not very caloric), students may rapidly feel hungry again afterwards. This dynamic is explored more in *Health and Nutrition.*

# PREDICTORS OF STUDENT ATTENTIVENESS

This section analyzes the relationship between various student-, household-, and school-level characteristics and student attentiveness. As the above analysis suggests that self-reported attentiveness is overestimated due to social desirability bias, we focus on predictors of working memory. We also utilize observed attentiveness in order to better understand predictors at the school level. Our analysis focuses mainly on the links between student hunger, teaching quality, and attentiveness as main areas of interest for HATUTAN.

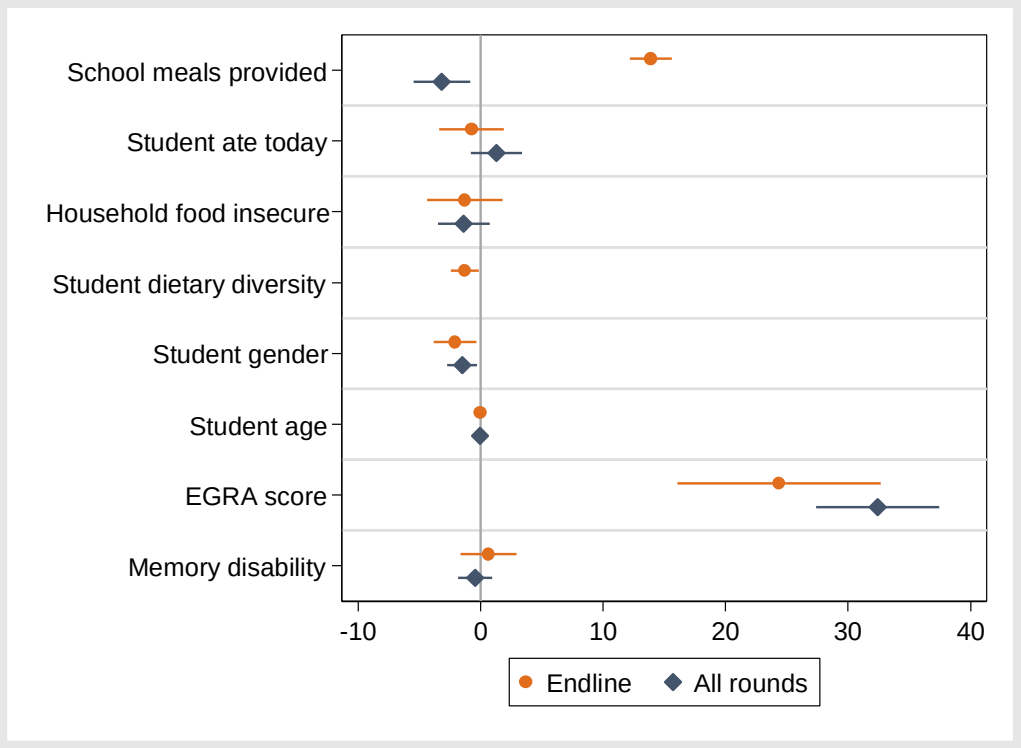
## Predictors of Working Memory

In the above analysis, we hypothesized that working memory may be affected by student learning outcomes as measured by overall EGRA scores, gender, age, and whether the caregiver reported that the student

has difficulty with memory. As such, we include these four predictors in our analysis. We also include indicators of student hunger, including whether the school served meals the day of the EGRA, whether the student had eaten on the day of the EGRA, whether the student's household experiences food insecurity, and the quality of the student's diet.[113] We also include school fixed effects to control for variables which vary at the school level.

Figure 25 shows predictors of working memory. At endline, we find a significant and positive correlation between whether the school provided meals and working memory; students' working memory scores were 13.9 percentage points higher in schools that provided meals at endline, all else held constant. However, when we expand the data to include all evaluation rounds, the relationship becomes negative and remains significant. We can only speculate about the reason for this; however, we note that school meals were provided far more frequently at midline than at baseline or endline. As such, this variable may in part serve as a proxy for evaluation round, and reflect the fact that working memory scores were lower at midline than at baseline or endline, likely in part due to the effect of COVID-19.

**Figure 25: Predictors of working memory**




Only one other indicator of student hunger is significant: We find a negative relationship between student dietary diversity and working memory scores at endline. The reason for this relationship is again unclear, as we would generally expect students with greater dietary quality to be more attentive due to the benefits of a healthy diet. We note, however, that there was generally little variance in dietary diversity for students, with 74% of students consuming one, two, or three food groups. Dietary quality was also generally low, even among those students who consumed multiple food groups; almost all students consumed grains, roots, and tubers and processed or sugary foods, while only 36% consumed any source of protein. As such, student dietary diversity scores may not accurately reflect actual dietary quality and its impact on

[113] i.e., student dietary diversity score, discussed more in *Health and Nutrition.* This variable is only included in the endline-specific analysis as data on students' diets was collected for the first time at endline.

attentiveness. Indeed, if we replace this variable with a variable for whether or not the student consumed protein, we find that the relationship with working memory remains negative but is no longer significant.

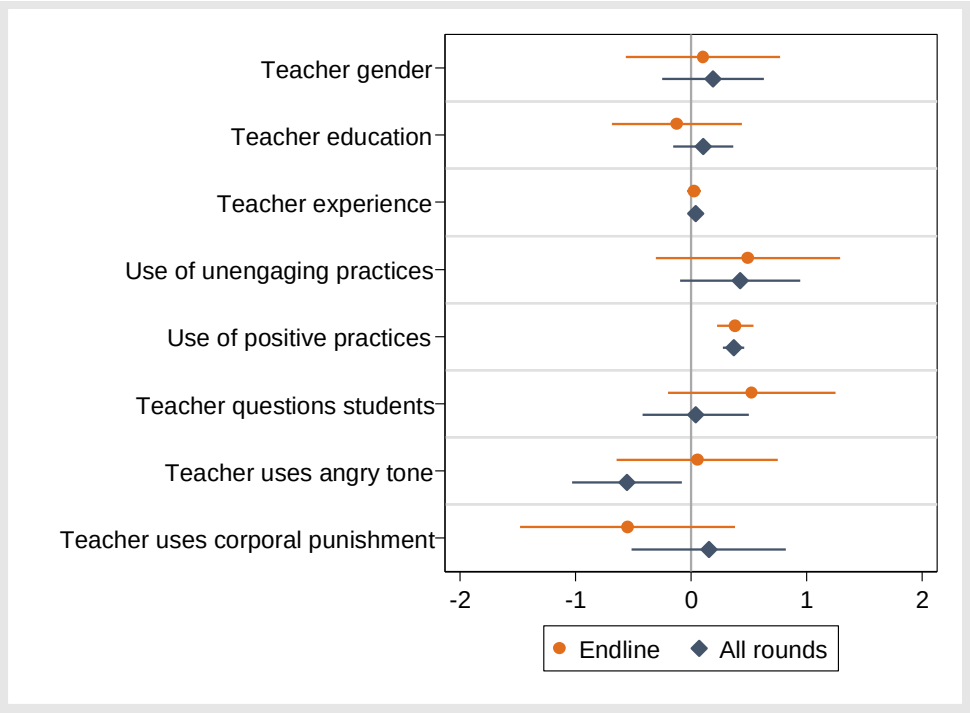
Outside of these markers of student hunger, we find that male students have significantly lower working memory scores, in line with gendered learning results in Timor-Leste. Additionally, we find a strong and significant positive relationship between EGRA scores and working memory. This result is unsurprising; we would expect students with better memory and better attentiveness to, on average, have better learning outcomes.

## Predictors of Observed Attentiveness

We now turn to classroom-level predictors of observed attentiveness. Figure 26 shows results of the analysis including eight predictive variables: teacher gender, education, and experience; the use of unengaging teaching practices, including copying from the board and repeating after the teacher; the number of positive teaching practices used;[114] whether the teacher asked questions to students; whether the teacher used an angry tone with students; and whether the teacher used corporal punishment.

We find only two significant predictors. First, the number of positive practices used is significantly correlated with higher student attentiveness. This finding reiterates the importance of the quality of instruction, discussed in the previous section.

**Figure 26: Predictors of observed student attentiveness**




Second, when data from all evaluation rounds is analyzed, we find a significant and negative relationship between teachers' use of an angry tone towards students and observed attentiveness: **Use of an angry tone was associated with a 5.5 percentage point decrease in the number of students observed paying attention**, all else held constant. While teachers may perceive the use of angry tones as necessary to ensure that students are listening in class, it appears the opposite may actually be the case, with harsh

[114] Including group or solo reading activities, the teacher reading aloud, the teaching using the reading corner or Lafaek magazine, the teacher using participatory games or exercises, the teacher asking open questions or for students' opinion, the teacher engaging students who are not participating, and group work.

tones serving to disengage students from class. Informing teachers of this relationship may help discourage the use of this negative practice, which may not only decrease attentiveness but may also harm students' mental wellbeing and feeling of safety in schools.

# 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. Finally, factors that may affect attendance are analyzed, from both individual and school perspectives.

Student attendance is highlighted in the results framework as a factor that may affect literacy scores. The program is expected to influence student attendance in 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

The overall attendance rate for grade 2 is reported in the table below. 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. Statistical significance for these findings is tested and presented in the table below, using a difference-in-differences regression model.

**Table 25: Change in average attendance rates for grade 2**

| | Intervention Schools | | Comparison Schools | | | |
|---|---|---|---|---|---|---|
| Grade 2 | BL | EL | BL | EL | DiD | P-value |
| n | 69 | 78 | 57 | 72 | | |
| Attendance rate | 67.4% | 62.1% | 63.4% | 71.9% | -13.8 | 0.02* |

**As shown in the table above, average attendance rate for grade 2 intervention students decreased by 5 percentage points since baseline, while the scores for comparison students increased by 8.5 percentage points.** Using the difference-in-differences regression model, we found a significant and negative effect for the intervention group relative to the comparison group. In other words, there is a significant difference of 14 percentage points difference in attendance rates between intervention and comparison groups from baseline to endline.

In order to gain a more in-depth perspective about attendance rates in intervention schools, we disaggregate attendance by gender among grade 2 students. Results are presented in the table below. **Average attendance decreased for both boys and girls with no statistically significant difference,** meaning that the drop in overall attendance is a widespread issue, without a gender group being more absent than the other. This may suggest that attendance decreases are driven more by school- or community-level factors, rather than home labor or other similar factors which may affect girls and boys unequally.

**Table 26: Average attendance by gender**

| Female | Male | Difference | P-value |
|---|---|---|---|
| **Grade 2 – Endline** | | | |
| 63.7% (n = 81) | 62.6% (n = 81) | 1.1 | 0.89 |
| **Grade 2 – Baseline** | | | |
| 66.3% (n = 73) | 68.1% (n = 72) | 1.8 | 0.44 |

A 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 as those that responded may not be representative of the whole sample. We must be cautious in attributing the results to program impact, and this finding must be analyzed while thoroughly considering contextual factors. Recent extreme weather conditions, and consequent increased difficulty for students to reach school, might be one of the key factors for decreased attendance. Socio-economic reasons might also be an important part of the decrease, with households needing an "extra hand" for chores or seasonal work. To shed more light on what is actually happening in those schools, we must not only look to attendance rates themselves, but to the wider related phenomena in the schools. As such, we also investigate dropout rates, cultural and economic incentives, and reasons for missing school, with the intent of finding patterns and highlighting possible issues to explain attendance variability.

To further investigate possible patterns of increase/decrease, we disaggregate attendance rates by municipality. In the intervention areas**, all municipalities experienced some degree of attendance decrease**. In Ainaro and Liquica the difference was trifling, at 2.3 and 1.3 percentage points respectively. The drop appears to be more substantial in Ermera (7 percentage points) and striking in **Manatuto, with a drop of 15.9 percentage points**. It is worth noting that a decrease in attendance in Manatuto has been registered between the baseline and 2021 midline, with an average rate of 78.2 percent at midline. It appears now at endline that the decrease in the area has probably not been the result of some sudden, particular event, but instead there is a steady pattern of decrease since baseline.

> ***Note from the program:*** Inquiries made to local staff and authorities in Manatuto suggest that there was an outbreak of eye infections ("pink eye") at the time of data collection. This information cannot, however, be consubstantiated by official records or program data and thus remains anecdotal. It may however offer an explanation to the increased absenteeism rate.

In contrast, among comparison municipalities, we note that there is an increase in average attendance for three out of four municipalities from baseline to endline (of minimum 5.6 percentage points and maximum of 18.3). However, **Covalima experienced a drop of 6 percentage points, while still remaining the municipality with the highest average attendance rate (80.3%).** Another fact worth noting is the sharp **increase in attendance in Manufahi**. While it registered the lowest rates at baseline (55.6%), at endline it had the second-highest attendance of all comparison municipalities (73.9%).

**Table 27: Attendance rates by municipality**

| | BL | EL |
|---|---|---|
| **Intervention Municipality** | | |
| Ainaro | 59.8% | 57.5% |
| Ermera | 64.4% | 57.4% |
| Liquica | 76.2% | 74.9% |
| Manatuto | 84.3% | 68.4% |
| **Comparison Municipality** | | |
| Aileu | 63.3% | 68.9% |
| Bobonaro | 59.1% | 69.7% |
| Covalima | 86.7% | 80.3% |
| Manufahi | 55.6% | 73.9% |

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 percentage of schools meeting this target was 38.2%, with 37% of comparison schools and 39.3% of intervention schools meeting this target. At endline, the overall percentage of schools meeting this target was 37%, with comparison schools at 47% and intervention schools decreasing to 27%. The quite stable overall percentage of schools reaching the target is the result of the increase in attendance in comparison schools, with the intervention school experiencing a substantial decline.

Reported reasons for missing school are presented in the table below. We first note the sharp decrease in "sickness" for both comparison and intervention groups. The decrease is more substantial in the comparison group (a difference of 26 percentage points) but it is also considerable among intervention students (18.7 points).

**Table 28: Change in reasons for missing school**

| | Intervention | | Comparison | |
|---|---|---|---|---|
| | BL | EL | BL | EL |
| n | 227 | 248 | 121 | 248 |
| Sick | 49.8% | 31.1% | 57.0% | 31.1% |
| Natural disaster | 11.5% | 22.9% | 4.1% | 22.9% |
| Did not want to go to school | 7.9% | 14.5% | 18.2% | 14.5% |
| Funeral, marriage, traditional ritual | 3.9% | 4.8% | 2.5% | 4.8% |
| Teacher did not attend | 3.5% | 4.8% | 0.8% | 4.8% |
| Other | 16.8% | 0.4% | 13.2% | 0.4% |

In contrast, the percentages of respondents reporting "natural disaster" as a reason to miss school increased noticeably both in treatment and intervention areas. While among comparison respondents it rose of 18.8 percentage points, among treatment students the increase is of 11.4 percentage points. A

sharp increase of the same reason was also registered at midline for both groups, due to heavy rains hitting the country when midline data was collected. At endline we see that the increase continues, **making natural disaster the second most reported reason for absence in the intervention group.** Many treatment schools are in more remote areas where students must cross rivers or walk across washed-out roads to reach schools. The increased occurrence of extreme weather phenomena in the country might be a key factor to explain, at least partly, the decrease in attendance rates in intervention schools.

The qualitative data provides further insight into reasons for student absence. In interviews, administrators, teachers, and parents frequently stated that children might not attend school due to inclement weather, especially during the rainy season when roads washed out and rivers flooded;[115] due to a lack of student desire to attend school and insufficient parental encouragement or oversight to ensure attendance;[116] and due to illness.[117] Another frequently cited reason was cultural ceremonies and celebrations; a teacher in Ainaro, for example, stated that "parents often take their children with them when they attend cultural events and ceremonies, which prevents their children from attending school."[118]

Regarding parental oversight, many respondents emphasized that parents' effective communication with schools and teachers was also vital to ensure student attendance, whether informally or through PTAs. A teacher in Ainaro, for example, stated that:

> *When a teacher does not tell parents about the condition of their students and vice versa, both parties will not know the condition of their students, because parents are at home and a teacher is at school every day.*
>
> - FGD with teachers, Ainaro, Int. 38

Other respondents stated that ensuring student attendance was the responsibility of the PTA.[119] Additionally, teachers were perceived by many to have an important role in ensuring student attendance and tracking down absent students. A story told by a teacher in Liquica provides a strong example of the important role teachers can play in this regard, as well as the obstacles faced by students:

> *Last year, I had a student who frequently missed class. I located her and learned that she is caring for her sick grandmother; if she attends school on a regular basis, her grandma would be left alone at home with no one to care after her. When I asked about her parents, I was told that her father had died and that her mother worked in Dili. She indicated that if she goes to school on a regular basis her grandmother would be left alone at home, and she is frightened that she will go hungry and no one will help feed her, and as a result she may die… When I asked if she had any other siblings at home who could help her with the chore, she said she's the only adult in the family, and her siblings are still little.*
>
> - FGD with teachers, Liquica, Int. 41

Outside of these dynamics influencing attendance, several other factors were mentioned less frequently by interviewees. These included distance, lack of transportation,[120] and work or housework responsibilities. A

[115] E.g., KII with administrator, Manatuto, Int. 7; FGD with fathers, Ermera, Int. 16; FGD with fathers, Manatuto, Int. 17; FGD with fathers, Liquica, Int. 18; FGD with teachers, Liquica, Int. 42

[116] E.g., KII with administrator, Ainaro, Int. 1; KII with administrator, Ermera, Int. 3; KII with administrator, Int. 6, Liquica; FGD with fathers, Ermera, Int. 15; FGD with mothers, Manatuto, Int. 32

[117] E.g., FGD with fathers, Ainaro, Int. 13; FGD with fathers, Manatuto, Int. 20; FGD with mothers, Liquica, Int. 29

[118] FGD with teachers, Ainaro, Int. 37. This sentiment was echoed in interviews including FGD with teachers, Ermera, Int. 39; FGD with teachers, Liquica, Int. 42; and FGD with teachers, Manatuto, Int. 43.

[119] E.g., FGD with fathers, Ermera, Int. 15; FGD with teachers, Liquica, Int. 41

[120] KII with administrator, Ermera, Int. 3; KII with administrator, Mantuto, Int. 7; FGD with fathers, Liquica, Int. 18

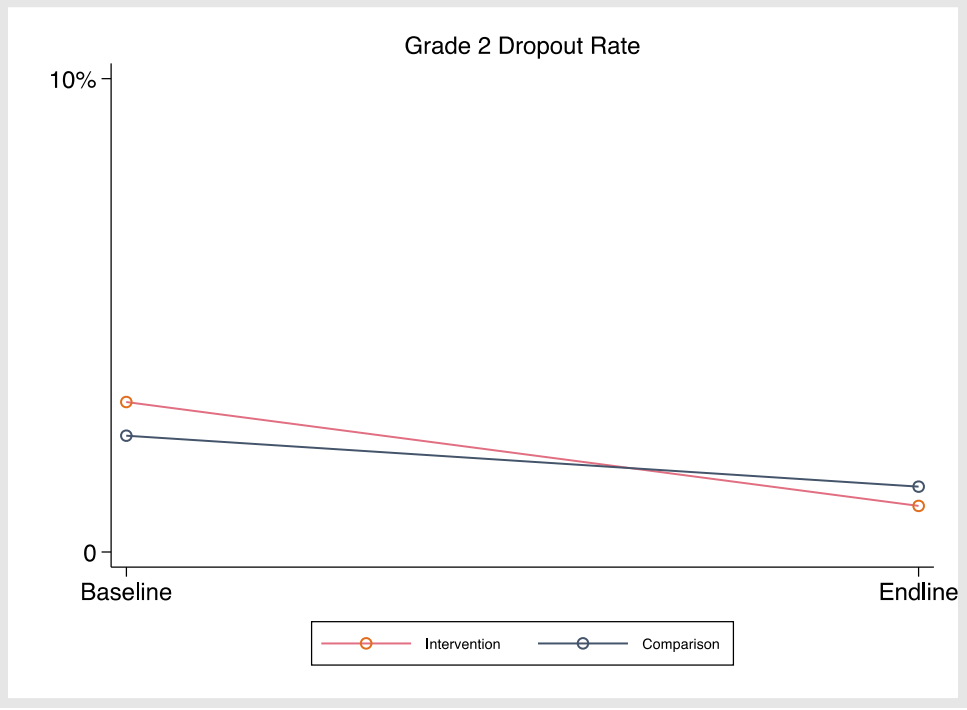
school administrator in Ainaro and teacher in Ermera gave pertinent examples of the latter factor: The administrator stated that "most of the parents here are farmers, [and] most of them go to the rice field and usually bring their children to the field to help them harvest,"[121] while the teacher stated that "when parents are busy with their work, they automatically give jobs to older brothers or sisters to take care of their younger siblings."[122]

Of particular relevance for HATUTAN, several respondents also mentioned that school meals acted as a key pull factor for students to attend school. For example, a teacher in Manatuto stated that, "When school meals are provided, practically all students attend; however, when they are discontinued, a portion of them no longer attend or are regularly absent."[123] Similarly, a father in Ermera stated that, "Sometimes [the children's] objective is to come depending on the availability of food or school feeding program."[124] These sentiments were echoed by respondents in Ainaro and Liquica.

## DROPOUT RATES

Next, dropout rates are analyzed. This is calculated as the number of dropouts (as recorded by the school) divided by the number of students still enrolled in each class as of the day data collectors arrived. Again, some sample sizes are small, especially 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 figure below: **The average dropout rates decreased for both comparison and intervention grade 2 students.** The difference was more substantial, though not significantly so, for intervention students; intervention schools at baseline reported higher dropout rates than comparison ones, while at endline the two average scores are almost the same.

**Figure 27: Change in dropout rates, grade 2 students**




[121] KII with administrator, Ainaro, Int. 1
[122] FGD with teachers, Ermera, Int. 39
[123] FGD with teachers, Manatuto, Int. 44
[124] FGD with fathers, Ermera, Int. 16

As previously done for the attendance rates, we disaggregate results by municipality of intervention schools. This is done to individuate geographical patterns that might help explain the findings. As can be seen in the table below, there are mixed results.

**Table 29: Dropout rates disaggregated by municipality**

| | BL | EL |
|---|---|---|
| **Intervention Municipality** | | |
| Ainaro | 3.9% | 2.6% |
| Ermera | 6.9% | 2.2% |
| Liquica | 0.0% | 0.0% |
| Manatuto | 0.1% | 0.7% |

In Ermera, average dropout scores decreased substantially, with a difference of 4.7 percentage points from baseline to endline. This result is interesting considering how the attendance rates instead significantly decreased from baseline to endline. Those two findings, when considered together, might be the sign of two things. First, it is possible that in Ermera all children are having more difficulty accessing schools. This might imply that the lowered dropout rates are not the sign of an improved system, but the result of less children even registering at school, consequentially reducing the average registered dropout rates. Second, it might be possible that, while many children at baseline dropped out completely for different reasons, there has been an increased sensitivity in the community about the importance of sending children to school. This, in turn, could have made parents request that their children do not drop out, and remain enrolled, and just do not go to school for longer period of time, hence the decreased attendance rates. Alternatively, if schools do not consistently remove students who have stopped attending from the register, this may also explain these results.

Given these results, it is also worth further investigating reported reasons of missing school in Ermera. A striking 58.2% of respondents said the children do not go to school because of some form of extreme weather conditions, against the 13.5% at baseline. Ermera seems to be strongly hit by floods, landslides, strong winds, and strong rains, with resultant road washouts; the consequences seem to still have a strong effect on overall school attendance, and possibly even accessibility. Indeed, in an FGD, one father emphasized the impact of rains on school attendance:

> *[Students don't come to school] because the rainy conditions don't let them... because the necessities that students need to use to attend school aren't there, such as umbrellas, such things as to cover their heads… As well, some are across on the other side of the river. The bridge, the water moves across and the water current is strong… this is the reason they're scared of the river and the rains.*
>
> - FGD with fathers, Ermera, Int. 16

In Manatuto, dropout rates have increased by 0.6 percentage points from baseline, which is a very small change. Nevertheless, this result, coupled with the drastic drop of attendance rates, might indicate a need for particular attention to the schools in Manatuto. The main reason for skipping school in the area remains sickness, at 55.2%.

Gender is a cross-cutting factor that cannot be overlooked in any context. For this reason, the table below presents average dropout scores for both girls and boys in intervention schools. Dropout rates from baseline to endline decreased for both groups, with a difference of 2.6 percentage points for both. **Given that, it is**

**evident how there is a pattern for boys, who, on average, drop out more than girls**. This is consistent in both baseline and endline.

**Table 30: Dropout rates disaggregated by gender**

| Female | Male | Difference | p |
|---|---|---|---|
| **Grade 2 – Endline** | | | |
| 0.9% (n=94) | 1.8% (n=94) | 0.9 | 0.07 |
| **Grade 2 – Baseline** | | | |
| 3.5% (n=73) | 4.4% (n=73) | 0.9 | 0.18 |

In addition to these dynamics, in the qualitative interviews, several respondents mentioned that students often dropped out due to economic difficulties. A father in Liquica, for example, mentioned the following:

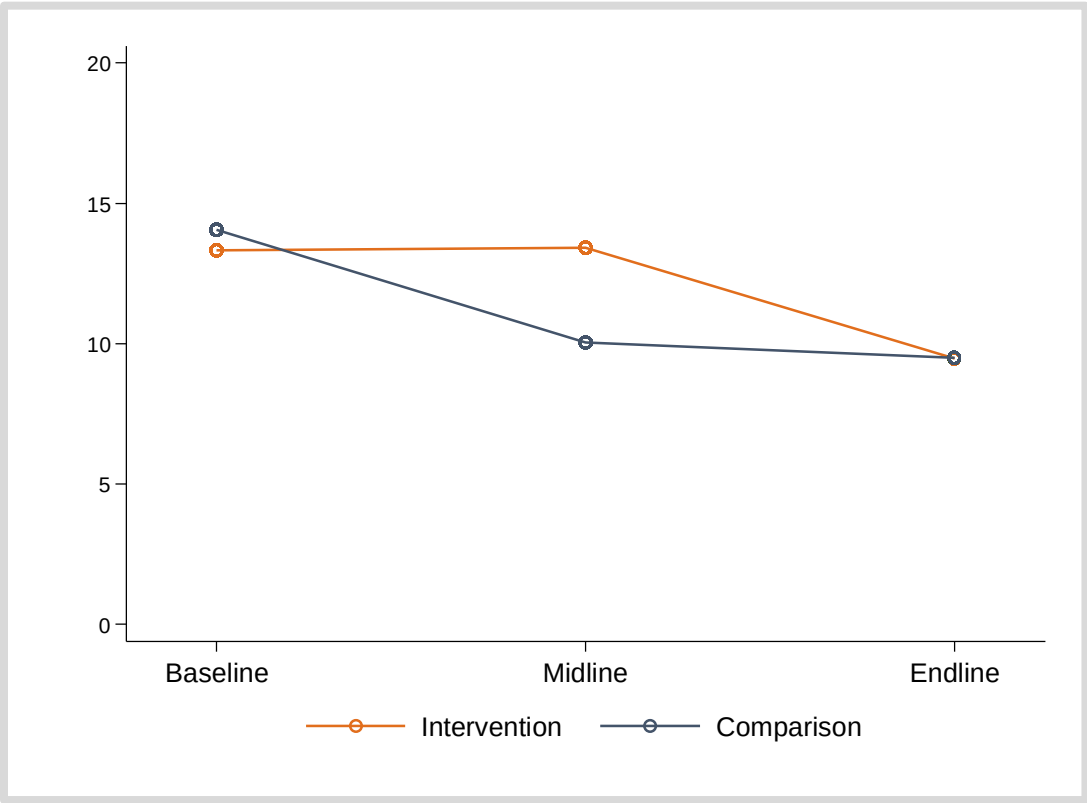
> *For the students, if they did not come, that might happen due to lack of economic capability of parents. The children might have no shoes and clothes or school uniform, so as result they don't want to come to school.*
>
> - FGD with fathers, Liquica, Int. 18

This sentiment was echoed by respondents in Ermera and Manatuto.[125]

In addition to this analysis, we also examine student dropout and re-enrollment as reported by caregivers, who were asked whether their grade 2 child had ever dropped out of school before later re-enrolling. The below figure shows the change in these rates across round and treatment group. First, we note that rates reported by caregivers are substantially higher than those discussed above from school records. This likely reflects several dynamics, including undercounting of dropout due to timing of data collection and potential inaccuracies in school record-keeping.

[125] FGD with fathers, Ermera, Int. 16; FGD with teachers, Manatuto, Int. 44

**Figure 28: Change in caregiver-reported student dropout and re-enrollment rates**




Second, the figure shows that caregiver-reported dropout rates have decreased from baseline to endline at relatively uniform rates for both intervention and comparison students. Dropout rates were around 14% at baseline, and have since dropped to around 9.5%. Therefore, while dropout rates appear to be higher than suggested through the analysis of school records above, the patterns found above seem to hold constant, with a decline in dropout at endline.

## 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 – at baseline the question referred to absences in the past month, while at midline the question referred to absences in the past week. At endline, the "weekly" wording of the question was retained from midline. As such, for comparability's sake, we present and compare midline and endline scores**. At endline, sickness is reported as a reason of absence by around 31% of the sample, both in comparison and treatment groups, which is a steady overall decrease from baseline**. At midline, the share was 35% for comparison schools and 33% for intervention schools.

The share of the sample who answered that no days were missed because of illness has remained similar across groups and rounds (midline versus endline). Additionally, it is the most common response. The same happened with 6 or more days, with a slight increase for comparison students and a slight decrease among intervention students. The share of respondents who answered "3 to 5 days" increased for comparison students and decreased by 1 percentage point for intervention students. The biggest increase for intervention students seems to be who skipped "1 or 2 days," with an increase of 3.3 percentage points from midline to endline.

**Table 31: Days missed because of illness**

| | Intervention | | Comparison | |
|---|---|---|---|---|
| | ML | EL | ML | EL |
| n | 733 | 739 | 623 | 601 |
| None | 76.1% | 76% | 82.7% | 80.7% |
| 1 to 2 days | 15.0% | 18.3% | 12.0% | 13.1% |
| 3 to 5 days | 5.7% | 4.7% | 4.5% | 5.2% |
| 6 or more days | 1.2% | 0.9% | 0.8% | 1.0% |

Using the difference-in-differences regression model, we find an insignificant and positive effect for the intervention group relative to the comparison group. Average days missed weekly because of an illness decreased among comparison children, while increasing slightly among intervention children. To further investigate possible patterns and to understand the full degree of the situation, missed days are disaggregated by gender; no significant difference in differences is found. **Additionally, there does not seem to be a difference between boys and girls in terms of quantity of days missed.**

**Table 32: Average days missed weekly because of illness, disaggregated by gender**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| Days missed | 0.7 | 0.8 | 0.7 | 0.5 | 0.3 | 0.79 |
| Days missed – female | 0.5 | 0.5 | 0.4 | 0.4 | 0.0 | 0.23 |
| Days missed – male | 0.5 | 0.5 | 0.4 | 0.5 | -0.1 | 0.49 |

The indicator related to absences due to illness is McGovern-Dole Custom Outcome #16. Average days missed in the treatment group are 0.5 at endline, an increase compared to the 0.3 missed days reported at midline.

## OTHER FACTORS AFFECTING ATTENDANCE

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 endline, **the majority of caregivers with both male and female children in intervention schools reported that it was safe for their child to walk to school** (85.4% of caregivers with male children and 89.5% of caregivers with female children). We note that the perception of safety for boys decreased from baseline to endline in both intervention and comparison groups. However, the difference was not significant.

The table below disaggregates results by comparison and intervention households at baseline and endline, and shows that intervention households saw a significant improvement in some indicators measuring access to public space and services for male students. **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 comparison households, which instead reported an increase in avoidance and fear towards school**. However, despite these positive results, it is still worth noting that more than a third of boys and a quarter of girls reported avoiding school, and children's reluctance to attend school was a relatively frequent refrain in FGDs and KIIs; this suggests a need for further improvement of this indicator.

**Table 33: Change in access to public spaces and services**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| **Male** | **BL** | **EL** | **BL** | **EL** | **DiD** | **P-value** |
| n | 248 | 398 | 186 | 285 | | |
| Safe to walk to school | 92.3% | 86.3% | 95.2% | 89.7% | -0.5 | 0.91 |
| Avoids school | 48.6% | 37.4% | 35.5% | 36.8% | -12.5 | <0.001* |
| Afraid of school | 17.9% | 10.3% | 10.8% | 15.1% | -11.9 | <0.001* |
| **Female** | **BL** | **EL** | **BL** | **EL** | **DiD** | **p** |
| n | 233 | 343 | 192 | 316 | | |
| Safe to walk to school | 85.4% | 90.6% | 97.9% | 88.4% | 14.7 | 0.03* |
| Avoids school | 41.0% | 29.4% | 31.9% | 26.3% | 5.9 | 0.48 |
| Afraid of school | 15.0% | 13.1% | 9.4% | 11.7% | -4.3 | 0.36 |

In contrast, for female students, no statistically significant difference was found among caregivers of female children regarding school avoidance or fear. Instead, we found a significant difference in differences in level of safety walking to school. **Comparison caregivers reported that their female children are less safe walking to school, while intervention ones reported greater levels of safety**.

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 one for girls). As mentioned above, we must practice the outmost caution when attributing results to the program, but the change in feelings of avoidance and fear towards school in intervention schools might hold particular significance, when remembering that boys are more likely to drop out than girls.

## PREDICTORS OF STUDENT ATTENDANCE

In this section, we test the relationship between various student- and school- level characteristics and attendance rates. We first look at individual-level variables, such as age and language,[126] and then analyze variables that are measured at the school level, such as class size. The outcome variable is 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 in the individual-level model. The only predictor that resulted in a significant coefficient is the level of education of the head of the household: Children whose heads of household are not educated miss on average 0.3 more days per week than children with educated parents, all else held constant. Having educated parents might mean that the income

[126] Individual level characteristics controlled for are: age, language, difficulty with self-care, mental disability, if the child is an orphan, level of education of the head of the household (HoH), experience of feeling of anxiety and depression, if the child takes more than 30 minutes to go to school.

of the household is medium-high, which in turn might mean that children live in areas with better access to school and might have the possibility of taking transportation when needed. Their parents may also put more importance on school, insisting that their children go to school more regularly than others.

**Table 34: Individual-level predictors of student absence**

| | Coefficient | p |
|---|---|---|
| Student age | -0.0 | 0.27 |
| Student language | -0.1 | 0.56 |
| Difficulty with self-care | 0.2 | 0.29 |
| Mental disability | -0.1 | 0.41 |
| Student is an orphan | 0.2 | 0.52 |
| HoH education | -0.3 | 0.01* |
| Experiences depression | 0.2 | 0.15 |
| Experiences anxiety | 0.0 | 0.92 |
| Within 30-minute walk | -0.3 | 0.06 |
| Spent at least half a day on chores | 0.1 | 0.58 |

Worth noting is that time it takes to reach school ("within 30-minute walk" in the table) also had a relatively large, negative coefficient, though the relationship with attendance was not significant. This might echo the fact, discussed above, that extreme weather phenomena and the consequent reduced road accessibility have had a strong effect on attendance rates at endline. However, overall, these results suggest that variability of attendance at endline is not satisfactorily explained by individual factors.

Secondly, we analyze the school-level model.[127] The table below shows the coefficient and p-value for each variable controlled for. As evident from the table above, only treatment status (i.e., whether the school is an intervention or comparison school) is a significant predictor of student absence at the school level, with children in comparison group attending more days of class on average than treatment children. As stressed throughout this section, we must exercise caution when claiming direct causality to the intervention (or lack thereof). This finding is an example of how challenges may not be due to the intervention itself, but contextual characteristics of the intervention schools (such as many of those schools being in remote locations and more difficult to access than comparison ones).

**Table 35: School-level predictors of student absence**

| | Coefficient | p |
|---|---|---|
| Student-teacher ratio | -0.0 | 0.75 |
| Reading materials available | -0.0 | 0.83 |
| School feeding everyday | 0.3 | 0.06 |
| Treatment | 0.3 | 0.01* |
| Toilets available at school | 0.0 | 0.75 |

[127] School-level characteristics controlled for: student-teacher ratio, reading materials available, everyday provision of meals for students and treatment.

Additionally, we should not overlook the factor concerning the availability of meals provided to students everyday (in the table as “School feeding everyday”). This factor has a large coefficient, though not significant. It shows that, in schools where meals are not provided every day, students miss on average 0.3 more days per week than students in whose schools meals are provided every day, all else held constant.

# SCHOOL MANAGEMENT

Effective school management is critical for promoting quality education and improving student outcomes. While teacher quality is an essential component of education, school administrators also play a crucial role in creating an environment conducive to learning. In particular, experienced and knowledgeable school administrators can provide training to teachers and ensure that classroom practices are effective, ultimately improving the quality of instruction. Effective administrators can also organize better conditions for the SFP, including through advocacy with the government and school budget management. Additionally, an active and effective PTA can be vital to strengthen school outcomes, including learning, school infrastructure, and the SFP. As such, in this section, we focus on the importance of school management, including administrators and PTAs, and explore its impact on student learning and teacher effectiveness.

## QUALITY OF SCHOOL MANAGEMENT

At endline, it was found that school administrators had an equivalent amount of experience[128] compared to baseline. However, there were some noteworthy differences in the educational qualifications: At endline, there was a 21 percentage point increase in the number of administrators who held a Bacharelato while there was a 16 percentage point decrease in the number of administrators who held a teacher training institute or Faculty of Education degree compared to baseline.

Aside from these statistics on experience and education, the school survey also gathered data on whether administrators provided coaching to teachers. The findings at endline indicate that only 7% of administrators stated they had never provided coaching. Moreover, compared to the baseline, more administrators claimed to provide weekly coaching, while fewer administrators claimed to provide monthly coaching. The difference-in-differences analysis further suggests that administrators in intervention schools reported a lower rate of having never provided coaching than administrators in comparison schools at the endline. Specifically, within treatment schools, administrators were approximately 6 percentage points less likely to report never providing coaching, while within comparison schools, administrators were 5 percentage points more likely to have never provided coaching (see Table 36[129]).

While the difference-in-differences results were not statistically significant, there was a substantial increase in the provision of coaching every trimester in intervention schools compared to comparison schools. These results imply that **the HATUTAN program may have successfully encouraged school administrators to provide training to teachers, but the frequency of coaching remains low.**

[128] Experience in this context is equivalent to the number of years that director/coordinator has been in his/her position.
[129] Percentages may not sum to 100 due to a “don’t know” option excluded from the table.

**Table 36: Change in administrator provision of coaching**

| | Intervention | | | Comparison | | | DiD (BL-EL) | P-value |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | | |
| n | 98 | 93 | 94 | 41 | 85 | 84 | | |
| Weekly coaching | 42.3% | 28.0% | 52.1% | 39.0% | 37.6% | 44.1% | 4.7 | 0.75 |
| Monthly coaching | 22.5% | 31.2% | 14.9% | 26.8% | 25.9% | 26.2% | -7.0 | 0.30 |
| Coaching every trimester | 22.5% | 36.6% | 25.5% | 31.7% | 30.6% | 20.4% | 14.3 | 0.19 |
| Never provided coaching | 12.2% | 3.2% | 6.4% | 2.4% | 5.9% | 7.1% | -10.5 | 0.18 |

In the qualitative data, few school administrators/coordinators mentioned explicitly providing training to teachers. Instead, many administrators described assigning teachers to classes based on their skills and experience with teaching various topics and ages. One of the teachers mentioned the following about the relationship with the school administrator and the role of the school administrator in regulating classroom issues:

> *The coordinator is always there to support and assist us. If children are not sitting quietly, the coordinator will admonish them. However, the children hold on to their so-called freedom. Even though the coordinator has warned them to sit quietly in class and listen to the teacher, they continue to come and go. This problem occurs because of freedom; we do not beat children since it is a crime. As a result, they are free to come and go as they choose. But the coordinator is always there for us.*
>
> - FGD with teachers, Ainaro, Int. 37

In addition, several school administrators mentioned meeting with teachers who were observed using corporal punishment or yelling at students. The purpose of these meetings was to discourage such methods and encourage more positive forms of discipline.[130]

## PARENT-TEACHER ASSOCIATIONS

The majority of schools (98% at endline) reported having a PTA. At endline, the average number of PTA members was six in treatment schools and four in comparison schools. However, **although there was an improvement in the frequency of PTA meetings in intervention schools at endline compared to baseline, almost half of them still reported not having held any meetings during the current school year.** At baseline, most comparison (71%) and treatment (53%) schools reported that their PTA did not hold any meetings during the year (see Table 37[131]); at endline, this percentage declined slightly for both intervention and comparison areas, to 48% and 69% respectively. Additionally, overall, the percentage of intervention and comparison schools that reported their most recent PTA meeting as having been held last week or last month increased to 33% and 21%, respectively, at endline, compared to 13% and 8% at baseline.

[130] KII with school coordinator, Ainaro, Int. 1; KII with school coordinator, Ainaro, Int. 2
[131] Percentages may not sum to 100 due to a "don't know" option excluded from the table.

**Table 37: Frequency of PTA meetings**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 91 | 92 | 38 | 77 | | |
| Last week | 2.2% | 8.7% | 2.6% | 2.6% | 6.5 | 0.21 |
| Last month | 11.0% | 23.9% | 5.3% | 18.1% | 0.1 | 0.99 |
| More than a month | 34.1% | 19.6% | 21.1% | 10.4% | -3.8 | 0.71 |
| Did not meet this year[132] | 52.8% | 47.8% | 71.1% | 68.8% | -2.7 | 0.81 |

Moreover, **at endline, PTAs were significantly more involved in monitoring learning quality, improving school infrastructure, and monitoring safety and security** in intervention schools than in comparison schools (see Table 38). Additionally, the percentage of PTAs that did not engage in any activities significantly decreased in intervention schools compared to comparison schools at endline. Building on this progress in PTA involvement may further enhance program effectiveness and sustainability.

**Table 38: PTA involvement**

| | Intervention | | | Comparison | | | DiD (BL - EL) | P-value |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | | |
| n | 93 | 93 | 94 | 38 | 83 | 81 | | |
| School budget management | 40.9% | 21.5% | 23.4% | 36.8% | 31.3% | 16.0% | 3.3 | 0.82 |
| Learning quality | 50.5% | 45.2% | 52.1% | 73.7% | 31.3% | 29.6% | 45.7 | 0.01* |
| Improve school infrastructure | 72.0% | 87.1% | 88.3% | 89.5% | 66.3% | 67.9% | 37.9 | <0.001* |
| Oversee SFP | 81.7% | 91.4% | 84.0% | 81.6% | 59.0% | 65.4% | 18.5 | 0.26 |
| Monitor safety and security | 75.3% | 73.1% | 62.8% | 81.6% | 54.2% | 48.1% | 21 | 0.04* |
| Monitor student attendance | 59.1% | 71.0% | 53.2% | 73.7% | 34.9% | 42.0% | 25.8 | 0.10 |
| Monitor teacher attendance | 65.6% | 71.0% | 57.4% | 63.2% | 31.3% | 39.5% | 15.5 | 0.26 |
| Monitor dropout | 48.4% | 47.3% | 33.0% | 60.5% | 34.9% | 29.6% | 15.5 | 0.40 |
| Does not do anything | 6.5% | 1.1% | 3.2% | 0% | 8.4% | 14.8% | -18.1 | <0.001* |

The household survey showed that 36% of caregivers at endline reported that PTAs were either active or somewhat active, 16% indicated that PTAs were inactive, and almost half of the caregivers were uncertain about the PTA's level of activity. Along these lines, the level of participation in PTAs remained low, suggesting limited reach and potential influence in school activities. At baseline, 27% of comparison

[132] **Note from the program:** The survey was conducted in February – March, and the school year started at the end of January.

households and 29% of treatment households reported having a member who participated in the PTA, while the corresponding figures at endline were 19% and 30%, respectively.

However, in contrast to the school survey, the household survey analysis showed a significant positive change in the frequency of PTA meetings. In particular, there was a 10-percentage point increase in the frequency of weekly PTA meetings for intervention households compared to comparison households. Furthermore, similar to the school survey, at endline, PTAs were significantly more involved in improving school infrastructure, and the percentage of PTAs in intervention households that did not engage in any activities significantly decreased compared to comparison households.

Overall, these results present a mixed picture of PTA activity. While there appears to have been some improvement in PTA meeting frequency since baseline, many PTAs remain inactive. Furthermore, many caregivers appear to be uninvolved in PTAs. However, PTAs focus on many areas of interest to HATUTAN, suggesting that they are a highly relevant body for program activities.

Qualitative data indicated that there are varying levels of coordination between the PTA and parents in the schools:

> *We don't have it [a PTA] now, but from 2022 going back to the school's establishment, there was one. We have not yet formed one for this year, as we need to have a meeting with parents and the previous association to determine if it will happen. Since the establishment of the school, the Parent and Teacher Association has consisted of five members. For our school, we chose three members, and CARE helped us with the other two members. The names of the members sitting in each section were posted, and it involved all parents, particularly those living near the school, to look after the school and its fence. We recently returned from a school holiday that lasted for more than a month, and we discovered that the school fence had completely broken due to the presence of animals in [location name]. The parents have been busy with their private activities such as farming, making it challenging to control the school and call for a meeting.*
>
> - FGD with teachers, Ermera, Int. 39

According to several respondents, providing compensation to PTA members is essential to increase their involvement in school activities:

> *The problem could be addressed by providing some financial support, such as money for transportation, to the PTA so that we can all work more willingly.*
>
> - KII with administrator, Manatuto, Int. 8

> *The primary problem is that sometimes, they are unable to respond when we call upon them to do something. When they arrive, they explain that, although it is their responsibility to assist the school, they are unable to do so due to pressing family matters. We must consider the fact that PTA members volunteer their time and do not receive compensation. Therefore, even though we need them, they won't come to school if they have pressing home tasks to complete or family obligations to attend to. This is one of the difficulties we encounter when working with them.*
>
> - KII with administrator, Liquica, Int. 6

Many respondents also highlighted the crucial role of PTAs in monitoring student attendance:

> *If a child is absent, the PTA will directly contact the parents to ask for the reason for the absence before the teacher writes a letter. If the discussion with the parents is not*

> *positive, they will ask the coordinator or the grade teacher to write a letter to the parents, stating "We have contacted the parents without success, so please send them a letter." Therefore, we send a letter to the parents.*
>
> - KII with administrator, Ermera, Int. 3

> *We have been working together until now. If the students do not come for two or three consecutive days, I make a decision and call the Parent and Teacher Association (PTA) to search through their parents' names and villages, so that we can determine the causes of their absences.*
>
> - KII with administrator, Manatuto, Int. 7

Several school administrators and directors observed that men tend to participate more actively in PTA meetings and other activities than women.[133] However, in general, organizing PTA meetings is a challenge due to low attendance, which is often attributed to the lack of incentives.

# SCHOOL FEEDING PROGRAM

One of the primary aims of the HATUTAN program was to assist the Government of Timor-Leste in the successful implementation of the school feeding program in all basic education and preschools throughout the school year. The program targets a total of 440 schools[134], and 435 schools have been able to provide daily meals which were prepared using USDA commodities. Additionally, the program supports local farmers through purchasing their produce for the SFP, which aims to increase local production for a sustainable source of nutritious food for local schools.

## PROVISION OF SCHOOL MEALS

Improving the implementation of the school feeding program is a key goal of HATUTAN. A small proportion of schools reported having a program supporting school meals at baseline (9%), which slightly increased at endline (12%). In addition, the survey found that schools have become more familiar with HATUTAN's support for the provision of meals to students: Among schools who reported having a program supporting school meals, the majority of schools at endline (82%) reported having a school feeding program through HATUTAN, compared to only 6% in the baseline.

Schools were also asked to report on the availability of a school feeding program on the day of the survey. The majority of comparison schools reported that there was no feeding program. However, during the endline survey, it was found that 16% of the intervention schools had meals prepared for the day, in contrast to 1% at baseline (see Table 39). The difference-in-differences coefficient was extremely large and significant, which suggests that **the HATUTAN program may have filled the gaps in the government's implementation of the SFP, as per its design. These findings also support the conclusion that the program has increased familiarity about the SFP**.

[133] KII with school coordinator, Manatuto, Int. 8; KII with school coordinator, Liquica, Int. 5

[134] 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.

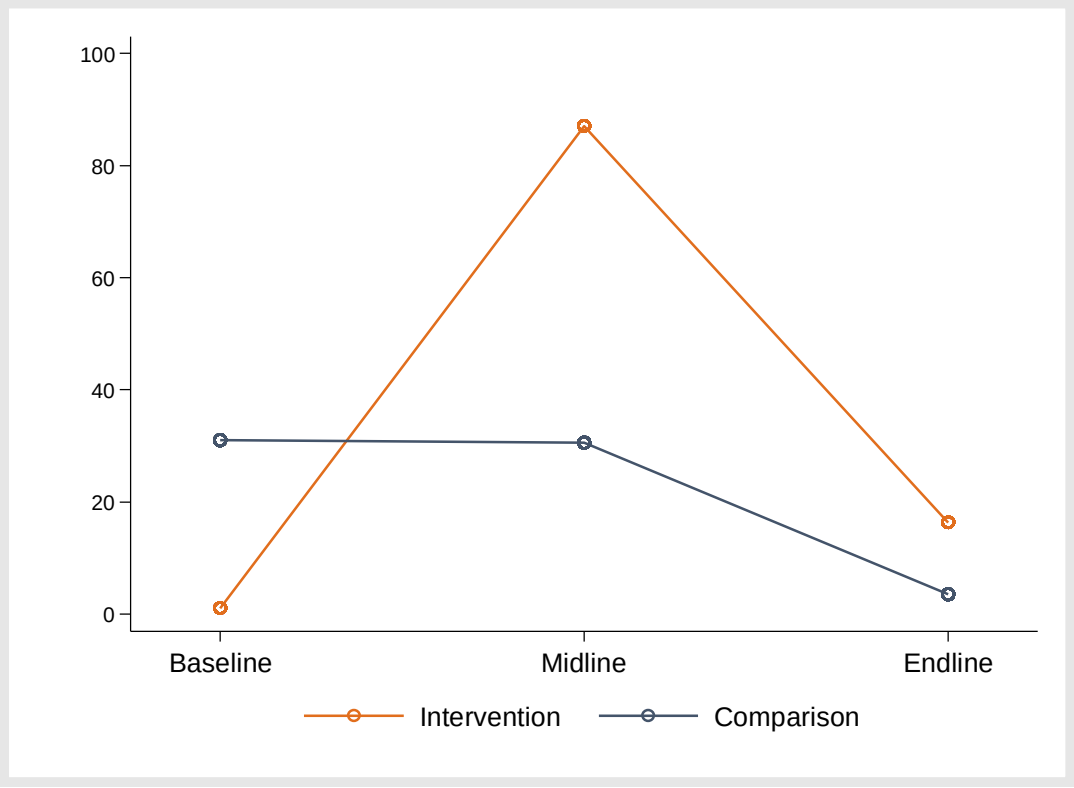
**Table 39: Meals provided to students, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 98 | 98 | 87 | 84 | | |
| Meals provided on day of survey | 1.0% | 16.3% | 31.0% | 3.6% | 42.8 | 0.01* |

These results were corroborated by the household survey, where more respondents from the intervention group reported an ongoing feeding program in schools compared to those in the comparison group: Households in the intervention areas reported a 17-percentage point increase in the provision of school meals at endline compared to baseline, while households in comparison areas reported a 30-percentage point decrease at endline. The difference-in-differences results were significant.

While these results show improvement in school meal provision among HATUTAN intervention schools, it is important to note that meal provision in the first period of the school year has declined drastically since midline among both intervention and comparison schools, as shown in the figure below. At midline, 87.1% of intervention schools and 30.6% of comparison schools were providing school meals; at endline, this declined by 70.8 percentage points for intervention schools and 27.0 percentage points for comparison schools. As discussed in the introduction, this significant decline is due to delays in government distribution of SFP funding to schools; HATUTAN has also faced some delays in commodity provision at the start of the 2023 school year. Overall, this emphasizes the need for continued support to and advocacy with the Government of Timor-Leste, as much work remains to be done to improve the consistency and reliability of SFP funding.

**Figure 29: Change in school meal provision**



# MANAGEMENT AND QUALITY OF THE SFP

HATUTAN also seeks to strengthen management of the school feeding program and the quality of school meals. In addition to information on meal provision, the school survey collected data on the individuals in charge of leading SFP implementation in each school and the food items included in the menu. The household survey provided supplementary information on the characteristics of the meals, such as quantity, taste, and hygienic food preparation.

School administrators were reported to lead the implementation of the feeding program in most schools (75% baseline, 82% endline), with no significant change in the person responsible for the SFP among intervention schools at endline (see Table 40). Additionally, in 46% of treatment schools and 20% of comparison schools, the PTA was reported to be responsible for program oversight.

**Table 40: School feeding program responsibility, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 98 | 98 | 87 | 84 | | |
| Director or coordinator | 68.4% | 88.8% | 81.6% | 75.0% | 27.0 | 0.11 |
| Deputy director | 7.1% | 7.1% | 6.9% | 3.6% | 3.3 | 0.50 |
| PTA | 37.8% | 45.9% | 21.8% | 20.2% | 9.7 | 0.32 |
| Teachers | 30.6% | 30.6% | 13.8% | 11.9% | 1.9 | 0.78 |
| Other[135] | 23.5% | 28.6% | 29.9% | 35.7% | -0.7 | 0.98 |

According to the school survey, the majority of schools reported having a menu, suggesting a degree of organization and preparedness. While there was a slight increase of 2 percentage points in the intervention schools reporting to have menus at endline, the percentage of comparison schools reporting to have menus decreased by 4 percentage points (see Table 41). **This could indicate that the HATUTAN program had a positive impact on the development and implementation of school menus.**

**Table 41: School feeding program menu available, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 98 | 98 | 87 | 84 | | |
| Yes | 92.9% | 94.9% | 94.2% | 90.2% | 6.0 | 0.02* |

Regrettably, due to data collection challenges at the baseline,[136] we lack extensive information on school menus. Consequently, we solely present data on the foods served in school meals from midline and endline; at endline, we only present data for intervention schools, as only three comparison schools served meals.[137]

[135] Among those who responded with "other" at endline, 36% of all schools said that a service provider is responsible for the feeding program.
[136] Note from the project: Since most schools did not serve meals at the baseline, there was no information on the menus.
[137] Disaggregating endline schools by treatment status shows no differences.

At midline, schools predominantly served three food groups: grains, roots, and tubers; legumes and nuts; and dark green vegetables such as spinach, lettuce, and mustard greens. These findings were generally consistent with the results at endline; both midline and endline data revealed a lack of fruit and eggs in school meals. However, there was a significant increase in the number of schools serving animal protein products at endline, driven by an increase in menus including fish (see Table 42).

***Note from the program:*** At the time of the midline, HATUTAN provided rice, beans, and oil to schools for the preparation of meals with parents often contributing leafy greens and other vegetables. At the endline, HATUTAN II (overlapping with HATUTAN) piloted the provision of locally and regionally procured (LRP) commodities to schools, namely rice, beans, and either peanuts or fish powder, again with parental contributions of vegetables. The meal has, by design, a higher protein content to compensate for the low consumption of protein at home and its impact on cognitive development. Due to the limited availability of LRP commodities meeting quality standards, the program was unable to provide them to all schools from the beginning of the school year.

During midline, the majority of comparison schools (69%) identified that only one food item from the list provided in Table 42 was included in school meals, while most intervention schools mentioned two (31%) or three (27%) items. However, at endline, 90% of all schools identified zero items from the list, indicating that the food program was not functioning as intended.

**Table 42: Menu items by study group among schools with meals served**

| | Midline | | Endline |
|---|---|---|---|
| | Comparison | Intervention | Intervention |
| n | 26 | 81 | 16 |
| Grains, roots, and tubers | 88.5% | 93.8% | 87.5% |
| Legumes and nuts | 15.4% | 77.8% | 75.0% |
| Dairy products | 7.7% | 0.0% | 0.0% |
| Eggs | 0.0% | 0.0% | 6.3% |
| Meat and meat products | 11.5% | 1.2% | 56.3% |
| Vitamin A-rich dark leafy greens | 15.4% | 40.7% | 18.8% |
| Other vitamin A-rich vegetables and fruits | 11.5% | 18.5% | 31.3% |
| Other fruits and vegetables | 0.0% | 0.0% | 6.3% |
| Other[138] | 69.2% | 8.6% | 6.3% |

According to these results, schools served food items belonging to a maximum of five food groups out of nine. At midline, meals provided in intervention schools had a higher average dietary diversity score of 2.4 compared to those served in schools in comparison areas, at 2.2. At endline, among intervention schools that served meals, meals had a higher average dietary diversity score of 2.9. Notably, about 19% of schools at endline that served meals scored a 4 on the dietary diversity scale (see Table 43).

[138] Including condensed milk.

**Table 43: School menu dietary diversity score, by study group**

| | Midline | | Endline |
|---|---|---|---|
| | Comparison | Intervention | Intervention |
| n | 26 | 81 | 16 |
| 1 | 0.0% | 7.4% | 0.0% |
| 2 | 84.6% | 48.2% | 43.8% |
| 3 | 11.5% | 40.7% | 31.3% |
| 4 | 3.9% | 3.7% | 18.8% |
| 5 | 0.0% | 0.0% | 6.3% |

The household survey also collected information on the availability, quantity, preparation, and taste of meals served to children in schools. Results showed that at both baseline and endline, most parents either completely or partially agreed that food was available every day (92% baseline, 87% endline), quantity was sufficient (79% baseline, 88% endline), meals were prepared in a hygienic manner (76% baseline, 89% endline), and they were tasty (70% baseline, 85% endline). The difference-in-differences analysis indicated that households in intervention areas at endline were more likely to disagree completely or partially with statements regarding school meals—that food was available every day, food was tasty, food was prepared hygienically, and that the quantity of food was sufficient—than would be expected based on results in comparison areas.

Based on the qualitative data collected, it was found that school coordinators and directors are primarily responsible for overseeing the school feeding program, with support from teachers and the PTA. When asked about the most difficult aspects of implementing the food program, many respondents cited challenges related to accessing clean water,[139] purchasing firewood,[140] and limited funds.[141]

The qualitative data also revealed that PTAs play a key role in monitoring the SFP. They ensure that meals are served as planned, oversee the cleanliness of the kitchen and utensils, identify missing items or food shortages, and determine budget needs. One school administrator mentioned that they contacted parents through their children to encourage them to provide any missing ingredients, as a way to encourage parents' and PTAs' participation in the school feeding program:

> *In this school feeding program, anyone can [provide] support. For example, the CARE staff provided some food items. If we didn't have certain items like vegetables, we informed the parents through their children to contribute some vegetables so that the supplier could cook them with the beans and provide food to our children every day.*
>
> - KII with administrator, Ainaro, Int. 1

Children reported receiving meals, and parents expressed appreciation for the government-provided school meals. Parents discussed the food quality served to children based on their observations or their children's feedback. Some of the recurring issues that parents mentioned were the lack of diversity in meals, with children often being served beans, and low food quality. Efforts are being made to improve dietary diversity

[139] KII with school coordinator, Ainaro, Int. 2; KII with school coordinator, Liquica, Int. 6
[140] KII with school coordinator, Liquica, Int. 5; KII with school coordinator, Liquica, Int. 6
[141] KII with school coordinator, Ainaro, Int. 2

in schools, and mobilization of government funds may help schools purchase more diverse and nutritious foods available locally:

> *The children eat red beans one day, and then a combination of banana buds, red beans, and chicken meat the next day. The menu is determined by the funds provided to the cooks, who are responsible for acquiring culinary components. The funds they are provided with are limited, neither more nor less. Children can consume fish and poultry meat throughout the season when they are available, although this is subject to availability. Meanwhile, children cannot consume the same food daily.*
>
> - FGD with mothers, Manatuto, Int. 31

> *School feeding is all about the school meal. When our children came home, we asked them, "Did you eat good food? Good rice? Good meal? Good side dishes?" and they responded that they did not eat today. They had meals sometimes only. School feeding provides good meals and good side dishes, but they only had them when there was a meal. That's what happened here, based on our observation and what we understood from our children.*
>
> - FGD with fathers, Manatuto, Int. 17

> *According to me, the children say they eat well. They say, "The food is tasty, but they serve very little."*
>
> - FGD with mothers, Ainaro, Int. 25

A father from Liquica municipality also raised concerns about school menus not adequately meeting the needs of students. He mentioned that teachers had to hold meetings with parents to address this issue:

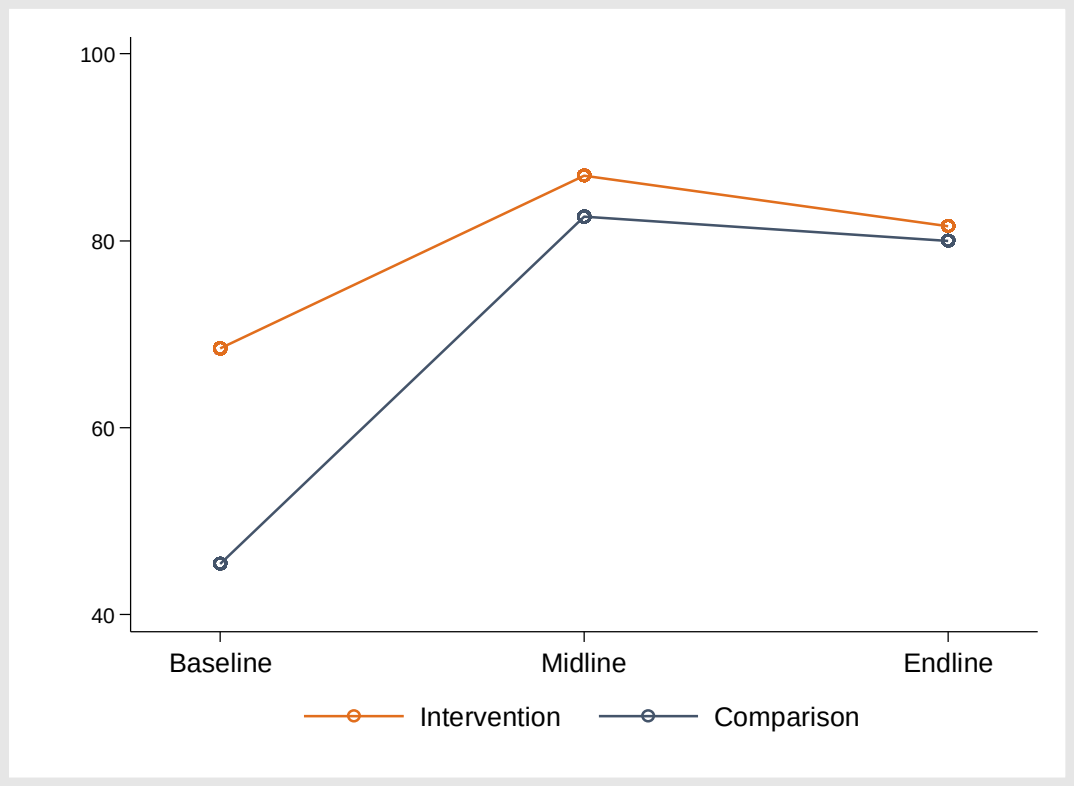
> *Sometimes the school feeding program did not respond to all the needs of the students. It has already come with its specific menu. If teachers wanted to add more on the menu, teachers could conduct a meeting with parents to contribute [some other menu items] so that the students might have various menu instead of only having the beans everyday. Those items contributed include the vegetables.*
>
> - FGD with fathers, Liquica, Int. 18

## FOOD PREPARATION AND STORAGE

Most parents felt that school meals were prepared in a hygienic manner, which is an important consideration to ensure that children do not get sick and miss classes. The majority of schools that reported having their own kitchen (88% baseline, 81% endline) indicated that they maintain clean kitchens by using detergent (79% baseline, 78% endline). At endline, more schools with their own kitchen reported having access to clean water for meal preparation compared to baseline. However, all schools with a kitchen used wood stoves and only a small proportion had a scale in the kitchen at endline. Overall, these findings suggest that **while some improvements were made in the availability of clean water and cleanliness of kitchens, there is still a need to address the use of wood stoves and the lack of scales in the kitchen.** It should be noted, however, that the extremely high cost of natural gas in country and the non-availability of gas stoves would result in dependency from external assistance to implement the SFP. Gas stoves were provided by WFP at the onset of the SFP, but largely abandoned later on due to the inability to afford gas. Electrical stoves are not available in country either and the cost of electricity is considerably high.

Figure 30 shows the change in access to clean water to prepare school meals[142] across evaluation rounds and treatment groups. We find that access to clean water for meals has increased since baseline in both intervention and comparison schools. However, the increase in access was substantially—though not significantly—higher for comparison schools than intervention schools; access in intervention schools increased by 13 percentage points from baseline to endline, while access in comparison schools increased by 35 percentage points. Furthermore, access to clean water declined slightly in both intervention and comparison schools from midline to endline, at similar rates.

**Figure 30: Change in access to clean water to prepare school meals**




**While this analysis does not show clear evidence of HATUTAN program impact, it is notable that at endline, over 80% of schools had access to clean water for meals.** This represents important progress, as using clean water in meal preparation is vital to reduce the risk of water-borne illnesses, including diarrheal illnesses.

According to the difference-in-differences analysis, changes in hygienic food preparation occurred at similar rates in both intervention and comparison schools. There were no significant relative differences in usage of detergent at endline in intervention schools compared to comparison schools (see Table 44): At endline, the daily use of detergent declined by approximately 13 percentage points in intervention schools and around 9 percentage points in comparison schools.

Regarding the frequency of daily cleaning of school storage areas, comparison schools exhibited progress by improving their score by 20 percentage points, thus reaching a level comparable to intervention schools. In contrast, intervention schools improved by approximately 18 percentage points. Notably, the percentage

[142] Data was also collected on school access to water and specific water sources; however, some aspects of this data were contradictory. For example, across all evaluation rounds, 108 enumerators stated that the school did not have water but then selected a source where the school obtains water, while 3 enumerators stated that the school did have water but then stated that no water source was available. As such, we focus only on access to clean water for school meals.

of schools in intervention areas that reported having no clean storage has significantly decreased relative to the comparison schools at endline (see Table 44). **Overall, these findings suggest that hygiene practices related to meal preparation have improved to some extent in both intervention and comparison schools.**[143]

**Table 44: Frequency of detergent use and availability of food storage, by treatment group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| **Frequency of detergent use** | | | | | | |
| n | 92 | 68 | 22 | 43 | | |
| No | 20.7% | 22.1% | 22.7% | 20.9% | 3.2 | 0.47 |
| Sometimes | 34.8% | 52.9% | 40.9% | 51.1% | 7.9 | 0.62 |
| Often | 6.5% | 0.0% | 0.0% | 0.0% | -6.5 | 0.06 |
| Every day | 38.0% | 25.0% | 36.4% | 27.9% | -4.5 | 0.80 |
| **Availability of food storage** | | | | | | |
| n | 98 | 71 | 30 | 40 | | |
| No | 27.8% | 5.6.% | 6.7% | 12.5% | -28.0 | <0.001* |
| Somewhat | 38.8% | 33.8% | 40.0% | 42.5% | -7.5 | 0.81 |
| Mainly | 13.3% | 22.5% | 43.3% | 15.0% | 37.5 | 0.24 |
| Yes | 20.4% | 38.0% | 10.0% | 30.0% | -2.4 | 0.82 |

The enumerators were also instructed to monitor the presence of animals in school kitchens during both midline and endline surveys. Interestingly, there was a significant disparity in the observations made between comparison schools and intervention schools. The percentage of schools where animals were observed increased from 20% to 47% in comparison schools, while remaining constant at 22% in intervention schools.

Regarding school feeding facilities, we analyze midline to endline results due to insufficient sample size at baseline. At endline, the intervention group had a higher percentage of schools with all observed facilities than the comparison group, with the exception of kitchen plates/cutlery (see Table 45). However, when subjected to a difference-in-differences regression, none of the differences were statistically significant. As a result, we are unable to confidently conclude that the program positively impacted the effectiveness of the SFP through the implementation of better facility standards.

According to data collected from the school survey, the majority of schools (64% at baseline, 76% at endline) reported having some or enough storage space available, and intervention schools were slightly more likely to have storage space compared to comparison schools. However, further analysis did not show any significant differences in the availability of storage spaces between intervention and comparison schools.

[143] Two of the comparison municipalities, Bobonaro and Manufahi, receive support from WFP and WHO for school feeding.

**Table 45: Availability of school facilities, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | | |
| School canteen | 5.4 (n=93) | 20.4 (n=98) | 3.5 (n=85) | 11.9 (n=84) | 6.6 | 0.56 |
| School kitchen | 97.8 (n=93) | 93.9 (n=98) | 81.2 (n=85) | 65.5 (n=84) | 11.8 | 0.35 |
| School kitchen plates | 84.9 (n=93) | 80.6 (n=98) | 83.5 (n=85) | 88.1 (n=84) | -8.9 | 0.28 |
| School kitchen handwashing | 35.2 (n=91) | 23.9 (n=92) | 33.3 (n=69) | 12.7 (n=55) | 9.3 | 0.31 |
| School food storage | 78.0 (n=91) | 77.2 (n=92) | 58.0 (n=69) | 72.7 (n=55) | -15.5 | 0.33 |

Most of the storage spaces at endline had cement floors (60% treatment, 40% comparison) and brick walls (62% intervention, 50% comparison). All of the intervention schools had an aluminum roof at the endline. About 18% of intervention schools had storage spaces with a leaking roof; results show that the percentage of intervention schools with a leaking roof decreased at endline by more compared to comparison schools, but the result is insignificant.

At baseline and endline, 79% and 72% of schools, respectively, reported that their storage spaces had adequate ventilation. The proportion of storage spaces without any method for raising food off the ground decreased from 23% at baseline to 12% at endline. Table 46 shows that most of the schools used pallets for raising food off the ground during both baseline (58%) and endline (80%). These findings suggest that efforts have been made to ensure that food served to children is stored in a sanitary space to prevent potential illness and absence from school.

**Table 46: Storage sanitation, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| n | 98 | 71 | 29 | 40 | | |
| Shelves | 9.2% | 7.0% | 20.7% | 7.5% | 11.0 | 0.38 |
| Pallets | 55.1% | 84.5% | 69.0% | 72.5% | 25.9 | 0.09 |
| None | 25.5% | 8.5% | 13.8% | 17.5% | -20.7 | 0.08 |
| Other | 18.4% | 9.9% | 27.7% | 10.0% | 9.2 | 0.39 |

According to school coordinators/directors and parents, the school feeding program faces challenges in terms of access to adequate facilities, materials, and cooks. In particular, respondents mentioned that old kitchen facilities and limited storage space are major obstacles to implementing the program. Additionally, the absence of some kitchen materials and clean water were cited as concerns:

> *Talking about the school feeding in this school, I think it was going well. The school students have good food, but the only concern was that they did not have enough plates, so the students needed to bring their own from home. If a student did not bring a plate, they would have to borrow from others so that they could have the food. There was still a lack of plates and clean water for the students to wash their hands.*
>
> - FGD with mothers, Manatuto, Int. 32

> *We must first assess the kitchen, dining hall, the lack of eating tables and seats, and the availability of drinking water in order to enhance the school feeding program and find a place to keep equipment and tableware. Regarding the latter, we began drilling a well last year, and once it is completed, we will have access to pure water at school going forward. Since we have hand-washing facilities, we are required to use them because kids need to cleanse their hands before entering the classroom and after using the restroom. When it comes to school lunch, we sometimes struggle to get a receipt from the stores where we make purchases because some can give us one while others can't.*
>
> - KII with administrator, Liquica, Int. 6

> *We need to help the school with some more plates. They did not have enough plates and spoons. Most of the plates and spoons were brought from home by each student in their bag and used them for their school meals. This should be the responsibility of the school, and as you are from CARE, you should attend to this kind of need. That was the weakness. The school needs to be responsible for providing plates, and it should be the responsibility of the state/government. We do not agree if the children have to bring plates from home. You should support the school with plates.*
>
> - FGD with fathers, Manatuto, Int.17

> *Here, we faced difficulty accessing clean water. It was challenging when we didn't have water. If we had water like other schools, I think there would be no problem. The hygiene was always good because we always made sure to have enough water to wash things like vegetables, dishes, etc.*
>
> - KII with administrator, Ainaro, Int. 2

> *I am concerned about the kitchen. We need to fix it and provide a space to put the food. We should never put the food for our children on the floor. We need to provide a space for food and a sink.*
>
> - FGD with mothers, Manatuto, Int. 32

***Note from the program:*** HATUTAN did provide basic kitchen equipment (such as large cooking pots and measuring scoops) to participating schools. The Government (MEYS and later Municipal Administrations) were to provide plates, cutlery, pots, and other basic equipment for the delivery of the SFP. In 2023, the service providers of school meals must purchase these items out of the administrative fund portion for the SFP.

Some parents expressed that schools are experiencing a shortage of cooks, and in some cases, the available cooks are working without rotation:

> *Here it's good, a little bit good. The service is good, except that the "marinir" (cooks) are the ones who are sometimes always active. They don't rotate and always work.*

> *During this period, they are always active, and they don't substitute for each other, but the service is good.*
>
> - FGD with fathers, Ermera, Int. 16

***Note from the program:*** There is no planned rotation of cooks.

Furthermore, according to some parents' answers, students are occasionally requested to bring vegetables or make contributions to the school. Parents expressed their willingness to contribute in order to ensure that their children have well-balanced meals instead of solely consuming rice and beans. This is especially the case when schools inform them about budget constraints and the limited availability of food options:

> *In the past, the school feeding was provided by the WFP. Sometimes, they would only bring beans, rice, and oil. So, I coordinated with parents and the PTAs. We held meetings, and I explained that the state and NGOs have supported their food. If children eat the same food every day, they will surely get bored. So, I asked them to support the addition of vegetables. Whether someone would bring pumpkin bud vegetables or pumpkin fruit for our kids to eat.*
>
> - KII with administrator, Manatuto, Int. 7

> *Rice and oil are always available, but the community will support with vegetables when the school does not have a budget. As of now, the school does not have any budget and has informed us to wait.*
>
> - FGD with mothers, Liquica, Int. 29

> *It was hard to get good vegetables in the rainy season because we lived near the riverside and could not cross the river. The community members who grew watercress, cassava leaves, etc. contributed, but those who did not grow anything contributed nothing.*
>
> - FGD with fathers, Ainaro, Int. 13

Finally, according to many parents, the main challenges related to the school feeding program are associated with weather conditions and the remote locations of the schools:

> *The biggest concern was the rain. When the government delivered the school feeding items, the car could not reach the school area. Sometimes, they delivered the items on the other side of the river, and parents had to carry them across. During the crossing, parents would sometimes slip and fall, and some food items got wet and damaged.*
>
> - FGD with fathers, Ainaro, Int. 13

> *It is difficult to properly implement the school feeding program because we have problems with accessing the menu due to the poor condition of the roads. This condition of the roads is due to climate change.*
>
> - KII with administrator, Manatuto, Int. 8

## PURCHASE OF LOCAL FOODS

In order to promote local production and ensure a sustainable food source for the SFP, schools were encouraged to purchase their produce locally. At baseline, most schools reported that they would purchase

food locally some of the time (53%) or all of the time (39%). By endline, most school administrators stated that they buy goods locally. At endline, less than one fifth of the schools (14%) said that they do not buy produce from local farmers, with half of these being intervention schools. Interestingly, among the schools that provided meals, comparison schools at endline were more likely to purchase local produce from farmers than intervention schools (see Table 47). However, since only one intervention school provided meals at baseline, a difference-in-differences analysis cannot be conducted due to the low sample size.

Among intervention schools, the primary reasons for not purchasing produce from farmers at endline were not having the budget to buy produce (50%) and the insufficient amount of farmer's produce (42%). During both baseline and midline, the government had not yet transferred money to schools for the SFP, which may be the main reason for this finding. **Unfortunately, purchasing patterns and dietary diversity continue to be constrained by the reality that the average cost of a nutritional diet using nutritious, locally available food items is higher than the SFP budget.**

**Table 47: Schools buying local produce from farmers, by study group**

| | Intervention | | Comparison | |
|---|---|---|---|---|
| | BL | EL | BL | EL |
| n | 1 | 16 | 27 | 3 |
| No | 0.0% | 37.5% | 3.7% | 0.0% |
| Yes, sometimes | 0.0% | 62.5% | 59.3% | 100.0% |
| Yes, all the time | 100.0% | 0.0% | 37.0% | 0.0% |

***Note from the program:*** Since the vast majority of schools were not providing meals at the baseline, the data indicating purchases from local farmers does not refer to the ongoing practice during the first period of the 2023 school year, but probably to previous years.

According to the data collected from school surveys, most schools purchased their produce locally during the baseline and endline periods. Among schools that purchased local produce, the types of produce purchased appeared consistent. 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%) were the main food items bought during both baseline and endline periods (see Table 48). However, this does not match what most schools reported as part of the meals served to children in school, which are primarily composed of carbohydrates, starchy food, beans, legumes and nuts, and dark green vegetables. This suggests that **food supplies purchased from local farmers may not make up a major portion of school meals.** The low consumption of fruits may also be explained by the limited number of schools that purchased fruits locally.

***Note from the program:*** Since schools had not yet received the funds to purchase local produce, the types mentioned are likely to refer to the purchases from school year 2022.

At endline, a larger number of intervention schools reported buying various protein sources such as meat, eggs, and seafood compared to baseline. However, changes between baseline and endline for intervention schools should be interpreted with caution due to the fact that schools had not yet received the SFP funds and were therefore referring to previous practice; the inability to cross-verify results against the actual meal provided may result in overreporting due to desirability bias. Additionally, there is a large disparity in sample

size, with data collected in only one intervention school at baseline. In the case of rice, maize, bread, legumes, beans, and nuts, a lesser number of schools purchased these items. 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 48: Local produce schools bought from farmers, by study group**

| | Intervention | | Comparison | |
|---|---|---|---|---|
| | BL | EL | BL | EL |
| n | 1 | 10 | 26 | 3 |
| Rice, maize, bread and foods prepared with rice, maize and wheat | 0.0% | 20.0% | 61.5% | 66.7% |
| Pumpkin, carrot, purple sweet potato | 100.0% | 40.0% | 69.2% | 33.3% |
| Potato, taro, yellow sweet potato, cassava, sago | 100.0% | 40.0% | 69.2% | 66.7% |
| Dark green vegetables (e.g., spinach, lettuce, pumpkin leaves, cassava leaves) | 100.0% | 100.0% | 88.5% | 66.7% |
| Other vegetables (e.g., cucumber, tomato, cabbage, eggplant) | 100.0% | 10.0% | 38.4% | 0.0% |
| Yellow fruits (e.g., mango, papaya, honeydew melon, passionfruit) | 0.0% | 20.0% | 0.0% | 0.0% |
| Other fruits (e.g., watermelon, tamarind, jackfruit) | 0.0% | 0.0% | 3.8% | 0.0% |
| Meat (beef, pork, sheep/goat meat, chicken, duck) | 100.0% | 10.0% | 57.7% | 0.0% |
| Seafood (e.g., fresh or dry fish, shrimp) | 100.0% | 10.0% | 26.9% | 0.0% |
| Legumes, beans, and nuts (e.g., beans, peas, soybeans or peanuts) | 0.0% | 30.0% | 50.0% | 0.0% |
| Fresh milk | 0.0% | 0.0% | 15.4% | 0.0% |
| Coconut oil | 0.0% | 0.0% | 7.7% | 0.0% |
| Condiments | 0.0% | 0.0% | 50.0% | 0.0% |
| Tofu or tempe | 0.0% | 20.0% | 15.4% | 0.0% |
| Eggs | 0.0% | 30.0% | 38.5% | 0.0% |

In intervention schools, there was an increase in the number of schools in Liquica that purchased goods from local farmers at endline compared to baseline. **However, there was a decrease in the number of schools in Ainaro and Ermera intervention municipalities that purchased goods from local farmers at endline compared to baseline,** with the decline in Ainaro being particularly noteworthy as shown in

Table 49. A possible explanation for this trend is the ongoing food security challenges faced by Timor-Leste, especially in rural areas where subsistence agriculture is a primary source of livelihood. Ainaro is one such rural municipality where the local population relies heavily on agriculture for sustenance. Other factors, such as the absence of government funds and the timing of the survey during the rainy season, may also explain the decrease in the number of schools buying goods from local farmers in some municipalities.

***Note from the program:*** As previously noted, the delayed transfer of SFP funds to schools means that responses refer to previous school years and it is not possible to verify their accuracy vis-a-vis the actual meals served. It is likely that responses are also affected by desirability bias.

**Table 49: Schools buying local produce from farmers, by municipality and study group**

| Municipality | Group | n | BL | EL | Change (BL to EL) | P-value |
|---|---|---|---|---|---|---|
| Ainaro | Intervention | 27 | 100.0% | 74.1% | -25.9 | <0.001* |
| Ermera | Intervention | 41 | 95.1% | 90.2% | -4.9 | 0.40 |
| Manatuto | Intervention | 19 | 100.0% | 100.0% | 0.0 | - |
| Liquica | Intervention | 11 | 81.8% | 90.9% | 9.1 | 0.56 |
| Aileu | Comparison | 23[144] | 78.3% | 90.9% | 12.6 | 0.25 |
| Covalima | Comparison | 10 | 70.0% | 70.0% | 0.0 | - |
| Bobonaro | Comparison | 32[145] | 100.0% | 80.6% | -19.4 | 0.01* |
| Manufahi | Comparison | 22[146] | 90.9% | 90.5% | -0.4 | 0.96 |

We note that the data cannot be analyzed any further for schools that reported having a feeding program, as the sample size per municipality and round of data collection is too small to conduct further analysis.

When the school feeding program is in operation and budget is available, schools try to purchase a range of vegetables from local farmers, as noted by most of the interviewed school administrators. However, many respondents mentioned that they are unable to purchase the produce due to insufficient supply from local farmers:

> *The only problem is the readiness or continuous supply of local products. We need to order them in advance before buying. This also means that they cannot provide the same product continuously. As a result, we need to buy from other farmers in other communities. The school snack is every day, so we require more stock to buy.*
>
> - KII with administrator, Ermera, Int. 4

> *We want to buy in a big quantity, but the products they have are not sufficient, so we need to buy little by little from different groups or individuals in the community until the products are enough for school feeding. We can only have an agreement with those who produce sufficient products. Currently, there is no such producer, so we only buy from the communities who have the products.*

[144] n = 22 for endline sample.
[145] n = 31 for endline sample.
[146] n = 21 for endline sample.

- KII with administrator, Liquica, Int. 5

*One of the difficulties here is that the conditions of the communities do not allow for many farms. Another difficulty is that farmers only grow one variety of vegetables… The vegetables are not enough, and when we need more, for example, I recently contacted the owner from Viqueque who bought land here in Natarbora and grew vegetables. I asked him to supply vegetables to our school. He supplied them, but he called me and said he could not supply again because Natarbora and Fatuberliu also needed them. The vegetables are not enough to supply to our school. I have a plan to ask my parents to plant vegetables in many plots.*

- KII with administrator, Manatuto, Int. 7

The typical items that the schools purchase for the feeding program include vegetables (e.g., cassava leaves, pumpkins), fruits, and meat (e.g., chicken, beef):

*It was a weekly purchase where we bought not only vegetables and fruits, but also meats like beef and pork. We purchased those meats in bulk from the local [farmers], not by the kilogram as in the market.*

- KII with administrator, Ainaro, Int. 2

*In the previous months, we made an agreement with local farmers to buy vegetables, beans, and pumpkins for school snack food preparation. They buy the local produce in the afternoon and cook them the following morning.*

- KII with administrator, Ermera, Int. 4

*Yes, we purchased vegetables such as banana buds. I encouraged them [farmers] to form vegetable-growing groups, but vegetables are no longer available due to the end of the planting season. However, they [farmers] do supply cassava leaves, pumpkin stalks, pumpkins, papaya fruits and leaves, and banana buds that are purchased by the school and vendors.*

- KII with administrator, Liquica, Int. 6

A school administrator from Liquica municipality noted that even when farmers are able to supply enough produce, the problem of food storage remains a challenge:

*We used to buy veggies for the school for a whole week. We paid them in advance after deciding on a price at the field, and the farmers brought the veggies to the school every day. Because the school lacks a fridge to keep the veggies for five straight days, the vegetables may rot or become damaged. Because the suppliers have their own farms, they can collect enough veggies every day to prepare for one day's meal.*

- KII with administrator, Liquica, Int. 6

## COMMUNITY CONTRIBUTION TO SCHOOL FEEDING PROGRAM

The government allocated funding for the SFP, each child in basic education is supposed to receive a meal based on 42 cents per day of which 35 cents is for food and seven cents for the payment of cooks, fuel, firewood, among others; at the time of the endline (25 cents per day at baseline/ midline). The responsibility of selecting and monitoring cooks who prepare the rice and purchase other local produce required for meal preparation was supposed to be with the parents although in practice the school administrator often made

the selection.[147] This section examines the role of schools and households in the SFP, with a focus on the involvement of PTAs in implementation.

The surveys conducted at baseline and endline indicated that the majority of schools had a PTA overseeing the school feeding program (78% baseline, 76% endline). The endline data showed an improvement in PTA involvement in school feeding in intervention schools compared to comparison schools, which reported lower rates of PTA oversight. There was also a significant increase in the percentage of respondents who believed that the PTA in their children's school was engaged in activities to improve school feeding (40% baseline, 72% endline). The endline data further showed that PTAs in the intervention areas had made more progress in improving school feeding than those in comparison areas, indicating that **PTAs in areas exposed to HATUTAN programming may have experienced relatively greater gains in effectiveness monitoring school feeding than those in areas without program activities**.

**Table 50: PTA involvement in school feeding, by study group**

| | Intervention | | Comparison | | DiD | P-value |
|---|---|---|---|---|---|---|
| | BL | EL | BL | EL | | |
| **PTA oversees school feeding** | | | | | | |
| n | 93 | 98 | 83 | 81 | | |
| Yes | 81.7% | 84.7% | 74.7% | 65.4% | 12.3 | 0.28 |
| **PTA improves school feeding** | | | | | | |
| n | 383 | 313 | 259 | 132 | | |
| Yes | 39.2% | 76.0% | 42.5% | 61.4% | 17.9 | 0.01* |

The school administrators confirmed that PTAs play a significant role in monitoring the school feeding program. They ensure that there is an adequate supply of produce for meals, check that meals are prepared in a hygienic manner, and oversee meal serving:

> *First, when school feeding has been distributed to this school, the PTA is present to ensure that school feeding really exists at this school. Second, they also wanted to know if the chefs were actually cooking according to the plans set by HATUTAN.*
>
> - KII with administrator, Ainaro, Int. 1

> *The PTA was successful in organizing groups last year to help with the school's meal program by supplying veggies and cooking oil that are shared among them. One group brings vegetables, rice, and cooking oil to the classroom from local stores. They are the ones in charge of monitoring this activity, making sure that the children eat their meals and follow the prescribed menus. The PTA and other parents visited this location and occasionally objected, saying the kids shouldn't consume what has been made. "My kids used to eat corn or cooked corn at home and then continue to eat them at school," they claimed. When we showed them the menu, they objected, saying, "No, our children are not accustomed to eating rice at home, so prepare rice instead". Children in towns prefer to consume less yam or sweet potatoes.*
>
> - KII with administrator, Liquica, Int. 6

[147] In 2023, the school groupings formed school feeding management teams to operate the SFP.

> *Talking about school feeding, up until now the school feeding came with its own menu for the school to implement. If they did not implement based on the rules, the Parent and Teacher Association and coordinator would give them warning and exercised direct observation to suppliers.*
>
> - FGD with fathers, Liquica, Int. 18

However, as mentioned earlier, the lack of sufficient funding and difficulty in maintaining facilities persist as challenges in some schools:

> *The government has recently allocated funds for school feeding. CARE HATUTAN did not support the first period, but in the previous year (2019), if I'm not mistaken, CARE HATUTAN facilitated the distribution of rice, beans, and oil, and we used some of these supplies. The government's money went out, and the third support from CARE also stopped, so we implemented the government's program.*
>
> - KII with administrator, Manatuto, Int. 7

> *The most difficult aspect was management. For example, if the supplier managed the money well, the school feeding program went well. However, if the money was not managed properly, problems could occur in the middle of the process. Nonetheless, the school continued to provide support to ensure that the school feeding program was managed according to the government's expectations.*
>
> - KII with administrator, Ainaro, Int. 2

> *We have a plan for the school feeding program. We want to have a school garden, but the challenge is that this school does not have a fence yet. Having water is not enough. We continue to ask the community to grow vegetables and fruits as part of their contribution to the school feeding program. With this, when there is a school feeding program, we can buy from them.*
>
> - KII with administrator, Liquica, Int. 5

> *Our main worry is the school fence; the school should construct a fence around the building to safeguard the plants, flowers, and vegetables that the children grow, so that when parents are unable to contribute veggies, the school can harvest from its own garden. Animals will invade the garden if there is no fence, destroying everything that students grow. At least, purchase some iron mesh to place around the school.*
>
> - FGD with fathers, Ermera, Int. 15

# HEALTH AND NUTRITION

A key focus of HATUTAN is improving nutrition and water, sanitation, and hygiene (WASH) knowledge and practices. Additionally, one of the McGovern-Dole strategic objectives is to increase the use of positive health, nutrition, and dietary practices. These practices are tied to broader program outcomes: The results framework posits that healthy practices decrease health-related absences in school, which improves students' school attendance and contributes to improved literacy.

This section focuses first on nutrition and health practices. To better understand drivers of adoption of good health and nutrition practices, we analyze knowledge of nutrition, health, and hygiene (including safe food

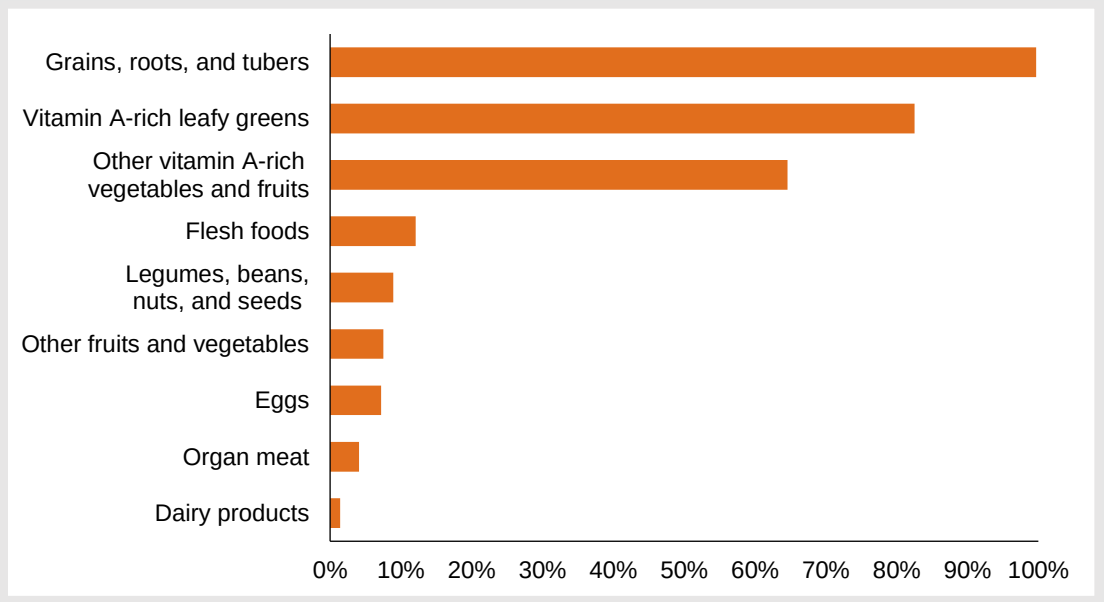
preparation and storage methods); access to clean water and sanitation; and access to preventative health interventions. To better understand the impact of nutrition practices, we also analyze the Body Mass Index (BMI) of students; as this data was not collected in previous evaluation rounds, we provide only a snapshot of results at endline. Lastly, we conduct a predictive analysis of the impact of household- and community-level factors on health and nutrition knowledge and practices.

## NUTRITION PRACTICES

In this section, we analyze data on changes in the food consumption of caregivers and children under two. Within the household survey and farmer's group survey, caregivers reported on the types of food they consumed during the previous day. Mothers of children under the age of 2 also reported on foods consumed by this child during the previous day. Additionally, at endline only, caregivers were asked about foods eaten by their grade 2 child (i.e., the child assessed with the EGRA) during the previous day. Within the EGRA, students were also asked about foods they had eaten at school the previous day. This provides a full snapshot of items eaten by students the previous day, including both items eaten at school and at home. Food items were categorized into groups; the full list of food items by groups is included in Annex 4.

We first discuss food consumption of female caregivers of child-bearing age (15 to 49); we focus on these caregivers as the nutrition of women of child-bearing age plays a fundamental role in their baby's development and nutrition. The below figure shows consumption of food groups for intervention caregivers at endline.

**Figure 31: Consumption by food group, female caregivers of child-bearing age**




We find that almost all caregivers consumed grains, roots, and tubers; 83% consumed Vitamin A-rich leafy greens; and 65% consumed other Vitamin A-rich vegetables and fruits. However, consumption of food groups drops off sharply after this, with only 12% of caregivers consuming flesh foods, the next highest category. This result is noteworthy as it implies very low consumption of protein-rich foods, including meat, animal products, nuts, beans, and legumes. In other words, the data suggests that **most households live on a carbohydrate-based diet with little access to protein sources**. This is consistent with the findings

of the 2020 Timor-Leste Food and Nutrition Survey, which also found that diets in Timor-Leste were dominated by carbohydrates and green leafy vegetables.[148]

Looking now at change over time in caregivers' nutrition practices, the below table shows percentages of caregivers who reported consuming each food group at baseline, midline, and endline, as well as the difference-in-differences results from baseline to endline. We find **no significant relative changes in food consumption between intervention and comparison groups from baseline to endline**.

However, it is worth noting differences in food consumption that have occurred across evaluation rounds, even if they have occurred similarly in both comparison and intervention areas. Compared to baseline, at endline, we find substantial decreases in consumption of vitamin A-rich vegetables and fruits (excluding leafy greens); other fruits and vegetables; legumes, beans, nuts, and seeds; and eggs. Consumption of vitamin A-rich leafy greens and flesh foods has meanwhile increased slightly, though in the latter case, from a very low baseline level. Overall, this may suggest a **widespread decrease in dietary quality over the past five years**, and increasing reliance on grains, roots, tubers, and dark leafy greens for calories and nutrition.

**Table 51: Change in consumption by food group, female caregivers of child-bearing age**

| | Intervention | | | Comparison | | | DiD (BL to EL) | P-value |
|---|---|---|---|---|---|---|---|---|
| Food group | BL | ML | EL | BL | ML | EL | | |
| n | 378 | 607 | 587 | 304 | 518 | 481 | - | - |
| Grains, roots, and tubers | 99.7% | 98.7% | 99.7% | 100.0% | 98.8% | 99.8% | 0.1 | 0.75 |
| Vitamin A-rich leafy greens | 78.3% | 69.7% | 82.5% | 79.9% | 76.8% | 83.8% | 0.3 | 0.97 |
| Other vitamin A-rich vegetables and fruits | 84.7% | 58.8% | 64.6% | 86.8% | 56.0% | 57.8% | 9.0 | 0.36 |
| Legumes, beans, nuts, and seeds | 16.1% | 11.2% | 8.9% | 8.2% | 8.3% | 2.9% | -2.0 | 0.70 |
| Other fruits & vegetables | 23.0% | 11.2% | 7.5% | 24.0% | 12.2% | 7.9% | 0.6 | 0.94 |
| Eggs | 11.9% | 9.1% | 7.2% | 10.5% | 5.4% | 4.4% | 1.4 | 0.76 |
| Flesh foods | 11.4% | 8.1% | 12.1% | 12.5% | 8.1% | 13.1% | 0.1 | 0.97 |
| Dairy products | 2.9% | 2.5% | 1.4% | 0.7% | 0.6% | 0.4% | -1.3 | 0.45 |
| Organ meat | 4.5% | 2.5% | 4.1% | 4.9% | 2.1% | 2.7% | 1.8 | 0.57 |

We now analyze the food groups eaten by children under the age of 2, as reported by caregivers. As with caregivers themselves, we first find that children's diets are dominated by the consumption of grains, roots, and tubers. Children also consume vitamin A-rich vegetables and fruits, although less frequently than caregivers, and regularly have their diet supplemented with breastmilk.

Looking at the difference-in-differences results in Table 52, we again find **no significant relative difference in the types of foods consumed by children under 2 in intervention areas compared to comparison areas**. However, we note that as with caregivers, consumption of several food groups declined from baseline to endline, especially vitamin A-rich vegetables and fruits and eggs. Overall, these patterns

[148] Timor-Leste Food and Nutrition Survey 2020, Government of Timor-Leste. See https://www.unicef.org/timorleste/press-releases/timor-leste-food-and-nutrition-survey-2020-preliminary-results-steady-progress-made.

suggest that **young children's diets are dominated by carbohydrates, especially grains, roots, and tubers, with few protein sources**. We do note that at endline, 29% of intervention caregivers and 30% of comparison caregivers reported that their child aged 6 months to 23 months had consumed breastmilk during the previous day, a positive sign as breastmilk is a highly useful supplement to a diverse diet for young children.

**Table 52: Change in consumption by food group, children under age 2**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| Food group | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| n | | | | | | | - | - |
| Grains, roots, and tubers | 94.9% | 93.9% | 90.4% | 97.4% | 92.5% | 95.0% | -2.1 | 0.58 |
| Vitamin A-rich vegetables and fruits | 44.9% | 44.8% | 34.7% | 50.7% | 30.6% | 35.5% | 4.9 | 0.51 |
| Legumes and nuts | 7.1% | 4.7% | 1.8% | 2.6% | 0.7% | 1.7% | -4.4 | 0.08 |
| Other fruits and vegetables | 6.1% | 2.4% | 1.8% | 3.9% | 3.4% | 0.8% | -1.2 | 0.76 |
| Eggs | 10.2% | 12.3% | 8.7% | 15.6% | 6.8% | 6.6% | 7.4 | 0.15 |
| Flesh foods | 4.1% | 5.7% | 3.7% | 6.5% | 4.1% | 3.3% | 2.8 | 0.46 |
| Dairy products[149] | - | 3.8% | 3.7% | - | 4.1% | 5.8% | -1.8 | 0.55 |
| Breast milk[150] | - | 26.4% | 29.2% | - | 34.0% | 29.8% | 7.1 | 0.47 |

In FGDs, mothers and fathers generally demonstrated knowledge of the value of protein-rich foods. A father in Manatuto, for example, stated that, "Here, when we talk about nutritious food, we refer to fish, meat, tofu, and tempeh."[151] Similar sentiments were expressed by parents in Ermera, Ainaro, and Liquica.[152] These foods—and a quality diet more generally—were considered particularly important for children's health:

> *[We must provide children with] high-quality food, so that they are healthy and their nutrition is adequate, and that they can eventually learn better… If the children's health is bad, they will lack motivation to study.*
>
> - FGD with fathers, Liquica, Int. 18

However, a common refrain was that economic challenges limit the quality of diets consumed by caregivers and children. This challenge was summarized in the following quote from a father in Ermera and echoed by parents in all other municipalities:[153]

> *We can't give nutritious food to our children as farmers since it costs money.*
>
> - FGD with fathers, Ermera, Int. 15

[149] DiD is midline to endline due to comparability issues with baseline data.
[150] DiD is midline to endline due to comparability issues with baseline data.
[151] FGD with fathers, Manatuto, Int. 20
[152] See FGD with fathers, Ainaro, Int. 14; FGD with mothers, Ermera, Int. 26; FGD with fathers, Liquica, Int. 19
[153] E.g., see also FGD with fathers, Liquica, Int. 18; FGD with fathers, Manatuto, Int. 19; FGD with mothers, Ainaro, Int. 25

Several parents also mentioned that limited availability of nutritious foods, including protein-rich foods, posed a challenge to dietary quality. In Ermera, for example, a mother stated that tofu and tempeh were not available in the village market; instead, families had to travel to the municipal capital to purchase these foods.[154] Similarly, when asked whether households in his community provide nutritious foods for their children, a father in Ainaro stated the following:

> *It depends on the season and the capacity of the community. We are far away from the market. We had tofu, tempeh, and eggs on a few occasions only… We also buy fish from fishermen and vegetables from the market, when we have money.*
>
> - FGD with fathers, Ainaro, Int. 13

Lastly, it is worth noting that in Ainaro and Ermera in particular, parents stated that some healthy foods were considered taboo for young children, pregnant women, or lactating mothers. These foods include fish, considered dangerous for children due to the risk of swallowing bones;[155] fruits such as watermelon,[156] mangoes, and lemons[157] for pregnant women in Ermera; and moringa for pregnant or lactating women in Ermera.[158] Strong understanding of food taboos is important for nutrition programming in order to ensure that communities are not insensitively pressed to eat forbidden foods.

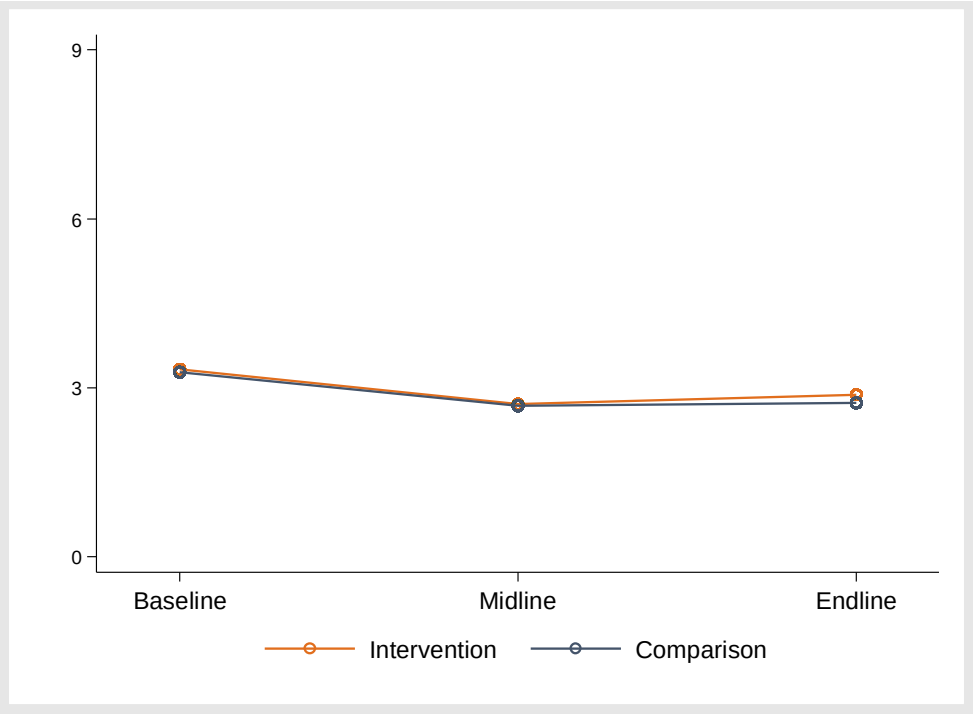
***Note from the program:*** Conversely, it is necessary to sensitize households on the risks of some food taboos which may put children at risk, including non-consumption of colostrum among some groups, or the safe ways of consuming nutritious food (such as fish for children). This approach is in line with the GoTL's activities.

To better understand nutrition practices and quality of nutrition, we calculate dietary diversity scores (DDS) for women and children. These dietary diversity scores reflect the diversity of diet based on the number of food groups consumed (up to nine for women of childbearing age and seven for children).

Average scores for women by treatment group and round are shown in the figure below. **We find very little difference between the average DDS of women in intervention and comparison groups, and no significant change in the intervention group relative to the comparison group**. Among both groups, dietary diversity scores declined slightly from baseline to midline, from around 3.3 to around 2.7, and then rose slightly at endline, to around 2.8 for intervention caregivers and 2.7 for comparison caregivers. In other words, we find that caregivers tend to consume, on average, around three food groups, and that there has been little change over time in this nutrition dynamic. Particularly in light of the finding above that most diets are dominated by carbohydrates, limited dietary diversity may indicate weak nutrition practices, which may be a result of the "hungry season" in Timor-Leste limiting food consumption, financial challenges which limit food purchases, limited knowledge of good nutrition practices, cultural attitudes towards food, or other dynamics.

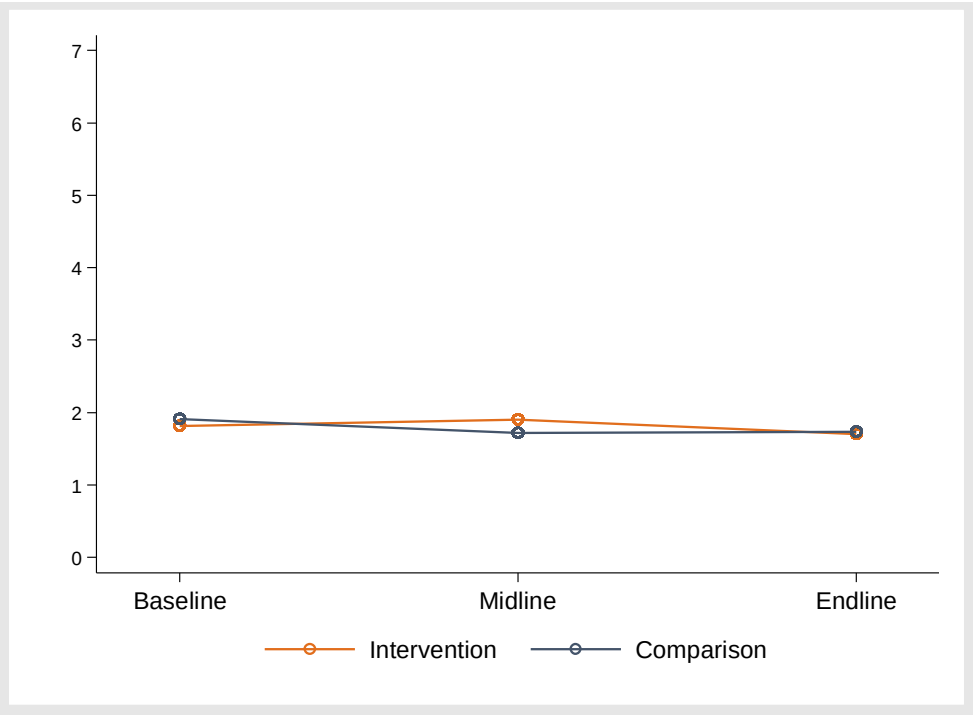
[154] FGD with mothers, Ermera, Int. 26
[155] FGD with fathers, Ainaro, Int. 13; FGD with mothers, Ainaro, Int. 25; FGD with fathers, Ermera, Int. 16
[156] FGD with mothers, Ermera, Int. 28
[157] FGD with fathers, Ermera, Int. 16
[158] FGD with fathers, Ermera, Int. 14; FGD with mothers, Ermera, Int. 27

**Figure 32: Change in dietary diversity score, women of childbearing age**




Children's DDS are shown in the below figure. We find **lower overall DDS for children than for caregivers**, and similarly little variance across intervention and comparison groups. Across all rounds, children consumed an average of 1.7 to 1.9 food groups; there was little difference from baseline to endline, and no significant relative difference in consumption across intervention and comparison children. As with caregivers, this finding again suggests a relatively **weak dietary diversity for children**.

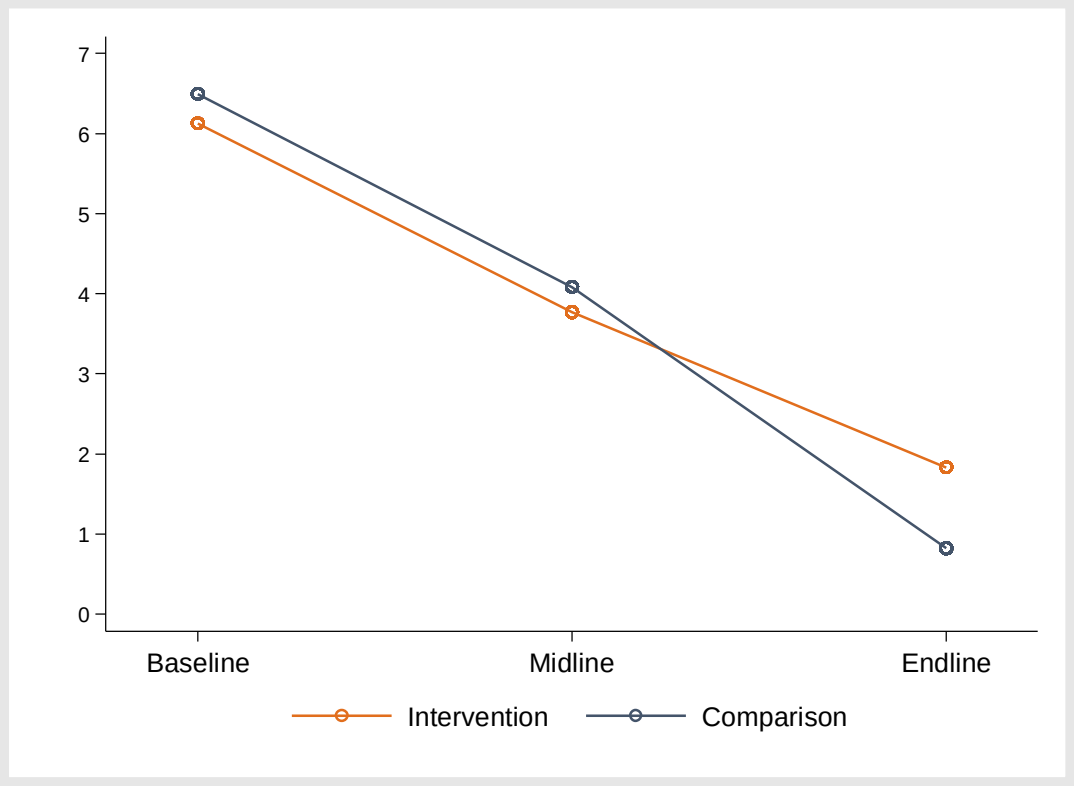
**Figure 33: Change in dietary diversity score, children ages 6 months to 23 months**



DDS are calculated to determine if children meet the minimum acceptable diet (MAD) for children ages 6 months to 23 months, which requires that children have consumed at least four of the seven food groups during the previous day. Figure 34 shows that attainment of a MAD has consistently declined since baseline from an already-low starting point. At baseline, the percent of children ages 6 months to 23 months who consumed at least four food groups was only around 6%; at endline, this had dropped to only 1.8% of intervention children and 0.8% of comparison children.

While there was no significant difference in the relative decline in intervention areas compared to comparison areas, this is a discouraging finding, as it suggests a **weakening of dietary diversity—and thus quality of nutrition—across Timor-Leste in the past five years.** This is an important area for future interventions, as child nutrition has long-ranging effects on health, success in education, and other factors which allow children to succeed in their future. Given that children's DDS have declined during the HATUTAN implementation period, we recommend revising program activities addressing this component of nutrition to improve efficacy.

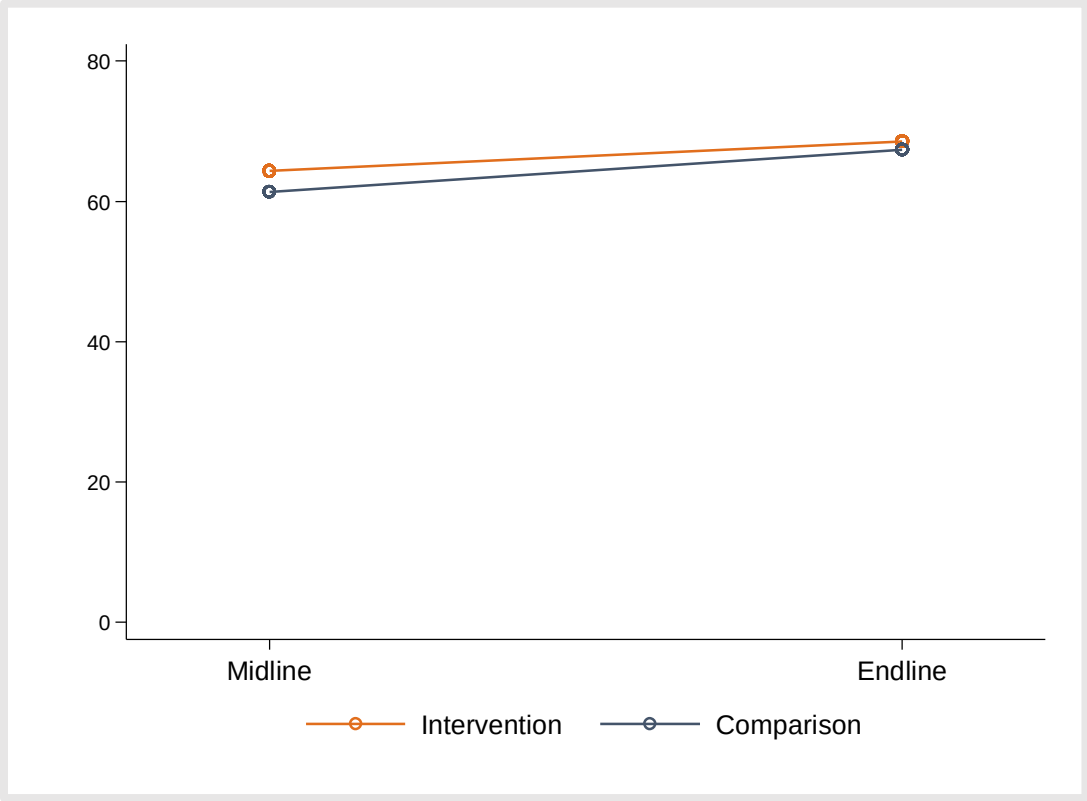
**Figure 34: Change in attainment of minimum acceptable diet, children ages 6 months to 23 months**




Looking now at nutrition practices for babies (under the age of 6 months), we analyze the prevalence of exclusive breastfeeding as well as the use of formula milk.[159] Figure 35 shows that while the majority of mothers practice exclusive breastfeeding, more than 30% of babies were not exclusively breastfed at endline. While prevalence of exclusive breastfeeding has increased since midline, the increase is relatively small and was similar in both intervention and comparison groups. This suggests, **the HATUTAN program may not have had an impact on improving the prevalence of exclusive breastfeeding for children under 6 months of age.**

[159] We exclude baseline data due to comparability issues.

**Figure 35: Change in use of exclusive breastfeeding, children under 6 months of age**




The data suggests that within comparison areas, exclusive breastfeeding is being supplanted by the use of formula milk. At endline, we find that 16.3% of comparison mothers reported using formula milk to feed their baby, compared to only 4.2% at baseline. In contrast, in intervention areas, there was a slight decline in the use of formula milk, from 7.0% at baseline to 5.5% at endline. When combined with data on breastfeeding, these findings suggest that the vast majority of babies under the age of 6 months are being fed either breastmilk or formula milk. However, it is important to note that 49 comparison caregivers and 73 intervention caregivers reported that their baby under the age of 6 months ate foods that are not breastmilk or formula milk, while four comparison caregivers and one intervention caregiver reported that their baby consumed neither breastmilk nor formula. Given the potential negative implications of these nutrition practices on babies' health, it is important to continue emphasizing the importance of exclusive breastfeeding .

Lastly, we analyze foods consumed by students. This data was collected for the first time at endline; as such, we provide only a description of findings at a single point in time in Table 53. We note that **70.2% of students reported that they did not eat anything at school the previous day** (66.5% of intervention students and 75.5% of comparison students); these students' (non-) responses are not included in the below table. In other words, this table includes student-reported data only for the 29% of students who said that they ate at school the prior day.

This table first shows that, as with caregivers and children under two, students' food consumption within the household (i.e., as reported by caregivers) was dominated by grains, roots, and tubers and vitamin A-rich leafy greens. Consumption of protein sources was relatively uncommon, with less than 20% of students in intervention areas consuming legumes or nuts and only 12.2% consuming meats. Caregivers were also asked about processed and sugary foods and drinks consumed by their grade 2 child; consumption of these

foods and drinks was extremely widespread, with nearly 100% of all caregivers stating that their grade 2 child had consumed such food/drink the previous day.

Looking at the foods students self-reported consuming during school the prior day, we again see that students most frequently reported consumption of grains, roots, and tubers. However, notably, around one-third of intervention students reported consuming legumes and nuts and one-quarter reported consuming meat. This level of consumption was significantly higher in intervention areas than comparison areas. Additionally, we note that processed and sugary foods were, on average, the second most frequently reported food consumed by students at school.

**Table 53: Foods consumed by students at endline**

| | Caregiver-reported | | Student-reported | |
|---|---|---|---|---|
| **Food group** | **Intervention** | **Comparison** | **Intervention** | **Comparison** |
| n | 739 | 601 | 482 | 251 |
| Grains, roots, and tubers | 99.6% | 99.5% | 73.2% | 63.4% |
| Vitamin A-rich leafy greens | 73.6% | 76.5% | 12.5% | 13.6% |
| Other vitamin A-rich vegetables and fruits | 4.7% | 4.0% | 2.3% | 1.2% |
| Other fruits and vegetables | 9.7% | 9.2% | 7.3% | 3.2% |
| Legumes and nuts | 18.2% | 14.7% | 32.4% | 1.6% |
| Eggs | 4.2% | 3.0% | 3.7% | 2.4% |
| Flesh foods and organ meats | 12.2% | 11.0% | 24.9% | 8.8% |
| Dairy products | 8.9% | 7.0% | 0.2% | 1.2% |
| Processed or sugary foods | 99.7% | 100.0% | 31.3% | 41.8% |

Overall, these findings suggest a **limited dietary quality for grade 2 students**. Diets are predominantly composed of carbohydrates, including processed and sugary foods with limited nutritional value that may contribute to health issues such as diabetes. Consumption of protein sources and vitamin-rich foods, outside of dark leafy greens, is limited. Furthermore, many students do not eat at all during school hours, further limiting caloric consumption and, potentially, consumption of nutrient-rich foods.

## BODY MASS INDEX

To further explore the implications of low dietary quality on children's health, at endline, grade 2 students' height and weight[160] was recorded. From these measurements, we calculate students' Body Mass Index (BMI) as weight in kilograms divided by height in meters squared. At an aggregate level, BMI provides a useful method to screen for weight categories that may lead to health problems, including both underweight and overweight. We utilize the United States Center for Disease Control categories for BMI, where a BMI below the 5$^{th}$ percentile by age is underweight, between the 5$^{th}$ and 84$^{th}$ percentile is healthy weight, between the 85$^{th}$ and 94$^{th}$ percentile is overweight, and above the 95$^{th}$ percentile is obese. For a male child

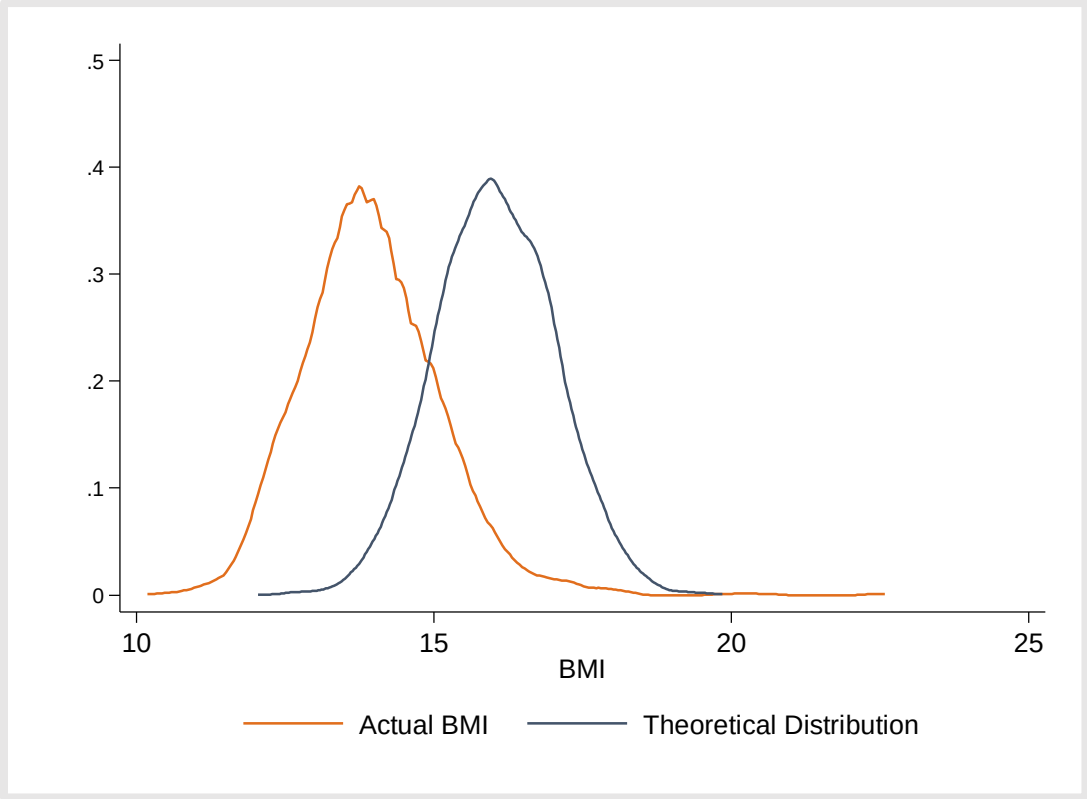
[160] Due to limitations of the scales used, weight was only measured to the nearest kilogram. As such, our measures of BMI are not completely accurate; however, this is unlikely to substantially bias findings.

aged 8, this means that a BMI below 13.8 is underweight, between 13.8 and 18.0 is healthy weight, and above 18.0 is overweight or obese.[161]

**We find extremely high underweight prevalence among grade 2 students**. The average BMI was only 13.9 kg/m$^2$ for intervention students and 13.6 kg/m$^2$ for comparison students. Fifty-six percent of comparison students had underweight BMIs, while 44.1% of intervention students had underweight BMIs. Almost all remaining students had normal BMIs, with only 0.5% of intervention students and 0.2% of comparison students classified as overweight. Average BMI was extremely similar across male and female students.

The below figure demonstrates the massive extent to which these measured BMIs deviate from a theoretical "ideal" distribution which centers upon the middle of the normal range for the distribution of student ages in the endline sample, a BMI of 16.0 kg/m$^2$. There is relatively little overlap between the actual and theoretical BMI distributions; in other words, many intervention students have BMIs under that which would be desirable for health outcomes.

**Figure 36: Actual and theoretical BMI distributions, intervention students**




These findings emphasize that the weak dietary diversity and nutritional practices found above may have a tangible impact on students' health outcomes, as measured at an aggregate level through BMI. Increasing the prevalence of SFPs, as well as (ideally high quality) calories provided in school meals, may be a vital intervention to improve student health.

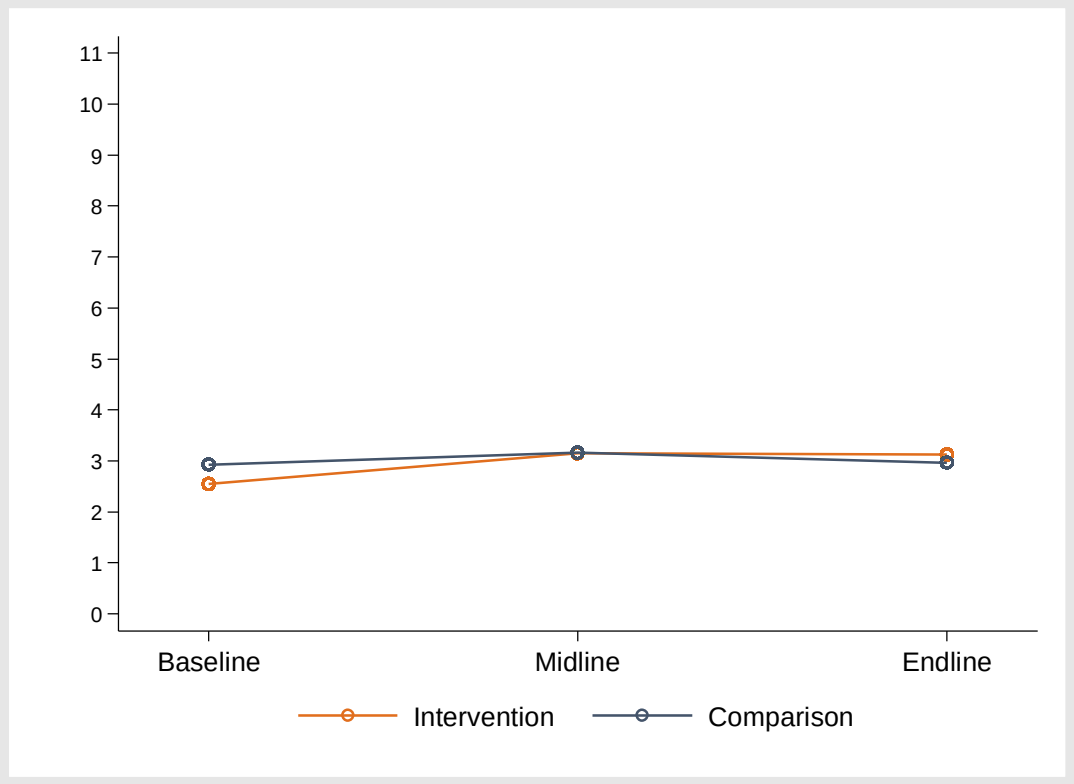
[161] See https://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm for further BMI cutoffs by age and gender.

# KNOWLEDGE OF NUTRITION

Knowledge of nutrition may be one factor contributing to nutrition practices, alongside economic constraints, cultural attitudes, and other factors. Within the household survey, caregivers were asked to provide examples of important maternal childcare practices to gauge their nutritional knowledge.[162] Eleven nutrition-related practices[163] were included in the tool; enumerators marked practices if mentioned by caregivers.

The figure below shows the change in the number of nutrition practices identified by caregivers across rounds. We first find that **nutrition knowledge was low across all rounds and treatment groups**; at endline, on average, intervention caregivers mentioned only 3.1 healthy nutrition practices and comparison caregivers mentioned only 3.0 practices. This represented a slight increase from baseline for both comparison and intervention caregivers, with no significant relative change in nutrition knowledge in intervention areas.

**Figure 37: Change in nutrition knowledge**




Looking further at the specific types of nutrition practices identified, at endline, we find that intervention caregivers were most likely to identify feeding children a variety of nutritious foods (56.7%), exclusively breastfeeding for 6 months (44.5%), initiating breastfeeding within one hour of childbirth (40.1%), breastfeeding frequently (38.8%), introducing safe foods to children at 6 months of age (37.9%), and feeding children iron-rich foods (31.9%) as healthy nutrition practices. In contrast, caregivers infrequently

[162] Caregivers were not prompted with behaviors and asked if they were important nutrition practices; they were rather asked to name practices without any prompting.
[163] Feeding children a variety of nutritious foods, exclusively breastfeeding for 6 months, initiating breastfeeding within one hour of childbirth, breastfeeding frequently, introducing safe foods to children at 6 months of age, feeding children iron-rich foods, continuing breastfeeding for one to two years, feeding children frequently, ensuring that pregnant and lactating women eat sufficient food, breastfeeding when ill, and providing expressed breastmilk if babies are unable to suckle.

identified feeding babies expressed breastmilk if unable to suckle (3.9%) and continuing to breastfeed when ill (7.6%) as healthy practices.

We find no significant relative changes in the rate at which caregivers identified various practices within intervention and comparison areas from baseline to endline. However, for intervention caregivers, we find a significant increase from baseline to endline in the frequency with which frequent breastfeeding and exclusive breastfeeding were identified as healthy nutrition practices. At baseline, around 25% of caregivers identified both practices; at endline, this had increased to 38% for frequent breastfeeding and 45% for exclusive breastfeeding.

While not a main topic of discussion in qualitative interviews, within FGDs, parents mentioned several hygienic nutrition practices. In Ainaro, Ermera, and Liquica, for example, parents stated that it was important to prepare food hygienically and ensure food was clean when cooking.[164] In Manatuto, a mother stated that, "Whatever food we have prepared on the table must be covered… because the flies may land on food anytime."[165] Lastly, in Liquica, mothers stated that breastfeeding was beneficial for babies' health.[166]

McGovern-Dole Custom Outcome #23 on nutrition knowledge asks that mothers identify at least three important nutrition or dietary recommendations. At endline, 53.7% of caregivers met this standard. The below table shows changes across round and treatment group; we find that achievement of Outcome #23 improved substantially, but not significantly, more among intervention caregivers than comparison caregivers from baseline to endline. However, achievement of the outcome declined for both intervention and comparison caregivers from midline to endline.

**Table 54: Change in achievement of McGovern-Dole Custom Outcome #23**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| n | 482 | 982 | 739 | 378 | 625 | 601 | - | - |
| Identified at least three nutrition practices | 40.0% | 55.2% | 53.7% | 47.1% | 60.0% | 50.4% | 10.4 | 0.09 |

Overall, this section suggests that **while caregivers do have some knowledge of healthy nutrition practices, more work is needed to expand knowledge**. Furthermore, we emphasize that knowledge does not necessarily translate into practice; substantial barriers to healthy nutrition practices may remain even for caregivers with high levels of nutrition knowledge, including economic circumstances, cultural barriers, or factors such as preference for processed or sugary foods. Holistic interventions which target not just knowledge but also this wide range of factors may be needed to instill lasting change in nutrition.

## HEALTH PRACTICES

We now discuss health practices, focusing on two practices: handwashing and vaccinating children. Table 55 shows the change in reported handwashing frequency from midline to endline by treatment group. Due to social desirability bias, it is likely that these numbers may be somewhat inflated, as caregivers are likely to report that they wash their hands more frequently than they actually do because they know it is a

[164] FGD with mothers, Ainaro, Int. 25; FGD with fathers, Ermera, Int. 16; FGD with mothers, Liquica, Int. 29
[165] FGD with mothers, Manatuto, Int. 32
[166] FGD with mothers, Liquica, Int. 30

desirable practice. However, this bias should affect both intervention and comparison areas equally, and thus should not bias our difference-in-differences analysis.

We find that among both intervention and comparison groups, **the frequency with which caregivers "always" wash their hands decreased at endline, while the frequency with which caregivers say they wash their hands "most times" increased**. These responses changed more acutely in intervention areas than comparison areas, but the relative difference was not significant. Furthermore, as will be discussed further below under *Access to Clean Water and Sanitation,* less than 10 percent of households had a handwashing station with soap at endline. As such, it is likely that handwashing practices are overreported by caregivers.

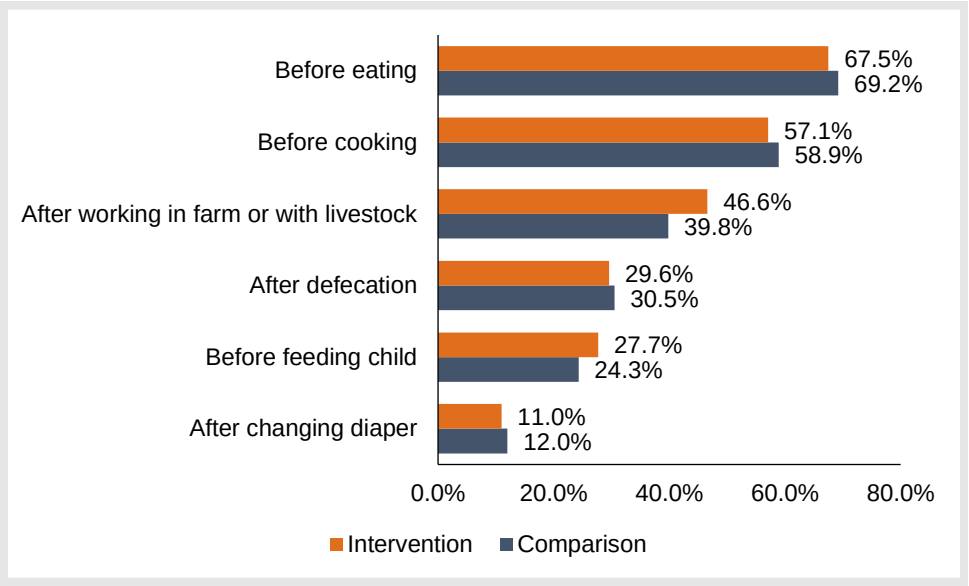
Given that the midline evaluation occurred during the COVID-19 pandemic, these results are likely a reflection of higher than average levels of handwashing due to COVID-19. At endline, then, it is likely that the below results reflect a reversion to more normal levels of handwashing. While it is discouraging that positive handwashing practices adopted to prevent the spread of COVID-19 have not continued, it is also perhaps not surprising, as other healthy practices—such as wearing a mask while ill—have also become less prevalent as the impact of COVID has declined.

**Table 55: Change in handwashing frequency**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| Handwashing frequency | ML | EL | Diff. | ML | EL | Diff. | DiD | P-value |
| n | 981 | 1,025 | - | 624 | 601 | - | - | - |
| Always | 46.7% | 35.7% | -11.0 | 41.7% | 36.4% | -5.3 | -5.8 | 0.43 |
| Most times | 17.9% | 28.7% | 10.8 | 21.0% | 26.8% | 5.8 | 4.9 | 0.48 |
| Occasionally, never, or don't know | 35.4% | 35.6% | 0.3 | 37.3% | 36.8% | -0.5 | 0.8 | 0.81 |

Caregivers were also asked to list all of the times the previous day when they had washed their hands. Figure 38 shows the frequency with which caregivers mentioned specific situations when they washed their hands. Caregivers were most likely to have washed their hands before eating and cooking; 67.5% and 57.1% of intervention caregivers mentioned washing their hands in these respective situations. In contrast, only 11.0% of intervention caregivers mentioned washing their hands after changing a baby's diaper. This percentage increases to only 17.8% when limited to caregivers with a baby under the age of 2 years, suggesting that caregivers are infrequently washing their hands after changing babies' diapers.

**Figure 38: Situations when caregivers washed hands**




We additionally note that caregivers infrequently reported washing their hands before feeding a child. This is a highly important health practice, as feeding a child with unwashed hands may introduce bacteria and diseases to the child. As such, it may be useful to emphasize the importance of this practice to caregivers, especially those with young children who are particularly vulnerable to diarrheal diseases that can be introduced through unsanitary feeding practices.

The qualitative data sheds further light on handwashing practices. Parents generally understood the importance of handwashing. However, mothers and fathers in Ainaro, Ermera, and Manatuto expressed that they did not have sufficient money to purchase soap, and as a result, they frequently washed their hands with only water.[167] In addition to this barrier, parents in Ainaro also stated that limited access to water reduced the frequency of handwashing.[168]

Other than economic constraints, several parents stated that there was a need to better reinforce the habit of handwashing; for example, a father in Liquica stated the following:

> *It is necessary for parents to teach the children to wash their hands every day in order for them to become accustomed to doing so. If we only remind them one day and then forget to remind them the next, they will stop doing it. As a result, teachers and parents should constantly remind them to wash their hands so that children become accustomed to washing their hands both at home and at school.*
>
> - FGD with fathers, Liquica, Int. 18

Similarly, fathers in a focus group in Ainaro stated the following:

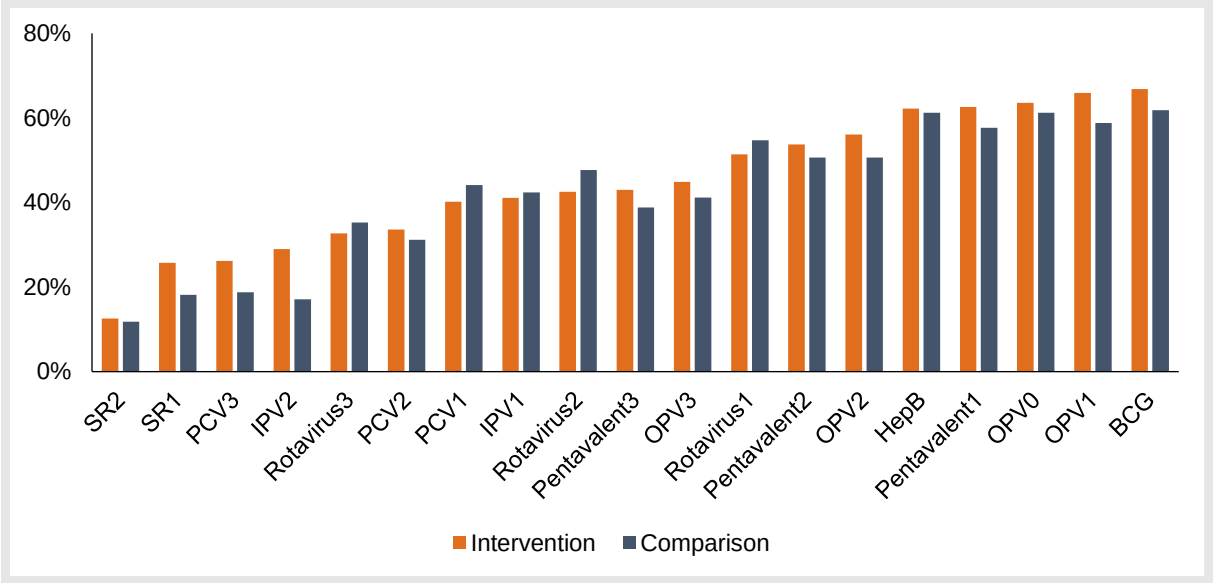
> *[Father 1:] It's not difficult [to wash hands with soap], but people don't want to do it. [Father 2:] It's difficult because people are not in the habit.*
>
> - FGD with fathers, Ainaro, Int. 14

[167] E.g., FGD with fathers, Ainaro, Int. 13; FGD with fathers, Ermera, Int. 16; FGD with fathers, Manatuto, Int. 19; FGD with mothers, Manatuto, Int. 32
[168] FGD with fathers, Ainaro, Int. 14

A mother in Ainaro echoed this, stating that "failing to wash hands is usually the norm."[169] These sentiments emphasize the importance of behavior change communication that instills handwashing as a regular habit, as well as the need to strengthen households' economic status to improve the use of soap in handwashing.

Lastly, at endline, mothers with children under the age of 2 were asked to show the vaccination record for their child. Figure 39 shows the prevalence of a variety of vaccinations. The most frequent vaccination was Bacille Calmette-Guérin (BCG) (tuberculosis) vaccine, which had been administered to 67% of intervention children and 62% of comparison children. Oral poliovirus vaccine (OPV) was also common among both intervention and comparison children, although only 45% of intervention children had received a full course of four oral poliovirus vaccines.

**Figure 39: Child vaccination prevalence**




On average, intervention children had received 8.5 vaccinations and comparison children had received 8. This falls below the approximately 25 vaccinations recommended by the United States Center for Disease Control for children under the age of 2; however, it also suggests that **most mothers are actively vaccinating their children**. From this analysis, two recommendations emerge. First, we recommend improving accessibility to vaccinations, as many mothers appear to view vaccinating children as a desirable practice but may only do so if vaccinations are convenient and available in an accessible location. Second, around 25% of children had no vaccinations. This may occur if mothers are either ignorant of the benefits of vaccinations, misinformed about their risks, or have very limited access to vaccinations, including due to long distances traveled on foot. As such, it may be useful to conduct further research to determine why some Timorese children do not receive any vaccinations, and then address these barriers.

## KNOWLEDGE OF HEALTH AND HYGIENE

Knowledge of good health and hygiene behaviors is one of six factors that are posited to contribute to increased use of health, nutrition, and dietary practices in the results framework. To measure knowledge of health and hygiene, in the household survey, caregivers were read lists and shown images of healthy and unhealthy hygiene practices and were asked to correctly identify healthy practices.

During fieldwork, some caregivers were observed to state that every practice (including irrelevant practices, such as playing soccer) was hygienic due to survey fatigue, as this group of questions came near the end

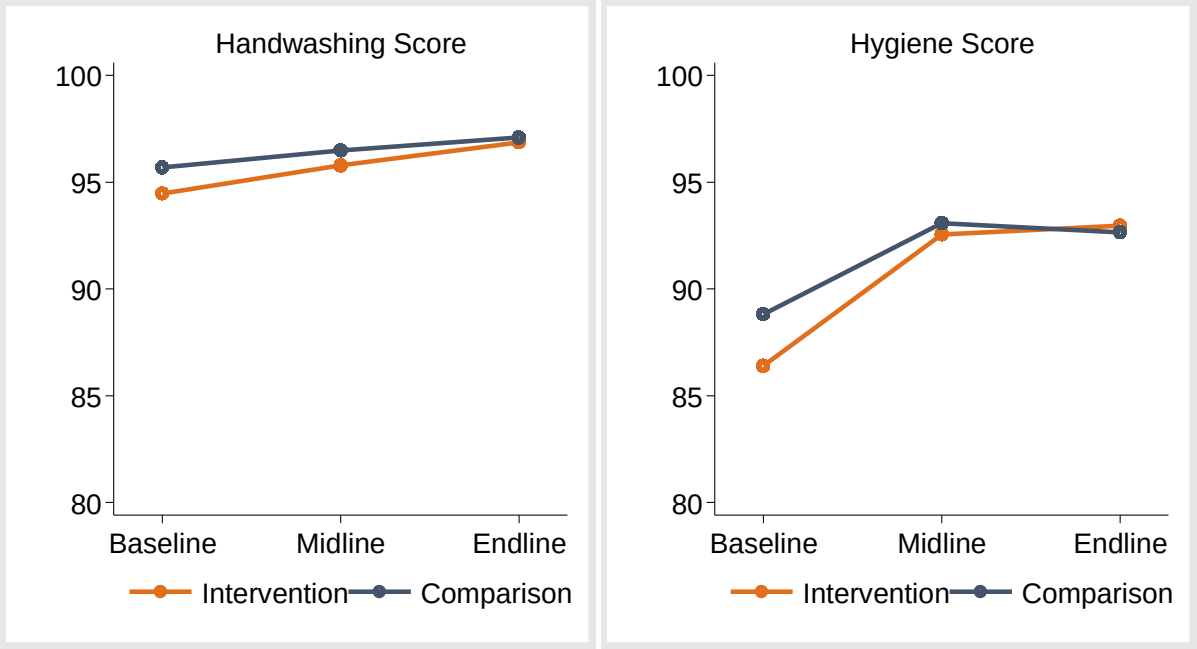
[169] FGD with mothers, Ainaro, Int. 25

of the survey. As such, we exclude caregivers who responded positively to every hygiene question from the sample. This excludes 41 baseline caregivers, 220 midline caregivers, and 139 endline caregivers.

We first create and analyze handwashing and hygiene scores. The handwashing score is based on caregivers' correct responses to eight questions asking them to identify situations before or after which they should wash their hands.[170] The hygiene score is based on caregivers' correct responses to 13 questions asking them to identify hygienic behaviors.[171]

The below figures show the change in handwashing and hygiene scores over time. We find that overall, **across all rounds, handwashing and hygiene scores have been high**—over 85% on average—suggesting strong knowledge of good hygienic practices. Scores have increased slightly over time; for handwashing, the average score for intervention caregivers increased by 2.4 percentage points between baseline and endline to an average of 96.9%, and for hygiene, the average score for intervention caregivers increased by 6.6 percentage points to an average of 93.0%.

**Figure 40: Change in handwashing and hygiene scores**




Analyzing the relative change in scores across rounds, we find no significant difference in handwashing score improvement among intervention caregivers compared to comparison caregivers. While the relative improvement in hygiene score is also not significant, we find substantial improvement among intervention caregivers for the hygiene score. For this score, we find that from baseline to endline, intervention caregivers' average scores had improved 2.7 percentage points more than expected given the results of comparison caregivers. These dynamics may, in part, reflect a ceiling effect, as around 85% of caregivers had a 100% handwashing score and around 36% had a 100% hygiene score at endline. However, overall,

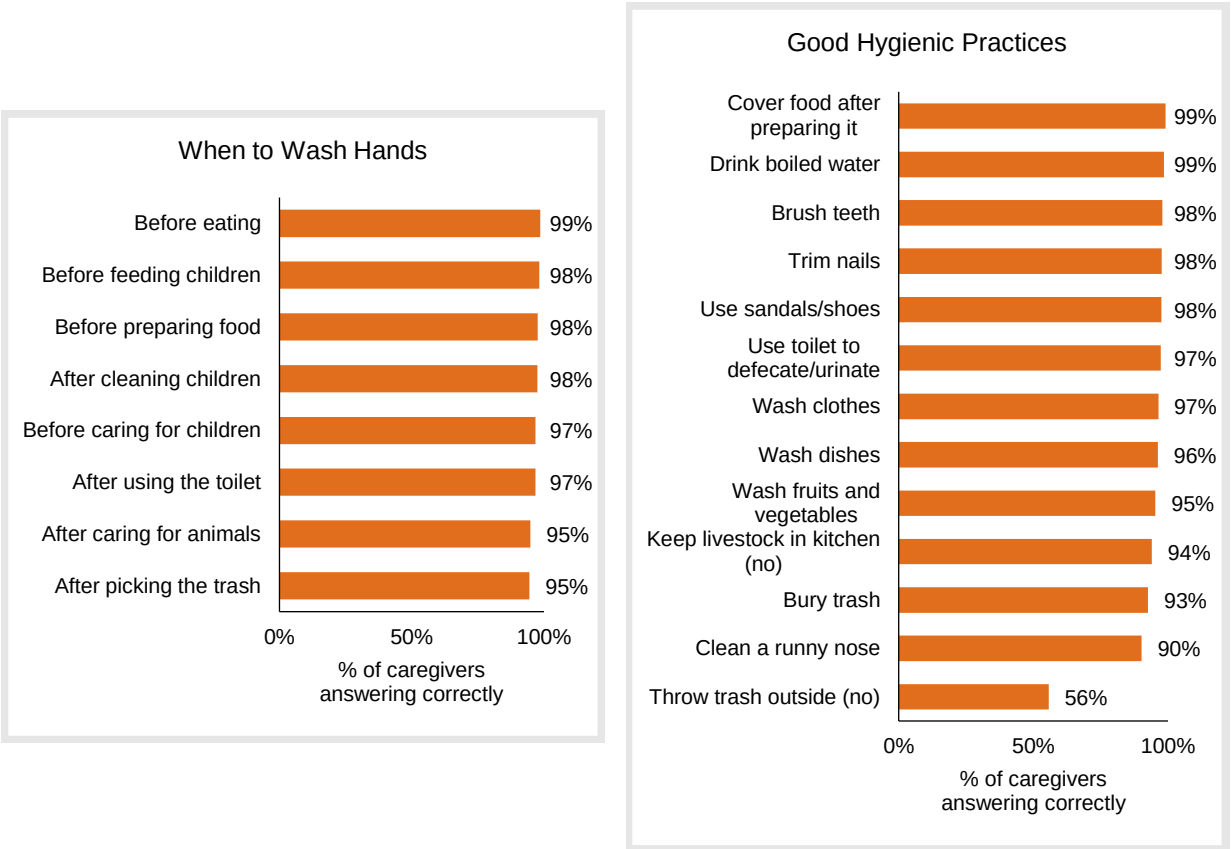
[170] Before caring for children, before eating, before feeding children, before preparing food, after caring for animals, after using the toilet, after picking the trash, and after cleaning children. Several "neutral" questions were also included that are not incorporated into the overall score; these include after reading books, after using the phone, and before work. If a caregiver answered "don't know," that was considered an incorrect answer.

[171] Drinking boiled water, brushing teeth, burying trash, covering food after preparing it, (not) keeping livestock in the kitchen, trimming nails, cleaning a runny nose, using sandals/shoes, using the toilet to defecate/urinate, (not) throwing trash outside, washing clothes, washing dishes, and washing fruits and vegetables. Several "neutral" questions were also included that are not incorporated into the overall score; these include writing, playing, and playing soccer.

they suggest both that **knowledge of hygienic practices is high and that there may have been a slight improvement in knowledge due to the HATUTAN program.**

Looking at specific hygienic practices, the below figure shows that for handwashing, the percent of intervention caregivers answering questions correctly was uniformly high at endline. Correct response rates were similarly high for most hygienic practices, with the exception of throwing trash outside, which only 56% of intervention caregivers stated was not hygienic. In a difference-in-difference analysis for the period from baseline to endline, we also find no significant changes in correct response rates to any question for intervention caregivers relative to comparison caregivers.

**Figure 41: Handwashing and hygiene knowledge, intervention caregivers at endline**




While qualitative interviews focused mainly on handwashing, analyzed in the above section, several parents also mentioned other hygienic practices. A mother in Manatuto, for example, stated that, "We need to put rubbish in a rubbish bin. Do not litter. This practice is good for our health."[172] Mothers in Ainaro also mentioned that it was important to use the toilet to defecate or urinate.[173]

Overall, these results suggest that knowledge is not a major barrier to healthy hygiene practices. We note, for example, that over 95% of caregivers stated that handwashing should occur before feeding children and after cleaning children. However, in the section above, we find that only 27.7% of intervention caregivers stated that they had washed their hands before feeding children the previous day, and only 11% stated that they had washed their hands after changing a diaper. This suggests that **while knowledge of handwashing practices is strong, actual behaviors remain weak.** As a result, it may be more effective

[172] FGD with mothers, Manatuto, Int. 32
[173] FGD with mothers, Ainaro, Int. 25

to pivot future programming away from a focus on knowledge and towards other potential barriers to behavior change. The following quote from an FGD in Manatuto provides support for this recommendation:

> *In this community, we have heard a lot about washing hands, and just recently from CARE. With this, in rural areas like this, we are aware that before having our meal we must wash our hands, but in reality, sometimes we forgot. When we are too hungry, we eat directly and we tend to forget washing our hands.*
>
> - FGD with fathers, Manatuto, Int. 20

Lastly, McGovern-Dole Custom Outcome #21 asks that caregivers identify at least 17 out of 19 healthy hygiene and handwashing practices. Table 56 shows that at endline, 91.5% of intervention caregivers met this standard, a large improvement since baseline and marginal decrease since midline.

**Table 56: Change in achievement of McGovern-Dole Custom Outcome #21**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| n | 156 | 865 | 953 | 328 | 522 | 534 | - | - |
| Identified at least 17 hygiene practices | 79.5% | 91.7% | 91.5% | 87.8% | 93.7% | 90.3% | 9.6 | 0.04 |

Analyzing the relative change in achievement of this outcome, we find significantly greater improvement among intervention caregivers from baseline to endline relative to comparison caregivers. Around 12 percentage points more intervention caregivers met this standard at endline than at baseline, while only around 2.5 percentage points more comparison caregivers met this standard. This suggests that **the HATUTAN program had a significant impact on achievement of McGovern-Dole Custom Outcome #21.**

# ACCESS TO CLEAN WATER AND SANITATION

In addition to health knowledge, the results framework identifies access to clean water and sanitation as a factor contributing to health, nutrition, and dietary practices. The HATUTAN program sought to increase access to clean water and sanitation, including access to an improved water source and improved sanitation facilities. In this section, we thus analyze the change over time in access to water and sanitation sources at the household and school levels.

## Households

At the household level, we first examine access to clean drinking water, followed by access to a toilet and handwashing station. Table 57 shows the change in access to an improved water source[174] by round and treatment group. We find that among both intervention and comparison households, access to an improved water source has remained high across all rounds, with over 90% of households stating that they obtain their water from an improved source in all rounds except midline (comparison households, 86.6%). There has been little change in access to an improved water source over the HATUTAN implementation period, however, and access among intervention households has slightly decreased from baseline to endline relative to comparison households.

[174] Piped water, public tap, borehole, protected dug well, rainwater harvesting, or trucked water.

**Table 57: Change in access to improved water source**

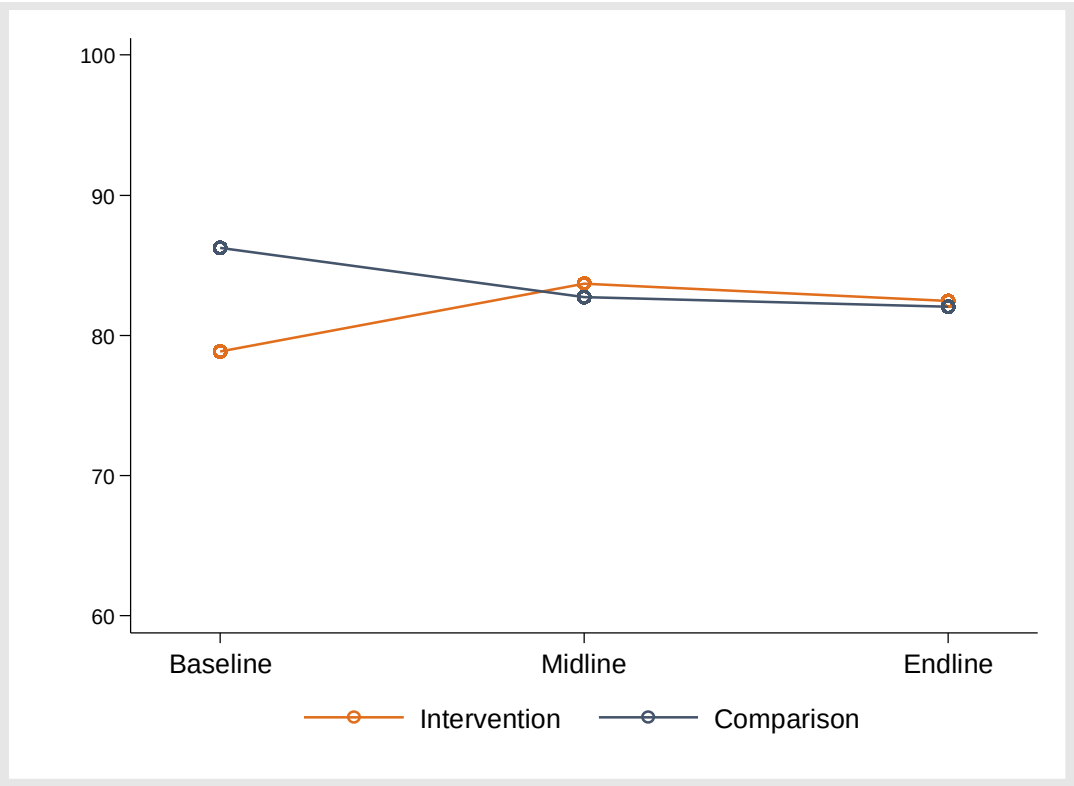
| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| n | 482 | 981 | 1,023 | 378 | 625 | 600 | - | - |
| Access to improved water source | 94.8% | 90.7% | 91.3% | 91.3% | 86.6% | 90.7% | -2.9 | 0.37 |
| Drinking water available all year | 34.7% | 44.9% | 47.1% | 51.3% | 41.9% | 55.5% | 8.3 | 0.26 |

**Although most households have access to an improved water source, reliable access to water all year round remains a challenge.** Less than half of intervention households and 55.5% of comparison households stated that drinking water was available all year. This indicator, however, improved from baseline to endline, and improved more substantially (though not significantly) for intervention households than comparison households. While these findings are not definitive, they suggest possible HATUTAN impact on access to drinking water; however, much more work is needed to further improve this statistic. Indeed, in an FGD, a father in Liquica stated that access to clean water could be a barrier to healthy practices:

> *[Children's] health is dependent on their parents. Parents should discipline their children and keep them clean, which means bathing, brushing their teeth, and wearing clean shoes and clothes to school. Our primary worry in this village, however, is the lack of clean water at certain times of the year, especially when it rains.*
>
> - FGD with fathers, Liquica, Int. 18

Looking now at toilet access, the below figure shows that **across all rounds, the majority of households had access to some type of toilet** (including a flush toilet or pit latrine). Access has increased for intervention households, from 78.8% at baseline to 82.4% at endline, and has decreased for comparison households, from 86.2% at baseline to 82.0% at endline. This relative improvement was substantial—intervention households' toilet access improved by 7.8 percentage points relative to comparison households—but not significant.

**Figure 42: Change in household access to a toilet**



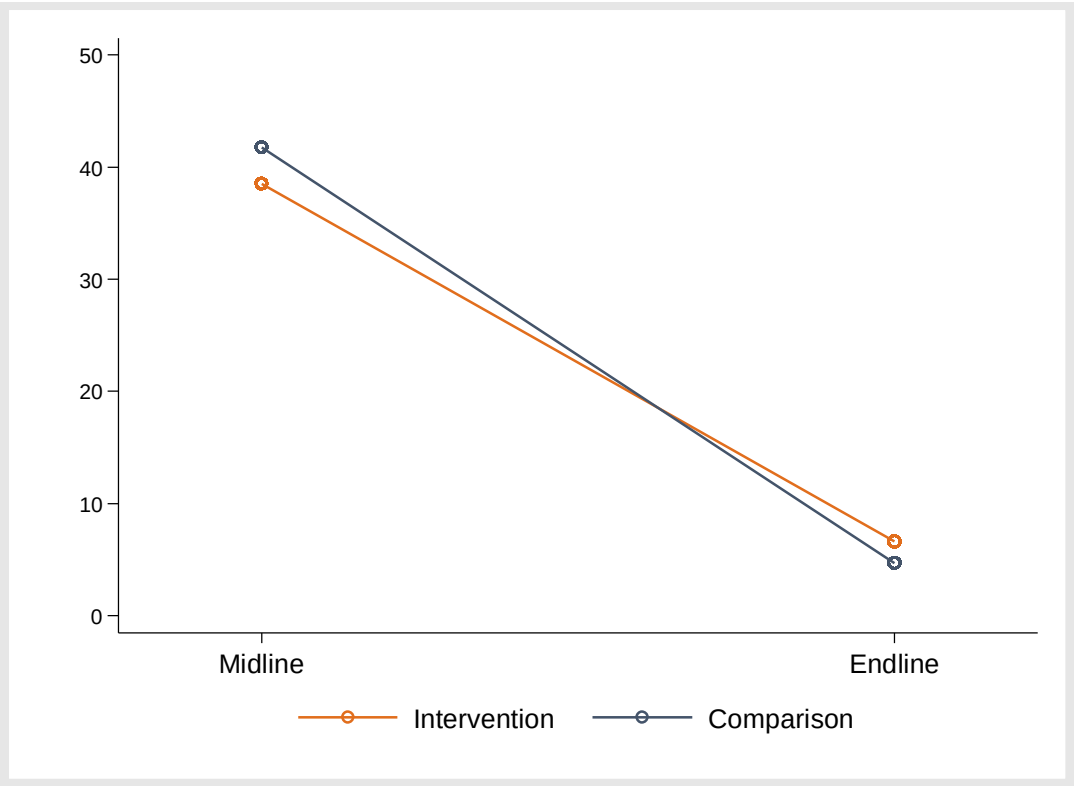

We also note that the most common toilet type across all rounds was pit latrines with slabs followed by uncovered pit latrines at baseline and endline. Use of "improved" toilet types, including improved pit latrines, composting latrines, and flush toilets, was relatively low; these three toilet types comprised only 17.3% of toilets used by intervention households at endline. While access to toilets is relatively high, there is thus further room to improve the quality of toilets used. In particular, **it may be helpful to target reduction in the use of uncovered pit latrines**; these latrines can result in the spread of faecal-borne diseases, such as cholera and diarrheal diseases, if flies are able to enter the latrine.

Lastly, we analyze households' access to handwashing stations and soap. This data was not collected at baseline; the below figure thus shows changes in access from midline to endline. **We find a significant decrease in household access to handwashing stations from midline to endline** across both intervention and comparison households. The decrease was somewhat more acute for comparison households; handwashing station access declined by 31.9 percentage points for intervention households and 37.1 percentage points for comparison households. This difference, however, was not significant.

In households with a handwashing station, enumerators were also asked to observe whether there was soap at the station.[175] There was a decline in soap availability from midline to endline, from 65.1% to 44.1% for intervention households and from 62.8% to 46.4% for comparison households. For intervention households, at endline, this meant that only 30 total households had a handwashing station with soap.

[175] Rather than ask, as answers would be highly vulnerable to social desirability bias.

**Figure 43: Change in household access to handwashing station**




These results generally point to a reduction in handwashing access from midline to endline, and corroborate findings regarding a decline in handwashing practices discussed in the above section *Health Practices.* As in that section, we hypothesize that the observed reduction in handwashing is likely due to a reduced emphasis on that practice as the COVID-19 pandemic has waned. In part, this is a positive finding, suggesting that COVID-19-induced initiatives to strengthen health practices were effective. However, it also shows a strong need to sustain such initiatives in order to ensure that handwashing, and other healthy practices, are instilled as life-long habits rather than short-term initiatives.

## Schools

We now look at toilet and handwashing station availability in schools.[176] Table 58 shows the change in the percent of schools with at least one toilet, the average number of toilets, and the average number of toilets available for female students. From baseline to endline, we find a substantial, though not significant, increase in the percent of intervention schools with at least one toilet available relative to comparison schools. Indeed, at endline, 86.7% of intervention schools had at least one toilet available for students, while only 69.1% of comparison schools had available toilets.

However, while the percent of schools with at least one toilet increased at endline, the average number of available toilets decreased. At endline, intervention schools had only an average of 2.2 toilets, a decline from an average of 2.5 toilets at baseline and midline. Comparison schools, meanwhile, had only an average of 1.5 toilets available at endline, a decline from an average of 2.4 toilets at baseline. While toilet availability declined less acutely in intervention schools than comparison schools, this relative difference was not significant. In other words, **we do not find evidence that the HATUTAN program had a significant impact on school toilet availability.**

[176] Access to clean water and kitchen handwashing stations are discussed in the *School Feeding Program* section.

**Table 58: Change in school toilet availability**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| n | 98 | 93 | 98 | 86 | 85 | 84 | - | - |
| At least one school toilet | 71.4% | 79.6% | 86.7% | 68.6% | 70.6% | 69.1% | 14.9 | 0.07 |
| Average number of school toilets | 2.5 | 2.5 | 2.2 | 2.4 | 1.9 | 1.5 | 0.6 | 0.10 |
| Average number of toilets for girls | 1.1 | 1.2 | 0.9 | 0.8 | 0.7 | 0.6 | -0.02 | 0.94 |

It is also important that schools have toilets available specifically for girls, as many girls may feel uncomfortable using toilets also used by boys or may face harassment to do so. At endline, in both intervention and comparison schools, we find less than one toilet available for girls on average. Indeed, at endline, 45.9% of intervention schools and 60.7% of comparison schools had no functional toilets available for female students. This represented a slight decline in toilet availability since baseline for intervention schools and a slight increase for comparison schools, but with no significant relative difference. Overall, this suggests that more efforts are needed to improve toilet availability, especially for female students, within schools.

Looking now at handwashing stations in schools, the below table shows change from midline to endline.[177] We find that availability of handwashing stations at schools has sharply declined at endline, similar to the decline found for households. There has been a similarly sharp decline in the percent of schools with soap available for at least 50% of handwashing stations. Declines in both variables were more severe for comparison schools than intervention schools, although there was no significant difference in the relative change in handwashing station availability by treatment group.[178]

**Table 59: Change in handwashing infrastructure at schools**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | ML | EL | Diff. | ML | EL | Diff. | DiD (ML to EL) | P-value |
| n | 95 | 98 | | 85 | 84 | | - | - |
| Handwashing station(s) for students | 86.3% | 59.2% | -27.1 | 87.1% | 40.5% | -46.6 | 18.7 | 0.18 |
| …with soap available (at least 50% of stations) | 74.4% | 34.5% | -39.9 | 83.8% | 17.7% | -66.1 | - | - |

[177] Data on handwashing stations was not collected at baseline.
[178] We do not calculate the difference-in-differences for soap availability due to low sample size.

As with households, this decline is likely driven by the end of COVID-19 campaigns encouraging frequent handwashing. While these campaigns were clearly effective, given the very high handwashing station availability at midline, the rapid decline in handwashing post-COVID is a clear sign that these campaigns were insufficient to instill handwashing as a life-long habit.

In qualitative interviews, school administrators were asked about students' handwashing practices at school and challenges facing handwashing. Several administrators stated that student habits were the primary challenge; for example, in Manatuto, one administrator stated the following:

> *The biggest difficulty [to handwashing] is that the students do not want to. There was water, a water tap, and soap, but the students did not want to wash their hands. When the teacher told the students to wash their hands before eating, to wash their hands before entering the classroom, the students did not want to wash their hands.*
>
> - KII with administrator, Manatuto, Int. 8

However, administrators in Ainaro stated that access to water was also a challenge to handwashing.[179] Additionally, in Ermera and Manatuto, purchasing soap was described as a challenge. An administrator in Ermera, for example, stated, "Even now there is no soap. Which teacher or school coordinator would spend their money to buy soap?"[180]

***Note from the program:*** The schools' concessional budget allows for purchase of hygiene supplies, including soap.

## ACCESS TO HEALTHCARE

Within the household survey, caregivers were asked about their ability to access healthcare and use of savings and loans for healthcare. Figure 44 shows the change in affordability of healthcare by evaluation round and treatment group. Healthcare was considered affordable if caregivers stated that they could afford the associated cost all or most of the time, or if they said there were no costs to access healthcare.[181]

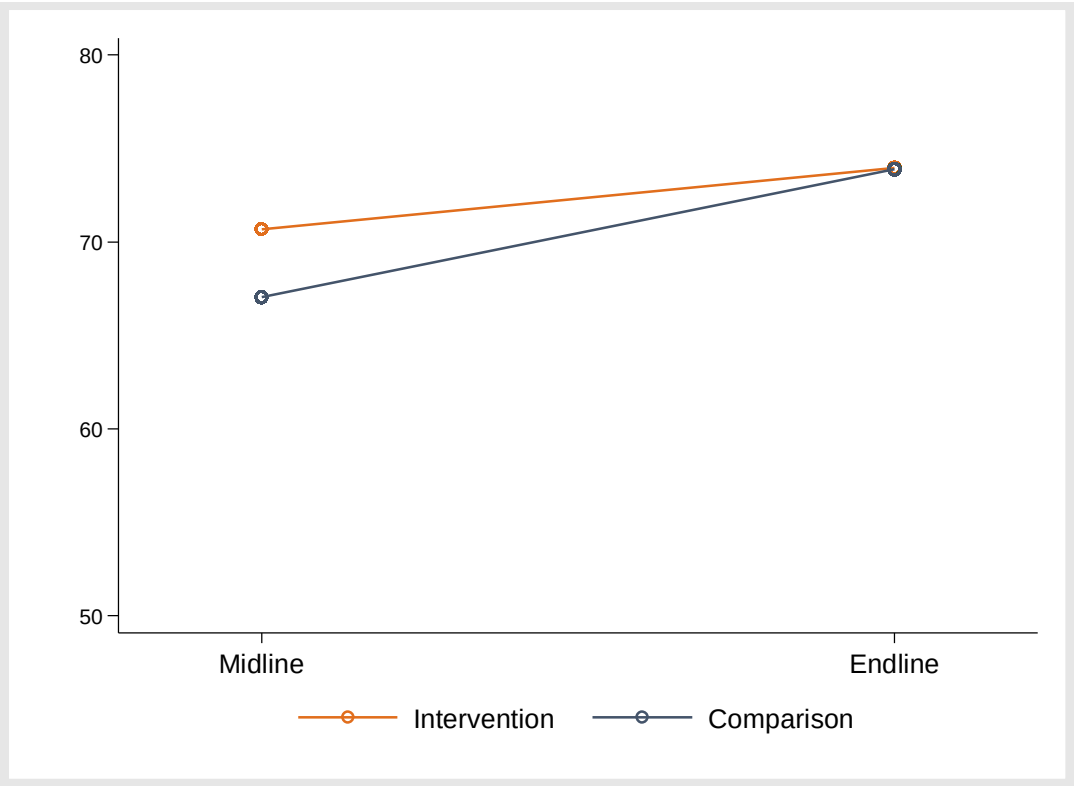
We first find that caregivers generally reported high levels of healthcare affordability. This was driven by a large percentage of caregivers who stated that there was no need to pay for healthcare; at endline, this option was selected by 63% of intervention caregivers and 67% of comparison caregivers. As medical care is free in Timor-Leste, the rate at which this option was selected is unsurprising; caregivers who did express challenges affording healthcare may visit private clinics or may be including costs such as transportation in their calculations.

[179] KII with administrator, Ainaro, Int. 1

[180] KII with administrator, Ermera, Int. 3; see also KII with administrator, Manatuto, Int. 7

[181] At baseline, this latter option was not included; we thus do not make comparisons to baseline here.

**Figure 44: Change in affordability of healthcare**




Healthcare affordability, however, saw relatively little change since midline for intervention caregivers, with only a 3.3 percentage point increase. Change was more substantial for comparison caregivers, but not significantly so. In other words, **we do not find evidence of a significant impact of HATUTAN on healthcare affordability.** We note, however, that affordability is likely not the only barrier to healthcare access; in Timor-Leste, the distance of health facilities from remote areas, access to transportation, and road accessibility all may significantly affect healthcare accessibility, and were not measured in HATUTAN evaluations. As such, the above analysis likely does not fully represent healthcare accessibility, nor HATUTAN's impact on it.

***Note from the program:*** Healthcare is provided free of charge in Timor-Leste's public clinics. The low expenditure in healthcare is likely to reflect its free provision.

To further analyze affordability of healthcare, we look at use of savings and loans as reported by heads of household (HoHs).[182] We find that at endline, intervention and comparison HoHs reported similar levels of savings and loans use for healthcare: Around 10 percent reported using savings for healthcare, and around 4 percent using loans. This was a reduction since baseline for both intervention and comparison households; the reduction was slightly greater for comparison households, but not significantly so.

[182] Other dimensions of savings and loans use are included in the section *Economic Empowerment.*

**Table 60: Change in use of savings and loans for healthcare**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| **Savings** | | | | | | | | |
| n | 233 | 646 | 671 | 202 | 377 | 316 | | |
| Used savings for healthcare | 15.0% | 19.7% | 9.1% | 19.3% | 16.5% | 10.1% | -3.2 | 0.75 |
| **Loans** | | | | | | | | |
| n | 134 | 178 | 244 | 78 | 50 | 46 | | |
| Used loan for healthcare | 14.9% | 3.9% | 4.5% | 18.0% | 6.0% | 4.4% | -3.2 | 0.56 |

Several factors may drive this pattern. First, this may be due to an increase in the affordability of healthcare; Figure 44 suggests that this likely is a factor. It may also, however, be due to decreasing prioritization of healthcare in HoH spending. This possibility is explored more in the section *Economic Empowerment.*

# PREDICTORS OF HEALTH AND NUTRITION

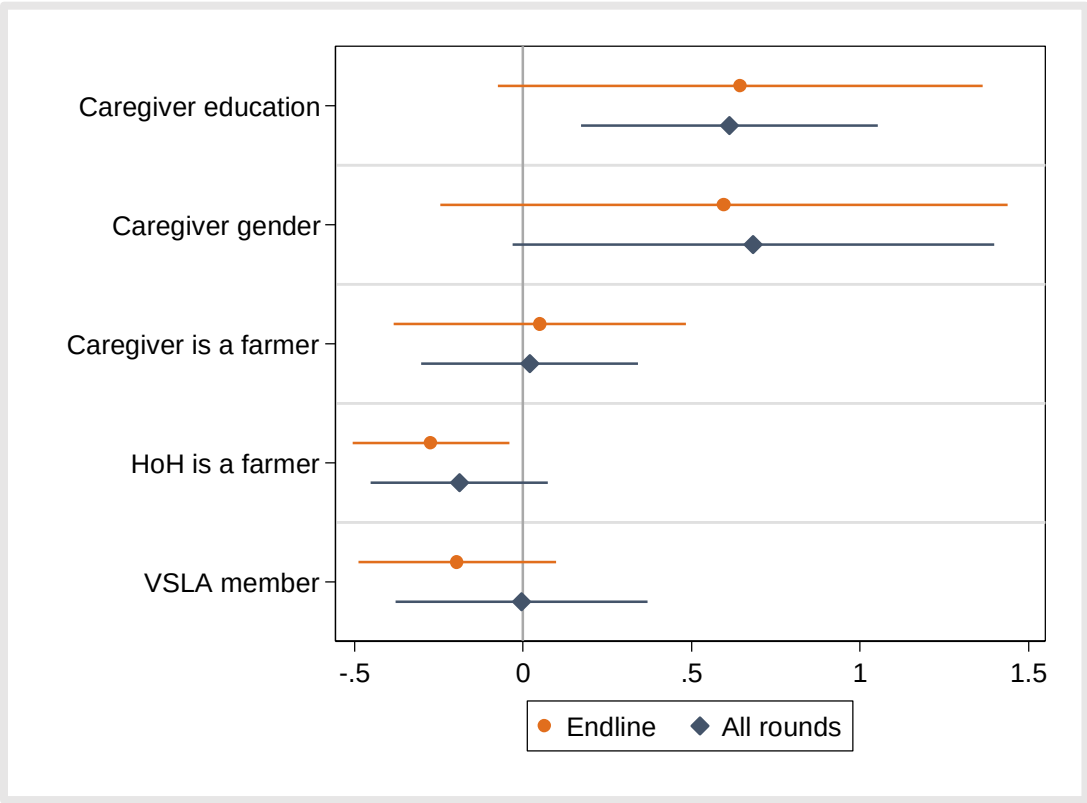
This section analyzes the relationship between various household characteristics and caregivers' nutrition knowledge, dietary diversity, and health knowledge. Within each predictive model, we include variables for caregiver education and gender. We also include additional variables for factors that may affect—or be affected by—each outcome of interest, discussed more below.

## Predictors of Nutrition Knowledge

To better understand dynamics around nutrition knowledge, in addition to caregiver education and gender, we include variables to control for whether the caregiver or HoH is a farmer and whether the household participates in a VSLA (as health and nutrition information was disseminated to caregivers through VSLAs). We also include municipality fixed effects to control for variables (such as treatment status) that vary at the municipality level. We hypothesize that farmers may have relatively more knowledge of nutrition due to their knowledge of agriculture. We also hypothesize that households who participate in VSLAs may be more exposed to HATUTAN trainings and also have more savings; as a result, they may have more nutrition knowledge as wealthier households are likely to have more ability to make nutritional choices, and thus more need for knowledge of good nutrition practices.

Figure 45 shows results for predictors of nutrition knowledge. At endline, we find only one significant relationship: Whether the HoH is a farmer is negatively correlated with nutrition knowledge. While this goes against our hypothesis, it may be due to cross-correlation between HoH occupation and economic status; indeed, we find that whether the HoH is a farmer is significantly and negatively correlated with household savings. However, when the HoH occupation variable is removed from the regression, we still do not find a significant relationship between VSLA participation or savings and nutrition knowledge, suggesting that some other dynamic may also be at play.

**Figure 45: Predictors of nutrition knowledge**




When we expand our analysis to include all rounds of data collection, meanwhile, we also find a significant relationship between caregiver education and nutrition knowledge. The above figure shows that more highly educated caregivers have significantly higher nutrition knowledge, all else held constant. This result, unlike that for HoH occupation, are unsurprising; we expect educated caregivers to also be more educated about nutrition, either through learning about nutrition in school or because they are more likely to read.

Lastly, we note that we do not find a significant relationship between VSLA membership and nutrition knowledge. This suggests that VSLAs may not be an effective platform for learning about nutrition, or that caregivers also learn about nutrition through mechanisms other than VSLA participation.

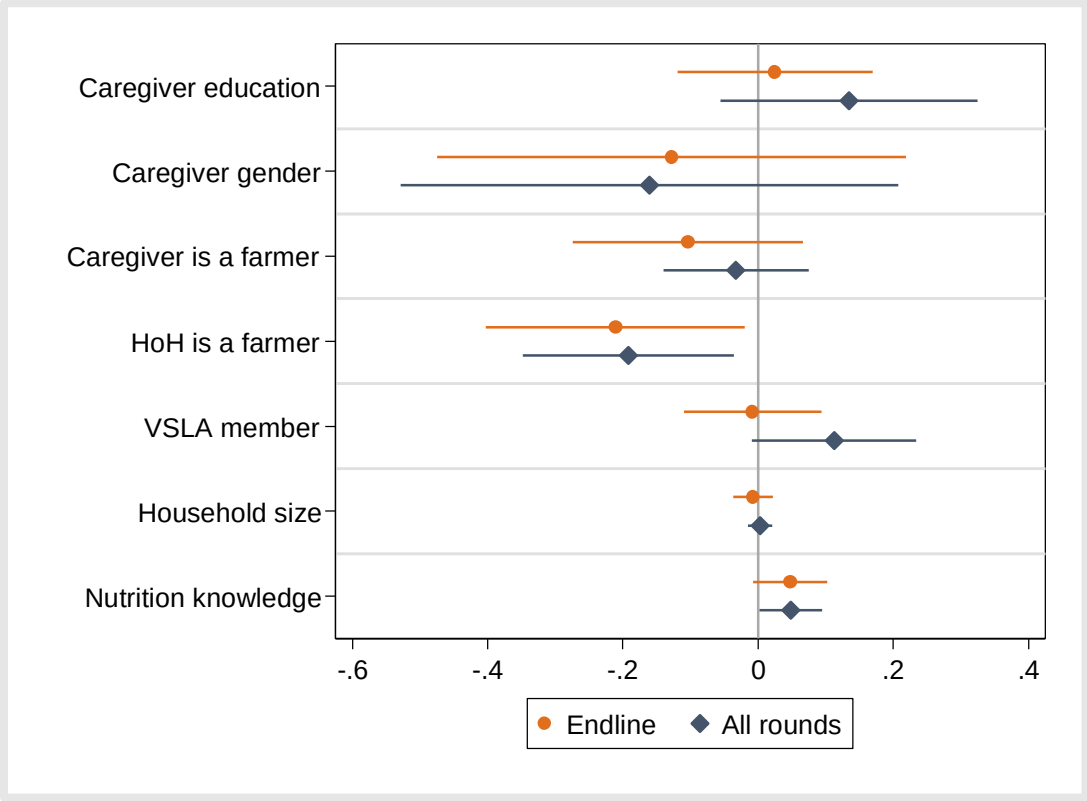
## Predictors of Dietary Diversity

Looking now at caregiver dietary diversity, we include all variables in the above regression (including municipality fixed effects) and hypothesize similar relationships. We also include variables controlling for the number of household members and nutrition knowledge. For household members, we hypothesize that larger households will have lower dietary diversity due to the need to feed more people, and thus the need to purchase mainly low-cost foods. For nutrition knowledge, we hypothesize that greater knowledge will correlate with higher dietary diversity due to the increased importance placed on achieving a nutritional diet.

At endline, in contrast to our hypotheses, we find only one significant relationship between any of the included variables and caregiver dietary diversity: whether the HoH is a farmer is correlated with lower caregiver dietary diversity. When we include data from all rounds, this result remains significant, and we also find a small but significant and positive correlation between nutrition knowledge and dietary diversity. The latter relationship is not surprising, as more knowledgeable caregivers may be more likely to prioritize dietary diversity. However, the former relationship—as in *Predictors of Nutrition Knowledge*—contradicts our hypothesis; we would expect that households involved in farming would have higher dietary diversity,

as they would be able to supplement their diets with ingredients produced on their farm. However, this relationship may be due to the timing of data collection during the Timor-Leste "lean season." As above, we note that removing this variable from the regression does not result in a significant relationship between VSLA membership or savings (i.e., economic status) and dietary diversity, suggesting the relationship between HoH occupation and dietary diversity is not solely explainable by a correlation with economic status.

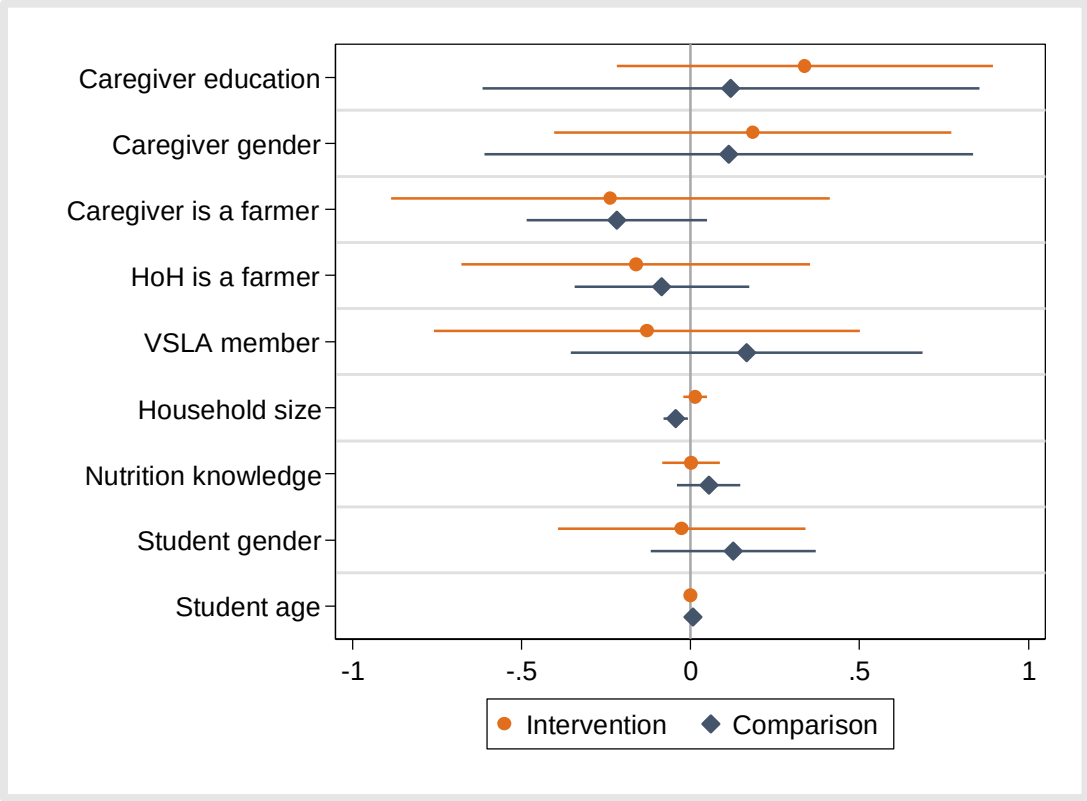
**Figure 46: Predictors of caregiver dietary diversity**




To supplement this analysis, we also analyze predictors of student dietary diversity. We create a student dietary diversity score based on both caregiver- and student-reported food consumption, excluding processed and sugary foods from the total score. In our regression, we then include the above variables as well as variables for student gender and age.

The below figure shows that at endline, we find no significant relationships between the included variables and student dietary diversity for intervention students. For comparison students, we find only one significant relationship: Larger household size is significantly correlated with lower student dietary diversity. This finding is unsurprising, given that we would expect resources to be more thinly spread within larger households.

**Figure 47: Predictors of student dietary diversity**



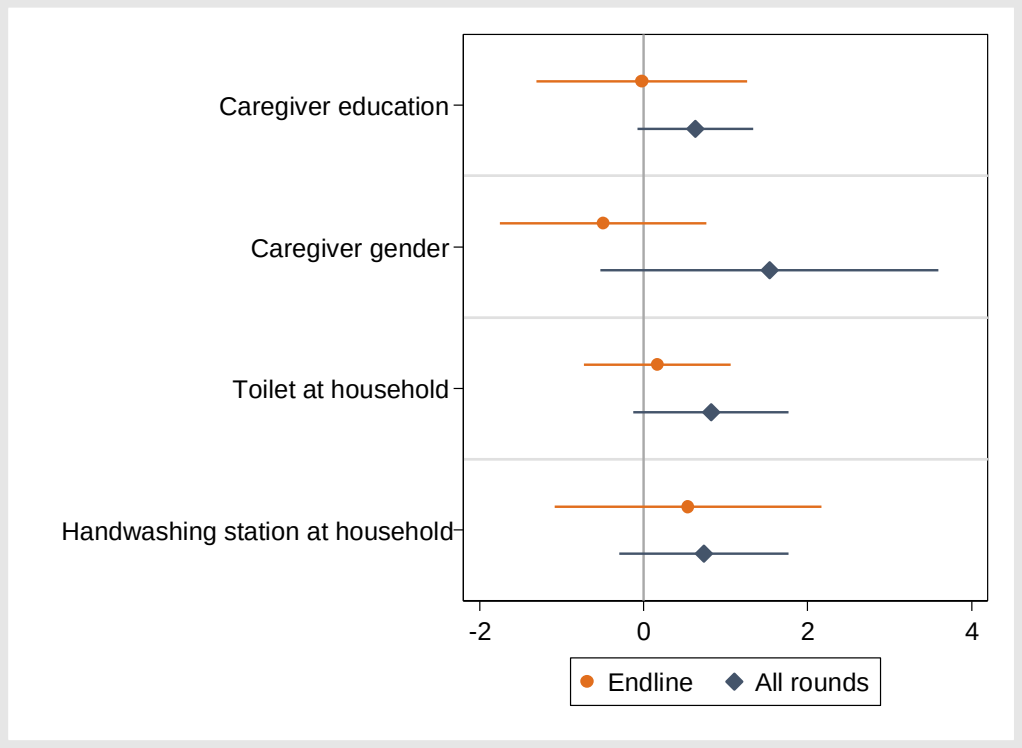

Overall, these findings suggest that economic well-being, nutrition knowledge, and household size may affect dietary diversity of households and students. However, we also note that we find no significant correlations between other variables of interest, including VSLA membership, and dietary diversity.

## Predictors of Health Knowledge

Lastly, we examine predictors of health knowledge, combining calculated handwashing and hygiene scores into one overall health knowledge score. In addition to caregiver education and gender, we include variables for whether the household has savings, whether there is a toilet available at the house, and whether there is a handwashing station available at the house. For these latter two variables, the relationship may operate in reverse; in other words, if we find a significant relationship between toilet availability and health knowledge, it may actually be the case that higher levels of health knowledge predict toilet access, as caregivers with more health knowledge may place greater priority on having a toilet at home.

The figure below shows that both for endline and all rounds, we find no significant predictors of health knowledge. We note, however, that this may be due to the fact that there is relatively little variance in health knowledge scores, with 81% of caregivers across all rounds scoring 90% or above. As such, the (lack of) statistically significant relationships found in this analysis should not be taken as a definitive sign that relationships do not exist.

**Figure 48: Predictors of health knowledge**




# AGRICULTURAL PRACTICES

In this section we focus on agricultural practices. The main practice of interest is the cultivation of a permagarden: We aim to assess if the program has made an impact in permagarden cultivation from midline to endline, what challenges farmers face in cultivating permagardens, and the effect that cultivating a permagarden may have on nutritional and economic outcomes. The HATUTAN program aimed to support farmers through trainings in adopting these techniques to boost yields and create sustainable sources of food.

No data was collected on permagarden cultivation at baseline, so the comparison over time used is from midline to endline. This may have the effect of making any potential impacts more difficult to detect, as there has been less time for practices to be taken up and for their effect to be felt. The sample is made up of households of grade 2 students that had farmers, as well as a separate group of farmers sampled independently from the household survey. The sample size overall is still relatively small for many of the indicators of interest, particularly in the comparison group, which limits the analyses that are possible.

## PERMAGARDEN USE

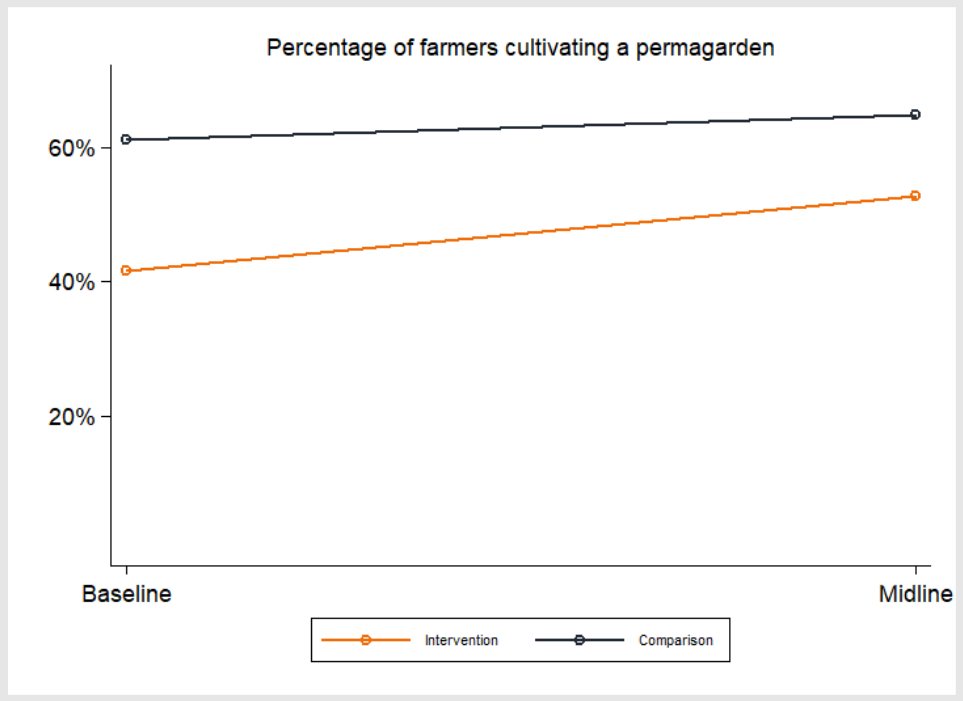
Table 61 shows change across rounds and treatment group in permagarden usage. For the intervention group receiving HATUTAN training 52.7% of the sample were currently cultivating a permagarden. This is up from 41.6% at the midline, an 11.1 percentage point increase overall. Compared to a rise of 3.6 percentage points in the comparison group, this gives a reasonably high difference in difference estimate of 7.4 percentage points, primarily due to how much more permagarden use rose in the intervention group compared to the comparison group.

**Table 61: Change in permagarden use**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| n | 30 | 78 | 18 | 22 | - | - |
| Permagarden usage (%) | 41.6 | 52.7 | 61.1 | 64.7 | 7.4 | 0.64 |

We can see this effect more clearly in Figure 49; the line tracking use of permagardens over time rose for the comparison group, but rose more steeply for the intervention group. However, since the finding is not statistically significant, we cannot decisively conclude that the program was effective in supporting farmers to take up permagardening. Given that HATUTAN is one of several interventions in Timor-Leste supporting permagarden adoption, this result is not surprising.

**Figure 49: Change in permagarden use**




To better understand the rise in permagarden cultivation in the intervention group, we disaggregate the overall increase by municipality. This will allow us to see whether the rise is evenly spread and broad-based, or whether much of the rise is account for by a single municipality. We can see that most municipalities saw an even increase in cultivation of permagardens; Ainaro, Ermera, and Manatuto saw rises of approximately 20 percentage points, with a fall in Liquica of 5 percentage points. Overall, the relatively even rise can be taken as some evidence of program impact: The increase wasn't accounted for by particular circumstances in one location but was seen independently across several. However, the fall in Liquica, from a low starting point, is a cause for concern.

**Table 62: Change in permagarden use by municipality**

| | Midline | Endline | Change |
|---|---|---|---|
| Ainaro | 58.8% | 78.5% | 19.7 |
| Ermera | 33.3% | 54.5% | 21.2 |
| Liquica | 27.3% | 22.4% | -4.9 |
| Manatuto | 43.5% | 66.7% | 23.2 |

## CHALLENGES FACED BY FARMERS

In this section we analyse the challenges that farmers faced while cultivating permagardens. This question was only asked of those cultivating a permagarden. In addition, questions about specific challenges were only asked of those with permagarden who said that they faced challenges. Because of this, and because of the large number of categories of interest, the sample size is quite small, reducing our ability to draw conclusions about statistical significance.

A higher proportion of farmers in the intervention group faced challenges compared to those in the comparison group, at 62.3% compared to 43.1%; the difference is statistically significant. Of the specific challenges, across both groups the common main challenges were pests, natural disasters, limited production, and lack of tools and materials. Rarer challenges cited were poor quality produce, limited amount of land, and personal issues.

**Table 63: Challenges faced by farmers at endline**

| | Comparison | Intervention | Difference |
|---|---|---|---|
| n | 51 | 212 | - |
| Face any challenges (% yes) | 43.1% | 62.3% | 19.2* |
| **Type of challenges** | | | |
| n | 22 | 132 | - |
| Natural disaster | 27.3% | 21.2% | -6.1 |
| Limited production | 18.2% | 13.6% | -4.6 |
| Poor quality of seeds | 4.6% | 8.3% | -3.7 |
| Damaged / stolen produce | 9.1% | 3.8% | -5.3 |
| Unable to sell | 13.6% | 1.5% | -12.1 |
| Lack of tools, materials, seeds | 18.2% | 16.7% | -1.5 |
| Personal issues | 4.5% | 3.0% | -1.5 |
| Poor quality of produce | 0.0% | 4.6% | 4.5* |
| Limited amount of land | 0.0% | 6.0% | 6* |
| Lack of technical support | 18.2% | 9.9% | -8.3 |
| Lack of money to invest | 18.2% | 3.8% | -14.4 |
| Pests | 36.4% | 35.6% | -0.8 |
| Other | 9.1% | 31.8% | 22.7* |

There were some differences in specific challenges across the groups. The comparison group were much more likely to be unable to sell produce, at 13.6% compared to 1.5% in the intervention group. Farmers in the intervention group were more likely to suffer from poor quality produce and a limited amount of land and

the results were statistically significant, although the number facing this challenge was still quite small. Farmers in the intervention were also less likely to face the challenges of lack of technical support and lack of money to invest, which may be a result of the support for permagardens offered by HATUTAN.

In general, in the intervention group farmers are less likely to have faced most of the individual challenges outlined in the survey, despite being more likely to have faced challenges overall. This is captured in the "Other" challenges, which intervention farmers are much more likely to have faced. Therefore there is some factor facing farmers benefitting from the HATUTAN program that is not yet fully understood from the data.

## ASSOCIATION WITH NUTRITIONAL AND ECONOMIC OUTCOMES

In this section we analyse whether permagarden usage is associated with various outcomes of interest. The two domains are nutritional outcomes and economic outcomes. HATUTAN aims to encourage and support permagarden use based on the assumption that it should lead to a more nutritious and secure food supply, as well as opportunities to sell excess produce for profit. In this case, we might expect to see improved nutritional and economic outcomes for those cultivating permagardens compared to those who do not cultivate permagardens.

It is important to note that in this section we are unable to fully assess causality; due to the sample size and the structure of the questions the analyses possible are limited. If differences are found between the groups it can only be taken as an indicative finding, as there may be other characteristics between the groups that we can't control for that may account for the difference – for example, if average income happens to be lower in one of the groups we might expect worse economic outcomes, unrelated to the practice of cultivating a permagarden.

In the table below we analyse how permagarden usage is associated with nutritional outcomes. The two indicators of interest are whether the student in the household had eaten on the day of the EGRA test, and whether anyone in the household had gone a day without eating in the previous 30 days. We note that sample size for the former indicator is very low, with only 53 intervention respondents at endline who both answered the question on permagardens and had a grade 2 child.

The results show a small but positive potential impact of permagardening. Students in households cultivating a permagarden were slightly more likely to have eaten on the day of the test, at 78.1% compared to 76.2% in those who lived in households not cultivating a permagarden. Households cultivating a permagarden were also less likely to have had anyone in the household go without food in the past 30 days, which may point to a steadier supply of food. It is important to note here, however, that neither result was statistically significant.

**Table 64: Difference in nutritional outcomes by permagarden use**

| | Didn't have permagarden | Did have permagarden | Difference |
|---|---|---|---|
| Student eaten on test day | 76.2% | 78.1% | 1.9 |
| Household not eaten | 8.6% | 6.4% | -2.2 |

Table 65 below summarises the results of how the economic outcomes vary by permagarden cultivation. Households cultivating permagardens were less likely to have savings (by 10.5 percentage points), more likely to use savings for investment (by 3.6 percentage points), and significantly more likely to use savings on debt (by 6.5 percentage points). Overall, this paints a mixed picture of households cultivating permagardens – on these indicators they are less likely to have savings than houses without permagardens but more likely to use those savings for investment or to pay off debt. Overall, then, it is possible that some potential positive economic impacts of permagarden cultivation have not materialised in the program.

Alternatively, it may be that households facing worse economic conditions are more likely to cultivate permagardens or be targeted by the HATUTAN program. Along these lines, we do find that farmers cultivating permagardens were relatively more likely to have been affected by natural disasters than farmers who were not cultivating a permagarden; this may explain some of the results found in this section.

**Table 65: Difference in economic outcomes by permagarden use**

| | Didn't have permagarden | Did have permagarden | Difference |
|---|---|---|---|
| Household has savings | 90.0% | 79.5% | -10.5 |
| Uses savings on investment (household or business) | 25.4% | 29.0% | 3.6 |
| Uses savings on debt (household or business) | 0.0% | 6.5% | 6.5* |

## FARM SALES

We now focus on farm sales. At endline we aim to understand trends in farmers selling their crops for profit, including what percentage sell crops for profit rather than just for household consumption, what percentage of crops they sell, and whether any profit is made. The HATUTAN program aimed to support farmers in VSLAs directly through training and therefore a potential benefit is that farmers are more profitable while using improved, sustainable agricultural practices.

We first analyze whether households grew crops for household consumption and what proportion of those crops were sold (for those who did grow crops for consumption). The table below summarizes the results over time for whether or not a household sold any crops. This score rose for both the intervention group, from 69.2% to 72.2%, and the comparison group. This resulted in a difference in difference score of -0.8 percentage. As noted in the introduction, the sample size is very small and the rise seen in the comparison group over time in particular is likely more reflective of a chance change rather than a real trend, as reflected in the high p-value and the lack of statistical significance.

**Table 66: Change in households selling crops**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| n | 26 | 54 | 4 | 13 | - | - |
| Sold any crops | 69.2% | 72.2% | 50.0% | 53.9% | -0.8 | 0.98 |

Next, we look at the proportion of crops sold by farmers. This question was only asked of those who answered in the previous question that they had grown crops for sale and not just for household consumption. For the comparison group, sample size is too small for analysis. For the intervention group, the proportion of crops sold declined from an average of 50.6% at midline to 45.4% at endline. This may suggest a weakening of food security, with farmers relying more on their own crops for sustenance rather than for income.

Lastly, we analyze whether farmers made any profit from the crops they sold in the previous six months. As above, sample size within the comparison group is too small for analysis. Among intervention farmers, the first thing to note is that the vast majority reported making a profit, ranging from 94.4% at midline to 94.7% at endline. There was, however, almost no change in profitability from midline to endline.

To gain further understanding of the intervention group and the flat trend, the change is disaggregated by municipality. This will allow us to check whether there are any trends in the different locations that are masked when aggregated with other locations. The results are summarised below in Table 4. It is clear that there are no trends specific to any of the municipalities. Two municipalities – Ermera and Liquica – saw no change and stayed at 100%, while there were very small changes in Ainaro and Manatuto.

**Table 67: Change in percentage making a profit from crop sales, by municipality**

| | ML | EL | Difference |
|---|---|---|---|
| Ainaro | 100.0% | 92.9% | -7.1 |
| Ermera | 100.0% | 100.0% | 0.0 |
| Manatuto | 83.3% | 90.0% | 6.7 |
| Liquica | 100.0% | 100.0% | 0.0 |

# ECONOMIC EMPOWERMENT

In this section we analyze economic empowerment by looking at activity around savings and loan use. There are several ways in which the program is expected to be beneficial for economic empowerment, including spillovers from other aspects of the program, increased capacity of government institutions and government support, the promotion of gender equality and better intra-household decision-making, and through the promotion of VSLAs.

We first assess the impact on whether a household has savings, where savings are kept, and what savings are used for. We then focus on VSLAs, including how many members of VSLAs had taken loans and what the loans were used for. On both savings and loan use, there is a particular focus on use on education. Finally, we will study the association between VSLA membership and nutritional and economic outcomes.

## SAVINGS

Table 68 shows the results of difference in difference regressions, both for whether a household has savings, and for where those savings are kept. We focus on changes from midline to endline due to limited data from baseline. For the intervention group, the percentage of households with savings was effectively unchanged from midline to endline, falling 0.4 percentage points from 65.8% to 65.4%. However, the percentage with savings in the comparison group fell sharply from 60.4% to 52.6%, giving an overall difference in difference estimate of 7.5. This result is statistically significant. This can be interpreted as a positive finding for the HATUTAN program; taking the comparison group as representing what would have happened to the intervention group without the program we can see that they are in a better position by staying flat over the period.

Analyzing the results for where savings are kept, we can see that all results aside from VSLAs are small in effect size and not statistically significant. The percentage keeping their savings in VSLAs rose for both intervention and comparison groups, but rose more in the intervention group and from a higher base. The difference in difference estimate is 7.9 but is not significant. This can also be taken as a positive finding from the program, which explicitly aims to encourage VSLA membership. Increased savings in VLSAs may also partly explain the finding above of whether a household has any savings.

**Table 68: Change in prevalence and location of savings**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| Household has savings | 65.8% | 65.4% | 60.4% | 52.6% | 7.5 | 0.03* |
| **Location of savings** | | | | | | |
| VSLA | 43.3% | 56.3% | 26.0% | 31.0% | 7.9 | 0.08 |
| Microfinance group | 2.6% | 2.1% | 1.9% | 1.3% | 0.04 | 0.97 |
| Kept at home | 53.3% | 49.5% | 71.6% | 70.3% | -2.4 | 0.60 |
| Kept at a bank | 7.1% | 3.9% | 6.9% | 5.1% | -1.4 | 0.52 |
| Other | 0.3% | 0.6% | 0.5% | 1.6% | -0.7 | 0.31 |

The table below shows the results of difference in difference regression for uses of savings. This question was only asked of those who responded positively to having savings. We note that across all categories and for both treatment and comparison, the percentage using their savings fell from midline to endline. There are numerous possible causes for this but one might be a general improvement of post-COVID economic conditions, allowing households to hold on to their savings rather than spending savings. However, this contrasts with the finding above about fewer households having any savings.

**Table 69: Change in use of savings**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| Food | 90.7% | 87.5% | 94.2% | 92.1% | -1.1 | 0.68 |
| Healthcare | 19.7% | 9.1% | 16.5% | 10.1% | -4.2 | 0.19 |
| Education | 87.5% | 85.3% | 91.3% | 90.2% | -1.1 | 0.71 |
| Agriculture | 23.2% | 12.1% | 20.7% | 13.9% | -4.4 | 0.22 |
| Business | 22.6% | 20.6% | 19.9% | 16.8% | 1.1 | 0.78 |
| Debt | 9.8% | 3.7% | 8.5% | 3.8% | -1.3 | 0.57 |
| Investment | 9.9% | 5.3% | 9.6% | 4.4% | 0.5 | 0.82 |
| Building | 22.9% | 20.4% | 28.1% | 19.3% | 6.3 | 0.11 |

The difference in difference scores, however, are relatively small in effect size and are not statistically significant. Many savings uses, including both healthcare and agriculture, have negative difference in difference scores, meaning that the percentage using their savings on them fell further in the intervention group and the comparison group. In contrast, the difference in difference score for using savings on building work was 6.3, meaning it fell further in the comparison group than in those taking part in the HATUTAN program. Overall, it is difficult to glean useful findings from this analysis, aside from noting that the desired effect of increased use of savings on education did not occur, perhaps due to the "ceiling effect" whereby the majority of households already spent savings on education at midline. However, we note that at endline within intervention households, we find a significant and positive relationship between VSLA membership and student attendance: Students in households who participated in VSLAs reported missing 0.5 fewer days of school the previous week, on average. This suggests that VSLA membership does indeed have an impact on education, whether through use of savings for schooling or other mechanisms such as greater sensitization on the importance of education.

## VSLA LOANS

In this section we focus on VSLAs, and in particular on those taking out a loan from a VSLA. As noted above, being a member of a VSLA and VSLA activity is encouraged in the HATUTAN program. Table 70 summarizes the results of a difference in difference regression on whether household in a VLSA had taken out a VSLA loan. In the intervention group there was a 3.9 percentage point rise in taking out a loan from midline to endline, while in the comparison group there was a 4.6 percentage point drop over the same time period. This leads to a difference in difference estimate of 9.5 percentage points, which is reasonably high, although the finding is not statistically significant. This can be understood as a positive finding from the program as it does support the uptake of VSLA activity, although increased use of loans can also be seen as an indicator of financial distress.

**Table 70: Change in households taking loans from VSLAs**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| n | 291 | 375 | 96 | 99 | - | - |
| Taken a VSLA loan | 61.2% | 65.1% | 52.1% | 46.5% | 9.5 | 0.23 |

In the table below we can see the results for difference in difference regressions on the use of VSLA loans. As with the use of savings in the previous section, the percentage of households using loans on nearly all of the items saw a fall over the period from midline to endline, and the difference in difference scores are mostly small and statistically insignificant. There are a few exceptions to this, however. While the percentage using loans for business fell in the intervention group, this same percentage rose sharply in the comparison group, leading to a large and statistically significant score of -20.8. The particularly high score is likely to be due to some unexplained factor in the comparison group, rather than reflecting particularly poor performance for those in the HATUTAN program. However, it might still be of concern that there was a fall in the intervention group, as business investment can be seen as a sign of financial health – a falling score might suggest a worsening financial situation. Another exception is the rise in households spending loans on funerals from 4.5% to 11.9% in the intervention group, leading to a statistically significant score of 16.3. The reasons for this are unclear, but again it is unlikely to reflect on anything to do with the program. The score for loan use on education was small and insignificant.

**Table 71: Difference in difference regression for VSLA loan use**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| | ML | EL | ML | EL | DiD | P-value |
| Food | 62.4% | 60.7% | 66.0% | 60.9% | 3.4 | 0.76 |
| Healthcare | 3.9% | 4.5% | 6.0% | 4.4% | 2.2 | 0.64 |
| Education | 71.9% | 65.6% | 70.0% | 58.7% | 5.0 | 0.64 |
| Agriculture | 12.9% | 7.4% | 4.0% | 4.4% | -5.8 | 0.36 |
| Business | 34.3% | 27.9% | 16.0% | 30.4% | -20.8 | 0.04* |
| Funeral | 4.5% | 11.9% | 22.0% | 13.0% | 16.3 | 0.02* |
| Debt | 3.4% | 0.8% | 4.0% | 3.9% | 1.5 | 0.64 |
| Building | 1.7% | 2.1% | 6.0% | 4.4% | 2.0 | 0.57 |
| Investment | 3.4% | 5.7% | 8.0% | 10.9% | -0.5 | 0.92 |

## VSLA MEMBERSHIP, NUTRITION, AND INVESTMENT

In this section we analyze whether VSLA membership is associated with various outcomes of interest. The two domains are nutritional outcomes and economic outcomes. HATUTAN aims to encourage VSLA membership with the expectation that it should lead to increased financial security, with follow-on effects in other domains. Here we test the assumption of a link between VSLA membership and improved nutritional outcomes. It is important to note that in this section we are unable to fully assess causality.

In the table below we can see the difference between the outcomes of interest by VSLA membership. On the three nutritional outcomes—whether the student had eaten on the day of the EGRA test and whether any member of the household had gone without eating—the effect sizes were small and not statistically significant. We can therefore tentatively conclude that there is little relationship between nutritional outcomes and being a VSLA member.

**Table 72: Difference in nutritional and economic outcomes by VSLA membership**

| | Not a VSLA member | VSLA member | Difference |
|---|---|---|---|
| **Nutritional outcomes** | | | |
| Student eaten on test day | 88.1% | 87.6% | -0.5 |
| Household not eaten | 10.5% | 8.5% | -2.0 |
| Caregiver dietary diversity (score) | 2.9 | 2.8 | -0.1 |
| **Economic outcomes** | | | |
| Uses savings on investment (household or business) | 17.6% | 29.1% | 11.5* |
| Uses savings on debt (household or business) | 3.0% | 4.3% | 1.3 |

However, there is a finding of note for the economic outcomes. There is little difference in whether a household used savings on debt, but households in VSLAs were more likely to have used savings either business or household investment. This may suggest that VSLA households are more financially secure, as they are in a position to spend more savings on non-essential goods and to take risk on business investment.

# GENDER AND POWER

The endline assessment 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 males and females across both the public and private spheres in order to better understand the different priorities, needs, activities, and responsibilities of male and female caregivers, and children. This analysis will help assess the impact of HATUTAN programming on gender dynamics, power, inclusiveness, and equality. 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.[183]

# DIVISION OF LABOR

Within the household survey, caregivers were asked about household tasks their grade 2 child performs. This included caregiving, housework (e.g., cooking or cleaning), fetching water or firewood, agricultural work, and helping with a family business. Caregivers also reported on 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 limit the time the child spends studying or doing homework.

## Gendered Division of Labor

In this section, we analyze the gendered division of children's performance of household tasks. At endline, we find that **female students remain significantly more likely to perform housework tasks compared to male students**, with caregivers reporting that 60.4% of female students perform daily housework tasks, as compared to 34.0% of male students. In contrast, **male students remain significantly more likely to participate in agricultural work**, with caregivers reporting that 33.4% of male students perform daily agricultural work, as compared to 23.4% of female students. While at baseline female students were significantly more likely than male students to participate in caregiving, the endline assessment records no gender differences in this. Likewise, there were no significant differences between genders in fetching and helping with a family business; overall, fetching water remains the most common task performed by each gender, and assisting with a family business remains the least common.

**Table 73: Percentages of second-grade children performing various daily household tasks**

| | Baseline | | | Midline | | | Endline | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | P-value | Male | Female | P-value | Male | Female | P-value |
| n | 434 | 426 | | 682 | 677 | | 683 | 659 | |
| Caregiving | 82.5% | 89.6% | 0.02* | 71.0% | 78.0% | 0.002** | 70.6% | 70.3% | 0.85 |
| Housework | 50.5% | 75.1% | <0.001*** | 38.6% | 62.0% | <0.001*** | 34.0% | 60.4% | <0.001*** |
| Fetching | 85.9% | 88.0% | 0.33 | 79.9% | 83.3% | 0.14 | 76.0% | 74.7% | 0.49 |
| Agriculture | 54.8% | 42.7% | 0.002** | 37.8% | 24.7% | <0.001*** | 33.4% | 23.4% | 0.002** |
| Business | 24.0% | 22.3% | 0.53 | 10.1% | 13.0% | 0.07 | 12.2% | 11.5% | 0.72 |

Examining change over time in children's participation in daily household tasks, the table below shows percentages of caregivers in intervention and control households who reported that their second-grade child performs each daily household task at baseline, midline, and endline for each gender separately, as well as the difference-in-difference results from baseline to endline. We find that **overall participation in each household task decreased fairly consistently across comparison and treatment households for both male and female students.** Furthermore, examining the effects of treatment status, we find that the proportion of boys participating in agricultural tasks has seen a significantly larger reduction in the intervention households as compared to the reduction seen in the comparison households. Likewise, the proportion of girls participating in caregiving tasks has seen a significantly larger reduction in the intervention households as compared to the reduction seen in the comparison households. However, as

[183] See CARE International, *Good Practices Framework: Gender Analysis* (Geneva: CARE International, 2012).

stated above, gender differences remain large in children's participation in housework, with girls carrying a disproportionate workload.

While the mechanisms driving this consistent decrease in participation in household tasks between assessment rounds is not entirely clear, the results of our analysis suggest that **the HATUTAN program might have been successful in reducing gender differences in children's participation in tasks where one gender typically had a disproportionate workload - in both caregiving and agricultural tasks.**

**Table 74: Change in the gendered division of children's labor within households**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Male** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **P-value** |
| n | 248 | 361 | 398 | 186 | 321 | 285 | | |
| Caregiving | 81.8% | 72.5% | 71.5% | 83.3% | 69.5% | 69.5% | 3.61 | 0.58 |
| Housework | 53.2% | 41.0% | 35.9% | 46.8% | 35.8% | 31.2% | -1.75 | 0.87 |
| Fetching | 87.5% | 81.4% | 78.6% | 83.9% | 78.2% | 72.3% | 2.73 | 0.39 |
| Agriculture | 56.5% | 37.7% | 30.7% | 52.7% | 38.0% | 37.2% | -10.30 | 0.003** |
| Business | 24.6% | 10.0% | 13.1% | 23.2% | 10.3% | 11.0% | 0.79 | 0.94 |
| **Female** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **P-value** |
| n | 234 | 373 | 343 | 192 | 304 | 316 | | |
| Caregiving | 92.7% | 79.9% | 67.6% | 85.8% | 75.7% | 73.1% | -12.41 | 0.03* |
| Housework | 76.9% | 64.9% | 63.3% | 72.9% | 58.6% | 57.3% | 1.98 | 0.74 |
| Fetching | 89.3% | 85.8% | 77.0% | 86.5% | 80.3% | 72.2% | 1.96 | 0.56 |
| Agriculture | 45.7% | 28.2% | 25.4% | 39.1% | 20.4% | 21.2% | -2.50 | 0.76 |
| Business | 22.0% | 13.9% | 13.1% | 22.8% | 11.8% | 9.8% | 4.05 | 0.65 |

## Time Spent on Household Tasks

Caregivers were also asked about the time their second-grade child spends on various household tasks. The below table displays the time spent on tasks by the gender of the child (Table 75). **Caregivers generally report very similar amounts of time spent on household tasks for children of each gender**, with most caregivers estimating that their grade 2 child spends a quarter of a day on these tasks. This is a comparable pattern to that found at baseline and midline, suggesting that a relatively similar amount of time is dedicated to daily tasks for both girls and boys. We note, however, that social desirability bias likely means that children's housework responsibilities are underestimated, as many caregivers may be aware that it is undesirable for children to spend large portions of their time on housework.

**Table 75: Time spent by second-grade children on various daily household tasks**

| | Baseline | | | Midline | | | Endline | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | P-value | Male | Female | P-value | Male | Female | P-value |
| n | 425 | 414 | | 625 | 642 | | 620 | 594 | |
| Whole day | 0.7% | 0.0% | 0.11 | 1.4% | 0.3% | 0.12 | 0.3% | 0.8% | 0.26 |
| Half day | 15.5% | 18.1% | 0.40 | 7.2% | 8.9% | 0.18 | 10.0% | 9.6% | 0.81 |
| Quarter day | 59.5% | 59.2% | 0.94 | 63.5% | 62.9% | 0.81 | 70.0% | 72.2% | 0.50 |
| An hour a day | 14.1% | 17.4% | 0.15 | 21.8% | 22.7% | 0.66 | 14.0% | 13.8% | 0.90 |
| Does not do chores | 10.1% | 5.3% | 0.06 | 5.8% | 4.5% | 0.50 | 4.7% | 2.9% | 0.22 |

Looking at change over time in the amount of time children spend on daily household tasks, Table 76 below shows the amount of time caregivers report that their second-grade child spends on household tasks at baseline, midline, and endline; for each gender separately; as well as the difference in difference results from baseline to endline. **We find no significant relative difference in the amount of time children spend on daily household tasks in intervention areas compared to comparison areas.** Nevertheless, a noteworthy exception to this, although statistically non-significant, is a minor increase in the number of girls who are reported to spend the whole day on household tasks in the intervention group, from 0.0% at baseline to 1.3% at endline, whereas a decrease is seen for boys.

Last, it is worth noting differences in time spent on household tasks that have occurred across evaluation rounds, even if they have occurred similarly in both comparison and intervention areas. We find that the amount of time children spend on daily household tasks has seen an overall decline since baseline, as suggested by: 1) a significant decrease in the number of caregivers reporting that their children spend a whole day or half a day on daily household tasks, a decrease that is seen fairly consistently across comparison and treatment households for both male and female children; and 2) a significant increase in the number of caregivers reporting that their children spend a quarter of a day on household tasks, which has seen a consistent increase across intervention groups and gender of the child. Overall, this pattern suggests that **children seem to be spending less time on household tasks in both intervention and comparison areas.**

**Table 76: Change in time spent on household tasks and impact on schooling**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Male** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **P-value** |
| **Time spent on tasks** | | | | | | | | |
| n | 244 | 339 | 365 | 181 | 286 | 255 | | |
| Whole day | 0.8% | 0.3% | 0.3% | 0.6% | 2.8% | 0.4% | -0.39 | 0.72 |
| Half day | 16.4% | 5.0% | 10.7% | 14.4% | 9.8% | 9.0% | -0.36 | 0.95 |
| Quarter day | 60.2% | 64.6% | 69.0% | 58.6% | 62.2% | 71.4% | -4.01 | 0.68 |
| An hour a day | 19.9% | 23.3% | 15.1% | 12.7% | 15.2% | 12.5% | 0.06 | 0.99 |
| Does not do chores | 7.4% | 6.2% | 3.8% | 13.8% | 5.2% | 5.9% | 4.39 | 0.37 |
| **Impact on schooling** | | | | | | | | |
| Late often | 0.4% | 0.6% | 0.3% | 2.2% | 0.7% | 0.0% | 2.08 | 0.02* |
| Late sometimes | 13.8% | 7.1% | 6.8% | 14.9% | 10.1% | 6.7% | 1.28 | 0.83 |
| Never late | 85.8% | 92.0% | 92.6% | 82.9% | 88.5% | 92.9% | -3.24 | 0.57 |
| Limits studying | 18.3% | 7.8% | 5.5% | 18.6% | 13.0% | 9.9% | -4.02 | 0.54 |

| Female | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
|---|---|---|---|---|---|---|---|---|
| **Time spent on tasks** | | | | | | | | |
| n | 229 | 358 | 310 | 185 | 284 | 284 | | |
| Whole day | 0.0% | 0.3% | 1.3% | 0.0% | 0.4% | 0.4% | 0.94 | 0.07 |
| Half day | 15.7% | 8.4% | 10.6% | 21.1% | 9.5% | 8.5% | 7.56 | 0.20 |
| Quarter day | 62.4% | 66.8% | 72.9% | 55.1% | 58.1% | 71.5% | -5.89 | 0.21 |
| An hour a day | 25.4% | 20.7% | 13.9% | 17.3% | 17.5% | 13.7% | -0.03 | 0.99 |
| Does not do chores | 4.4% | 3.4% | 0.3% | 6.5% | 6.0% | 5.6% | -3.19 | 0.31 |
| **Impact on schooling** | | | | | | | | |
| Late often | 1.3% | 0.8% | 0.6% | 1.1% | 1.1% | 0.4% | 0.06 | 0.91 |
| Late sometimes | 16.4% | 7.8% | 10.6% | 9.0% | 10.2% | 4.6% | -1.27 | 0.76 |
| Never late | 82.3% | 91.3% | 88.4% | 89.9% | 88.7% | 94.4% | 1.59 | 0.72 |
| Limits studying | 17.5% | 9.0% | 8.4% | 9.3% | 9.5% | 6.0% | -5.89 | 0.35 |

## Household Tasks and Impact on Schooling

We also asked caregivers about the impact the time spent on household tasks has on their child's schooling. At each round, we find no significant gender differences in the perceptions of caregivers of whether tasks made their male or female children late for school (see Table 77). At baseline, caregivers with male children were moderately, though non-significantly, more likely than caregivers with female children to report that their children had less time to study or do homework due to household tasks. At the endline, this gap between male and female children has decreased, with caregivers of male and female children reporting close-to-equivalent rates at which household tasks took time from studying. In general, at the endline, **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**, in both comparison and intervention households.

**Table 77: Caregivers' reports on the impact household tasks on their child's schooling**

| | Baseline | | | Midline | | | Endline | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | P-value | Male | Female | P-value | Male | Female | P-value |
| n | 427 | 420 | | 625 | 642 | | 620 | 594 | |
| Tasks make student late often | 1.2% | 1.2% | 0.97 | 0.6% | 0.9% | 0.46 | 0.2% | 0.5% | 0.17 |
| Tasks make student late sometimes | 14.3% | 13.1% | 0.59 | 8.5% | 8.9% | 0.62 | 6.8% | 7.7% | 0.63 |
| Tasks do not make student late | 84.5% | 85.7% | 0.66 | 90.4% | 90.2% | 0.85 | 92.7% | 91.3% | 0.46 |
| Student has less time to study | 18.4% | 13.9% | 0.06 | 10.1% | 9.2% | 0.72 | 7.3% | 7.2% | 0.99 |

Examining whether the extent to which daily household tasks interfere with children's schooling has changed over time, Table 76 above displays the percentages of caregivers reporting whether household tasks make their child late to school or limits their time studying, at baseline, midline, and endline; for each gender separately; as well as the difference-in-difference results from baseline to endline. The analysis finds a significant difference only for the percent of male students reported to be often late to school, with

a significantly larger decrease in the comparison households as compared to the reduction seen in the intervention households.

However, in line with the findings of children spending overall less time on household tasks, we note that there seems to be an overall decline in the negative impact of children's participation in household tasks on their schooling, as suggested by a significant decrease in the number of caregivers reporting that household tasks sometimes or often make their children late or that children have less time to study due to household tasks, decreases that are seen fairly consistently across comparison and treatment households for both male and female children; and a significant increase in the number of caregivers reporting that household tasks do not make their children late, which has seen a consistent increase across intervention groups and gender of the child. Overall, this pattern suggests **that children have become less likely to be late to school due to household tasks in both intervention and comparison areas.**

While these results are overall positive, it remains true for many children that time spent on household tasks affects their schooling. In the qualitative data, several teachers raised their concerns about the excessive time children spend on household tasks and its negative effect on their schooling. For example, a teacher in Ainaro stated:

> *If children's parents continue to expect them to do chores at home, they will struggle in school. Parents assign their children tasks such as fetching water, collecting firewood, and other household duties. This results in children knowing little and forgetting what they learn at school. They get up in the morning, grab their notebooks, and go to school. They always say "Oh." when asked about something at school.*
>
> - KII with teachers, Ainaro, Int. 37

## HOUSEHOLD DECISION-MAKING

We asked caregivers within the household survey about who makes decisions regarding their 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," and "elders/grandparents." At midline, these options were expanded to include "my husband/spouse alone." While this expanded list of choice options provides for a 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 from baseline to end-line. As such, we present the results from a difference-in-differences analysis comparing the endline values to those at midline.

Comparable to at midline, the vast majority of respondents in both intervention and comparison areas at endline reported that either they themselves or in conjunction with their spouse made decisions on their child's or baby's eating and hygiene practices (Table 78). Further comparable to the midline data, caregivers report little involvement of elders in household decision-making at the endline.

The difference-in-differences regression analysis shows that caregivers surveyed in the intervention households have become less likely to report making decisions on their baby's eating practices alone, and more likely to report taking these decisions together with a spouse, whereas no changes are seen in the comparison households (Table 78). A similar pattern of results is seen regarding who makes decisions on their child's eating practices as well as child's hygiene practices, although these effects do not reach statistical significance. All in all, the results of the endline analysis suggest that spouses have become more involved in household decision-making, and more so in intervention households than in comparison households, suggesting that the **HATUTAN programming may have contributed to a greater degree of shared household decision-making between spouses.**

**Table 78: Change in household decision making**

| Decisions on… | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Child's eating** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (ML to EL)** | **P** |
| n | 482 | 982 | 1,025 | 378 | 625 | 601 | | |
| Myself alone | 50.6% | 46.3% | 38.6% | 49.5% | 48.8% | 47.8% | -6.7 | 0.11 |
| Myself and my spouse | 46.7% | 45.4% | 53.1% | 48.4% | 46.2% | 46.1% | 7.8 | 0.18 |
| Myself/spouse with elders | 2.1% | 2.0% | 2.1% | 1.9% | 1.4% | 2.7% | -1.2 | 0.12 |
| Elders | 0.6% | 0.9% | 2.3% | 0.3% | 1.8% | 1.0% | 2.2 | 0.08 |
| Spouse alone | - | 3.3% | 3.9% | - | 1.8% | 2.5% | -0.1 | 0.97 |
| **Child's hygiene** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD** | **P** |
| n | 482 | 982 | 739 | 378 | 625 | 600 | | |
| Myself alone | 58.5% | 60.9% | 50.9% | 55.8% | 51.4% | 55.2% | -6.7 | 0.33 |
| Myself and my spouse | 38.0% | 40.7% | 47.0% | 42.1% | 45.6% | 41.8% | 10.0 | 0.22 |
| Myself/spouse with elders | 2.7% | 1.7% | 6.8% | 2.1% | 1.6% | 1.3% | -0.8 | 0.42 |
| Elders | 0.8% | 0.9% | 0.3% | 0.0% | 0.8% | 1.0% | -0.8 | 0.25 |
| Spouse alone | - | 0.9% | 1.2% | - | 0.6% | 0.7% | 0.3 | 0.62 |
| **Baby's eating** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD** | **P** |
| n | 51 | 87 | 80 | 31 | 50 | 50 | | |
| Myself alone | 58.8% | 70.1% | 43.8% | 51.6% | 60.0% | 52.0% | -18.4 | 0.07 |
| Myself and my spouse | 35.3% | 27.6% | 51.3% | 48.4% | 36.0% | 44.0% | 15.7 | 0.03* |
| Myself/spouse with elders | 3.9% | 1.1% | 1.3% | 0.0% | 2.0% | 2.0% | 0.1 | 0.68 |
| Elders | 2.0% | 1.1% | 1.3% | 0.0% | 0.0% | 0.0% | 0.1 | 0.95 |
| Spouse alone | - | 0.0% | 2.5% | - | 2.0% | 2.0% | 2.5 | 0.47 |

## CONTROL OF PRODUCTIVE ASSETS

Caregivers were asked who makes decisions related to productive assets, including large and small household purchases, loans, gardens, produce[184] and livestock, and household businesses. Moreover, respondents were asked about garden-related decision-making. At baseline, all caregivers were asked to state who makes 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. At endline, respondents were only asked who makes decisions regarding permagardens. As a result, the baseline and midline data are not directly comparable. We therefore present the results from a difference-in-differences analysis comparing the endline values to those at midline, and only for permagardens.

Looking at change over time in the control of productive assets, the table below shows caregivers' reported involvement in decision-making in their household, at baseline, midline, and endline; for each asset type separately; as well as the difference results from baseline to endline across intervention and comparison groups. **We find no significant relative changes in control of productive assets between intervention and comparison groups from baseline to endline.**

[184] In the household survey, only 26 respondents were asked who makes decisions regarding the sale of produce at midline, and 63 at endline, so this indicator is not included in the analysis.

**Table 79: Change in control of productive assets**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Large purchases** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **p** |
| n | 482 | 982 | 739 | 378 | 625 | 601 | | |
| Myself | 11.6% | 12.3% | 9.6% | 10.1% | 7.2% | 11.2% | -3.2 | 0.13 |
| My spouse | 28.6% | 24.2% | 31.5% | 25.4% | 18.2% | 29.5% | -1.2 | 0.94 |
| Myself and my spouse | 50.6% | 53.1% | 51.0% | 57.9% | 63.5% | 50.3% | 8.1 | 0.63 |
| Others | 9.2% | 10.4% | 7.9% | 6.6% | 11.1% | 9.2% | -3.8 | 0.22 |
| **Small purchases** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **p** |
| n | 482 | 982 | 739 | 378 | 625 | 601 | | |
| Myself | 50.2% | 56.5% | 40.5% | 52.1% | 50.9% | 43.8% | -1.4 | 0.83 |
| My spouse | 11.6% | 8.8% | 12.0% | 6.9% | 4.2% | 10.5% | -3.2 | 0.42 |
| Myself and my spouse | 33.8% | 30.0% | 44.0% | 38.1% | 41.3% | 41.8% | 6.5 | 0.48 |
| Others | 4.4% | 4.7% | 3.5% | 2.9% | 3.6% | 4.0% | -1.9 | 0.42 |
| **Sell chicken** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (ML to EL)** | **p** |
| n | 448 | 982 | 1,024 | 363 | 625 | 600 | | |
| Myself | 25.2% | 22.8% | 20.3% | 21.2% | 21.3% | 21.7% | -2.9 | 0.31 |
| My spouse | 17.4% | 15.4% | 17.6% | 15.4% | 14.6% | 20.2% | -3.4 | 0.63 |
| Myself and my spouse | 52.7% | 52.8% | 55.4% | 59.5% | 57.9% | 52.2% | 8.4 | 0.21 |
| Others | 4.7% | 9.0% | 6.7% | 3.9% | 6.2% | 6.0% | -2.1 | 0.40 |
| **Sell livestock** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (ML to EL)** | **p** |
| n | 447 | 982 | 1,024 | 370 | 625 | 600 | | |
| Myself | 10.3% | 10.8% | 13.2% | 7.3% | 6.6% | 10.2% | -1.2 | 0.50 |
| My spouse | 22.4% | 22.5% | 23.7% | 18.4% | 18.2% | 26.3% | -6.9 | 0.59 |
| Myself and my spouse | 58.6% | 56.0% | 54.3% | 66.5% | 65.1% | 54.2% | 9.2 | 0.54 |
| Others | 8.7% | 10.7% | 8.8% | 7.8% | 10.1% | 9.3% | -1.2 | 0.71 |
| **Start business** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (ML to EL)** | **p** |
| n | 401 | 982 | 1,024 | 335 | 625 | 600 | | |
| Myself | 21.0% | 22.5% | 23.9% | 14.3% | 14.6% | 23.2% | -7.2 | 0.34 |
| My spouse | 16.5% | 11.2% | 10.0% | 9.0% | 12.3% | 12.5% | -1.4 | 0.69 |
| Myself and my spouse | 60.0% | 50.0% | 55.4% | 73.1% | 60.6% | 53.5% | 12.5 | 0.13 |
| Others | 2.5% | 16.3% | 10.7% | 3.6% | 12.5% | 10.8% | -3.9 | 0.43 |

At midline and endline, respondents who had received training on permagardens were asked who makes decisions about these gardens. Most respondents at endline stated that they make the decision together with their spouse (59.0%), followed by them alone (25.0%), or just the respondent's spouse (11.0%). Looking at change over time, we find no significant relative changes in control of permagardens between intervention and comparison groups from midline to endline.

At midline and endline, respondents who participated in VSLAs were also asked who in their family made the final decision about how to use the VSLA loan. Most respondents at endline stated that they make the decision together with their spouse (53.1%), followed by them alone (33.5%) or just the respondent's spouse (10.3%). This response pattern is comparable to that registered in the midline assessment. As with the other assets studied, we find no significant relative changes in who makes decisions on VSLA loans between intervention and comparison groups from midline to endline.

## 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 (more difficult criteria to meet). Additionally, at baseline, respondents were not asked about participation in VSLAs. As such, we present the results from a difference-in-differences analysis comparing the endline values to those at midline.

### Training and Participation

At endline, among all respondents, men reported receiving AES at higher rates than women – 18.4% compared to 9.4% respectively. Although statistically non-significant, this might point towards an increased gender gap since midline, where these values were 19.3% and 15.3% respectively. Likewise, women at endline were more likely to state that they had not received support from AES (89.1% compared to 80.3%). For garden training, men were significantly more likely to state that they had received training at endline – 10.8% compared to 6.2% of women. In comparison, at baseline, men also reported somewhat higher rates of access to AES, membership in a farmer's group,[185] and ability to attend agricultural training. Thus, while there are no statistically significant differences in support from AES for male and female respondents, female respondents are 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.**

Examining change over time in whether respondents receive support from AES and training on keyhole gardens, the table below shows the percentage of respondents stating that they had received support or training, at midline, and endline; for each gender separately; as well as the difference in difference results from midline to endline. **We find that males in the intervention groups have become significantly more likely to receive support from AES at endline than they were at midline.** No other relative differences are found in respondents' answers in intervention groups compared to comparison groups.

[185] At midline and endline, 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.

**Table 80: Change over time in whether respondents receive support from AES or training on keyhole gardens, by gender**

| | Intervention | | Comparison | | | |
|---|---|---|---|---|---|---|
| **Female** | **ML** | **EL** | **ML** | **EL** | **DiD** | **P-value** |
| Support from AES | 14.5% (n = 690) | 9.9% (n = 493) | 16.4% (n = 438) | 8.6% (n = 382) | 3.3 | 0.44 |
| Garden training | 9.3% (n = 875) | 7.0% (n = 689) | 1.8% (n = 610) | 5.2% (n = 557) | -13.8 | 0.07 |
| **Male** | **ML** | **EL** | **ML** | **EL** | **DiD** | **P-value** |
| Support from AES | 20.0% (n = 105) | 31.0% (n = 42) | 14.3% (n = 14) | 2.9% (n = 34) | 21.8 | 0.04* |
| Garden training | 21.5% (n = 107) | 10.0% (n = 50) | 6.7% (n = 15) | 11.6% (n = 43) | -30.3 | 0.16 |

### Participation in VSLAs

Respondents at midline and endline were also asked about the number of male and female household members participating in a VSLA. Around 48.0% of households kept their savings in VSLA at endline compared to 37.8% at midline. At endline, intervention households were 9 percentage points more likely to participate in VSLA than comparison households, at 55.9% as compared to 45.1% at ML. However, using the difference-in-differences regression model to examine change over time in whether respondents participate in VSLA, we find no relative differences in respondents' answers in the intervention group compared to the comparison group. Among households with members participating in VSLAs, on average, 1.6 household members participated in a VSLA at endline. Women were more likely to participate in a VSLA than men; on average, endline respondents reported that 71.0% of the household members participating in a VSLA were female. This is a moderate increase from midline, where 63.0% of household members participating in VSLA were female.[186] Furthermore, 54.6% of households at endline reported that all VSLA members were female (compared to 43.8% at midline), while only 13.3% of households reported that all VSLA members were male (compared to 19.2% at midline). **These results suggest that VSLAs may be effectively targeted towards women as compared to men, and more so at endline than at midline.** This dynamic may help increase women's financial independence and ability to claim rights.

## CONTROL OVER ONE'S BODY

HATUTAN program activities included trainings for PTA and VSLA members on gender awareness and sexual and GBV prevention in Ainaro, Ermera, and Liquica. Within the household survey, we asked caregivers 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 intervention respondents who were asked about at least one scenario,[187] at the endline 49.3% believed that a husband is justified in beating his wife in at least one scenario, compared to 55.2% at baseline.

We conducted further analysis to examine change over time in respondents' views of whether a husband is justified in beating his wife. The below table shows percentages of respondents who consider beating

[186] Note: This effect was only marginally significant with $p = 0.07$

[187] Respondents were not asked these questions if another person was present; 472 were asked at least one question at midline, 564 at midline and 520 at baseline. Male caregivers were not intended to be asked these questions, and as such, we exclude 28 male caregivers/respondents who were asked at least one of these questions at endline.

justified in each type of scenario; at baseline, midline, and endline, as well as the difference-in-differences results from baseline to endline. **No significant changes are found in respondents' views of whether gender-based violence is justified in each of the circumstances given, between intervention and comparison groups from baseline to endline.** On a municipality level, we find the most substantial reduction in the proportion of respondents in Ermera stating that domestic violence is justified in at least one of the scenarios, from 62.2% to 52.6%. We see less change in other intervention areas.

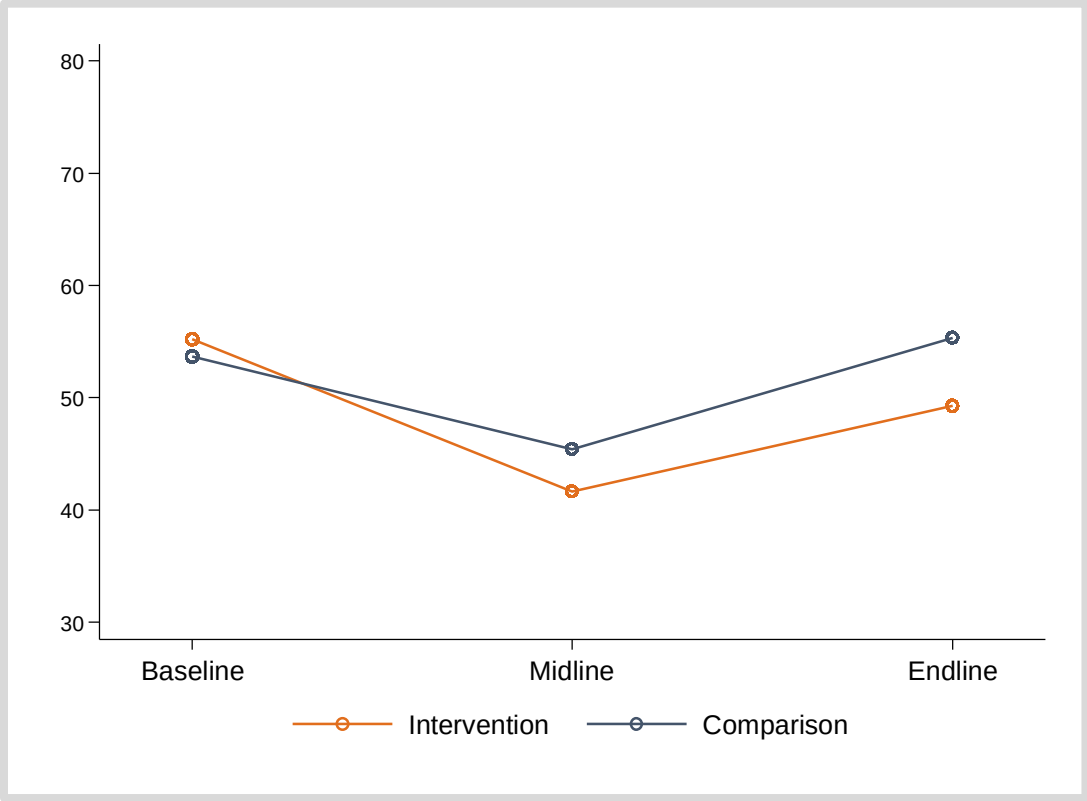
**Table 81: Change in attitudes towards gender-based violence**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| Outcome | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P |
| She goes out without telling him | 39.8% (n = 246) | 23.6% (n = 327) | 32.4% (n = 210) | 42.2% (n = 199) | 31.6% (n = 193) | 39.7% (n = 146) | -5.2 | 0.65 |
| She neglects the children | 48.2% (n = 251) | 31.8% (n = 330) | 42.4% (n = 210) | 48.0% (n = 198) | 31.8% (n = 195) | 45.2% (n = 146) | -3.5 | 0.76 |
| She argues with him | 25.2% (n = 246) | 26.5% (n = 328) | 35.9% (n = 212) | 31.5% (n = 200) | 32.6% (n = 193) | 39.2% (n = 148) | 2.4 | 0.80 |
| She burns the food | 19.0% (n = 248) | 12.6% (n = 326) | 21.3% (n = 211) | 18.4% (n = 196) | 16.6% (n = 193) | 26.7% (n = 146) | -6.2 | 0.49 |
| At least one of the above | 55.2% (n = 259) | 41.7% (n = 336) | 49.3% (n = 215) | 53.7% (n = 205) | 45.5% (n = 196) | 55.3% (n = 150) | -7.6 | 0.51 |

However, it is worth noting a change that has occurred across evaluation rounds that occurred similarly in both comparison and intervention areas. The data suggest that, compared to baseline, at endline **respondents in both intervention and comparison households were significantly more likely to consider beating justified when a woman argues with her husband,** with a 9.9 percentage point increase in the number of respondents justifying domestic gender-based violence in this scenario.

It is worth highlighting that at midline the data showed promising results, with fewer respondents believing that a husband was justified in beating his wife in most scenarios as compared to baseline (Figure 50), but, as shown above, this trend seems to have reversed at endline, with more respondents believing that a husband is justified in beating his wife than at midline. The similarity in trends across comparison and intervention households suggests that this pattern might be a result of a broader change in social norms or attitudes towards domestic gender-based violence, rather than a result of any specific program interventions.

**Figure 50: Change in attitudes towards gender-based violence**




Further analyses by the age of caregiver find a significant relationship between age and attitudes towards domestic gender-based violence among endline respondents, **with younger respondents being more likely to consider domestic violence justified in each of the scenarios.** This suggests that the change in attitudes may be generational, and programming might benefit from directly targeting younger caregivers. No relationship is found between caregivers' level of education and views toward domestic violence.

# VIOLENCE AND RESTORATIVE JUSTICE

Caregivers were asked about the ways in which teachers handle misbehavior in class. They were asked about positive/neutral forms of discipline (giving a verbal warning, informing parents, and having a conversation with the child), and negative forms of discipline (shouting at the child, using corporal punishment, and assigning chores). Moreover, caregivers, as well as school directors, were asked about avenues for reporting abuse at school.

## Teachers' Discipline Practices

Among all respondents at endline, caregivers were most likely to state that teachers had a conversation with the child (40.8%), followed by giving a verbal warning (37.9%), shouting at the child (35.9%), and using corporal punishment (32.6%). Respondents were least likely to state that teachers informed parents (18.6%) or that the teacher assigned chores to the child (10.3%).

Examining changes over time in how teachers handle misbehavior in class, the table below shows percentages of caregivers reporting each discipline practice at baseline, midline, and endline, as well as the difference-in-differences results from baseline to endline. **No significant changes are found in any of the discipline practices reported by intervention groups as compared to comparison groups from baseline to endline.**

However, some changes have occurred across evaluation rounds for both treatment status groups. We find statistically significant decreases in the reported prevalence of giving verbal warnings, informing the parents, and assigning chores. Similarly, we find substantial decreases in the reported prevalence of teachers having a conversation with the child and shouting at the child – although this is statistically non-significant. No meaningful changes are seen in the reported use of corporal punishment by teachers. Overall, **this may suggest a widespread overall decrease in the use of teachers' discipline practices over the last few years, with the exception of corporal punishment.**

**Table 82: Change in teachers' discipline practices**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Outcome** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (BL to EL)** | **P-value** |
| n | 482 | 734 | 741 | 378 | 625 | 601 | | |
| Gives verbal warning | 43.2% | 49.5% | 37.4% | 51.6% | 47.5% | 38.4% | 7.38 | 0.16 |
| Informs parents | 30.3% | 21.5% | 17.4% | 29.9% | 22.2% | 20.1% | -3.12 | 0.76 |
| Has conversation with child | 46.3% | 45.6% | 41.3% | 48.2% | 46.2% | 40.1% | 3.08 | 0.79 |
| Shouts at the child | 40.3% | 39.5% | 35.9% | 39.4% | 44.3% | 33.9% | 1.12 | 0.89 |
| Uses corporal punishment | 36.1% | 31.7% | 34.0% | 30.4% | 30.1% | 31.0% | -2.62 | 0.69 |
| Assigns chores | 26.4% | 15.1% | 10.4% | 28.8% | 13.9% | 10.2% | 2.73 | 0.73 |

Disaggregating by the gender of the child, we find that, while the reported use of verbal warnings has decreased for both boys and girls, **caregivers of boys have become significantly more likely to report that teachers use verbal warnings** (see Table 83). Likewise, at baseline, midline, and endline, caregivers of male students were more likely to report that teachers use corporal punishment – although this difference does not reach statistical significance. Furthermore, at baseline, midline, and endline and in both treatment groups on aggregate, caregivers of both male and female children were more likely to say that corporal punishment was justified against boys than against girls. No marked gender differences are observed in the reported use of other discipline methods.

The use of corporal punishment and other negative discipline practices is concerning as it is likely to decrease student attendance and motivation. This concern was raised by several caregivers during the qualitative interviews.[188] For example:

> *Sometimes, before parents are contacted to come to the school, the children have already been subjected to physical abuse, which creates fear and reluctance to attend school.*
>
> - KII with mothers, Manatuto, Int. 31

Simultaneously, several respondents also mention that, at home, children might also be beaten by their parents for not going to school: "Parents have informed that even though they hit their children, they still refuse to go to school".[189]

[188] FGD with fathers, Manutoto, Int. 20; KII with mothers, Ermera, Int 28; FGD with mothers, Manutoto, Int. 31;
[189] KII with coordinators, Ainaro, Int. 2

**Table 83: Teacher's use of the different discipline practices, by student's gender**

| | Baseline | | | Midline | | | Endline | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | P-value | Male | Female | P-value | Male | Female | P-value |
| n | 434 | 426 | | 682 | 677 | | 683 | 659 | |
| Gives verbal warning | 49.3% | 44.4% | 0.13 | 48.1% | 49.0% | 0.73 | 39.4% | 36.3% | 0.003** |
| Informs parents | 32.3% | 27.9% | 0.10 | 22.1% | 21.6% | 0.80 | 19.3% | 17.9% | 0.42 |
| Has conversation with child | 48.6% | 45.5% | 0.25 | 44.4% | 47.4% | 0.29 | 40.3% | 41.3% | 0.67 |
| Shouts at the child | 39.9% | 39.9% | 0.99 | 44.7% | 38.7% | 0.03* | 36.3% | 33.7% | 0.42 |
| Uses corporal punishment | 36.9% | 30.3% | 0.06 | 33.6% | 28.4% | 0.04* | 35.4% | 29.7% | 0.06 |
| Assigns chores | 28.6% | 26.3% | 0.62 | 14.8% | 14.3% | 0.79 | 10.1% | 10.5% | 0.73 |

## Teachers' Discipline Practices and Student Disability

We further analyze the reported use of discipline practices by whether the caregiver reports their second-grade child to have a physical or mental disability. At endline, we find no significant relationship between physical disability and teachers' discipline practices, though we do note that only a small number of households report their child to have a physical disability (8%).[190] However, we do find significant relationships between cognitive and mental health disabilities and reported discipline practices.[191] In intervention households, we find that caregivers that report their child to have at least one cognitive/mental health disability at endline are significantly less likely to report that teachers shouted at the student than caregivers of students without a mental disability (31.0% as compared to 42.2%). Conversely, in comparison households, caregivers of children with at least one cognitive/mental health disability are significantly more likely to report that teachers shouted at the student than caregivers of students without a mental disability. This is a reversal from a pattern seen at midline, in which students with cognitive and mental health disabilities were reported to be more likely to be shouted at. We also find that in comparison households, caregivers of students with cognitive and mental health disabilities were significantly less likely to report that verbal warnings are given to their children than caregivers of children without these disabilities. No differences in this are observed in intervention households. We don't find any significant differences between students with and without cognitive and mental health disabilities in any of the other discipline practices.

The results at midline found that various discipline practices were disproportionally used on children with mental disabilities, with various "negative" discipline practices more commonly used on children with mental disabilities than those without. **The endline results suggest that this pattern has been reversed, with students with mental disabilities either significantly less or equally likely to experience teacher violence than their counterparts with no mental disabilities.** The only exception here is that children with mental disabilities in comparison areas remain more likely to be shouted at than children without mental disabilities. Overall, these findings point towards children with mental disabilities facing decreased challenges in schools and are now being less marginalized. However**,** given that the HATUTAN does not specifically address the intersection of disability and discipline practices it is uncertain whether these

[190] Students are classified as having physical disability if they were reported to have an issue with at least one of the following: eyesight, hearing, or mobility.
[191] 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).

endline findings represent a sustainable change or have an ephemeral nature, warranting further monitoring of the intersection between disability and power in HATUTAN II.

## Avenues for Reporting Abuse

Caregivers in the household survey were also asked who they could report abuse to if their second-grade child was abused or harassed at school. As shown in the table below, the vast majority of caregivers would report the abuse to the headteacher (63.5%). Notably, at baseline, 9.4% of respondents answered that they would not be able to report abuse; this percentage increased at midline, to 14.9%, and only decreased slightly at endline to 12.5%. This suggests that **there may remain barriers to accessing restorative justice for children abused at school, with little change observed over the last five years**.

Using a difference-in-difference regression analysis to examine changes over time in caregivers' perceptions of avenues for reporting abuse in intervention as compared to control areas, we find no significant relative changes in who respondents would report abuse to between baseline and midline.

Within the school survey, school directors were asked about avenues children can report abuse occurring at the toilets or abuse by a teacher. The majority of directors, at all rounds, stated that abuse could be reported to the director/coordinator of the school or to a teacher (Table 84). Only a very small percentage of directors stated that abuse could not be reported (0.0 - 0.6% at each round). As highlighted in the midline evaluation, the contrast in proportions of respondents stating that abuse cannot be reported across the household and school survey suggests that while directors may believe that there are effective avenues for reporting abuse and obtaining justice, caregivers might not agree or are not aware of these avenues for reporting.

**Table 84: Reporting of abuse at school**

| | BL | ML | EL |
|---|---|---|---|
| **Household survey: Reporting abuse of child at school** | | | |
| n | 860 | 1,359 | 1,340 |
| Head teacher | 79.3% | 67.2% | 63.5% |
| Police | 1.9% | 1.8% | 2.7% |
| Social services | 0.5% | 0.2% | 0.7% |
| Local authorities | 3.4% | 2.3% | 3.4% |
| Cannot report | 9.4% | 14.9% | 12.5% |
| Other | 9.8% | 19.8% | 23.8% |
| **School survey: Reporting abuse of child in toilets** | | | |
| n | 181 | 178 | 182 |
| Family/relatives | 10.5% | 6.7% | 4.4% |
| Director/coordinator | 45.3% | 47.2% | 55.5% |
| Teacher | 32.6% | 25.8% | 24.7% |
| PTA | 2.8% | 9.5% | 1.7% |
| Cannot report | 0.6% | 0.0% | 0.6% |
| Other | 8.1% | 8.0% | 0.6% |
| Don't know or did not respond | 0.0% | 3.3% | 8.3% |

| | BL | ML | EL |
|---|---|---|---|
| **School survey: Reporting abuse of child by teacher** | | | |
| n | 180 | 178 | 182 |
| Family/relatives | 17.2% | 7.3% | 5.5% |
| Director/coordinator | 44.4% | 49.4% | 58.8% |
| Teacher | 19.4% | 22.5% | 20.9% |
| PTA | 1.7% | 7.9% | 2.2% |
| Cannot report | 0.6% | 0.6% | 0.6% |
| Other | 16.8% | 8.6% | 1.7% |
| Don't know or did not respond | 0.0% | 3.8% | 6.6% |

# ASPIRATIONS AND STRATEGIC INTERESTS

Within the household survey, caregivers were asked about whether boys and girls were equally skilled at math and reading as well as a variety of questions about whether girls and boys had equitable experiences and opportunities at schools. At midline and endline, the response option “don’t know” was included to these questions, while this option was not included at baseline. Due to the inclusion of this response at midline, baseline and endline results are not directly comparable. Therefore, we conduct a difference-in-differences analysis of change over time from midline to endline.

## Perceived Gender Gaps in Literacy and Numeracy Skills

Among intervention respondents at endline, the majority of caregivers stated that they thought boys and girls had the same capacity for reading and writing (60.3%), and for math (57.9%). However, **intervention caregivers were more likely to state that girls have more capacity for reading and writing or math than boys**: 16.2% of intervention respondents stated that girls have more capacity than boys for reading and writing and 17.5% stated that girls had more capacity for math, whereas only 5.4% of intervention respondents stated that boys have more capacity than girls for reading and writing and 6.6% stated that boys had more capacity for math. This trend also emerged from the qualitative interviews, with several respondents noting that girls are often considered having higher capabilities than boys in the classroom. However, this is not attributed to inherited gender differences in intelligence but rather gender differences in self-discipline and/or time spent studying, or lack of skills of the teacher.[192] As one teacher described it:

> *Some people say that the challenge faced is that male students cannot read. In my opinion, it's not true that they can't read, but rather that they're misbehaving and don't want to listen. On the other hand, female students are more disciplined in character and listen better; probably because they are afraid.*
>
> - KII with teachers, Ainaro, Int. 38

Looking at whether perceptions of gender gaps in literacy and numeracy skills have changed over time, the below table shows percentages of caregivers stating whether they believe boys or girls have the same literacy and numeracy capacities at baseline, midline, and endline, as well as the difference-in-differences results from midline to endline. **We find no significant relative changes in respondents' perceptions of gender gaps in literacy and numeracy skills**, reported by intervention groups compared to comparison groups from midline to endline.

[192] KII with teacher, Ainaro, Int. 38; KII with teacher, Manatuto, Int. 44

However, it is worth noting some changes that have occurred similarly in both comparison and intervention areas between evaluation rounds. Compared to midline, at endline, we find a substantial increase in the proportion of respondents reporting that girls and boys are equally skilled at math, from 51.8% to 57.9% for intervention respondents and from 53.4% to 57.4% for comparison respondents. We see the same pattern for literacy skills, with the percentage of respondents reporting that girls and boys are equally skilled at math rising from 53.7% to 60.3% for intervention respondents and from 55.8% to 58.4% for comparison respondents. However, this effect does not reach statistical significance. Overall, **this may suggest a widespread increase in perceived gender equality over the past two years**.

**Table 85: Change in perceptions of gendered literacy and numeracy skills, as well as perceptions of gendered experiences at school**

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| **Literacy skills** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** | **DiD (ML to EL)** | **P-value** |
| n | 419 | 723 | 742 | 333 | 625 | 601 | | |
| Equal | 71.4% | 53.7% | 60.3% | 71.8% | 55.8% | 58.4% | 4.1 | 0.48 |
| Girls better | 19.6% | 18.0% | 16.2% | 21.9% | 18.2% | 14.0% | 2.5 | 0.26 |
| Boys better | 9.1% | 5.0% | 5.4% | 6.3% | 6.4% | 7.2% | -0.4 | 0.77 |
| Neither | 0.0% | 0.1% | 0.5% | 0.0% | 0.2% | 0.7% | -0.1 | 0.81 |
| Don't know | 0.0% | 23.2% | 17.5% | - | 19.4% | 19.8% | -6.1 | 0.12 |
| **Math skill** | | | | | | | | |
| n | 417 | 734 | 741 | 332 | 625 | 601 | | |
| Equal | 72.4% | 51.8% | 57.9% | 73.8% | 53.4% | 57.4% | 2.2 | 0.56 |
| Girls better | 18.2% | 15.4% | 17.5% | 16.0% | 14.4% | 13.1% | 3.4 | 0.33 |
| Boys better | 9.1% | 7.5% | 6.6% | 9.9% | 9.0% | 5.5% | 2.6 | 0.09 |
| Neither | 24.0% | 0.4% | 0.9% | 0.3% | 1.6% | 1.3% | 0.8 | 0.18 |
| Don't know | 0.0% | 24.9% | 17.0% | - | 21.6% | 22.6% | -9.0 | 0.02* |
| **Encouraged to participate** | | | | | | | | |
| n | 416 | 734 | 741 | 331 | 625 | 601 | | |
| Equal | 92.1% | 62.3% | 70.5% | 95.6% | 67.8% | 71.2% | 4.8 | 0.38 |
| Girls more | 3.9% | 4.1% | 6.2% | 2.5% | 3.8% | 4.3% | 1.6 | 0.52 |
| Boys more | 3.1% | 3.4% | 1.9% | 1.6% | 2.9% | 1.0% | 0.4 | 0.83 |
| Neither | 1.0% | 0.8% | 0.5% | 31.0% | 1.3% | 0.2% | 0.8 | 0.24 |
| Don't know | 0.0% | 29.4% | 20.9% | - | 24.2% | 23.3% | -7.6 | 0.09 |
| **Able to ask questions** | | | | | | | | |
| n | 387 | 734 | 741 | 331 | 625 | 601 | | |
| Equal | 87.9% | 55.3% | 64.1% | 91.0% | 55.7% | 63.7% | 0.7 | 0.88 |
| Girls more | 7.8% | 4.8% | 7.2% | 7.4% | 6.9% | 3.2% | 6.1 | 0.05 |
| Boys more | 3.9% | 3.4% | 2.8% | 1.3% | 2.6% | 1.5% | 0.5 | 0.79 |
| Neither | 0.5% | 1.6% | 1.2% | 0.3% | 1.8% | 1.5% | -0.2 | 0.89 |
| Don't know | 0.0% | 34.9% | 24.7% | - | 33.1% | 30.1% | -7.2 | 0.21 |

| Able to get help | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| n | 400 | 734 | 741 | 331 | 625 | 601 | | |
| Equal | 91.8% | 61.6% | 71.5% | 92.9% | 63.8% | 69.4% | 4.4 | 0.09 |
| Girls more | 4.3% | 4.0% | 5.3% | 4.8% | 5.0% | 3.0% | 3.3 | 0.12 |
| Boys more | 3.3% | 2.9% | 1.8% | 1.6% | 2.6% | 1.2% | 0.3 | 0.75 |
| Neither | 0.8% | 1.0% | 0.4% | 0.6% | 1.4% | 1.2% | -0.3 | 0.82 |
| Don't know | 0.0% | 30.7% | 21.1% | - | 27.2% | 25.3% | -7.7 | 0.08 |

## Gendered Experiences and Opportunities at School

In general, at the endline the vast majority of respondents either believed that schools treated students equitably by gender (as measured by the 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 endline. This was echoed by the qualitative respondents, who unanimously stated that boys and girls receive the same treatment and opportunities at school.[193] For example:

> *I feel that teachers don't treat them differently, because a teacher's real job is to teach all students, not treat them differently based on gender or age.*
>
> - KII with teachers, Manatuto, Int. 44

Although these are positive results, it is important to note that social desirability bias in respondents' answers or unconscious gender biases that affect the way teachers respond to and treat children are likely and cannot be ruled out. Furthermore, although respondents state that girls and boys are treated equivalently when asked about this explicitly, respondents' answers to other questions suggest that they consider that boys and girls have different characteristics and behavior for example with girls being shyer than boys,[194] and this might lead to differential treatment.

Examining change over time in whether respondents believe children have similar experiences at school, the table above (Table 85) shows percentages of caregivers stating whether they believe boys and girls are treated equitably at school in each round, as well as the difference-in-differences results from midline to endline. **We find no significant relative changes in respondents' perceptions of gender differences in children's experiences at school**, reported by intervention groups compared to comparison groups from midline to endline.

However, looking at changes that occurred in intervention areas between midline and endline, we find a significant 8.8 percentage point increase in the number of intervention respondents reporting that girls and boys are equally able to ask questions at school and a significant 9.9 percentage point increase for their equal ability to get help. We see the same pattern for respondents' perceptions of whether boys and girls are equally encouraged to participate; however, this effect does not reach statistical significance. We also find the age of caregivers to be a significant predictor of whether they report girls and boys to be treated equitably at school, **with younger caregivers being more likely to report that both genders receive equitable treatment and opportunities at school.**

While these are positive results, as above, it is important to note that perceptions of gender equality in school opportunities has declined across both intervention and comparison groups from baseline to endline.

[193] KII with administrator, Ainaro, Int. 1; KII with administrator, Ainaro, Int. 2; KII with administrator, Liquica, Int. 5; KII with administrator, Manatuto, Int. 7; FGD with teachers, Manatuto, Int. 43; FGD with teachers, Manatuto, Int. 44
[194] KII with administrator, Ainaro, Int. 1; KII with administrator, Ermera, Int. 3

This drop was observed most acutely from baseline to midline, and has since rebounded slightly, though not to baseline levels, at endline.

Caregivers were also asked about the maximum level of education they could support for their second-grade child. At all rounds, the majority of caregivers stated that they could support their boys and girls through university (at baseline: 79.7% and 76.6% respectively; at midline: 72.9% and 70.3% respectively; and at endline: 63.4% and 61.2% 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 boys and girls through school.

# LEARNING AGENDA

The McGovern-Dole Learning Agenda aims to answer questions related to school meal program implementation and education. For HATUTAN specifically, the learning agenda focuses on literacy, health, nutrition, agriculture, and gender-based violence. The learning agenda also includes a sustainability component; this is discussed more under *Program-specific Outcomes.*

## EDUCATION AND LITERACY

### Relationship Between School Feeding and Literacy

A key area of interest for HATUTAN and the McGovern-Dole Learning Agenda is the effect of SFPs on learning, including student literacy. We first seek to understand ways in which the combination of school meal interventions and educational interventions improve education and literacy levels.

At endline, we found that due to delays in government distribution of SFP funding, only 16% of intervention schools reported providing meals to students on the day of data collection, a substantial decline from midline when 89% of intervention schools reported providing meals. This limited prevalence of school meal provision reduces our ability to draw clear linkages between school feeding and literacy; however, in order to better understand potential effects, we limit our analysis to two intervention municipalities, Ainaro and Manatuto. Eleven schools in Ainaro and five in Manatuto—41% and 26% of schools in these municipalities, respectively—reported providing meals to students on the day of data collection.

In a simple regression with overall EGRA scores as the outcome of interest and provision of school meals as the predictor, we find a positive but not statistically significant relationship. In other words, school provision of meals is associated with a 5.7 percentage point increase in literacy scores, but this is not significant. However, when we include control variables for student gender, age, and school fixed effects in this model, we find a significant and negative relationship between school meals and literacy scores: Provision of school meals is significantly associated with a 3.4 percentage point decrease in overall EGRA scores, all else held constant. We note that this runs counter to findings in the predictive analysis of literacy outcomes; however, as this predictive analysis included all municipalities, its results for school meal provision may be a reflection of limited school feeding prevalence at endline rather than the actual impact of school meals.

Given the limited prevalence of school feeding at endline, we run these regression models for midline data as well to test robustness of findings. We note that midline data also does not provide a "perfect" sample for this analysis due to the severe impact of COVID-19 during midline; however, it allows us to validate results found above. In contrast to results above, for the model including control variables, we find a significant and positive relationship between school feeding and literacy: **School meal provision is associated with a 1.3 percentage point increase in literacy scores, all else held constant.**

Overall, this analysis provides a mixed picture of the impact of school feeding on literacy; we do not suggest that the above findings are definitive, however, due to limitations in the data as a result of SFP funding delays and the COVID-19 pandemic. A variety of confounding factors—including those affected by the HATUTAN program—may also influence both school feeding and literacy. We now explore these factors to better understand the linkages between school feeding interventions, education interventions, and literacy.

One key factor which could affect both school feeding and literacy is training of school administrators. HATUTAN sought to improve the capacity of school administrators to advocate for SFP funding and manage the SFP. Better-trained school administrators are also more likely to coach teachers, thus improving teaching practices and the quality of education.

Directors were not asked about the specific training provided to them by HATUTAN in the school survey; however, they were asked whether they provided coaching to teachers, which we use a proxy for both administrator involvement and whether administrators were trained. Interestingly, at endline within Ainaro and Manatuto, we find no significant relationship between administrator coaching and EGRA scores and a significant but negative relationship between coaching and provision of school meals. Along these lines, adding coaching as a control variable has little effect on the regression between school feeding and overall literacy scores (for both the endline and midline groups).

A second key factor is PTA involvement, as PTAs may both assist with SFP management and advocate for improved student education outcomes. Almost all schools had a PTA at endline; however, reported levels of PTA activity varied widely, from inactive to very active.

At endline, for Ainaro and Manatuto, we again find a significant and negative relationship between level of PTA activity and whether school meals were provided, and no significant relationship between PTA activity and overall EGRA scores. However, when PTA activity is added as a control to the regression between school feeding and overall literacy, we find that there is no longer a significant relationship between these latter two variables (for either the endline or midline data). This suggests that some of the relationship between school meal provision and literacy may be explained by PTA activity.

Overall, these findings are inconclusive. While we find a negative relationship between school feeding and literacy at endline, it is highly unlikely that school feeding *causes* lower EGRA scores. Rather, especially given the limited number of schools providing meals at endline, it may be that the schools supported by HATUTAN which were able to provide meals at endline were among the most vulnerable; in other words, schools providing meals at endline may have also been those serving more disadvantaged students, who would thus also generally perform worse on the EGRA. The opposite direction of the relationship when midline data is used suggests that this may indeed be the case.

However, the analysis does suggest that PTAs may play an important role in strengthening linkages between school feeding and literacy, potentially by providing support for both. As such, interventions which target PTA efficacy and activity may help improve overall program efficacy, as well as improving the cost effectiveness of SFP interventions.

## Relationship Between Nutrition, School Feeding, and Literacy

The learning agenda also seeks to understand differences in school feeding program-literacy linkages for mal- and undernourished children. We note that at endline, almost half of the intervention students and over half of the comparison students were underweight, suggesting that undernourishment is extremely prevalent. However, to better understand nutrition, education, and school feeding dynamics, we utilize students' dietary diversity scores, with students consuming less than three healthy food groups (i.e., not processed or sugary foods) considered to be undernourished. At endline, based on this metric, 64.3% of intervention students were undernourished.

At endline, for Ainaro and Manatuto, we find a positive though not significant relationship between whether school meals were provided and whether students are adequately nourished. Two dynamics may influence this: First, we would expect that school meals would help improve student nourishment by providing them with an additional meal including at least rice and beans (two food groups). However, as discussed above, if school meals were provided in more deprived areas at endline, then we might expect a negative linkage between school feeding and nourishment due to higher levels of deprivation. The combination of these two dynamics could produce the result found here.

Looking at all intervention students at endline, meanwhile, we again find a positive but not significant relationship between student nourishment and overall literacy scores. This finding is somewhat more surprising than the linkage between school feeding and nutrition, as we would expect better-nourished students to be more attentive and to likely come from better educated or wealthier families, all of which should correlate to improved educational outcomes.

Lastly, we add student nourishment as a control variable into the regression examining the relationship between school feeding and overall literacy scores. When nourishment is added as a control, we find no significant relationship between school meal provision and literacy scores. This suggests that the mechanism by which school meals improve literacy may indeed be through improved student nutritional outcomes. Overall, these findings suggest that student nourishment may be an important dynamic influencing educational outcomes from school meal programs, although the extent of the relationship is not clear.

## Relationship Between School Feeding and Teaching

The McGovern-Dole learning agenda seeks to understand the impact of school meal interventions on the resources available to teachers, such as classroom time on task, teacher motivation, and teachers' use of engaging teaching practices. 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.

In the section *Student Attentiveness*, our predictive analysis found few clear relationships between school feeding and attentiveness as measured by working memory. At endline, we found a significant and positive relationship between school meal provision and working memory, but a significant and negative (though small) relationship between student dietary diversity and working memory. Furthermore, when data was expanded to include all rounds, we found a significant and negative relationship between school feeding and literacy scores. This unclear relationship is likely, at least in part, to be a function of the limited prevalence of school feeding at baseline and endline compared to midline. Indeed, we note that at midline—when school feeding was most prevalent—whether a student had eaten was indeed a significant predictor of attentiveness. As such, while we cannot fully validate the Theory of Change's hypothesis that school feeding may improve classroom time on task, this relationship remains plausible.

If students become more attentive due to school feeding, teacher attendance may increase as teachers observe positive changes in student behavior and become more motivated. Improved teacher attendance may then lead to stronger education outcomes for students. Indeed, across all rounds, we find a positive and significant relationship between teacher attendance and overall EGRA scores within intervention schools (controlling for student gender, age, and school fixed effects). However, we note that at endline, we find a significant and *negative* relationship between teacher attendance and overall literacy scores. The reason for this is not entirely clear; however, this finding suggests that teacher attendance is not sufficient to improve educational outcomes. The quality of instruction is also a vital component for student learning, discussed more below.

Looking at the relationship between school feeding and teacher attendance, as hypothesized, we find a significant, positive relationship between school feeding and teacher attendance within Ainaro and Manatuto at endline. This relationship remains positive and significant when we expand the data to include all intervention schools across all evaluation rounds. As proposed in the Theory of Change, this relationship could be explained by improved teacher motivation. However, the relationship may also be due, at least in part, to improved school administrator or PTA functioning, which may influence both the provision of school meals and teacher attendance. Indeed, we find positive, though not significant, relationships between provision of coaching to teachers, PTA activity, and teacher attendance, suggesting that these factors may also play a role.

Lastly, we examine potential linkages between school feeding and the use of positive teaching practices. We utilize data from all evaluation rounds for this analysis, as sample size at endline is small. Our results are less conclusive than those above; we find a significant but negative relationship between provision of school meals and the number of positive teaching practices observed, and no significant relationship between provision of school meals and use of corporal punishment or angry/harsh language towards students. In other words, these results do not show a clear link between school feeding and improved teaching. However, it is important to note that teaching practices may be expected to change slowly, and only as a result of continued training, feedback, and positive classroom results. As such, more work may be needed to reinforce any progress made during HATUTAN implementation.

# SCHOOL FEEDING PROGRAM IMPLEMENTATION

## Community-Level Systems

The McGovern-Dole learning agenda seeks to investigate the local governance and management systems necessary for effective implementation of school meal programs. This is relevant to the HATUTAN program's learning agenda, which focuses on identifying successful partnership models and exit strategies to ensure program sustainability. The majority of schools (75% at baseline and 82% at endline) reported that the school director or coordinator was responsible for overseeing the feeding program, with no significant changes observed among intervention schools as compared to comparison schools. At endline, 46% of treatment schools and 20% of comparison schools reported that the PTA was responsible for program oversight. Another 36% of respondents reported that a service provider was responsible.

In spite of some improvement in the frequency of PTA meetings in treatment schools at endline, about half of them still reported not having any meetings during the current school year, and most comparison schools (69%) also did not hold any meetings. The household survey revealed that the level of participation in the PTA remained low, suggesting limited participation and potential influence in school activities. At baseline, 27% of comparison households and 29% of intervention households reported having a member who participated in the PTA, while the corresponding figures at endline were 19% and 30%, respectively. However, the majority of schools (94% at baseline and 98% at endline) reported having a PTA and that those PTAs are doing activities to improve school feeding (76% intervention, 61% comparison). 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 schools and provide oversight of the school feeding program.

## Food Production, Procurement, and Preparation

In both the McGovern-Dole and HATUTAN program-specific learning agendas, the sustainability of meal program components such as food production, local procurement, and food preparation are important considerations. In terms of food preparation, improvements were observed in the number of schools with kitchen space and access to clean water. The majority of treatment schools reported having kitchen space (94%) and maintaining clean kitchens using detergent (78%). At endline, over 80% of all schools had

access to clean water for meals. However, all schools with a kitchen used wood stoves and only a small proportion had a scale in the kitchen. The progress made in improving kitchen infrastructure is significant, but further improvements are needed to ensure safe and sustainable food preparation.

Regarding food production and local procurement, the majority of school administrators reported purchasing goods from local farmers at endline. Only 14% of schools reported not buying produce from local farmers, with half of them being intervention schools. The types of produce purchased were consistent, with 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%) being the main items bought. Among treatment schools, the primary reasons for not purchasing from local farmers were insufficient budget (50%) and insufficient amount of produce from farmers (42%).

### Food Safety and WASH

Both the McGovern-Dole and HATUTAN program-specific learning agendas emphasize the importance of ensuring food safety and hygiene in school feeding programs. The HATUTAN program specifically looks at WASH interventions that can contribute to this goal.

The household survey conducted as part of the study revealed that the majority of schools with their own kitchen space (88% baseline, 81% endline) reported using detergent to maintain clean kitchens (79% baseline, 78% endline). However, there was a decline in the daily use of detergent in both intervention and comparison schools at endline.

At endline, the treatment group had a higher percentage of schools with observed facilities, such as canteens and handwashing stations, than the comparison group, with the exception of kitchen plates/cutlery. Additionally, access to clean water for food preparation in intervention schools increased by 13 percentage points.

In the baseline report, it was highlighted that parents expressed concern about food safety in schools due to unsanitary food preparation practices that could result in illness and cause children to miss school. However, during the endline survey, the majority of caregivers (82%) in treatment schools stated that the food served to children in school is prepared in a hygienic manner, although this figure had decreased from the baseline (92%). The findings from both the school and household surveys suggest that there were improvements in the level of hygiene maintained during food preparation in schools. However, it is unclear to what extent parents are aware of, or confident about, the hygienic preparation of food in schools.

In terms of maintaining clean storage spaces, the baseline report indicated that about 72% of treatment schools reported somewhat clean storage spaces. However, this result may have been biased as photos taken by enumerators suggested otherwise. By the endline, there was significant improvement with 94% of treatment schools reporting clean storage spaces. Additionally, at baseline, about 74% of treatment schools raised food off the ground with pallets, shelves or another method. This number increased to 94% at endline.

## HEALTH AND NUTRITION

### Relationship Between WASH Programs and Literacy

The HATUTAN learning agenda seeks to understand how WASH projects impact learning and literacy outcomes. As such, we analyze the relationships between handwashing behaviors and health/hygiene knowledge reported by caregivers and student literacy outcomes. We hypothesize that improved WASH knowledge and behaviors may lead to improved literacy outcomes if these behaviors reduce student illness and improve student attentiveness. We control for caregiver education in our analysis as a key confounding variable that may affect both student educational outcomes and health/hygiene behaviors and knowledge.

Looking first at handwashing behaviors, we analyze the relationship between caregivers who reported that they always wash their hands before preparing food and student-level outcomes of interest. For intervention areas at endline and across all evaluation rounds, we find no significant relationship between handwashing behaviors and overall EGRA scores in a variety of models controlling for caregiver education, student gender and age, and school fixed effects. Similarly, we find no significant relationship between student absences due to illness or working memory scores and handwashing behaviors. Overall, this suggests that caregiver-level handwashing behaviors may not have a direct effect on grade 2 students' learning outcomes; however, this does not suggest that caregivers' health behaviors are unimportant. These behaviors may have a greater impact on babies and toddlers, or may have an effect on children that is not measured through the data collected in this evaluation.

However, our analysis in *Health and Nutrition* suggested that caregivers likely substantially overreport their handwashing practices, as less than 10% of households had a handwashing station with soap at endline. As such, we also analyze the relationship between whether a household had a handwashing station and student outcomes. We note, however, that the limited number of households with a handwashing station with soap—only 30 intervention households at endline—reduces our ability to draw conclusions from the analysis. Potentially as a result of this limited sample, we find no significant relationships between whether a household had a handwashing station with soap and EGRA scores, absences, or working memory.

Moving on to health and hygiene knowledge, as with handwashing behaviors, we find no significant relationship between health knowledge and literacy outcomes. However, we do find a significant, though small, relationship between health knowledge and whether a student was absent due to illness: A one-point improvement in health and hygiene knowledge was associated with a 0.4 percentage point higher likelihood that the student was absent due to illness in the week preceding data collection. This finding is somewhat counterintuitive, as we would expect better knowledge of health and hygiene to be associated with less student illness. However, it is possible that student illnesses are, in this case, driving caregivers to gain more knowledge of hygiene; a caregiver with an unwell child may have more incentives to learn about health and hygiene in order to care for their child.

Overall, this analysis shows an inconclusive relationship between health and hygiene behaviors and knowledge and student literacy, attendance, and attentiveness. While we do not find a direct link between WASH interventions and literacy, however, we emphasize that this does not mean that a link does not exist.

## Relationship Between Nutrition and Student Outcomes

In the above analysis under *Relationship Between Nutrition, School Feeding, and Literacy*, we found potential positive relationships between school meal provision and student nutrition, as well as between student nutrition and literacy outcomes. This suggests that school meals may play an important role in the outcomes of students, especially malnourished students, although findings were inconclusive. To further examine these relationships, in this section, we analyze the relationship between student nutrition and student attendance and illness.

Our analysis found that at endline, around 24% of intervention students were absent due to illness on at least one day of school during the week preceding data collection. Among students who were absent due to illness, an average of around 2 school days were missed.

However, we find no significant relationship between student dietary diversity and whether the student was absent due to illness or the number of days missed due to illness. We also find no significant relationship between student undernourishment (measured as students who ate fewer than three healthy food groups) and illness-related absences. We do, however, find a significant relationship for endline intervention students between BMI and whether the student was absent due to illness: **A one point increase in BMI**

**was associated with a 2.7 percentage point lower likelihood that the student missed school due to illness,** controlling for factors varying across municipality.

Overall, these results suggest a possible, though inconclusive, relationship between nutrition and student attendance and illness. Given the possible positive relationship between school meals and nutrition, this suggests that school meals may have an effect on reducing absences due to illness. It is important to note, however, that many confounding factors may also affect this relationship, including hygiene practices (at both the student and community levels), public health measures, and access to healthcare. As such, it is important for future interventions to continue targeting health and hygiene behaviors to reinforce the benefits gained through improvements to nutrition.

## AGRICULTURE

The HATUTAN learning agenda seeks to establish how local procurement during harvest time can be supplemented with international food aid to promote sustainable school feeding. The extent to which this 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. 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. This question was analyzed in the midline evaluation for HATUTAN, and that analysis is updated here with endline figures.

The success of the school feeding program over the course of the intervention has been studied more extensively in previous sections and will be briefly summarized here. The percentage of schools in the comparison group providing meals on the day of the survey fell from 31.0% to 3.6%, while for intervention schools it rose from 1.0% to 16.3%. This led to a large and highly statistically significant difference in difference score of 42.8 percentage points. However, the percentage of intervention schools providing meals fell compared to midline, and the percentage of 16.3% still leaves great room for improvement.

There was also a benefit for student nutrition. Students in intervention schools were more likely to have eaten on the day of the EGRA assessment and the difference in difference score of 5.8 percentage points was statistically significant. Therefore, the first part of the learning agenda for agriculture can be answered positively. International aid can be beneficial for school feeding programs.

Next we analyze the dynamics of local procurement of food. While the percentage of comparison schools buying local produce stayed mostly flat from baseline to endline, the percentage of intervention schools buying local produce for school feeding fell slightly from 92.2% to 87.8%. This resulted in a statistically significant negative difference in difference score of 5.3 percentage points, meaning the intervention may have had a weakly negative impact on buying local produce. There are multiple possible explanations for this trend and there are likely multiple causes. Intervention schools were chosen in part on the basis of greater deprivation and remoteness; as such, the lasting effects of COVID restrictions and supply pressures may have impacted both groups differently. Therefore, the picture at present based on the limited quantitative data to address the question appears to be an effective school feeding program driven by international food aid rather than both sources complementing each other. It is worth noting however that most schools still purchase produce locally, at nearly 90%.

***Note from the program:*** Since schools had not yet received SFP funds to conduct local purchases at the time of the baseline, midline, or endline, their responses referred to the practice during previous school years. Since the responses refer to past periods, it is not possible to verify the actual occurrence of purchases or the content of meals. Therefore, responses may be severely affected by desirability bias, particularly in comparison schools, which do not expect the level of scrutiny treatment schools are used to.

Finally, it is also of interest in the learning agenda to understand if there are different outcomes for schools relying on international aid against those relying on local procurement. However, since provision of government funds for the SFP was delayed at endline, schools were generally unable to source local foods due to a lack of money to do so. As such, our analysis suggests that schools relying on local procurement were much less likely to have been in a school that provided meals than those relying on international food aid; however, this is a reflection of inconsistent government funding.

# PROGRAM-SPECIFIC OUTCOMES

In addition to measures of program impact discussed in prior section, the endline evaluation also seeks to assess the performance of HATUTAN along Organization for Economic Cooperation and Development (OECD) Development Assistance Committee (DAC) criteria. These include seven key areas: design/relevance, management and coordination, effectiveness, efficiency, sustainability, impact, and gender and social equity.

## DESIGN/RELEVANCE

The HATUTAN program was designed to address a wide variety of factors known to affect literacy and health. Program activities pivoted at the onset of the COVID-19 pandemic to ensure that activities remained relevant and responsive to the changing context, and pivoted again following the easing of COVID restrictions. Below, we analyze the design and relevance of key activities, keeping in mind the constraints imposed mid-program implementation due to COVID-19. We analyze if interventions were relevant to the priorities of the Government of Timor-Leste (GoTL) and schools and responsive to the needs of students, households, farmers, and schools.

### Training School Administrators

The HATUTAN Theory of Change argues that strengthening school administrators' management skills and leadership is important to improve school outcomes and achieve sustainability. Administrators are expected to improve literacy outcomes through their involvement in coaching and monitoring teachers, manage school infrastructure to improve student attendance and SFP functioning, and advocate for and manage the SFP.

Throughout the program, HATUTAN has worked closely with school administrators, both directly and through the MEYS. Administrators have received trainings on developing student literacy, management and budgeting, and gender-based violence. Administrators were also provided with guidance during COVID-19 to adapt and respond to the pandemic.

At endline, we find no significant correlations between directors' years of experience, education levels, and provision of coaching to teachers and overall literacy scores and provision of school meals. We do, however, find a significant, negative correlation between directors' years of experience and the use of engaging teaching practices in schools, with more years of experience correlated with significantly less use of engaging teaching practices. This finding suggests that longer-tenured directors may be more familiar and comfortable with traditional teaching practices, emphasizing the need for continued training of directors to ensure they understand the value of engaging teaching.

Additionally, we find that at endline, 89% of respondents to the school survey stated that intervention directors were responsible for the SFP. Furthermore, the majority of respondents to the school survey stated that if a girl is harassed or abused, she should report the abuse to the school director/coordinator first. As such, despite the inconclusive link between administrator experience and quality and several outcomes of interest, administrators remained highly relevant to the goals of HATUTAN at endline. Training

of school administrators appears to have had particular relevance for the SFP and to improve avenues for reporting gender-based violence in schools.

## Strengthening PTAs

HATUTAN identified PTAs as important to monitor school infrastructure, teacher attendance and quality, school meal provision, and other key outcomes of interest. Establishing and strengthening PTAs also provides an important means of sustainability. Program activities thus sought to strengthen (or establish) PTAs, facilitate community meetings on school quality, and improve PTA capacity to monitor the SFP, teacher attendance, school budgets, and gender-based violence in schools, among other aspects.

At endline, the majority of schools had PTAs, and most PTAs reported active involvement in improving school infrastructure, overseeing the SFP, monitoring safety and security, monitoring teacher and student attendance, and ensuring learning quality. As such, activities targeting PTAs are clearly relevant to the overarching goals of HATUTAN. However, we also note that more than half of PTAs did not hold any meetings since the beginning of the current school year, and PTA involvement in areas such as school budget management and monitoring dropout rates was relatively low.

## Provision of Reading Materials

Students' access to reading materials at school and at home can substantially help strengthen reading abilities. As such, HATUTAN sought to develop content for the Lafaek student and teacher magazines, provide literacy teaching and reading materials during COVID-19, and support school administrators to manage their school's reading materials.

Strengthening access to reading materials remained a relevant program activity at endline, as only 66% of intervention schools were observed to have storybooks or magazines that could be used by grade 2 classes and only 65% of observed classrooms had a reading corner. Furthermore, only 54% of households had children's books or magazines.

However, provision of reading materials does not always translate into student access to reading materials. For example, 9% of schools which were observed to have reading materials for grade 2 students also reported that they did not lend storybooks for students to take home. Schools that did not lend books reported doing so out of concern that children were careless or would lose the books. As such, while provision of reading materials remained relevant at endline, more work is needed to ensure that reading materials are actually available to students.

## Provision of School Meals

Provision of commodities to support the government-run SFP was a primary activity of HATUTAN. At endline, we found significant linkages between school meal provision and EGRA scores and working memory, and a positive (though not significant) relationship between school meal provision and student attendance. The provision of school meals continues to be highly relevant for student in Timor-Leste, helping to improve nutrition, student attendance, student attentiveness, and overall literacy abilities. HATUTAN support for school feeding is further relevant due to challenges in government management and funding of the SFP.

## 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. 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.

At endline, most intervention schools stated that they sometimes purchased produce from local farmers for school meals. The primary reasons for not purchasing produce were insufficient budget and insufficient amounts of local produce. Support for farmers' groups may help mitigate both of these issues; increasing local production will clearly help deal with supply issues, but may also help reduce prices if farmers learn more efficient and cost-effective ways to grow produce. In other words, at endline, support for farmers' groups remained relevant for the overarching goal of improving the dietary quality of school meals.

### Forming VSLAs

The HATUTAN program sought to form VSLAs in order to serve as a foundation for other trainings on topics such as nutrition, agriculture, and gender. VSLAs also provide a useful source of monetary support for their members. As such, through Community Development Agents, HATUTAN established and provided support to VSLAs in intervention communities, reaching a total of 3,799 members.

Because VSLAs are the foundation for other trainings, their establishment and support was of clear relevance to HATUTAN's goals. However, provision of support solely through VSLAs may exclude households who are not included in these groups. To avoid this risk, the program has also provided training directly to PTAs, community health volunteers, and to teachers and administrators.

### Training on Health and Nutrition Practices

Improved health, hygiene, and nutrition practices among caregivers can help improve student health and dietary quality, a key goal for HATUTAN. These practices can also help improve education outcomes by increasing student attendance and attentiveness. HATUTAN program activities in this area included development of a behavior change strategy, provision of trainings to community members, and training of school cooks to improve safe preparation of nutritious school meals.

The endline evaluation found that health and hygiene knowledge was high, but that this knowledge may not be effectively translating into good health practices. As such, trainings may not have been appropriately designed to address the behavior change barriers to health and hygiene practices. Additionally, while training on nutrition appears to have been relevant given more limited knowledge of good nutrition practices, there appear to be further barriers to nutrition than those addressed by HATUTAN, including economic constraints and cultural norms. As such, while activities in this area were broadly relevant, they could have been more effectively tailored to the local context to enhance impact.

### Capacity Building and Advocacy

HATUTAN had a strong focus on capacity building and advocacy at the national level both during COVID-19 and outside of the context of the pandemic. Activities included advocating for policy changes related to the SFP and education and strengthening the government of Timor-Leste's ability to deliver the SFP. Unfortunately, as discussed in this evaluation report, government capacity to fund and manage the SFP remained limited at endline. As such, this workstream remains highly relevant; however, its design may not have been fully effective to reach program goals.

## MANAGEMENT AND COORDINATION

Effective management of HATUTAN was vital to ensure program success, as was coordination both across implementing partners and with external stakeholders, such as the GoTL. In the period following the midline evaluation, HATUTAN was able to retain many program managers, including its Chief of Party and Deputy Chief of Party (who shifted to a new role under HATUTAN II in October 2022). In one case, the program was able to promote HATUTAN staff to a higher-level managerial role. In another, continuity was ensured through overlap between a newly hired staff member and outgoing staff. These measures helped ensure

that program managers had a high level of familiarity with the program, improving its efficacy and coordination.

No internal coordination challenges were reported within program documents. Externally, the program coordinated effectively with government partners, including the Ministry of Education, Youth, and Sports (MEYS), Ministry of Health (MoH), Ministry of State Administration, and Ministry of Agriculture and Fisheries. A consistent challenge to this process, however, came from frequent turnover of ministry staff; in a one-year period, for example, three Director-General were replaced within the MEYS, in addition to other key staff within the MEYS and MoH. Additionally, coordination with the MEYS was often a lengthy process, and MEYS review and approval of reports and proposed materials was reported to take up to several months.

Despite these challenges, HATUTAN worked closely with the GoTL to involve ministries in program activities where appropriate and to advocate for systems-level change, including for more consistent and reliable school feeding support. Other development partners were also included in conversations, further raising the efficacy of coordination.

# EFFECTIVENESS

The effectiveness of HATUTAN—as with most other development programs worldwide—was affected by the COVID-19 pandemic. However, HATUTAN was able to pivot activities relatively successfully in response to the pandemic, such as by providing take-home rations to students and increasing hygiene-related trainings. Furthermore, given that the COVID-19 pandemic was an event that could not have been predicted, controlled for, or prevented by the HATUTAN program, in this section, we focus on analyzing other issues relevant to program efficacy. At baseline, three factors were identified as having a potentially large impact of effectiveness: school infrastructure, PTAs, and school administrators. We also analyze several factors identified at midline, including diversity of learning outcomes within classrooms, students' access to literacy materials, and cultural norms. Lastly, we identify systems-level issues as another key constraint to effectiveness.

**School infrastructure** was identified as a potential constraint to effectiveness because SFP implementation requires infrastructure, including a kitchen and clean water source, to support it. The below table shows the change in access to relevant school infrastructure across rounds for intervention schools. At endline, the majority of intervention schools—93.9%—had a school kitchen; this has, however, decreased since midline. Furthermore, 18.5% of schools did not have clean water to prepare meals, an increase since midline; only 23.9% had a handwashing station and 39.1% had enough food storage capacity, decreases from both baseline and midline; and only 20.4% had a canteen or a place for students to eat.

**Table 86: Key infrastructure within intervention schools**

| | BL | ML | EL |
|---|---|---|---|
| n | 98 | 93 | 98 |
| School kitchen | 93.9% | 97.9% | 93.9% |
| Clean water | 68.5% | 87.9% | 81.5% |
| Handwashing station | 41.8% | 35.2% | 23.9% |
| Sufficient food storage | 40.8% | 45.1% | 39.1% |
| Canteen | 17.4% | 5.4% | 20.4% |

Overall, these findings suggest that while the HATUTAN program aimed to address these key infrastructure-related issues, it had limited success. At endline, infrastructure remained a constraint to the establishment of effective and hygienic SFPs in all intervention schools.

> ***Note from the program:*** Since school meals were not being provided in a large proportion of intervention schools at the time of the endline, kitchens, storage, and canteen spaces were not in use. This may have artificially increased the reports of missing key infrastructure.

**PTAs** were identified at baseline as playing a critical role in the success of SFPs and overall school management. PTAs are discussed in detail in the section *School Management*; here, we note that the majority of schools reported having a PTA and that the frequency of PTA meetings improved among intervention schools at endline, although half of PTAs still reported that no meetings had been held during the current year. Furthermore, PTAs reported greater involvement in monitoring learning quality and improving school infrastructure at endline. This progress in PTA involvement likely helped boost program effectiveness, in addition to improving sustainability.

**School administrators** were also identified as crucial for efficacy due to many activities' reliance on the ability and motivation of administrators to assist with implementation. Several program activities thus focused on training school administrators, including in budgeting, procurement, human resources management, community mobilization, classroom management, and more. Quality of administrators is again discussed in more detail in *School Management*; a key finding from this section is that over half of administrators reported providing coaching to teachers on a weekly basis and only 6% stated that coaching had never been provided. This suggests that HATUTAN activities may have effectively targeted administrators, although there was still room for further improvement at endline.

Within classrooms, a **wide range of literacy abilities** among students was identified as a factor that may challenge program effectiveness due to the difficulty of teaching effectively for a diversity of skill levels. At endline, this challenge remains relevant: We find an average "skill gap" (i.e., difference between the highest overall EGRA score and the lowest overall EGRA score in a school) of 41.3 percentage points. Furthermore, at endline, 20% of intervention students came from schools that included students with no literacy abilities (i.e., students scoring 0% on the EGRA) as well as students with strong literacy abilities (50% or greater on the EGRA). Over half of students came from schools that included both students with very weak literacy abilities (5% or less on the EGRA) and relatively strong abilities (33% or greater on the EGRA). Given the acute challenges teachers face in any context to teach to a wide range of skill levels, this remains a constraint to effectiveness at endline.

While students' **access to literacy materials at school** increased across all evaluation rounds, 16.3% of intervention schools at endline did not lend story books to students—a decrease since midline, but still 16 schools in total. Among these schools, the majority reported that they did not lend books because students would lose them or were careless with them. As at midline, this suggests that administrators or teachers may mistrust students or believe that books are "too nice" for student use. As such, in future initiatives, it may be important to address administrator/teacher perceptions towards loaning reading materials in addition to providing materials.

Regarding **cultural norms**, the midline evaluation emphasized that the tendency for gender and power norms to change very slowly posed a challenge to effectiveness. Here, we expand upon this challenge to suggest that the slow-changing nature of many cultural norms may pose a similar challenge to efficacy. One key area is in norms around nutrition and health practices; typical diets, for example, are heavily influenced by culture and can be highly resistant to change. Similarly, health behavior change often happens very slowly, and is not only influenced by knowledge but by cultural dynamics, beliefs, the actions

taken by others within a community, and a number of other factors. As such, the HATUTAN program's emphasis on increasing knowledge—particularly of health dynamics, where knowledge appears relatively high—may have had limited efficacy in strengthening practices, as it only addresses one dynamic of behavior change.

Lastly, we note that **systems-level challenges with SFP implementation** posed a challenge to HATUTAN's effectiveness. Persistent delays in government provision of SFP commodities and support limited HATUTAN's ability to support the SFP in a timely manner, and resulted in limited measures of SFP success at endline compared to midline. It remains necessary to continue advocacy with the GoTL to improve the consistency and reliability of SFP support.

## EFFICIENCY

Efficiency refers to the extent to which program activities delivered results in an economic and timely way, including the program's value for money. If program activities were excessively costly and had only limited impact on outcomes of interest, the dedication of resources to these activities may not have been justifiable. Furthermore, even if value for money is high, better understanding of program efficiency will allow for further improvement in the use of limited resources. Understanding efficiency is also important from an operational perspective to better judge the feasibility of achieving outcomes given limited resources.

We do not analyze program budgets here; furthermore, calculating robust value for money measures can be difficult in the context of development programs due to the difficulty of costing outcomes.[195] Instead, we note some key considerations for an efficiency analysis that could be undertaken by CARE or the program funder.

First, resources should be considered holistically, not just in terms of monetary cost but also in terms of human, environmental, and time costs. Provision of commodities for school feeding or of literacy materials, for example, may have a relatively high financial cost but low human and time costs. Training to school administrators, meanwhile, may have a lower financial cost but require extensive staff support and a great deal of time. All of these factors must be considered in a robust analysis of program efficiency.

Second, efficiency should be considered not just in terms of the number of beneficiaries, but also in terms of higher-level impact on beneficiaries and potential trickle-down effects. For example, relatively few school administrators—503 in total—benefitted from HATUTAN, especially compared to the 90,000 school-aged children who were supported by the program. However, if activities targeting school administrators enable administrators to advocate for enhanced government support for school feeding, provide effective training to teachers, and engage with PTAs to support students, then the 503 direct beneficiaries may extend benefits to thousands of additional beneficiaries. In other words, outputs (e.g., number of trained administrators) and overall program impact may vary by type of beneficiary and program activity.

Third, it is important to again note that changes in knowledge do not always correlate with changes in behavior. As such, activities which seek to improve knowledge may not efficiently translate to behavior change, especially as these activities can be costly in terms of time and human resources.

Lastly, it is important to note that HATUTAN's design intentionally reduces some aspects of efficiency in order to reach the most at-need beneficiaries. The selection of highly remote, rural schools for intervention can significantly increase program costs; despite this, selection is justifiable in terms of development impact. Any efficiency analysis must therefore be grounded in this context.

[195] For example, it is difficult to calculate the specific monetary benefit of an improvement in literacy outcomes or in the proportion of students reporting that they ate a school meal.

## SUSTAINABILITY

In order to ensure that the benefits of HATUTAN persist after close of the program, sustainability was a prominent consideration during implementation. The HATUTAN work plan focused on sustainability through capacity-building and training activities at the local, regional, and national levels. These activities included training of school administrators, school cooks, and teachers; establishment of VSLAs and farmer's groups; and strengthening of PTAs, aspects discussed throughout the report.

The HATUTAN program incorporated **training** within most of its activities, with ten out of 12 activities relying on training as a critical element. Strengthening school administrators was expected to increase sustainability by enabling administrators to better manage schools, apply for funding, advocate for the SFP, and coach teachers. Training of school cooks was expected to increase sustainability by improving cooks' abilities to prepare nutritious and hygienic school meals. Training of teachers was expected to result in sustainable improvements to literacy and other learning outcomes through improved use of engaging teaching practices and reductions in ineffective or harmful practices. Training of farmers was expected to sustainably strengthen SFPs by improving schools' linkages with local producers; this represented a key success at endline, with 87.8% of intervention schools stating that they had purchased produce from local farmers for school feeding. Lastly, strengthening of PTAs was expected to provide sustainable support for schools and students by improving accountability of schools to the local community.

We note that a key challenge to this approach to sustainability is attrition of trained school personnel. Indeed, at endline, we found that administrators had, on average, the same amount of experience as at baseline, suggesting that there has been substantial attrition. Regardless, it is worth noting that while attrition may reduce program benefit to intended beneficiaries, it does not erase program impact. Instead, the benefits of training on administrators (or cooks and teachers) may accrue to a different group of students/schools than selected by the program.

**Advocacy and capacity-building** has also occurred at the systems level through engagement with national line ministries and municipal authorities, although the latter are highly constrained by lack of resources. Since midline, key activities along these lines have included development of a school feeding manual with involvement from the MEYS, supported by the Ministry of State Administration's preparation of a Decree Law on SFP funding; interaction with suco council members on gender awareness and prevention of gender-based violence; and advocacy for timely, predictable, and appropriate cash transfer to schools for the SFP. To further enhance partnerships with the government and ensure buy-in of government officials, HATUTAN has also included GoTL representatives in joint monitoring sessions, presented the results of the midline evaluation to GoTL officials, and engaged GoTL officials in the tool design and training processes for the endline evaluation.

These activities, however, faced challenges. Review and approval of the SFP funding by GoTL was substantially delayed. Meetings with suco councils revealed that following trainings on gender, no steps had been taken by most councils to implement action plans due to lack of time or financial resources and COVID-19-related difficulties. Furthermore, as noted in the *School Feeding Program* section, timely disbursement of government funds remains a major issue, especially from municipal administrations to schools.

Lastly, HATUTAN engaged community members, teachers, and parents in a **collaborative approach to policy implementation**. This was a crucial step for sustainability, as top-down policy implementation that involves, for example, only government officials or central school administrators is less likely to succeed, as it is less likely to have buy-in from those responsible for carrying out policies.

To measure progress towards sustainability and graduation from the program, HATUTAN developed a sustainability assessment plan. This included five targets listed in the below table. For each target, we include commentary on achievement at endline.

**Table 87: Summary of HATUTAN sustainability**

| Target | Achievement |
|---|---|
| Improved national and municipal support for school feeding | **Moderate.** HATUTAN has seen several advocacy successes in improving government support for the SFP, including supporting the Ministry of Finance's earliest-ever transfer of SFP funds to municipalities. However, challenges remain at both the national to municipal level and the municipal to school level; challenges at the latter level delayed SFP funding transfers to schools despite prompt national-municipal transfers. More work is needed to improve timely support for school feeding. |
| 100% of schools provided with daily school meals without USDA support | **Limited.** This target has not been achieved at endline; most schools were not providing students with meals due to delays in government funding transfers. Even when government transfers are provided, USDA support remains important to supplement SFP limitations. |
| Increased student literacy rates | **Moderate.** The endline evaluation found significant improvement in literacy for intervention students relative to comparison students, but overall literacy rates remain low, and few students are capable of reading and understanding a grade-appropriate passage. |
| Improved nutrition, health, hygiene, and gender equity practices | **Limited.** The endline evaluation found no significant improvements in health or nutrition practices. Dietary quality remains low and handwashing practices have declined since midline; while knowledge of health and nutrition is fairly high, this does not appear to have translated to behaviors. Gender equality practices have also seen limited change over the past five years. |
| Improved prevention and response to domestic, sexual, and gender-based violence | **Limited.** Gender-based violence remains a persistent issue. As described above, trainings with suco councils on this issue had limited tangible impact. It is important to note that attitudes towards gender and power tend to change very slowly over time. |

Overall, while the HATUTAN program has had valuable impact on students, households, and schools, its impact does not yet appear to be fully sustainable. Support for HATUTAN II, to be implemented over the next five years, should focus on strengthening these indicators in order to allow communities and the GoTL more broadly to graduate from the program.

# IMPACT

Program impact is the main focus of this evaluation report. The table below provides a summary of key findings.

**Table 88: Summary of HATUTAN impact**

| Outcome | Impact |
|---|---|
| Literacy | **Moderate.** The endline evaluation finds evidence that HATUTAN had a significant impact on improving overall literacy scores in intervention schools. The program appears to have had greater impact on students with some literacy abilities than on students with no literacy abilities (i.e., students who scored 0% on the EGRA). Despite this impact, overall literacy scores remain very low, at only 12.0% for intervention students at endline. Furthermore, only 20.4% of intervention students demonstrated that they could read and understand the meaning of a grade-level passage at endline. |
| Quality of instruction | **Moderate.** We find a substantial and significant decrease in the use of traditional teaching practices in intervention schools compared to comparison schools, and an increase in availability of reading corners and materials. Additionally, we find some suggestive evidence that HATUTAN may have successfully motivated male teachers to enhance their treatment of female students.<br><br>However, we find no evidence of a significant relative improvement in the use of most engaging, negative, or gender-biased teaching practices within intervention schools. Additionally, we find a significant, negative relative difference in teacher attendance across treatment groups, with comparison schools outperforming intervention schools. |
| Student attentiveness | **Moderate to high.** We find evidence that HATUTAN may have had a significant positive impact on student attentiveness as measured through classroom observations. Additionally, we find that the HATUTAN program may have had a significant impact on reducing student hunger at endline. |
| Student attendance | **Limited.** We find a significant decline in student attendance in intervention schools compared to comparison schools, and no significant difference in dropout rates across treatment groups. However, we do find a relative decline in intervention students reporting that they are afraid of, avoid going to, or are unsafe going to school at endline. |
| School management | **Limited to moderate.** We find evidence that HATUTAN may have improved the involvement of PTAs, particularly in monitoring learning quality, improving school infrastructure, and monitoring safety and security. However, we find no significant impact on school administrators' provision of coaching nor in the prevalence of PTAs or the frequency with which they meet. |

| | |
|---|---|
| School Feeding Program | **Moderate.** Evidence suggests HATUTAN had a strong impact on the SFP at midline. While there was a regression in school meal provision at endline due to delays in government funding, intervention schools were still more likely to provide school meals than comparison schools. Additionally, we find evidence that HATUTAN may have had a positive impact on the development of school menus. However, HATUTAN appears to have had limited impact on hygienic preparation of school meals or on purchase of local produce for school meals. |
| Nutrition knowledge and practices | **Limited.** We find weak dietary quality for caregivers and children, extremely high prevalence of underweight BMIs, and relatively limited nutrition knowledge among caregivers, with no evidence of HATUTAN impact. |
| Health knowledge and practices | **Moderate.** We find evidence that HATUTAN had a positive impact on health and hygiene knowledge, with significant improvement relative to comparison areas at endline. However, we find a decline in handwashing behaviors at endline and no evidence of impact on access to clean water and sanitation. |
| Agricultural practices | **Limited.** We find no evidence of program impact on use of permagardens, economic outcomes of farmers, or farm sales from midline to endline. However, we note that sample size of farmers was limited, reducing our ability to draw conclusions about impact. |
| Economic empowerment | **Moderate.** We find that HATUTAN may have had positive impact on household savings at endline. However, we find no evidence of an increased use of savings or loans on education. |
| Gender and power | **Limited.** We find some evidence of reduced gender gaps in children's participation in housework. However, we find little change in other areas of interest, including control of productive assets, participation in decision making, and indicators of gender-based violence. |

## GENDER AND SOCIAL EQUITY

The HATUTAN program included an explicit gender equity focus; evaluations also sought to better understand the social equity impact of the program by collecting data on disability status of students and other factors which may disadvantage students or households. Furthermore, the program's decision to target municipalities with the worst education and health indicators in the country represents a further focus on equity.

Broadly, the program successfully engaged girl students and female caregivers (including mothers). Consistent with patterns found across Timor-Leste, we found stronger education outcomes, including engagement and achievement in school, for grade 2 girls compared to boys. Teachers' attitudes towards girls and boys were also fairly equitable, and HATUTAN appears to have motivated male teachers to enhance their treatment of female students.

However, although girls' learning achievements and treatment in schools are on par with or above those of boys at young ages, girls still face many gendered challenges which may lead them to fall behind boys later in education and as adults. Girls bear many obligations for both academic performance and housework; at endline, while we find that girls and boys spend similar total amounts of time on housework (as reported by

caregivers), we also find that girls are responsible for significantly more tasks than boys.[196] As a result of these obligations, girls may eventually struggle to keep up with both academic and household responsibilities.

Furthermore, our analysis shows that adult women still have limited decision-making power in households, are often limited to working as caregivers or in low-paying jobs, and face gender-based violence. Gender norms often encourage young women to start families and stay at home to care for children, rather than pursuing higher education or a career which could provide the woman with greater ability to make decisions and advocate for herself in the household. Overall, despite the great potential shown among girls in school and HATUTAN interventions, gender norms and barriers continue to reduce the options available to girls as they grow older.

It is also important to note that in the context of Timor-Leste, some of the differential impacts observed on girls serve to widen the gender gap, particularly in learning outcomes where girls outperform boys at young ages. For example, as shown in the section *Literacy Outcomes*, girls not only consistently outperform boys in literacy assessments but also appear to have benefitted more from HATUTAN programming. One reason driving this may be teachers' greater likelihood of using negative teaching practices, such as a harsh tone or corporal punishment, on boys than on girls. Furthermore, caregivers were somewhat more likely to believe that girls were better at reading and math than boys, suggesting a gendered bias which may affect boys' learning. Overall, these findings suggest that particularly at young ages, there may be a need to increase focus on boy students to help them catch up to girls.

As noted in prior evaluations, these findings suggest a need for intervention to sustain girls' successes as they age, and potentially to improve boys' educational performance at young ages. However, gender norms tend to change slowly over time, and we found little change in gender and power as a result of HATUTAN interventions at endline. This is not necessarily evidence that interventions were ineffective due to the difficulty of influencing cultural norms and the length of time necessary to do so. However, it may be useful to undertake further investigation during HATUTAN II to enable the program to have greater influence over gender norms.

# CONCLUSIONS AND RECOMMENDATIONS

This endline evaluation has shown that HATUTAN had positive impact on a wide range of outcomes, including literacy, quality of instruction, student attentiveness and hunger, the school feeding program, health and hygiene knowledge, and household savings. In other areas, impact was more limited due to multiple factors including the mid-program interruption of the COVID-19 pandemic, the tendency for cultural norms to change slowly over time, and challenges faced by the Government of Timor-Leste.

Given that HATUTAN programming will continue over the next five years under the HATUTAN II program, in this conclusion, we focus on providing recommendations for HATUTAN II. In some cases, conclusions validate the approach to improving nutrition, health, and literacy for primary grade students. In others, they suggest gaps in impact that could be addressed with small changes to program activities.

## LITERACY AND EDUCATION

**The HATUTAN program had a positive impact on grade 2 students' literacy outcomes** as measured through the EGRA. Following a decline in literacy scores at midline due to the COVID-19 pandemic and resultant school closures, at endline, literacy scores had rebounded to levels above baseline for intervention

[196] Including varying for family members, cooking or cleaning, fetching water or firewood, helping with agricultural work, and helping with a family business or other work.

students. We find that the HATUTAN program may have both mitigated the negative impacts of COVID-19 on learning, as shown at midline, and allowed students to learn more quickly after returning to school post-COVID, although the program's impact on literacy post-COVID appears to have been more modest.

**However, literacy abilities remain very weak, and there is a major gap in literacy abilities between letter recognition and word recognition.** The average overall score on the EGRA was only 12.0% for intervention students, and only 20.4% of intervention students demonstrated the ability to read and understand the meaning of a grade-level passage. Scores were highest—though still low in absolute terms—for letter recognition, at 24.1% for intervention students. Scores then dramatically decreased for subsequent tasks requiring students to read words and passages or interpret the meaning of a text.

**HATUTAN appears to have had greater impact on students who were at least able to recognize letters, rather than students without any reading ability.** We find no significant impact on the percent of students with no literacy abilities (i.e., students scoring 0% on the EGRA). However, both for overall literacy scores and for most subtasks, we find a positive and significant improvement in scores among students scoring greater than 0%.

**Recommendation: Improve HATUTAN II's targeting of students with no literacy abilities.** These students appear to have been insufficiently targeted by HATUTAN activities, which were more effective at strengthening literacy among higher-performing students. Tailoring future activities to more effectively enable students to at least recognize letters will increase the program's reach and help achieve its goal of improving literacy among disadvantaged Timorese children.

**We find that HATUTAN had a significant impact on lessening the use of traditional teaching practices in intervention schools.** However, we find no significant impact on the overall use of engaging or negative teaching practices, although we find a possible reduction in the use of corporal punishment towards girls. As such, for HATUTAN II, it may be useful to adjust activities targeting engaging and negative teaching practices to enhance program impact on quality of instruction.

**HATUTAN also had a positive impact on access to literacy materials in grade 2 classrooms**, with a significant increase in reading corners. Furthermore, over 80% of intervention schools reported lending books for students to take home. However, we note that schools that did not lend books generally reported doing so out of fear that students were careless or may lose the books. As such, simply increasing the availability of reading materials may not always translate to increased student access to those materials. It may be helpful for HATUTAN to address barriers to students' access to literacy materials outside of just the number of materials available, including administrator attitudes towards book lending.

**Student attentiveness, as measured through classroom observations, appears to have been positively affected by the HATUTAN program.** This may have been through HATUTAN's relative impact on reducing student hunger; at endline, most students reported that they had eaten on the day of the EGRA, and most households were not acutely food insecure. However, as we will discuss further below, the quality of the diet eaten by students is generally low, which may lead students to rapidly feel hungry again after eating and thus impact attentiveness.

**At endline, we find a decline in student attendance in intervention schools,** while attendance in comparison schools increased. The decline in intervention schools occurred across all municipalities but was most acute in Manatuto and, to a lesser extent, Ermera. Natural disaster was frequently cited as a reason for student absences, a factor outside the control of the HATUTAN program.

## SCHOOL FEEDING

**HATUTAN had a positive impact on school meal provision at endline compared to baseline**. However, school meal provision substantially declined since midline in both intervention and comparison schools due to delays in government provision of SFP funding.

**Recommendation: Enhance the focus on advocacy activities with the Government of Timor-Leste during HATUTAN II.** Delays in SFP funding continue to have a widespread negative impact on school meal provision. While some improvements were seen in this area during HATUTAN, endline results show a clear need for continued enhancement of government capacity to provide funding regularly and consistently.

**School meals have a relatively low level of dietary diversity;** most schools serve carbohydrates, legumes and nuts, and dark green, leafy vegetables. Meals lack fruits, other vegetables, and protein other than legumes and nuts. Improving school meal dietary diversity through linkages with local farmers should be a continued priority during HATUTAN II.

**Among intervention schools providing school meals, a slight majority reported buying produce from local farmers.** Schools that did not buy local produce reported that they did not have the budget to do so or that local produce was insufficient. Unfortunately, purchasing patterns and dietary diversity of school meals continue to be constrained by the reality that the average cost of a nutritional diet using nutritious, locally available food items is too high. Continuing to support linkages between schools and local farmers will remain an important component of HATUTAN II.

**The majority of schools have a PTA involved in overseeing the SFP.** However, participation in PTAs and their levels of activity remain limited; in intervention areas, almost half of PTAs had not met during the current school year.

**Recommendation: Support PTAs to enhance their level of activity and increase household participation.** PTAs address many areas of relevance to HATUTAN and HATUTAN II, including not only the SFP but also learning quality, school infrastructure, student and teacher attendance, and safety and security. As such, PTAs are important bodies to enhance HATUTAN II's impact and sustainability.

## HEALTH AND NUTRITION

**HATUTAN appears to have had little impact on nutrition practices, including the quality of diets consumed by women of child-bearing age and children under the age of 2 years.** Indeed, we find evidence suggesting a widespread decrease in dietary quality over the past five years, with increasing reliance on grains, roots, tubers, and dark leafy greens for calories and nutrition. Consumption of protein-rich foods, in contrast, remains very low.

**Almost half of grade 2 students have underweight BMIs**, with an average BMI of just 13.9 kg/m$^2$ for intervention students. These findings emphasize that the weak dietary diversity and nutritional practices found above may have a tangible impact on students' health outcomes, and reiterate the importance of the SFP and other activities seeking to improve food consumption and dietary quality.

**Nutrition knowledge was relatively low across all rounds, with very limited change at endline.** While caregivers do have some knowledge of healthy nutrition practices, more work is needed to expand knowledge. However, we emphasize that knowledge does not necessarily translate into practice;

substantial barriers to healthy nutrition practices may remain even for caregivers with high levels of nutrition knowledge.

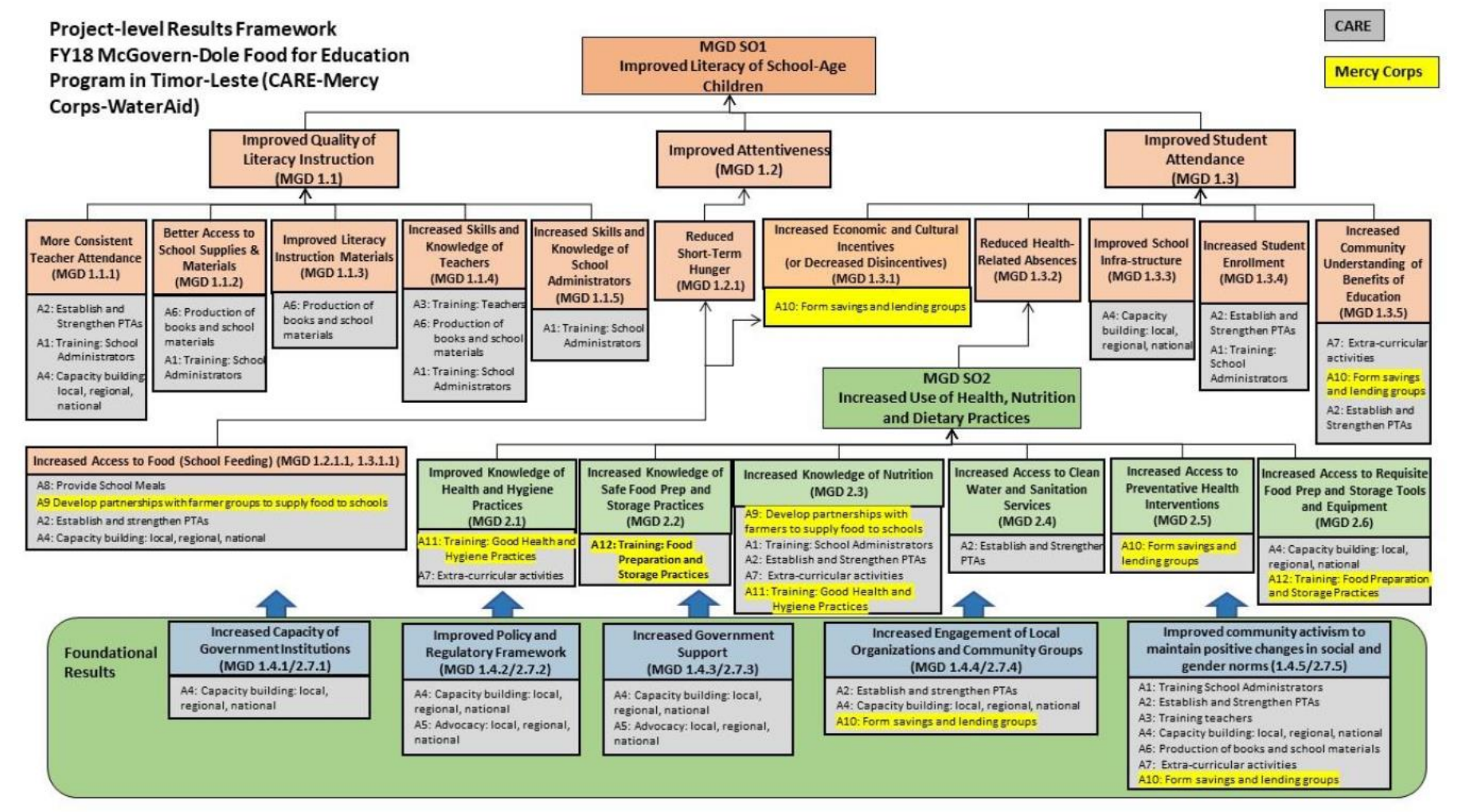
**Recommendation: Adjust activities seeking to improve the quality of nutrition in HATUTAN II.** The limited impact of HATUTAN on nutrition knowledge and practices is concerning given the impact of a poor-quality diet on children's health, attentiveness, school attendance, and general well-being. While this evaluation suggests that further improvement in nutrition knowledge is needed, our analysis of health knowledge and behaviors suggests that simply improving knowledge may not translate into improved behaviors. As such, it may be more effective to focus on other barriers to improved nutrition, such as cultural norms, economic circumstances, or preferences for processed and sugary foods.

**Handwashing practices deteriorated from midline to endline** for both intervention and comparison groups. We also find a decline in the prevalence of handwashing stations within households and schools. Given that the midline evaluation occurred during the COVID-19 pandemic, these results are likely a reflection of higher-than-average levels of handwashing due to COVID-19. At endline, then, it is likely that the below results reflect a reversion to more normal levels of handwashing, showing that behaviors adopted during COVID were not sustained.

**Across all rounds, knowledge of handwashing and hygiene practices was high** for both intervention and comparison groups. We find possible slight improvement in knowledge of hygienic practices due to the HATUTAN program.

**Recommendation: Address barriers other than knowledge which limit health and hygiene practices.** This evaluation suggests that while knowledge of handwashing practices is strong, actual behaviors remain weak. As a result, it may be more effective to pivot future programming away from a focus on knowledge and towards other potential barriers to behavior change.

# ANNEX 1: THEORY OF CHANGE



# ANNEX 2: INDICATORS AND OUTCOMES

## HATUTAN INDICATORS ASSESSED AT BASELINE, MIDLINE, AND ENDLINE

| | | | Results | | | Targets | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Indicator #** | **Indicator Description** | **Standard or Custom** | **BL (Intervention)** | **ML (Intervention)** | **EL (Intervention)** | **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<br><br>*Modified at baseline: Students were assessed 2-3 months after starting Grade 2* | Standard #1 | 1.8% can respond to 80% of questions correctly<br><br>1.3% male, 2.3% female | 1.5% can respond to 80% of questions correctly<br><br>0.8% male, 2.2% female | 1.7% can respond to 80% of questions correctly<br><br>1.3% male, 2.2% female | N/A | N/A | 2.3% (baseline +25%) | N/A | 2.7% (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<br><br>61% male, 68% female<br><br>63% rural, 72% urban | 69% of teachers<br><br>72% male, 67% female<br><br>68% rural, 78% urban | 79% of teachers<br><br>76% male, 81% female<br><br>80% rural, 75% 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 | 44% with 80% attendance recorded on the day of and day preceding the survey | 18% with 80% attendance recorded on the day of and day preceding the survey | N/A | 29% (baseline +10%) | 31% (baseline +20%) | 34% (baseline +30%) | 36% (baseline +40%) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 39% with 80% attendance recorded on the day of the survey<br><br>55% with 80% attendance recorded on the day before the survey<br><br>25% rural, 6% urban<br><br>18% Ainaro, 20% Ermera, 45% Liquica, 16% Manatuto | 76% with 80% attendance recorded on the day of the survey<br><br>55% with 80% attendance recorded on the day before the survey<br><br>39% rural, 69% urban<br><br>24% Ainaro, 44% Ermera, 55% Liquica, 73% Manatuto | 56% with 80% attendance recorded on the day of the survey<br><br>29% with 80% attendance recorded on the day before the survey<br><br>18% rural, 22% urban<br><br>11% Ainaro, 10% Ermera, 64% Liquica, 21% Manatuto | | | | | |
| MGD 1.1.2 | Percentage of schools with reading corners in grade 2 | Custom | 37% of schools<br><br>36% rural, 44% urban<br><br>36% Ainaro, 32% Ermera, 55% Liquica, 42% Manatuto | 56% of schools<br><br>55% rural, 61% urban<br><br>70% Ainaro, 49% Ermera, 73% Liquica, 42% Manatuto | 65% of schools<br><br>71% rural, 39% urban<br><br>63% Ainaro, 61% Ermera, 91% Liquica, 63% Manatuto | 39% (baseline +5%) | 41% (baseline +10%) | 45% (baseline +20%) | 48% (baseline +30%) | 53% (baseline +40%) |
| MGD 1.2 | Percent of students with working memory scores equal or above 50% | Custom | 29% of students<br><br>30% male, 28% female<br><br>29% rural, 30% urban<br><br>33% Ainaro, 26% Ermera, 35% Liquica, 27% Manatuto | 28% of students<br><br>26% male, 29% female<br><br>27% rural, 29% urban<br><br>24% Ainaro, 27% Ermera, 29% Liquica, 34% Manatuto | 30% of students<br><br>30% male, 30% female<br><br>31% rural, 28% urban<br><br>28% Ainaro, 36% Ermera, 35% Liquica, 18% Manatuto | N/A | N/A | 35% (baseline +20%) | N/A | 41% (baseline +40%) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MGD 1.2.1 | Percent of students who report that they did not consume any food during the school day | Custom | 14% of students<br><br>14% male, 13% female<br><br>12% age 5-7, 15% age 8-10, 20% age 11-13<br><br>13% rural, 16% urban<br><br>14% Ainaro, 13% Ermera, 7% Liquica, 19% Manatuto | 9% of students<br><br>10% male, 9% female<br><br>9% age 5-7, 9% age 8-10, 10% age 11-13<br><br>9% rural, 11% urban<br><br>7% Ainaro, 10% Ermera, 7% Liquica, 13% Manatuto | 13% of students<br><br>13% male, 13% female<br><br>13% age 5-7, 12% age 8-10, 9% age 11-13<br><br>13% rural, 13% urban<br><br>13% Ainaro, 13% Ermera, 8% Liquica, 18% Manatuto | N/A | N/A | 9.0% | N/A | 5.0% |
| MGD 1.3 | Average student attendance rate in USDA supported classrooms/schools | Standard #2 | 70% attendance rate<br><br>69% male, 68% female<br><br>69% rural, 74% urban<br><br>63% Ainaro, 70% Ermera, 74% Liquica, 78% Manatuto | 72% attendance rate<br><br>70% male, 75% female<br><br>71% rural, 75% urban<br><br>74% Ainaro, 69% Ermera, 66% Liquica, 79% Manatuto | 64% attendance rate<br><br>62% male, 67% female<br><br>64% rural, 65% urban<br><br>71% Ainaro, 56% Ermera, 78% Liquica, 69% 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)<br><br>51% Ainaro, 49% Ermera, 56% Liquica, 56% Manatuto | 58% of parents (44% involved in savings groups)<br><br>82% of those trained on VSLAs<br><br>71% of those who took loans<br><br>55% Ainaro, 52% Ermera, 79% Liquica, 57% Manatuto | 58% of parents (56% involved in savings groups)<br><br>88% of those trained on VSLAs<br><br>66% of those who took loans<br><br>57% Ainaro, 52% Ermera, | 41.0% | 50.0% | 60.0% | 70.0% | 80.0% |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 61% Liquica, 70% Manatuto | | | | | |
| 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<br><br>1.5 days male, 1.7 days female | 0.52 days missed due to illness per week<br><br>0.5 days male, 0.5 days female | 0.8 days missed due to illness per week<br><br>0.8 days male, 0.8 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% | 78% | 87% | 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 | 40%<br><br>47% male, 39% female | 55%<br><br>39% male, 57% female | 54%<br><br>34% male, 55% 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 | 75 intervention schools | 79 intervention schools | 87 intervention schools | 136 | 143 | 149 | 154 | 160 |
| MGD 2.4 | Number of schools using improved sanitation facilities (WaterAid) | Standard #28 | 70 intervention schools | 74 intervention schools | 85 intervention schools | 156 | 159 | 163 | 165 | 165 |

# MCGOVERN-DOLE STANDARD AND CUSTOM OUTCOMES

| | Intervention | | | Comparison | | | | |
|---|---|---|---|---|---|---|---|---|
| | BL | ML | EL | BL | ML | EL | DiD (BL to EL) | P-value |
| 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 | 1,447 | 1,490 | 1,468 | 1,014 | 1,123 | 1,041 | | |
| Achieved | 23.5% | 13.0% | 20.4% | 29.7% | 15.1% | 20.2% | 6.5 | 0.12 |
| Standard Outcome 2: Percent of schools with an average student attendance rate of at least 80 percent | | | | | | | | |
| n | 55 | 90 | 72 | 43 | 80 | 53 | | |
| Achieved | 40.0% | 33.3% | 26.4% | 34.9% | 46.3% | 32.1% | -10.8 | .39 |
| Standard Outcome 27: Number of schools with an improved water source | | | | | | | | |
| n | 98 | 93 | 98 | 87 | 85 | 84 | | |
| Achieved | 76.5% | 85.0% | 88.8% | 82.8% | 76.5% | 86.9% | 8.1 | 0.18 |
| Standard Outcome 28: Number of schools using improved sanitation facilities | | | | | | | | |
| n | 98 | 93 | 98 | 86 | 85 | 84 | | |
| Achieved | 71.4% | 80.0% | 86.7% | 68.6% | 70.6% | 69.1% | 14.9 | 0.06 |
| Custom 5: Percentage of teachers adhering to improved learning practices in schools (based on demonstrating four or more) | | | | | | | | |
| n | 98 | 98 | 94 | 43 | 84 | 84 | | |
| Achieved | 64.3% | 69.4% | 78.7% | 69.8% | 69.1% | 77.4% | 6.8 | 0.53 |
| 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 | 66 | 76 | 98 | 68 | 68 | 84 | | |
| Achieved | 31.8% | 39.5% | 18.4% | 30.9% | 47.1% | 27.4% | -9.9 | 0.30 |
| Custom 7: Percentage of schools with access to reading materials in classrooms | | | | | | | | |
| n | 98 | 93 | 98 | 87 | 85 | 84 | | |
| Achieved | 57.1% | 65.6% | 66.3% | 46.0% | 58.8% | 41.7% | 13.5 | 0.18 |
| Custom 12: Percent of students who report they are attentive in class | | | | | | | | |
| n | 1,409 | 1,410 | 1,459 | 1,004 | 1,054 | 953 | | |
| Achieved | 95.5% | 95.8% | 93.3% | 96.5% | 94.5% | 92.5% | 1.8 | 0.31 |
| Custom 13: Percentage of students who report that they did not consume any food during the school day | | | | | | | | |
| n | 1,442 | 1,485 | 1,466 | 1,012 | 1,122 | 1,040 | | |
| Achieved | 13.4% | 10.6% | 13.4% | 11.7% | 10.5% | 17.4% | 5.8 | 0.02* |
| Custom 16: Percentage of days of absence from school due to illness | | | | | | | | |
| n | - | 732 | 739 | - | 621 | 601 | | |
| Achieved | - | 10.1% | 10.1% | - | 7.8% | 8.9% | -1.0 | 0.56 |
| Custom 21: Percentage of participants who can identify important hygiene/sanitation practices | | | | | | | | |
| n | 163 | 733 | 1,025 | 362 | 623 | 601 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Achieved | 82.8% | 93.2% | 94.1% | 88.1% | 93.3% | 92.0% | 7.4 | 0.12 |
| 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 | 482 | 733 | 739 | 378 | 623 | 601 | | |
| Achieved | 40.0% | 56.1% | 53.7% | 47.1% | 60.2% | 50.4% | 10.4 | 0.03* |
| Ministry of Agriculture Indicator: Number of schools procuring nutritious foods from local producers/farmers | | | | | | | | |
| n | 98 | 93 | 98 | 87 | 85 | 84 | | |
| No | 4.1% | 54.8% | 12.2% | 11.5% | 8.2% | 15.5% | 4.2 | 0.49 |
| Yes (sometimes or all the time) | 95.9% | 45.2% | 87.8% | 88.5% | 91.8% | 84.5% | -4.2 | 0.49 |
| Ministry of Agriculture Indicator: Percentage of daily school feeding foods procured from Timorese farmers | | | | | | | | |
| n | 94 | 42 | 86 | 77 | 78 | 71 | | |
| Dark green vegetables such as water spinach, lettuce, mustard, pumpkin leaves, cassava leaves | 90.4% | 81.0% | 90.7% | 88.3% | 73.1% | 78.9% | 9.7 | 0.54 |
| Pumpkin, carrot, purple sweet potato | 76.6% | 54.8% | 50.0% | 68.8% | 51.3% | 40.9% | 1.4 | 0.92 |
| Potato, taro, yellow sweet potato, cassava | 75.5% | 61.9% | 54.7% | 62.3% | 50.0% | 59.2% | -17.7 | 0.11 |
| Beans, peas, soybeans, peanuts | 66.0% | 16.7% | 26.7% | 50.7% | 26.9% | 35.2% | -23.8 | 0.16 |
| Beef, pork, sheep, goat, chicken, duck | 63.8% | 14.3% | 12.8% | 46.8% | 7.7% | 14.1% | -18.4 | 0.23 |
| Eggs | 50.0% | 2.4% | 7.0% | 41.6% | 6.4% | 5.6% | -7.1 | 0.64 |
| Rice, maize, bread | 45.7% | 21.4% | 19.8% | 45.5% | 44.9% | 25.4% | -5.9 | 0.64 |
| Cucumber, tomato, cabbage, eggplant | 45.7% | 11.9% | 8.1% | 36.4% | 6.4% | 7.0% | -8.3 | 0.53 |
| Condiments | 24.5% | 4.8% | 0.0% | 33.8% | 2.6% | 0.0% | 9.3 | 0.51 |
| Fish (fresh or dried), shrimp, other seafood | 21.3% | 4.8% | 2.3% | 22.1% | 0.0% | 2.8% | 0.3 | 0.98 |
| Tofu, tempeh | 18.1% | 0.0% | 2.3% | 19.5% | 1.3% | 0.0% | 3.7 | 0.60 |
| Mango, papaya, honeydew melon, passionfruit, other yello fruits | 17.0% | 9.5% | 10.5% | 7.8% | 0.0% | 12.7% | -11.4 | 0.37 |
| Milk (not sweetened condensed milk) | 10.6% | 0.0% | 0.0% | 5.2% | 7.7% | 1.4% | -6.9 | 0.36 |
| Coconut oil | 4.3% | 2.4% | 0.0% | 3.9% | 0.0% | 0.0% | -0.4 | 0.94 |
| Watermelon, tamarind, jackfruit | 4.3% | 0.0% | 0.0% | 2.6% | 0.0% | 0.0% | -1.7 | 0.58 |

# ANNEX 3: METHODOLOGICAL ANALYSIS

## DEMOGRAPHICS OF ACHIEVED SAMPLE AND DIFFERENCES IN INTERVENTION AND COMPARISON SAMPLES

In this section, we describe the demographic composition of the endline sample and analyze differences across intervention and comparison samples which may pose a challenge to analysis. Any differences between baseline, midline, and endline cross-sectional samples (discussed in a subsequent section) do not necessarily pose a methodological challenge if characteristics vary uniformly across intervention and comparison groups. The method of analysis becomes problematic only if characteristics change over time within just the intervention group or just the comparison group. For example, if the government of Timor-Leste implemented a teacher training program in two intervention municipalities over the past two years, we might find that learning outcomes or the quality of instructions in intervention schools had improved relative to comparison schools and mistakenly attribute this to the impact of the HATUTAN program rather than the government initiative.

To attempt to control for potential differences between intervention and comparison groups, comparable sub-districts from comparison municipalities were matched with intervention sub-districts based on language, size, location, and typical livelihoods. Comparison schools were selected from these sub-districts by identifying a set of schools with a similar average “remoteness” score as that of the intervention schools in the matching sub-district. Both intervention 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 intervention and comparison areas. Furthermore, in some cases, the best matching comparison administrative post already had too many schools with similar interventions.

To better understand the potential impact of these issues, we analyze differences in demographic characteristics across intervention and comparison groups. We note that our analysis is necessarily limited to only those variables for which data was collected. Unobserved characteristics that we cannot control for may also pose a threat to inferential validity.

### Students

We first examine demographic differences across students assessed with the EGRA. We find fairly gender-balanced samples across intervention and comparison schools, with no significant differences in gender composition between intervention and comparison groups. However, the average age of comparison students was significantly (though not substantially) lower than that of intervention students, and comparison students were significantly less likely than intervention students to speak Tetum-Prasa as their native language. Both of these characteristics may bias our analysis; we would generally expect both younger students and minority language speakers to perform worse on learning assessments. As such, differences in student ages and native languages may mean that our difference-in-differences analysis tends to overestimate the impact of HATUTAN on learning outcomes. To reduce the potential impact of this issue, we control for student age and native language in some regression specifications.

**Table 89: Differences in demographic characteristics of students by treatment group**

| Characteristic | Intervention | Comparison | Difference | P-value |
|---|---|---|---|---|
| n | 1,468 | 1,041 | - | - |
| Male | 53.0% | 49.8% | -3.2 | 0.11 |
| Female | 47.0% | 50.2% | 3.2 | 0.11 |
| Average age (years)[197] | 7.7 | 7.5 | -0.2 | <0.001* |
| Native Tetum speaker | 65.0% | 58.0% | -7.0 | <0.001* |

Table 90 additionally shows information on the disability status of students, as collected in the household survey. Students were classified as having a physical disability (difficulty with eyesight, hearing, or walking), a mental disability (difficulty remembering/concentrating, with self care, or communicating), or difficulty making friends if caregivers reported that the child had some difficulty, a lot of difficulty, or could not do this task at all.

**Table 90: Students with disabilities by treatment group**

| Disability | Intervention | Comparison | Difference | P-value |
|---|---|---|---|---|
| n | 741 | 602 | - | - |
| **Physical disabilities** | | | | |
| Eyesight | 1.5% | 0.8% | -0.7 | 0.26 |
| Hearing | 4.9% | 3.5% | -1.4 | 0.21 |
| Walking or climbing | 1.2% | 0.8% | -0.4 | 0.49 |
| **Mental disabilities** | | | | |
| Remembering or concentrating | 28.2% | 26.7% | -1.5 | 0.55 |
| Self care | 27.0% | 29.5% | 2.5 | 0.31 |
| Communicating | 18.7% | 22.0% | 3.3 | 0.13 |
| Difficulty making friends | 3.5% | 2.5% | -1.0 | 0.28 |
| Experiences anxiety or worry monthly or more often | 15.0% | 15.3% | 0.3 | 0.89 |
| Experiences depression monthly or more often | 6.2% | 8.3% | 2.1 | 0.14 |

This table first shows that physical disabilities were reported relatively infrequently by caregivers. The most frequently reported physical disability was in hearing, with around 4% of all caregivers reporting that their child had a hearing disability. There were no significant differences in the prevalence of these disabilities across intervention and comparison groups.

[197] Excluding students with reported ages under 5 years or who did not know their age, a total of 63 comparison students and 237 intervention students.

Mental disabilities, in contrast, were generally reported with much greater frequency. In particular, around 27% of all caregivers reported that their child had difficulty remembering or concentrating, 28% difficulty with self care, and 20% with communication. However, there were no significant differences across comparison and intervention students.

## Households

Moving on to households, the below table shows differences in key demographic and livelihoods characteristics of heads of household (HoHs) and the size of households. We first find that a majority of HoHs speak Tetum, and there are substantial differences across treatment groups, with comparison HoHs significantly less likely to be native Tetum speakers. In contrast, around half of households in both intervention and comparison groups have savings, and education levels and occupations are very similar across treatment groups.

**Table 91: Differences in head of household and household characteristics by treatment group**

| Characteristic | Intervention | Comparison | Difference | P-value |
|---|---|---|---|---|
| n | 741 | 602 | - | - |
| Native Tetum speaker | 73.8% | 58.6% | -15.2 | <0.001* |
| Household has savings | 52.6% | 52.5% | -0.1 | 0.94 |
| **Education level** | | | | |
| Primary or less | 59.2% | 57.6% | -1.6 | 0.55 |
| Pre-secondary, secondary, or technical school | 35.8% | 37.7% | 1.9 | 0.46 |
| University | 4.9% | 4.5% | -0.4 | 0.75 |
| **Occupation** | | | | |
| Farmer (own consumption) | 36.3% | 40.2% | 3.9 | 0.15 |
| Farmer (sale and own consumption) | 29.3% | 27.7% | -1.6 | 0.53 |
| Unemployed | 6.2% | 6.6% | 0.4 | 0.75 |
| Other | 28.2% | 25.4% | -2.8 | 0.25 |
| **Household size** | | | | |
| Average number of household members | 7.3 | 6.9 | -0.4 | 0.008* |
| Average number of children under 3 | 0.6 | 0.5 | -0.1 | 0.63 |
| Average number of children ages 5-15 | 2.9 | 2.7 | -0.2 | <0.001* |

Lastly, we find that average household sizes are significantly (though not substantially) larger for intervention households than comparison households, and that comparison households have significantly

(though not substantially) fewer children aged 5-15 on average. Assuming that higher numbers of children and larger households reduce the resources available to the household and to each individual child (as resources must be spread amongst more children), this dynamic may dampen outcomes for intervention groups relative to comparison groups.

Looking now at caregivers, the below table shows differences in few key demographic and livelihoods characteristics. First, we note that caregivers were overwhelmingly female: Over 90% of caregivers in both intervention and comparison households were female, with no significant difference between comparison and intervention groups. Comparison and intervention caregivers also had very similar average ages, around 37 years, and had little difference in occupation or education level.

**Table 92: Differences in caregiver characteristics by treatment group**

| Characteristic | Intervention | Comparison | Difference | P-value |
|---|---|---|---|---|
| n | 741 | 602 | - | - |
| Female | 93.1% | 92.9% | -0.2 | 0.85 |
| Average age (years)[198] | 37.1 | 36.8 | -0.3 | 0.57 |
| Native Tetum speaker | 73.1% | 58.9% | -14.2 | <0.001* |
| **Education level** | | | | |
| Primary or less | 61.0% | 57.1% | -3.9 | 0.15 |
| Pre-secondary, secondary, or technical school | 36.6% | 39.6% | 3.0 | 0.26 |
| University | 2.4% | 3.3% | 0.9 | 0.33 |
| **Occupation** | | | | |
| Farmer (own consumption) | 29.8% | 29.8% | 0.0 | 0.99 |
| Farmer (sale and own consumption) | 23.5% | 21.8% | -1.7 | 0.46 |
| Unemployed | 24.0% | 26.0% | 2.0 | 0.42 |
| Other | 22.7% | 22.4% | -0.3 | 0.93 |

Caregivers' native language, however, varied substantially across intervention and comparison groups; intervention caregivers were significantly more likely to speak Tetum as a native language. Assuming that Tetum speakers would be more likely to be able to help their child with schoolwork, this dynamic may lead us to overestimate program impact.

Overall, it is important to note that this analysis suggests that there are significant differences between intervention and comparison households. However, the exact effect of these differences in our analysis of program impact is unclear, as some differences are likely to result in underestimates of impact while others are likely to lead us to overestimate impact. Given that we cannot estimate the exact effect size of

[198] Excluding caregivers who did not know their age (117 caregivers).

demographic differences on outcomes of interest, it is unclear if these differences would tend to "cancel out" or would have a systematic effect on our difference-in-differences calculations.

### Schools

Lastly, we analyze differences in school characteristics across treatment groups. Many school characteristics—such as access to water and electricity or book lending practices—represent outcomes of interest; as such, we do not analyze these characteristics here. However, the below table shows differences in student-teacher ratio and the types of teachers employed in schools, dynamics which are not specifically targeted through HATUTAN interventions but which may affect learning outcomes. The table also includes the percent of schools with multigrade classes; while HATUTAN sought to improve teaching approaches within multigrade classes through teacher trainings and coaching, the program did not have a direct impact on whether or not a school had multigrade classes.

**Table 93: Differences in school characteristics by treatment group**

| Characteristic | Intervention | Comparison | Difference | P-value |
|---|---|---|---|---|
| n | 98 | 93 | - | - |
| Has multigrade classes | 47.7% | 48.4% | 0.7 | 0.92 |
| Student-teacher ratio (average)[199] | 23.5 | 21.2 | -2.3 | 0.14 |
| Permanent teachers (% of total teachers) | 38.6% | 51.5% | 12.9 | <0.001* |

Across all schools, we find that around 48% of schools have multigrade classes (which tend to make teaching more difficult, as teachers must balance a wide range of learning levels in their lessons), the average student-teacher ratio is around 22 students to one teacher, and around 45% of teachers are permanent (rather than contracted or volunteer) teachers.

Only one of these characteristics varies significantly across treatment group: Comparison schools have, on average, significantly more permanent teachers as a percent of their teaching workforce than intervention schools. If we assume that permanent teachers are more capable of effectively teaching due to potentially higher levels of education, training, or stability than contract or volunteer teachers, then this may lead us to underestimate HATUTAN impact.

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 IN SAMPLES ACROSS ROUNDS

We now analyze any differences in observable characteristics among students, households, and schools in the cross-sectional sample. We focus on differences in the baseline and endline samples, as the focus

[199] Excluding three schools that reported 0 students enrolled.

of most of our analysis throughout the report. Given that the above section found some variance across comparison and intervention groups, any differences across baseline, midline, and endline samples may bias our analysis. For example, if endline students are, on average, older and more likely to be female than baseline students, we may overestimate our measures of program impact.

## Students

We first examine demographic differences across grade 2 students assessed with the EGRA. The below table shows that the gender composition and native language of students was similar across rounds: There was no significant difference in the percent of male/female students from baseline to endline, nor was there a significant difference in the percent of students speaking Tetum natively from baseline to endline.

However, we do find a significant difference in the average age of students between baseline and endline. Baseline students were slightly, but significantly, older than endline students, with an average age of 7.8 at baseline and 7.6 at endline. Given that we would expect older students to perform better on learning assessments, this may bias our estimates of program impact slightly downward.

Additionally, we note that from midline to endline, we find a significant difference in the percent of students speaking Tetum natively, with a higher percentage at midline. This may again bias estimates of midline-to-endline program impact downward. We find no significant differences in student age or gender.

**Table 94: Differences in demographic characteristics of students by round**

| Characteristic | Baseline | Midline | Endline | Diff. (BL to EL) | P-value |
|---|---|---|---|---|---|
| n | 2,461 | 2,613 | 2,509 | - | - |
| Male | 52.1% | 51.3% | 51.7% | -0.4 | 0.78 |
| Female | 48.0% | 48.7% | 48.4% | 0.4 | 0.78 |
| Average age (years)[200] | 7.8 | 7.6 | 7.6 | -0.2 | <0.001* |
| Native Tetum speaker | 64.2% | 65.7% | 62.1% | -2.1 | 0.12 |

Looking now at student disability status as reported by caregivers, the below table shows no significant differences in the prevalence of physical disabilities but large differences in the prevalence of mental disabilities from baseline to endline. We find that at endline, students were significantly less likely to have difficulty remembering or concentrating and to experience anxiety or depression, and were significantly more likely to have difficulty with self care and communication.

These dynamics may bias reported estimates of program impact in different ways. Both memory- and communication-related disabilities may affect literacy scores; however, the reduced prevalence of memory disabilities at endline is likely to bias our estimates of program impact upwards, while the increased prevalence of communication disabilities is likely to bias estimates downwards. Given that we do not know the exact extent to which these disabilities affect literacy scores, the overall impact on impact estimates is unclear. We note, however, that regression analysis finds that both memory and communication disabilities have a significant and negative impact on overall EGRA scores, and that the estimated effect size of communication disabilities is greater than that of memory disabilities (-4.2 for communication and -1.8 for memory). As such, it is likely that the relative differences in the prevalence of these disabilities may "cancel out" to some extent (although it is unlikely that they will perfectly cancel out).

[200] Excluding students with reported ages under 5 years or who did not know their age, a total of 63 comparison students and 237 intervention students.

Difficulties with self care, anxiety, and depression, meanwhile, may affect outcomes such as student attendance or attentiveness. As with memory and communication, however, the direction of differences varies, with an increase in prevalence of self care difficulties and a decrease in prevalence of anxiety and depression. The effects of these disabilities on our measurements of impact may thus also cancel out to a certain extent, although we cannot definitively state the extent to which this is the case.

It is also important to note that Table 90 showed no significant difference in the prevalence of disabilities across treatment groups. As such, variance in disability prevalence across rounds is unlikely to have a substantial effect on our estimates of program output, as this variance occurs relatively equally across intervention and comparison groups. Differences in average age, however, is more problematic, as this also varies across intervention and comparison groups.

**Table 95: Students with disabilities by round**

| Disability | Baseline | Misline | Endline | Diff. (BL to EL) | P-value |
|---|---|---|---|---|---|
| n | 860 | 1,356 | 1,341 | - | - |
| **Physical disabilities** | | | | | |
| Eyesight | 1.7% | 1.5% | 1.2% | -0.6 | 0.30 |
| Hearing | 5.1% | 4.3% | 4.3% | -0.9 | 0.35 |
| Walking or climbing | 0.7% | 2.1% | 1.0% | 0.3 | 0.38 |
| **Mental disabilities** | | | | | |
| Remembering or concentrating | 36.4% | 36.9% | 27.5% | -8.9 | <0.001* |
| Self care | 18.7% | 35.3% | 28.1% | 9.4 | <0.001* |
| Communicating | 14.7% | 25.2% | 20.2% | 5.4 | 0.001* |
| Difficulty making friends | 4.0% | 6.1% | 3.1% | -0.9 | 0.27 |
| Experiences anxiety or worry monthly or more often | 24.4% | 16.2% | 15.1% | -9.3 | <0.001* |
| Experiences depression monthly or more often | 11.6% | 8.1% | 7.2% | -4.5 | 0.001* |

## Households

Moving on to households, the below table shows differences in key demographic and livelihoods characteristics of heads of household (HoHs) and the size of households. We find significant differences across many HoH characteristics, including significantly fewer HoHs who speak Tetum natively, have only a primary education, and are a farmer for own consumption, and significantly more HoHs who are a farmer for sale and own consumption or who are unemployed. At the household level, we also find significantly more households with savings and significantly smaller household sizes at endline.

In Table 91, we found significant differences across treatment group for only native language and household size; as such, we focus on these characteristics as the most potentially problematic for analysis. Given that fewer HoHs reported speaking Tetum natively at endline than at baseline (and midline), this may potentially

bias estimates of program impact on literacy downwards, as we would expect native Tetum speakers to be more able to assist their grade 2 children to learn.

**Table 96: Differences in head of household and household characteristics by round**

| Characteristic | Baseline | Midline | Endline | Diff. (BL to EL) | P-value |
|---|---|---|---|---|---|
| n | 860 | 1,607 | 1,629 | - | - |
| Native Tetum speaker | 70.9% | 70.1% | 67.7% | -3.2 | 0.09* |
| Household has savings | 50.6% | 63.7% | 60.6% | 10.1 | <0.001* |
| **Education level** | | | | | |
| Primary or less | 62.2% | 59.9% | 57.7% | -4.5 | 0.03* |
| Pre-secondary, secondary, or technical school | 33.5% | 34.6% | 37.0% | 3.5 | 0.08 |
| University | 3.8% | 5.5% | 5.1% | 1.3 | 0.14 |
| **Occupation** | | | | | |
| Farmer (own consumption) | 45.7% | 39.8% | 35.1% | -10.6 | <0.001* |
| Farmer (sale and own consumption) | 19.5% | 27.8% | 30.6% | 11.0 | <0.001* |
| Unemployed | 3.3% | 6.1% | 7.2% | 3.9 | <0.001* |
| Other | 31.5% | 26.4% | 27.2% | -4.3 | 0.03* |
| **Household size** | | | | | |
| Average number of household members | 7.7 | 7.4 | 7.1 | -0.5 | <0.001* |
| Average number of children under 3 | 0.7 | 0.6 | 0.6 | -0.1 | <0.001* |
| Average number of children ages 5-15 | - | 2.8 | 2.7 | -0.1[201] | 0.005* |

The effect of smaller household sizes at endline, meanwhile, may operate in the opposite direction. As above, if we assume that larger households have, on average, worse outcomes of interest due to the need to share resources among a larger group of people, this may lead us to overestimate measures of program impact.

Looking now at caregivers, the below table shows differences in key demographic and livelihoods characteristics. We find that at endline, caregivers were significantly less likely to speak Tetum natively or be a farmer for own consumption and significantly more likely to be a farmer for sale and own consumption or be unemployed than at baseline.

[201] Not measured at baseline; reported difference is midline to endline.

**Table 97: Differences in caregiver characteristics by round**

| Characteristic | Baseline | Midline | Endline | Diff. (BL to EL) | P-value |
|---|---|---|---|---|---|
| n | 860 | 1,607 | 1,342 | | |
| Female | 90.8% | 92.4% | 93.0% | 2.2 | 0.07 |
| Average age (years)[202] | 36.9 | 36.2 | 37.0 | 0.1 | 0.85 |
| Native Tetum speaker | 71.1% | 69.8% | 66.8% | -4.3 | 0.03* |
| **Education level** | | | | | |
| Primary or less | 63.0% | 63.3% | 59.2% | -3.8 | 0.08 |
| Pre-secondary, secondary, or technical school | 35.0% | 34.7% | 37.9% | 2.9 | 0.16 |
| University | 1.9% | 1.9% | 2.8% | 1.0 | 0.13 |
| **Occupation** | | | | | |
| Farmer (own consumption) | 41.3% | 33.6% | 29.8% | -11.5 | <0.001* |
| Farmer (sale and own consumption) | 16.7% | 18.9% | 22.7% | 6.0 | <0.001* |
| Unemployed | 8.8% | 21.0% | 24.9% | 16.1 | <0.001* |
| Other | 33.1% | 26.6% | 22.6% | -10.6 | <0.001* |

In Table 92, we found differences in the native language spoken by caregivers across treatment groups. As such, differences in this trait may impact our analysis. Given that caregivers were less likely to speak Tetum natively at endline than at baseline (and midline), this may bias our estimates of program impact downwards. In particular, given that we also find lower prevalence of HoHs speaking Tetum natively at endline, students may have somewhat less household support to improve their literacy skills.

## Schools

Lastly, we analyze differences in school characteristics across treatment groups. As in the above section, the below table shows differences in student-teacher ratio, the types of teachers employed in schools, and whether the school has multigrade classes.

We find only one significant difference; at endline, schools had significantly fewer permanent teachers as a percent of total teachers than at baseline. Unfortunately, our above analysis also found a significant difference in this trait across treatment groups, meaning that this difference may impact analysis. Given that schools had fewer permanent teachers at endline, and assuming permanent teachers are more highly trained and experienced and thus more effective teachers, this would lead us to underestimate program impact. To mitigate the impact of this issue, however, we include school-level controls in some regression estimates.

[202] Excluding caregivers who did not know their age (117 caregivers).

**Table 98: Differences in school characteristics by round**

| Characteristic | Baseline | Midline | Endline | Diff. (BL to EL) | P-value |
|---|---|---|---|---|---|
| n | 188 | 209 | 182 | | |
| Has multigrade classes | 51.6% | 50.0% | 44.5% | -7.1 | 0.17 |
| Student-teacher ratio (average)[203] | 22.4 | 21.5 | 23.3 | 0.9 | 0.47 |
| Permanent teachers (% of total teachers) | 58.5% | 55.3% | 47.3% | -11.2 | <0.001* |

[203] Excluding three schools that reported 0 students enrolled.

# ANNEX 4: FOOD GROUP CATEGORIES

Within the household survey and farmer's group survey, caregivers reported on the types of food they consumed during the previous day. Mothers of children under the age of 2 also reported on foods consumed by this child during the previous day. Additionally, at endline only, caregivers were asked about foods eaten by their grade 2 child (i.e., the child assessed with the EGRA) during the previous day. Within the EGRA, students were also asked about foods they had eaten at school the previous day. Answers for foods consumed by the caregiver and child under 2 were categorized by food group as listed in the below table.

**Table 99: Food group categorization, caregivers and children under 2**

| Food Group | Food Item | Respondent |
|---|---|---|
| Grains, roots, and tubers | Maize, rice, bread, cereals/porridge, noodles, rice, mash/residue, or other foods made from grains such as maize or wheat | Caregiver and child under 2 |
| | White potatoes, white yams, white sweet potato, cassava, or any other foods made from roots | Caregiver and child under 2 |
| | Thin porridge | Child under 2 |
| Legumes and nuts | Any foods made from beans, peas, lentils, peanuts | Caregiver and child under 2 |
| | Any foods made from nuts and seeds such as pumpkin, sunflower seeds | Caregiver and child under 2 |
| Dairy products | Milk or food prepared with milk (not including condensed milk) | Caregiver and child under 2 |
| | Infant formula | Child under 2 |
| | Milk such as tinned, powdered, or fresh animal | Child under 2 |
| Eggs | Eggs | Caregiver and child under 2 |
| 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 under 2<br>*Note: Counted as "flesh food" for children* |
| Flesh foods | Any meat such as beef, pork, lamb, goat, chicken, or duck | Caregiver and child under 2 |
| | Any flesh from wild animals, such as game meat, bush rats, wild birds, deer, wild boar, wild goat | Caregiver and child under 2 |
| | Fresh or dried fish, shellfish, or seafood | Caregiver and child under 2 |
| | 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 under 2<br>*Note: Counted as "Vitamin A-rich fruits and vegetables" for child* |
| 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 under 2<br>*Note: Counted as* |

| | | |
|---|---|---|
| | | *"Vitamin A-rich fruits and vegetables" for child* |
| | Ripe mangoes, ripe papaya, melon, passionfruit, or other fruits that are dark yellow or orange inside | Caregiver and child under 2<br>*Note: Counted as "Vitamin A-rich fruits and vegetables" for child* |
| | Foods made with red palm oil | Caregiver |
| Other fruits and vegetables | Any other vegetables, like cucumbers, tomatoes, cabbage, eggplant, etc. | Caregiver and child under 2 |
| | Any other fruits like watermelon, tamarind, jackfruit, etc. | Caregiver and child under 2 |
| | Any indigenous/wild fruits | Caregiver and child under 2 |

The list of foods was updated at endline for new questions asked about food consumption of students; the revised categorization is included below.

**Table 100: Food group categorization, students**

| Food Group | Food Item |
|---|---|
| Grains, roots, and tubers | Rice, bread, noodles, porridge |
| | Corn/maize |
| | White potatoes, sweet potato, cassava, banana, taro, breadfruit |
| Legumes and nuts | Tofu, tempeh, beans, soybeans, mung beans |
| | Peanuts or cashews |
| Dairy products | Cheese |
| | Yogurt |
| | Powdered milk such as Dancow or Indomilk |
| Eggs | Eggs |
| Flesh foods and organ meats | Sausage, canned meat, or dried meat |
| | Beef or buffalo, goat, beef liver, or cow intestine |
| | Pig, wild pig, or dog |
| | Chicken, chicken gizzard, or chicken intestines |
| | Fish or seafood |
| Vitamin A-rich dark leafy greens | Cassava leaves, papaya leaf, spinach, water spinach, pumpkin leaves, sweet potato leaves, taro greens |
| | Mustard greens, chinese cabbage, bok choy, watercress, fiddlehead fern, moringa leaves, wild greens |
| Other vitamin A-rich vegetables and fruits | Carrot, orange sweet potato, yellow pumpkin |
| | Ripe mango, ripe papaya, passion fruit |
| | Orange or pomelo |
| Other fruits and vegetables | Tomatoes, eggplant, bitter gourd, green pumpkin, chayote, cucumber, cabbage |
| | Snake beans/long beans, French beans, seaweed, mushrooms, zucchini, lettuce |
| | Banana, custard apple, avocado, green mango, green papaya, jackfruit |

| | |
|---|---|
| | Pineapple, guava, star fruit, watermelon, strawberries, coconut flesh |
| Processed foods | Cakes, cookies, sweet biscuit, bolu |
| | Candies, beng-beng bars, ice cream |
| | Chips |
| | Instant noodles |
| | French fries, savory pancake, fried bananas, deep fried bread, deep fried tofu or tempeh, deep fried meat, fried fish, sweet doughnut |
| | Sweetened tea, sweetened coffee, Tehbotol, Energen, Dancow, Indomilk |
| | Fruit juice, fruit drinks such as Dellos |
| | Soft drinks such as Big Cola, Floridina, Coca-Cola, Sprite, energy drinks such as Krating Daeng |

# ANNEX 5: SAMPLE CHARACTERISTICS

The below table shows the disaggregated sample for all rounds and treatment groups by gender, disability status, native language, and caregiver education level.

| | Intervention | | | Comparison | | |
|---|---|---|---|---|---|---|
| **Characteristic** | **BL** | **ML** | **EL** | **BL** | **ML** | **EL** |
| **Student gender** | | | | | | |
| Male | 51.7% | 50.7% | 53.0% | 52.6% | 51.7% | 49.8% |
| Female | 48.3% | 49.4% | 47.0% | 47.4% | 48.3% | 50.2% |
| **Student disability** | | | | | | |
| Eyesight | 1.7% | 1.0% | 1.5% | 1.9% | 2.1% | 0.8% |
| Hearing | 4.8% | 4.1% | 4.9% | 5.6% | 4.5% | 3.5% |
| Walking or climbing | 0.6% | 2.5% | 1.2% | 0.8% | 1.8% | 0.8% |
| Remembering or concentrating | 38.1% | 38.2% | 28.2% | 34.3% | 35.3% | 26.7% |
| Self care | 18.1% | 35.3% | 27.0% | 19.6% | 35.0% | 29.5% |
| Communicating | 15.4% | 24.8% | 18.7% | 13.8% | 26.0% | 22.0% |
| Difficulty making friends | 4.6% | 4.9% | 3.5% | 3.2% | 7.4% | 2.5% |
| Experiences anxiety or worry monthly or more often | 23.7% | 13.1% | 15.0% | 25.4% | 19.7% | 15.3% |
| Experiences depression monthly or more often | 10.4% | 6.0% | 6.2% | 13.2% | 10.6% | 8.3% |
| **Student native language** | | | | | | |
| Tetum-Prasa | 66.6% | 67.5% | 65.0% | 60.9% | 66.8% | 58.0% |
| **Caregiver native language** | | | | | | |
| Tetum-Prasa | 75.5% | 71.7% | 72.9% | 65.1% | 67.7% | 58.6% |
| **Caregiver education** | | | | | | |
| Primary or less | 66.0% | 65.7% | 61.0% | 59.5% | 60.6% | 57.1% |
| Pre-secondary, secondary, or technical school | 32.6% | 32.0% | 36.6% | 38.1% | 37.9% | 39.6% |
| University | 1.5% | 2.3% | 2.4% | 2.4% | 1.4% | 3.3% |

consilient