



Laos McGovern-Dole Project

Impact Evaluation

March 2020

USDA McGovern-Dole International Food for Education and Child Nutrition – Learning and Engaging All in Primary School (LEAPS) II Project Impact Evaluation

The United States Department of Agriculture (USDA) McGovern-Dole International Food for Education and Child Nutrition (McGovern-Dole) Learning and Engaging All in Primary School (LEAPS) II project in Laos is implemented by Catholic Relief Services from 2016 to 2021. The primary objectives of LEAPS II are to improve the literacy skills of students and increase their use of improved health and dietary practices in 350 schools in seven educationally disadvantaged districts of Savannakhet province. This report presents the follow-up results of the impact evaluation.

Agreement Number: FFE-439-2016/009-00

Funding Year: Fiscal Year 2016

Project Duration: 2016-2021

Implemented by: Catholic Relief Services (CRS)

Evaluation Authored by: IMPAQ International, LLC

Elnaz Safarha

Uttara Balakrishnan

Jessica Li

Daniel Zaas

DISCLAIMER: The author's views expressed in this publication do not necessarily reflect the views of the United States Department of Agriculture or the United States Government.

USDA McGovern-Dole International Food for Education and Nutrition Program – Learning and Engaging All in Primary School (LEAPS) II Project, Lao PDR

Final Impact Evaluation Report

March 2020



SUBMITTED TO

Catholic Relief Services (CRS)

Savannakhet, Lao PDR

www.crs.org

ATTENTION

[REDACTED]

CRS

[REDACTED]

SUBMITTED BY

IMPAQ International, LLC

10420 Little Patuxent Parkway

Suite 300

Columbia, MD 21044

(443)256-5500

www.impaqint.com

PROJECT

USDA McGovern-Dole International Food for Education and Child Nutrition – Learning and Engaging All in Primary School (LEAPS) II Project (2016-2021), Award Number FFE-439-2016/009-00, Lao PDR – Final Impact Evaluation Report.

TASK & DELIVERABLE

Final Impact Evaluation Report

AUTHORS

Elnaz Safarha, IMPAQ International

Uttara Balakrishnan, IMPAQ International

Jessica Li, IMPAQ International

Daniel Zaas, IMPAQ International

ACKNOWLEDGEMENTS

We thank the U.S. Department of Agriculture (USDA) and Catholic Relief Services (CRS) for their financial support. We extend special thanks to Cornelia Sage, Patrick Williams, Inpeng Bounyavong, Vatvisa Keosalivong, and their colleagues at CRS for their support and advice. We also thank CRS' remarkable field staff and community mobilizers in Lao PDR for facilitating the planning and rollout of the impact data collection.

Special thanks go to SKO, our data collection partner, their team leader, Salika Khoonbarthao, and the 18 amazing enumerators for their professionalism. Without their support and hard work in the field, we would not have been able to collect data with such high quality. We appreciate Save the Children staff for being available to provide support and advice when needed. We thank Sara Borelli, Senior Research Associate at IMPAQ International, for her solid insights and technical support.

We also acknowledge the government of Lao PDR that provided us permission to conduct fieldwork in Lao and facilitated the logistics of our data collection. Finally, we thank all the local stakeholders and beneficiaries that freely shared their perspectives and enabled us to have an in-depth analysis for the report. The findings, interpretations, and conclusions expressed in this report are entirely those of the authors. They do not necessarily represent the views of IMPAQ International, CRS, or USDA.

TABLE OF CONTENTS

Acknowledgements.....	ii
Table of Contents.....	iii
List of Exhibits	v
List of Acronyms.....	vi
Executive Summary.....	vii
Methods.....	vii
Key Findings	viii
Recommendations	ix
Section 1. Introduction	1
1.1 Overview of Lao Context.....	1
1.2 Overview of Programmatic Implementation	3
1.2.1 Provision of School Meals	4
1.2.2 Literacy Interventions	4
1.2.3 Inclusive Education (IE)	5
1.3 Evaluation Background	6
Section 2. Evaluation Approach	7
2.1 Impact Evaluation Research Questions	7
2.2 Evaluation Methodology.....	7
2.3 Sampling Design	9
2.4 Data Sources	10
2.5 Data Analysis.....	11
Section 3. Fieldwork.....	12
3.1 Human Subjects Protection	12
3.2 Preparation for Data Collection	12
3.3 Field Activities	13
3.4 Field Challenges	13
Section 4. Evaluation Sample.....	14
4.1 Distribution of Students.....	14
4.2 Student Background at Follow Up	15
4.2.1 Student Characteristics	15
4.2.2 Household Environment	16

4.3 Attrition at Follow Up	19
Section 5. Results	22
5.1 School Feeding and Literacy Boost Interventions.....	24
5.1.1 Letter Recognition.....	24
5.1.2 Reading Comprehension	27
5.2 School Feeding Intervention	29
5.2.1 Letter Recognition.....	29
5.2.2 Reading Comprehension	31
5.3 Comparison of the Treatment Arms	33
Section 6. Limitations.....	36
Section 7. Conclusions	38
7.1 Summary of Key Outcomes.....	38
7.2 Recommendations	39
7.2.1 Quick Wins	39
7.2.2 Medium-Effort	39
7.2.3 Long-term Strategy	40
References	41
Appendices.....	42
Appendix A. LEAPS II Results Framework	A1
Appendix B. School Lists by group	B1
Appendix C. Additional Exhibits	C1
C.1 Additional Balance Checks	C1
C.2 Robustness Checks.....	C1
Appendix D. Inter-rater Reliability	D1
Appendix E. Survey Instruments	E1
Appendix F. Impact Evaluation SOW and TOR.....	F1

LIST OF EXHIBITS

Exhibit 1. LEAPS II Targeted Districts in Savannakhet Province	3
Exhibit 2. PSM-DID Impact Evaluation Design	9
Exhibit 3. Sample Distribution by District and Type of School.....	14
Exhibit 4. Students' Gender by Treatment Group	15
Exhibit 5. Students' Age by Treatment Group	16
Exhibit 6. Grade Repetition and Preschool Participation by Treatment Group.....	16
Exhibit 7. Main Language Spoken at Home by Treatment Group	17
Exhibit 8. SES by Treatment Group	17
Exhibit 9. Availability of Books at Home by Treatment Group	18
Exhibit 10. Home Literacy Practices by Treatment Group	19
Exhibit 11. Sample Distribution by Original Students	20
Exhibit 12. Reasons for Attrition by Treatment Arm	20
Exhibit 13. Balance Tests between Tracked and Replacement Students at the Follow Up.....	21
Exhibit 14. DID Effect on Literacy Outcomes	23
Exhibit 15. Impact of SF+LB on Letter Recognition.....	25
Exhibit 16. Impact of SF+LB on Letter Recognition by Gender	26
Exhibit 17. Impact of SF+LB on Letter Recognition by Main Language (Lao vs. non-Lao)	26
Exhibit 18. Impact of SF+LB on Reading with Comprehension	28
Exhibit 19. Impact of SF+LB on Reading with Comprehension by Gender	28
Exhibit 20. Impact of SF+LB on Reading with Comprehension by Main Language Spoken at Home	29
Exhibit 21. Impact of SF Intervention on Letter Recognition.....	30
Exhibit 22. Impact of SF Intervention on Letter Recognition by Gender	30
Exhibit 23. Impact of SF Intervention on Letter Recognition by Language.....	31
Exhibit 24. Impact of SF Intervention on Reading Comprehension.....	32
Exhibit 25. Impact of SF Intervention on Reading Comprehension by Gender	32
Exhibit 26. Impact of SF Intervention on Reading Comprehension by Language.....	33
Exhibit 27. Comparing Treatment Arms Letter Recognition.....	34
Exhibit 28. Comparing Treatment Arms - Reading Comprehension.....	35
Exhibit 29. LEAPS II Conceptual Framework	A1
Exhibit 30. Balance Tests between Tracked and Missing Students at Baseline.....	C1
Exhibit 31. Balance Tests between Missing at Baseline and Replacement Students at the Follow Up	C1
Exhibit 32. First Difference Estimator: SF+LB Intervention.....	C3
Exhibit 33. First Difference Estimator: SF Intervention	C4
Exhibit 34. Lagged Dependent Variable: SF+LB on Letter Recognition.....	C4
Exhibit 35. Lagged Dependent Variable: SF on Letter Recognition	C5
Exhibit 36. IRR by Literacy Skill Subtests for Impact Sample	D1

LIST OF ACRONYMS

ANOVA	Analysis of Variance
C	Comparison
CBR	Community-based Rehabilitation
CRS	Catholic Relief Services
DESB	District Education and Sports Bureau
DID	Difference-in-Difference
ECD	Early Childhood Development
EDP	Evaluation Design Plan
EGRA	Early Grade Reading Assessment
EMC	Emerging Markets Consulting
GoL	Government of Lao PDR
IE	Inclusive Education
IRB	Institutional Review Board
IRR	Inter-rater Reliability
ITT	Intention to Treat
LBRA	Literacy Boost Reading Assessment
LEAPS	Learning and Engaging All in Primary School
LB	Literacy Boost
M&E	Monitoring and Evaluation
MGD	McGovern-Dole
PESS	Provincial Education and Sports Service
PSM	Propensity Score Matching
SCI	Save the Children International
SES	Socioeconomic Status
SF	School Feeding
SILC	Saving and Internal Lending Committees
SKO	SKO, Co., Ltd.
SOW	Scope of Work
THR	Take Home Rations
TOT	Training of Teachers
UO	University of Oregon
USDA	United States Department of Agriculture
VEDC	Village Education Development Committee
WASH	Water, Sanitation, and Hygiene

EXECUTIVE SUMMARY

Learning and Engaging All in Primary School (LEAPS) II is a five-year project (2016–2021) funded by the U.S. Department of Agriculture McGovern-Dole (MGD) International Food for Education and Child Nutrition program. LEAPS II in Lao People’s Democratic Republic (PDR) is being implemented by Catholic Relief Services (CRS) together with government partners, Save the Children International (SCI), and the University of Oregon (UO). The primary objectives of LEAPS II are to improve the literacy skills of students and increase their use of improved health and dietary practices in 350 schools in seven educationally disadvantaged districts of Savannakhet province. LEAPS II programming consists of: provision of school meals; trainings for educators in inclusive education, as well as in Literacy Boost (LB) teaching methods; and improving water, sanitation, and hygiene (WASH) practices and facilities (see Section 1.2 Overview of Programmatic Implementation).

CRS selected IMPAQ International to design and conduct the performance and impact evaluations of LEAPS II. This report presents the follow-up results of the impact evaluation. In this impact evaluation, we assess whether we can attribute any observed improvements in the literacy of school-aged children to the two key LEAPS II components (LB and/or school feeding (SF) interventions), after two years of implementation on both Grade 1 and Grade 2 (first cohort of the LB) in Sepon and Atsaphone districts. The key goals of these LEAPS II interventions, particularly relevant for the impact evaluation, include:

- Improving quality of literacy instruction with an intensive model of training (i.e. providing teacher training), using SCI’s LB programming; and
- Maintaining strong attendance and attentiveness by ensuring a healthy and hunger-free learning environment for all students (i.e. providing SF).

To identify the overall impact of these LEAPS II interventions, the evaluation team studied the effects on children’s literacy outcomes by answering two questions:

- What are the program impacts on letter identification among Grade 2 students?
- What are the program impacts on reading and understanding of Grade 2 level text?

Methods

The IMPAQ team used a quasi-experimental design with a combination of Propensity Score Matching (PSM) and Difference in Differences (DID). We used the PSM technique to create a comparison group of schools that did not receive any interventions and used the DID method to study the project’s effect on students’ literacy outcomes after two years of exposure in each of the treatment groups (SF and SF+LB treatment arms).

In collaboration with our local data collection partners, we visited 159 schools (45 schools in the SF treatment arm, 59 schools in SF+LB treatment arm, and 55 Comparison schools) across the two districts of Atsaphone and Sepon. We administered the same student survey and LB Reading Assessment (LBRA) designed at baseline to assess student outcomes after two years of project exposure. For students who we were not able to track, we randomly replaced them with students present on the day of data collection who had not been sampled for the baseline study, but received the LB for two full years (i.e., students that started Grade 1 in 2017 and were in Grade 2 in 2019 in the same school). This process helped us maintain statistical power. However, we were only able to track 48 percent of students from baseline at follow up for different reasons, including grade repetitions, school drop-outs, and student absenteeism (Section 6. Limitations describes the evaluation’s limitations in more detail). To ensure that selective

attrition is not driving our results, we conducted our analyses for three sub-samples: **balanced sample** (students who we tracked from baseline to follow up); **unbalanced sample** (balanced sample plus students that we only surveyed at the baseline); and **full sample** (unbalanced sample plus students who we only surveyed at the follow up, i.e. replacements).

Key Findings

Below we provide a snapshot of the key impact evaluation outcomes for the SF and SF+LB interventions on reading comprehension and letter recognition, using data collected from the LBRA, developed and adapted to LBRA to the Lao curriculum in consultation with MoES before the start of LEAPS II. In our analysis, we control for any observed differences between each of the treatment arms and comparison group and across the three sub-samples (balanced, unbalanced, and the full sample).

Letter Recognition

- In the past two years, exposure to the SF+LB interventions increased students' proficiency in identifying letters by approximately 41 percentage points compared to the comparison group students, statistically significant at the 1 percent. However, the impact of the SF intervention alone on students' letter proficiency was minimal (3 percentage points) and not statistically meaningful.
- We also performed two sub-group analyses: girls vs. boys and Lao vs. non-Lao speakers. Impact results for the SF+LB intervention by gender are large, statistically significant, and consistent with the results for the overall sample ($p < 0.01$). Both girls and boys were more likely to identify letters by approximately 40 percentage points compared to their peers in comparison group. The variation of the SF+LB impact on boys was a bit higher than girls across different sample types, which ranged from 40 to 44 percentage points. With respect to main language spoken at home, both samples experienced positive and statistically meaningful impacts compared to Lao and non-Lao speakers in comparison schools. However, the magnitude of the effect on letter recognition was larger for non-Lao speakers (approximately a 47 percentage-point increase) than Lao speakers (approximately a 32 percentage-point increase) compared to comparison group students, ($p < 0.01$) in SF+LB schools. The sub-group analyses in SF also did not show any meaningful impact on letter proficiency.

Reading Comprehension

- We did not find any evidence that either of the treatment arms (SF and SF+LB) had any statistically meaningful impact on reading comprehension of Grade 2 students. The magnitude of the effect was also trivial for SF (approximately a 4 percentage-point increase) and SF+LB (a 2 percentage-point increase) on reading comprehension compared to comparison group students. Results for all three samples (balanced, unbalanced, and the full sample) were consistent across the SF and SF+LB interventions.
- While we did not find a statistically meaningful impact of the program on reading comprehension of either girls or boys, the magnitude of the coefficient was higher for girls than boys across SF and SF+LB. The effects of SF and SF+LB were higher for non-Lao speakers compared to Lao speakers, however, the impact was not statistically significant for either of the language groups.

Comparing the Treatment Arms (SF vs. SF+LB)

- We also determined the incremental impact of LB by comparing the SF and SF+LB with each other. We noted a large and statistically significant impact of LB on letter recognition, but the results show no significant additional impact of the LB intervention on reading comprehension. The LB

intervention has led to an additional impact of 39 to 42 percentage points across various samples ($p < 0.01$). Since the overall impact of SF+LB on letter recognition was approximately 41 percentage points, this implies that LB drives most of the impact of the SF+LB intervention on letter recognition.

Recommendations

We present the following recommendations to CRS based on lessons learned in the field and findings after analyzing the collected data. We divide this section into three parts:

- **Quick wins** are cost-effective recommendations for relatively simple problems, which can be implemented immediately and yield maximum impact.
- **Medium-effort** items require a modified evaluation budget and SOW, but are not necessarily as time or cost intensive as long-term strategy items. Medium-effort recommendations are important to consider for bridge or future projects.
- **Long-term strategy** recommendations may require more time, effort, and cost to implement, and therefore would be most useful to consider for future studies beyond LEAPS II.

Quick Wins

- **Reinforce the importance of education for all, especially for students from different ethnicities, in LB teacher training modules.** As mentioned in Section 1.1 Overview of Lao Context., gender disparity and ethnic group disadvantages are known education issues at the national level in Lao PDR. Responses from participants in the midterm performance evaluation's qualitative interviews indicated that students who do not speak Lao as the primary language at home, which is the official language of instruction, may feel discouraged to continue their schooling, which results in dropouts. However, in the impact evaluation, the subgroup analysis by language shows a statistically significant higher effect of SF+LB on letter recognition for non-Lao speakers than for Lao speakers. Even for reading comprehension, the non-Lao sample witnessed a marginally significant impact. Thus updating the LB training materials to emphasize training school staff and the community to make schools a more welcoming environment for *all* children (i.e. the objective of inclusive education component), but with the focus on marginalized students, non-Lao speakers, and girls, could result in even greater impact for those who usually fall behind.

Medium-Effort

- **Unpack the main driving force behind the impact of the program on letter recognition, and study the mechanisms responsible for no effects of the program on reading comprehension.** CRS could consider a study to better understand the mechanisms that hinders improvements in reading comprehension by speaking to teachers or teacher trainers and incorporate the results in implementation to the extent possible. In addition, since the impact evaluation estimates suggest the SF+LB interventions improved letter recognition but not reading comprehension, LEAPS II could consider unpacking specific LB activities to understand which LB activities have led to these changes. IMPAQ could study the effect of the various LB activities on additional outcome variables, such as teacher trainings, reading camps, and/or school supplies. IMPAQ could collect additional quantitative information from students and teachers on these outcomes. Further, this kind of study would help future programming by pointing out which aspects of LB activities are least impactful, and thus should be modified for future iterations.
- **Consider alternative outcome indicators to study the effect of SF on students' schooling.** School meals did not seem to affect letter recognition and reading comprehension directly. Further, all impacts of the SF+LB intervention on letter recognition came from LB activities. These results

reflect prevailing literature that indicate direct linkages between school meals and cognitive outcomes, including literacy gains, are rather complex; rather, SF may have more indirect effects on student cognition through increased school participation (attendance, one of the main outputs of LEAPS II according to the Theory of Change) and improved nutrition intake.¹ As well, SF programs are more likely to improve cognition in the presence of complementary inputs, such as teacher ability, instructional materials, and school infrastructure.² Qualitative findings and CRS' M&E data confirm that LEAPS II has progressed quite well on some of these mediating factors in the causal chain between SF and literacy, such as higher attendance. IMPAQ could conduct further investigation on the causal effect of SF on some of these intermediary outcomes to better understand improvements on the extensive margin as a result of school meals.

Long-term Strategy

- **Modifying the longitudinal nature of the study.** The current impact evaluation was originally designed as a longitudinal quasi-experimental study (following the same students over time). However, studies have shown that longitudinal studies suffer from high attrition. In this study too, we could only track 48 percent of students from baseline to follow-up. Consequently, we recommend a repeated cross-sectional design in future studies (sampling a fresh group of students at each point in time). This will have two main advantages: (1) the study will not suffer from high attrition because a new group of students will be sampled each time and (2) the study will be less expensive since tracking the same student can be quite costly.

¹ Buttenheim, A., Alderman, H., & Friedman, J. 2011. Impact evaluation of school feeding programmes in Lao People's Democratic Republic. *Journal of Development Effectiveness*. 3 (4), 520-542. Adelman, S., et al., 2008. The impact of alternative food for education programs on learning achievement and cognitive development in northern Uganda. Washington, DC: The World Bank.

² Chakraborty, T. & Jayaraman, R. 2016. School Feeding and Learning Achievement: Evidence from India's Midday Meal Program. IZA Discussion Papers, No. 10086, Institute for the Study of Labor (IZA), Bonn, Germany

SECTION 1. INTRODUCTION

The United States Department of Agriculture (USDA) has funded Catholic Relief Services (CRS) for two project phases of Learning and Engaging All in Primary School (LEAPS) Project as part of the McGovern-Dole (MGD) International Food for Education and Child Nutrition Program. The second phase, a five-year intervention (FY2016–FY2021) with a budget of \$27 million, extends and expands the previously implemented LEAPS I project. LEAPS II aims to improve the literacy skills and health and dietary practices of approximately 86,049 new and continuing students in 350 schools across seven educationally disadvantaged districts of Savannakhet province in Lao PDR, including Atsaphone, Nong, Outhoumphone, Phalanxai, Phin, Sepon, and Vilabouly.

This report assesses whether any observed improvements in the literacy of school-aged children can be attributed to the project’s Literacy Boost (LB) and/or School Feeding (SF) interventions after two years of LEAPS II implementation in Atsaphone and Sepon.

1.1 Overview of Lao Context

In recent years, Lao PDR has been on track to meet most of its Millennium Development Goals (MDGs) for education, and specifically, achieving universal education.³ Government enforcement of new education policies, along with support from international organizations, has helped build new schools, recruit and train new teachers, and improve accessibility to all levels of schools. Thus, across 24 years (1990-2014), the primary school net enrollment rate has increased from 65 to 98 percent.⁴ In addition, the attendance rate in primary schools has increased from 59% for girls and 64% for boys in 2000 to 72% and 74.2% in 2017.⁵

Although Lao PDR has expanded access to education, the country still faces many challenges. Despite a high net enrollment rate, a review of recent literature on Lao PDR’s primary education system revealed that it still suffers from high dropout rates and grade repetition. Early grade dropouts remained above 10 percent for students in Grade 1, slightly lower at around 5 percent for Grades 2 through 4, but overall Grade 5 completion rates also remain low at 70 percent.⁶ Repetition rates are greatest for Grade 1 primary school students, varying from 22 percent to 35 percent according to different estimates.⁷ The complexity of the Lao language, particularly for non-Lao speaking ethnic children, could account for some of these issues, which aligns with responses from participants in our qualitative interviews indicating that students who do not speak Lao as the primary language at home may feel discouraged to continue their schooling.

Recognizing the challenge of limited academic achievement in the early years of primary school and acknowledging that many students end up repeating grades, the Ministry of Education and Sports (MoES)

³ World Bank. 2016. Reducing Early Grade Dropout and Low Learning Achievement in Lao PDR: Root Causes and Possible Interventions. Washington, D.C.: World Bank Group

⁴ World Bank. 2016. Reducing Early Grade Dropout and Low Learning Achievement in Lao PDR: Root Causes and Possible Interventions. Washington, D.C.: World Bank Group

⁵ Laos Social Indicator Survey II. 2017. Accessed here: <https://dhsprogram.com/pubs/pdf/FR356/FR356.pdf>

⁶ World Bank. 2016. Reducing Early Grade Dropout and Low Learning Achievement in Lao PDR: Root Causes and Possible Interventions. Washington, D.C.: World Bank Group

⁷ Onphanhdala, P., and Suruga, T. 2008. Obstacles to Basic Education Development in Rural Lao PDR: Delayed Enrollment, Grade Repetition, and Dropout. Graduate School of International Cooperation Studies, Kobe University; Ministry of Education and Sports. 2015. Situational Analysis: Student Learning Outcomes in Primary Education in Lao PDR

introduced a “Progressive Promotion” policy in 2009, which moves students automatically from one grade to the next without consideration of performance. However, this policy has not necessarily reduced dropout rates or grade repetition because of limited enforcement, and the fact that it does not target some of the underlying causes of high primary school dropout rates/grade repetition.⁸ Based on the 2012 National Assessment of Student Learning Outcomes, 17 percent of students in Grade 3 have reached pre-functional levels of literacy and only 23 percent of students have attained sufficient proficiency in the Lao language to continue learning at the next level of schooling without remedial assistance.⁹

Among the numerous factors that could account for high dropout rates/grade repetition, Lao PDR still faces high absenteeism rates of students, which may reflect a mismatch between the academic school year and the agricultural cycle.¹⁰ For example, in rural communities that are often highly agricultural, households locate their land plots far from town centers and schools. In these communities, parents typically take their children with them to live at the field during specific periods, such as during planting or harvest.¹¹ With children absent for months at time in the academic year, their performance in the classroom suffers, which could account for the low levels of literacy in primary school students.¹² As discussed in the literature, irregular attendance at school can have detrimental effects on learning.¹³

LEAPS II has responded to some of these obstacles underlying the low attendance rate and literacy level in Lao PDR via its SF and SF+LB activities. In particular, SF responds to the prevailing literature that school meals can enhance cognitive outcomes via increased attendance while SF+LB underscores the evidence that SF programs are more likely to improve cognition in the presence of complementary inputs available through the school system.¹⁴ For example, in this context, the LB intervention addresses some of the more common factors cited in the literature as mediators in the causal link between SF and improved education gains. These factors include improving pedagogies in the classroom by the LB teacher training and enhancing access to instructional materials through reading corners (See Section 1.2 Overview of Programmatic Implementation for detailed description of the LEAPS II activities). However, it should be noted that a myriad of factors likely contribute to these education challenges in Lao PDR, and addressing certain problems falls outside the scope of LEAPS II.¹⁵

In this context, this evaluation studies the effect of SF and LB interventions of LEAPS II, together and separately, on students’ literacy outcomes (letter recognition and reading comprehension). The introduction of the LB approach and other literacy activities as part of LEAPS II will offer evidence on the effectiveness of the school meals and literacy interventions and help enhance sustainability of such programs in Lao PDR. IMPAQ anticipates that the findings from this evaluation could be useful to not only USDA, but a variety of external stakeholders, including the government of Lao PDR, other MGD implementers, and US policymakers.

⁸ Itthida, Ngnangnouvong. 2015. “Progressive Promotion”: Dropout Issue in Lao Primary Education and Misplaced Policy. *Procedia – Social and Behavioral Sciences*. Vol 197. P.2355-2361.

⁹ Ibid.

¹⁰ Ministry of Education and Sports. 2013. *Education Sector Development Plan (2011-2015), Review and Update: Final Report*.

¹¹ Ibid.

¹² UNICEF. 2015. *Situational Analysis: Student learning Outcomes in Primary Education in Lao PDR*. Accessed here: <https://www.unicef.org/laos/media/331/file/UNICEF%20Lao%20PDR.pdf>

¹³ Ibid.

¹⁴ Chakraborty, T. & Jayaraman, R. 2016. *School Feeding and Learning Achievement: Evidence from India's Midday Meal Program*. IZA Discussion Papers, No. 10086, Institute for the Study of Labor (IZA), Bonn, Germany

¹⁵ Cerdan-Infantes, P., Marshall, J., and Naka, E. 2016. *Reducing Early Grade Dropout and Low Achievement in Lao PDR: Root Causes Research and Possible Interventions*.

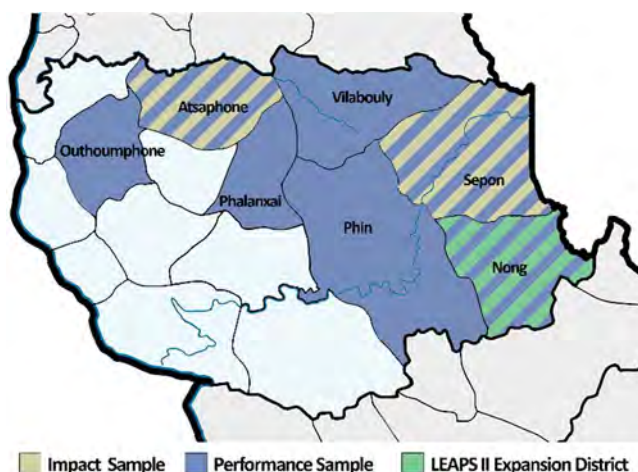
1.2 Overview of Programmatic Implementation

In response to low school attendance and performance in six educationally disadvantaged areas of Savannakhet province in Lao PDR, CRS implemented LEAPS I, in collaboration with local partners and the Ministry of Education and Sports (MoES), from 2012 to 2016. Supported by USDA, LEAPS I reached more than 36,000 students in 310 communities. It improved attendance and enhanced students' reading skills through approaches that empowered government, parents, teachers, and community leaders.¹⁶

To build on previous success and expand on LEAPS I, in September 2016, USDA awarded CRS \$27 million to implement the second phase of the LEAPS program in seven districts of Savannakhet province.¹⁷ CRS is leading the implementation of LEAPS II in partnership with the government, Save the Children International (SCI), and University of Oregon (UO). LEAPS II aims to improve the literacy skills, as well as health and dietary practices, of approximately 86,049 new and continuing students in 350 schools across the six districts of LEAPS I, and an additional district, Nong. Exhibit 1 provides a map of project areas in both phases. Although LEAPS II retains some of the same elements as the previous phase, such as school meals, take-home rations (THRs), and inclusive education (IE), LEAPS II has incorporated lessons learned and identified needs from LEAPS I to add several new activities.

According to the LEAPS II theory of change (see Appendix B. School Lists by group), providing school meals can both enhance attendance by keeping students in school for the full day and increase students' attentiveness by mitigating short-term hunger. These two factors, in complement with the Literacy component, should improve students' literacy skills, in keeping with MGD Objective 1 (the focus of this evaluation). In addition, the LEAPS II theory of change suggests that providing food preparers at school with training on safe food preparation and storage practices; improving teachers' knowledge of health and hygiene practices to transfer their knowledge to students; and building/rehabilitating wells and water stations at schools will help to improve health and dietary practices (MGD Objective 2). To address these objectives, LEAPS II implements four main interventions: 1) school meals; 2) literacy related activities; 3) inclusive education; and 4) Water, Sanitation, and Hygiene (WASH). Since the focus of this evaluation is on MGD Objective 1 (improvement in literacy outcomes), we elaborate on the first three activities. Please see midterm performance evaluation for a full detailed description of all the four interventions.

Exhibit 1. LEAPS II Targeted Districts in Savannakhet Province



¹⁶ Learning and Engaging All in Primary School (LEAPS). 2016. "Final Evaluation Report."

¹⁷ LEAPS II has expanded its project area to Nong district, except for LB intervention.

1.2.1 Provision of School Meals

Through LEAPS I, CRS provided nearly 12 million meals to more than 36,000 children. This support was instrumental in supporting students' attendance and attentiveness improvement in the classroom.¹⁸ LEAPS II expands upon this intervention and will provide throughout the duration of the intervention more than 32 million daily meals, with a daily ration of 100 grams of rice, 35 grams of lentils, and 10 grams of vegetable oil per student.¹⁹ Communities contribute to the school meals either through cash contributions so that schools can purchase additional ingredients, or directly via the provision of food items, such as eggs and condiments. Village Education Development Committees (VEDCs) mobilize this community support, and CRS provides training to VEDC members on how to record and track such contributions as part of its modules on VEDC roles and responsibilities. Cooks, storekeepers, and teachers also receive a monthly take-home ration (THR), but to move toward sustainability of operations, CRS intends to phase down the rice in the bundle of THRs for those communities with high-performing VEDCs, based on the results from the annual VEDC rating tool. Some other activities that take place as part of the school meals component of LEAPS II include:

- Rehabilitation and maintenance of storehouses constructed under LEAPS I, and provision of materials (e.g. roofing sheets, nails, and cement) to support the community to build new warehouses in Nong
- Establishment of school gardens and subsequent training of teachers and principals to use these gardens as a tool for working with children and to improve sustainability of school meals
- Refresher trainings for cooks and storekeepers in safe food preparation and hygiene practices

1.2.2 Literacy Interventions

CRS, SCI, and UO work jointly on this component of LEAPS II to improve literacy outcomes. SCI leads the implementation of their signature Literacy Boost (LB) intervention, which aims broadly to improve the reading skills of children in the first five years of primary school. The LB intervention contains the following components:

- *The LB reading assessment (LBRA).* The LB practitioners²⁰ assess children's reading levels and evaluate their literacy learning needs based on those assessment
- *Training.* The LB training centers on five core skills of reading acquisition: 1) automaticity in letter recognition, or letter knowledge; 2) phonemic awareness; 3) fluency (reasonable reading speed) and accuracy; 4) vocabulary development; and 5) reading comprehension, which is the ultimate goal of language learning, including literal, inferential, and evaluative. In addition to these five skills, a key component of the intervention is training teachers on important instructional techniques paired with other essential factors, such as children's motivation, access to text, home environment, and reading opportunities. SCI and CRS work collaboratively to train MoES and school administrators, and then co-facilitate with the government the rolldown of such trainings to the teachers, which encourages sustainability.
- *Provision of reading materials.* Room to Read as part of LEAPS I established school libraries, and LEAPS II enhances the availability of reading materials by setting up reading corners in Grade 1, 2, and 3 classrooms. To facilitate the summer LB activities, community literacy volunteers (CLVs)

¹⁸ Catholic Relief Services. 2016. Learning and Engaging All in Primary School (LEAPS): Final Evaluation Report.

¹⁹ Catholic Relief Services. 2018. LEAPS II FAQs.

²⁰ SCI staff trained PAs

take ownership of the reading corners and transfer the books from the school to the community. At this time, books may be housed in the home of the village head or CLV, or in a public space.

- *Community action.* Quality teaching and learning environments inside and outside of schools help all readers grow their reading skills. The LB program guides parents and communities to support children as they learn to read and foster their love of reading, including encouraging reading at home and holding school open houses. SCI conducts parent engagement workshops/trainings to encourage parents to create home literacy environments. As well, CRS and SCI staff, after receiving training from UO on the summer reading camps, work with CLVs to implement these activities that aim to further develop the literacy skills of children. These summer reading camps have now been extended to the entirety of the academic year in LB schools.

The LB intervention takes place in half of the schools receiving the SF intervention, with three cohorts of 60 schools each, for a total of 180 schools in 6 districts of LEAPS II.²¹ Each cohort will receive project support for Grades 1 and 2. The impact evaluation targets the two districts (Sepon and Atsaphone) that were included in the first cohort of LB schools. For this cohort, LB activities began in Year 2 of the project for incoming Grade 1 students in September/October 2017.

Complementing the LB approach, CRS conducted a special study to implement summer school readiness camps, and SCI led another to research effective teaching methods for minority language speakers. The summer school readiness camps provided literacy instruction to children outside of the school year to help them reach grade-level literacy in the upcoming year. CRS holds week-long trainings for CLVs to help run these camps. Finally, CRS trained teachers to engage in ‘learning circles’—meetings where teachers can reflect on the successes of and challenges encountered in using LB techniques—to foster a culture of continual learning and adaptation.

1.2.3 Inclusive Education (IE)

Working in partnership with UO since LEAPS I, CRS ensures that its IE approach falls in line with MoES IE policies to ensure all children have access to school regardless of disability, gender, or any other circumstance. In LEAPS I, UO collaborated with MoES to develop an IE curriculum for MoES to implement through the Teachers’ Training College, which would train all new teachers in IE practices to match the new IE laws adopted by the Government of Lao PDR (GoL) in 2011. In the design of LEAPS II, CRS and UO jointly developed an online curriculum and TOT model to engage a diverse group of stakeholders in IE practices. The following IE trainings were conducted during LEAPS II:

- Training school administrators to understand IE policy, IE management, IE concepts of differentiated instruction, and how to create a welcoming learning environment, and then supporting them in the roll down of this IE training to teachers, which includes coaching teachers on working with students who have different disabilities and modifying teaching practices and learning resources to accommodate these students
- Training PESS and DESB on community-based rehabilitation (CBR), which aims to improve awareness of disabilities in the community, and then working with the government to flow down these lessons to the VEDCs
- Supporting VEDCs in their community mapping exercises to ensure that school management focuses on students with disability
- Training DESB how to enroll disabled children into school

²¹ Schools in Nong (the LEAPS II expansion district) were not targeted for the LB intervention.

- UO online IE trainings on classroom management techniques for inclusion of disabled children given to high-performing teachers

1.3 Evaluation Background

CRS selected IMPAQ to design and conduct both the performance and impact evaluations of LEAPS II. The performance and impact evaluations followed slightly different timelines during baseline (baseline performance in March 2017 and baseline impact in September 2017). However, the impact follow-up converged with the midterm performance evaluation and thus we collected data for both evaluations at the same time in March 2019. This report presents the impact evaluation, which measures the causal effect of the LEAPS II LB and SF interventions, together and separately, on the literacy of school-aged children after two years of exposure to the program from 2017 to 2019. The impact evaluation covers Atsaphone and Sepon districts where LEAPS II initiated its first cohort of LB activities during the 2017-18 school year. Specifically, this follow-up impact evaluation seeks to:

- Assess the impact of the LB and SF interventions, together and separately, on literacy outcomes;
- Discuss any necessary mid-course corrections to effectively and efficiently meet the project's stated goals and objectives; and
- Provide recommendations aimed at improving the quality of programming and/or resource management.

To address these objectives, we used a similar sampling strategy and survey instrument (student survey and LBRA) as baseline. We visited the same 159 schools (45 schools receiving SF only, 59 schools receiving SF and LB, and 55 schools that have not, and will not, receive either of the interventions in LEAPS II).²² Schools in treatment groups (SF and SF+LB) were all part of LEAPS I program and received the SF during the first phase, however, comparison schools were not even part of LEAPS I. While we conducted the baseline evaluation at the beginning of the 2017-2018 school year (September 2017), we collected follow-up data from the same respondents, as available, at the end of the 2018-2019 school year (March 2019), resulting in a panel with two separate temporal data points. We then compared literacy outcomes for students in treatment schools and the comparison schools from baseline to midline.

Section 2. Evaluation Approach describes our evaluation approach in detail, including research questions, sampling design, and data sources and analysis. Afterward, Section 3. Fieldwork describes the data collection fieldwork, including human subject protection procedure. We then move to describe the key characteristics of the sampled students (Section 4. Evaluation Sample). Finally, this report concludes with a discussion of limitations in Section 6. Limitations and then the key outcomes and recommendations in Section 7. Conclusions.

²² Comparison schools did

SECTION 2. EVALUATION APPROACH

This section describes the quantitative methodology implemented for the LEAPS II impact evaluation after two years of LB and SF implementation. The evaluation methodology follows a longitudinal quasi-experimental evaluation design, which we detail below in 2.2 Evaluation Methodology.

We also provide relevant findings from qualitative data collection and analysis given the overlap between areas selected for performance and impact evaluation in Sepon and Atsaphone. Although, per the agreed upon SOW, we designed our qualitative component specifically to enhance the performance evaluation, the key informant interviews (KIIs) and focus group discussions (FGDs) conducted in these two districts provide interesting and useful information that could be helpful in understanding the impact analysis.

2.1 Impact Evaluation Research Questions

The impact evaluation measures the causal effect of the LB and SF interventions, together and separately, on primary school-aged children in Atsaphone and Sepon districts of Savannakhet province, to answer two key research questions:

- What are the program impacts on letter recognition among Grade 2 students?
- What are the program impacts on reading and understanding of Grade 2-level text?

2.2 Evaluation Methodology

To address the above-mentioned research questions, we used a quasi-experimental design combining Propensity Score Matching (PSM) and Difference in Differences (DID) methods. In this design, we followed three key steps from baseline to follow up to assess the causal effect of the SF and SF+LB interventions on students' literacy skills.

Step 1. In February 2017, prior to data collection for the impact evaluation,²³ we grouped all schools that previously received SF support during LEAPS I into clusters of schools (on average four schools in each cluster) based on their proximity, access to roads, and access to electricity.²⁴ We then assigned randomly the clusters of schools to the two treatment arms:²⁵ one to receive SF during LEAPS II (SF only); the other, to receive the LB, in addition to SF (SF+LB).

This random assignment allowed us to calculate the incremental two-year effect of the LB intervention. Specifically, comparing the literacy outcomes at the student level provided the following unbiased estimate of the incremental two-year effect of the LB intervention:

$$\Delta_{LB} = E [(Y_1^{SF+LB} - Y_1^{SF})],$$

where Y_1 denotes outcomes at follow up, and $SF+LB$ and SF denote the two within-district treatments, respectively.

²³ We finished this step before the baseline data collection for performance evaluation (February 2017) and prior to impact baseline evaluation (September 2017).

²⁴ We completed this step during the preparation for the baseline data collection of the performance evaluation in the spring of 2017.

²⁵ The term “treatment-arm” in the rest of this document refers to the treatment schools only, and not to the comparison schools.

Step 2. In August 2017, we received a list of potential comparison schools in Atsaphone and Sepon from CRS that were not on the list of LEAPS II target schools and these schools were not slated to receive any of the LEAPS II program activities. We used PSM to create a list of comparison schools that look as similar as possible to the treatment schools based on observable characteristics. To implement PSM, the evaluation team followed three steps:

- IMPAQ received from CRS a list of 79 potential comparison schools in 20 clusters (averaging about 4 schools per cluster, as noted) in the Atsaphone and Sepon districts (34 schools in 9 clusters in Atsaphone, and 45 schools in 11 clusters Sepon).
- The IMPAQ team generated cluster-level characteristics for the comparison clusters.
- The IMPAQ team estimated a propensity score based on the cluster-level characteristics²⁶, and identified 13 comparison clusters (52 schools) that were nearest matches to the treatment clusters.

Based on our calculations in Steps 1 and 2, we sampled 159 schools with 45 SF schools in 12 clusters, 59 SF+LB schools in 14 clusters, and 55 comparison schools in 19 clusters. We visited the same schools at baseline (September 2017) and at follow up (March 2019).²⁷

Step 3. After completing follow-up data collection in March and April 2019, we used DID to compare the changes in literacy outcomes from baseline to follow up between each of the treatment arms and comparison group:

$$\Delta_{SF} = [(Y_1^{SF} - Y_0^{SF}) - (Y_1^C - Y_0^C)]$$

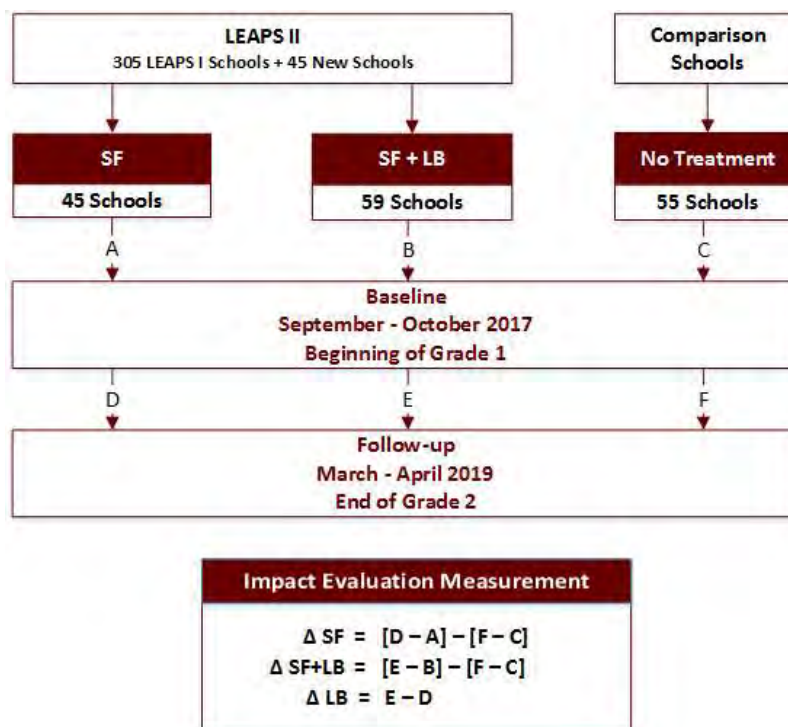
$$\Delta_{SF+LB} = [(Y_1^{SF+LB} - Y_0^{SF+LB}) - (Y_1^C - Y_0^C)],$$

where Y_0 and Y_1 denote outcomes at baseline (i.e. beginning of Grade 1 in September – October 2017) and follow up (i.e. end of Grade 2 in March 2019), respectively, and SF , $SF+LB$, and C denote the respective treatment arms and the comparison group. At the student level, this is an intent-to-treat (ITT) estimate, as it does not account for whether individual students received SF or took a literacy assessment, but only whether they attended a participating school. Exhibit 2 illustrates the overall evaluation design approach.

²⁶ IMPAQ used the following cluster level characteristics for PSM that were available in the schools lists provided by CRS: total students, total teachers, teacher-student ratio, average number of students per school, average number of teachers per school, percent Lao ethnicity, total schools, percent of complete schools (i.e. schools that contain all Grades 1 – 5). IMPAQ did not include gender weights as one of the selection criteria in sampling schools because the school lists did not have that information.

²⁷ We originally planned to visit 55 schools in 13 clusters. However, because of some logistical challenges in the field during baseline, we ended up visiting 55 schools in 19 clusters to replace the inaccessible schools. The additional 6 clusters were not selected based on matching techniques (i.e. there is no propensity score associated to them) but they were selected based on other criteria like accessibility and size of schools, in order to maximize overall sample size. See baseline impact report 2017 for more detail.

Exhibit 2. PSM-DID Impact Evaluation Design



2.3 Sampling Design

As described above, we selected 159 schools for inclusion in the DID sample. For the initial power calculations conducted at baseline in 2017 considering a two-year incremental effect of LB, IMPAQ estimated that a sample size of 1,456 students can detect an effect size of 0.39 standard deviations in literacy with a power of 80 percent.²⁸ For the two-year net effect of the LB intervention using a comparison group, the initial power analysis confirmed the sample size of 2,912 students would be sufficient to detect an effect size of 0.32 standard deviations in literacy for a power of 80 percent.^{29,30}

²⁸ IMPAQ used 27 treatment clusters of an average of 4 schools and estimated a sample size of 14 students per school. IMPAQ assumed variation between clusters of 0.05 and within clusters of 0.24. We also had standard assumptions of 80 percent power and 0.05 significance level. The minimum detectable effects were determined using Optimal Design software.

²⁹ At baseline, IMPAQ assumed a total of 40 clusters (27 treatment and 13 comparison) with an average of 4 schools per cluster and 18 students per school (a rounded average between 14 students in treatment schools and 28 students in comparison schools). IMPAQ assumed variation between clusters of 0.05 and within clusters of 0.24 and was assumed no correlation over time. IMPAQ also had standard assumptions of 80% power and 0.05 significance level. The minimum detectable effects were determined using Optimal Design software.

³⁰ IMPAQ followed the power calculations for sampling provided by CRS through the TOR. We confirmed their validity through our recalculations. In the absence of electronic student lists, IMPAQ was not able to take into account any gender weights in sampling students.

During baseline data collection in 2017, we selected a sample of girls and boys at random from the first-grade class to participate in a baseline survey.³¹ We sampled a total of 840 first-grade students (372 girls and 468 boys) from the 55 comparison schools, 743 first-grade students (354 girls and 389 boys) from the 59 SF+LB schools, and 601 first-grade students (264 girls and 337 boys) from the 45 SF schools, for a total sample size of 2,184 students (i.e. about 14 students per school, on average).

For the impact evaluation follow up in 2019, we visited the same schools to survey the selected students, who were in first grade at baseline (September 2017) and in second grade at follow up (March and April 2019). However, with baseline students repeating grade one, dropping out of school, being absent, moving from the area, and/or other reasons, we only tracked 48 percent of students from baseline to follow up. To maintain power, for those students that we could not track, we replaced them with available students who we did not sample at baseline. We followed the sampling strategy used at baseline to replace students. Although, we could not reach the total target of 2,184 from baseline, our replacement strategy allowed us to survey 1,657 students across all treatment and comparison groups. According to our power calculations shared with CRS in August 2017 (described at the beginning of this subsection), this sample of 1,657 students sampled at follow up (2019) still resulted in sufficient power (80 percent) to detect an effect size of 0.39 standard deviations on literacy outcomes.³² Section 4. Evaluation Sample provides details on the attrition rate and the number of students surveyed and sampled in each school.

2.4 Data Sources

To measure the impact of the two interventions on children’s literacy outcomes and answer the research questions, the IMPAQ team collected data using the same student survey and LBRA as the baseline impact evaluation, which is a short version of the survey used for the performance evaluations.³³ We translated and adapted the student survey to the Lao context through cognitive interviews³⁴ at baseline for the performance evaluation (March 2017), and we did not change them for the impact evaluation at baseline or follow up.³⁵

To measure the reading skills of students at the end of Grade 2, enumerators also administered the LBRA as part of the student survey. SCI developed and adapted the LBRA to the Lao curriculum in consultation with MoES before the start of LEAPS II. The GoL has defined Grade 2 reading competencies³⁶ as follows:

- Show knowledge of the objectives and content of the text
- Read long text such as letters, notification, reports and explanation
- Use information from the text in different context and objectives

³¹ To identify the n^{th} student for random selection, the enumerators physically lined up students in class and used a simple rule. In the sampling rule, they divided the total number of girls or boys present in class by total number of targeted girls or boys to be surveyed. We planned to select seven girls and seven boys in each of the treatment arm schools and 14 girls and 14 boys in each of the comparison schools.

³² Power calculations (shared with CRS in August 2017) were done taking reading comprehension as the main indicator.

³³ The student survey used for the impact evaluation was a shorter version of the performance evaluation, by excluding the key performance indicators.

³⁴ A cognitive interview is an individual, face-to-face, in depth interview that aims to understand how respondents comprehend and respond to questions.

³⁵ For the impact evaluation, we added only a couple of questions at follow up to capture when they started in their school and which grades they repeated.

³⁶ However, given the complexity of the Lao language, the government is developing new competencies to match the new curricula and match literacy expectations to the learning realities in the classroom.

- Identify different structure and characteristics of the text such as: heading, topic, paragraphs, section and punctuations

Considering these benchmarks for reading competencies, SCI, in collaboration with CRS and IMPAQ, field tested the LBRA before performance baseline data collection in February 2017. During this process, IMPAQ also conducted a “passage equating” exercise with four passages from story books considered appropriate for Grade 2 level³⁷ in Lao. IMPAQ administered this field test to strong readers in Grade 3 — the best benchmark for the reading proficiency of students at the end of Grade 2 — in two schools with Lao and non-Lao skills. After analyzing the test data, the evaluation team collaborated with SCI to select the most appropriate passage to assess second-grade students’ literacy. IMPAQ repeated this process before collecting data at baseline for the impact (September 2017) and at midterm for the performance and impact (March 2019) evaluations.

Using the same instruments and measuring the key outcomes of interest in the same way as the baseline impact evaluation enabled us to capture relevant changes for the same respondents over time. Using the same instrument between impact and performance also helped maximize comparability in literacy outcomes and findings between the two evaluations.

2.5 Data Analysis

After transferring the downloaded data from our programming server, Kobo, to a secure platform, we started our data analysis with an exhaustive assessment of quality checks. We then proceeded with data management to check for data completeness, duplicate entries, skip pattern logic, and data cleaning.

Descriptive Analysis. Given our high attrition rate (52 percent), we explored the data more in depth to analyze the cleaned data descriptively and compared mean outcomes between different groups by running t-tests and using p-values. More specifically, we looked at the differences in means between tracked students and replacements at the follow up; missing students from baseline and replacements; as well as between each of the treatment arms and comparison group at midterm for the entire sample, highlighting any emerging patterns. These statistics helped the evaluation team assess the similarity of average characteristics and outcomes considering attrition. We included any observable differences that our regression analysis needed to control when performing the final DID analysis. Section 4. Evaluation Sample presents the descriptive summary statistics for background characteristics of the students in the sample.

DID Analysis. For the final impact evaluation IMPAQ estimated the program effects using regression analysis. To estimate the two-year effect of the SF only and the combined SF+LB, we used a DID specification. In addition, we conducted sub-group analysis by key demographic characteristics, such as gender and main language spoken at home (i.e. Lao speakers vs. non-Lao speakers). We describe the DID specification and the sub-group analysis in further detail in Section 5. Results .

Regression Analysis for LB effect. Additionally, because the two interventions SF and SF+LB are independent interventions, up to now we have estimated their treatment effects separately (for SF and for SF+LB). However, to test whether the two treatment arms significantly differ we estimated them in the same DID specification as outlined in Section 5.3 Comparison of the Treatment Arms in more detail.

³⁷ The SCI LB coordinator and advisor in 2015 conducted this process for another evaluation of LB in Bolikhamxay in Lao PDR.

SECTION 3. FIELDWORK

This section describes the activities that the evaluation team conducted before, during, and after data collection, from human subject protection to field preparation, data collection, and quality assurance. It also discusses some of the challenges that our team faced in the field when collecting data, and how we addressed these unexpected obstacles.

3.1 Human Subjects Protection

Prior to collecting data for the baseline evaluation, IMPAQ sought Institutional Review Board (IRB) approval to ensure that the proposed evaluation complied with international rules and procedures. IMPAQ submitted protocol documents to Chesapeake IRB, now merged with Schulman IRB as Advarra, on February 28, 2017 (Pro00020840). The IRB submission (at baseline) ensured that the research is ethically sound and safeguards the rights, safety, and well-being of children and other respondents. We used the IRB-approved evaluation instruments for collecting data at midterm, including student survey with LBRA, as well as informed consent and assent forms. The qualitative data presented in this report is based on the performance evaluation findings; as such, please refer to that report for more details on the design and approval of the protocols to ensure compliance with ethical research guidelines and human subjects protection.

Before administering the survey, we trained enumerators on procedures for interviewing respondents, protecting respondents' privacy and confidentiality, and securing the data. We also reprised the training from SCI on safeguarding children at school. During the data collection, the survey team first obtained teachers' and/or principals' written consent to survey their students. The team then asked for students' verbal assent, assuring children that their participation was voluntary and that they could terminate the survey at any point. After data collection, the evaluation team protected the privacy and confidentiality of respondents by storing the data on secure servers and separating personally identifiable information from the survey data.

3.2 Preparation for Data Collection

The data collection for impact follow up converged with midterm performance data collection at the end of academic school year 2018-2019 in March 2019. From February 27 to March 1, 2019, we trained 18 enumerators, recruited by our new data collection partner SKO, for a total of four days. All new contracted enumerators had extensive experience in collecting data from young children and conducting reading assessments in the same project areas.

The training consisted of two days of theoretical training with the 18 enumerators practicing the student survey and reading assessment. Enumerators learned 1) the intention of each survey question; 2) how to ask those questions to vulnerable respondents; 3) how to measure students' literacy; and 4) how to use tablets to implement the in-person surveys offline without an Internet connection. Similar to enumerator's training at baseline, SCI also participated in training activities conducted by IMPAQ to support clarifying the correct pronunciation for each of the LBRA subtests in the rural areas.

After the initial training, we provided an opportunity for enumerators to practice with real respondents for one day in two non-LEAPS schools in Phin district whose student populations spoke Lao and minority languages. This pilot mimicked real data collection to give enumerators hands-on practice. Afterward, enumerators regrouped with the IMPAQ team to debrief and discuss any issues they encountered.

For information on the preparation undertaken for our qualitative fieldwork, including our approach to sampling and selecting the seven schools as part of the performance evaluation, please refer to that report.

3.3 Field Activities

As mentioned above, we started the data collection on March 4, 2019 for impact and performance evaluations at the same time. SKO organized the enumerators into two teams of four and two teams of five individuals. The SKO team leader followed the teams of enumerators to oversee data quality and provide technical support under supervision of IMPAQ's team. The CRS community mobilizers (CMs) and a Provincial Education and Sports Service (PESS) representative accompanied the fieldwork team to facilitate informal meetings between SKO and the District Education and Sports Bureaus (DESBs) and to coordinate school visits.

All enumerators regrouped with their supervisors in their village every day to debrief, submit data collection logs, submit electronic surveys, and plan for the next days of data collection. The SKO team leader was responsible for regularly updating IMPAQ's project director on challenges faced and decisions made. The team completed fieldwork in six weeks before the Lao New Year started on April 11.

3.4 Field Challenges

Overall, during the six weeks of data collection, both performance and impact evaluations did not face any major challenges that could delay the data collection activities and data collection was completed in a timely manner. However, tracking the same respondents from baseline did pose a large challenge for the impact evaluation.

We had provided to the enumerators a list of students from baseline with some background information so that the enumerators could identify the students through the school principals. However, for various reasons, such as grade repetition, dropout, and relocation to another school, we could not find all 2,184 students from baseline. Consequently, in these cases, enumerators replaced the students randomly with other students present at school to maintain the power. However, this solution was not feasible in some schools due to their low attendance rate. In some schools, on the day of our school visit, we learned that our target grade had been canceled because of teacher absenteeism.³⁸ When this happened, enumerators asked other teachers or the school principal to call students to come back to school for the survey. As mentioned in 2.3 Sampling Design, our replacement strategy allowed us to survey 1,657 students across all treatment and comparison groups. Even so, differential attrition was not a major concern since impact evaluation estimates from different sub-samples of students (e.g. including only tracked students vs including tracked and missing students vs. including tracked, missing, and replacement students) were very similar as discussed in Section 4. Evaluation Sample and Section 5. Results.

³⁸ This incident happened in 9 schools in impact evaluation sample: two schools in Atsaphone because of a village festival and funeral ceremony; and seven schools in Sepon for various reasons, including teacher had not attended school for a week (one school), school has been closed for three days (one school), family emergency (one school), and celebrations for weddings or festivals (four schools).

SECTION 4. EVALUATION SAMPLE

This section describes the impact evaluation sample for LEAPS II in Atsaphone and Sepon.

First, in Section 4.1 Distribution of Students, we present the distribution of students across the three evaluation arms: SF, SF+LB, and comparison (C) schools. In Section 4.2 Student Background at Follow Up, we provide descriptive statistics of key student and household characteristics (as reported in the student survey). We report information separately by the three groups of schools (treatment and comparison groups) and indicate whether the differences between each of the treatment arms (SF or SF+LB) and comparison schools are statistically significant. These equivalence checks help verify and control for the differences between each of the treatment arms and comparison group in the regression analysis. Finally, in the last Section 4.3 Attrition at Follow Up, we discuss the response and attrition rates across the three groups, as well as the differential attrition (i.e. the difference in the rates of attrition between each of the treatment and comparison groups).

4.1 Distribution of Students

At baseline, IMPAQ reached a sample of 2,184 students in 159 schools with 45 SF schools in 12 clusters, 59 SF+LB schools in 14 clusters, and 55 comparison schools in 19 clusters. As explained in 2.3 Sampling Design, we visited the same schools in March and April 2019 to administer a follow-up survey to students surveyed at baseline. However, due to baseline students in all study arms repeating grades, dropping out of school, being absent, relocating to another school, and/or other reasons, we only tracked 48 percent of students originally sampled in the baseline.

Given the challenges of locating students surveyed at baseline and to ensure a large enough sample to maintain the power (as calculated at baseline), we replaced the students who could not be tracked with available students at the time of the follow-up survey. That is, for students whom we could not track, we replaced them with available students not surveyed at baseline (616 out of 1,657 students). In replacing students, we ensured to keep the gender ratio balanced to the extent that the sex-ratio of present students allowed.³⁹ Teachers provided written consent for the children in their school. We also asked each student for his or her assent to participate in the survey; no student refused.

Based on our sampling strategy, we defined three types of students in our sample:

1. *Tracked students* – students surveyed at both baseline and follow up
2. *Missing students* – students surveyed at baseline but not at follow up
3. *Replacement students* – students not surveyed at baseline but surveyed at follow up.

Exhibit 3 shows the distribution of students in each evaluation group at the baseline and follow up.

Exhibit 3. Sample Distribution by District and Type of School

District	Students Surveyed at Baseline			Students Surveyed at Follow Up		
	SF	SF+LB	Comparison	SF	SF+LB	Comparison
Atsaphone	248	345	298	211	288	213
Sepon	353	398	542	284	219	342
Overall	601	743	840	495	607	555
	2,184			1,657		

³⁹ If a girl was not available at follow up, we randomly selected another girl from present students. However, if there were not enough girls available on the day of school visit, we replaced her with a boy in the same class to maintain the power.

Source: Student survey; authors' calculations.

To ensure comparability across treatment and comparison groups for the follow-up evaluation, IMPAQ completed tests of balance based on students' backgrounds; we present those results in the next section.

4.2 Student Background at Follow Up

Including the replacements, we surveyed a total number of 1,742 students. However, we excluded 85 students from our impact analysis because they had not been exposed to the program for the full two years. We refer to them as new comers in the report.⁴⁰ The balance tests reported here reflect the overall sample of 1,657 students surveyed at follow up, regardless of whether they were originally part of the baseline sample (i.e. original students) or replacements at follow up, and excluding the 85 newcomers.

⁴¹We also checked the sensitivity of the data by excluding replacements and highlighting any differences that emerged.

4.2.1 Student Characteristics

This section describes students' key demographic characteristics, including gender and age, as well as their experience in school (grade repetition and participation in pre-school), between each treatment arm and comparison group at the follow up.

Each exhibit in this section compares the means of outcomes between each of the treatment arms (SF or SF+LB) and comparison group at the follow up. We added total number of observations (N) for each percentage in a column next to the means, or as a note in a graph. We show the t-test results, if any, with an asterisk above each treatment percentage and a table/graph note explaining each asterisk.

Gender. Of the 1,657 students sampled, 784 were female (47 percent) and 873 were male (53 percent). The gender ratio was balanced between SF and comparison schools where both had around 46 to 45 percent girls. However, the proportion of girls was significantly higher in the SF+LB group (51 percent) compared to the comparison group (45 percent), at the 5 percent level.

Exhibit 4. Students' Gender by Treatment Group

Variable	SF	N	Comparison	N	SF+LB	N	Total
Female	46%	227	45%	247	51%**	310	784

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. N = 495 for SF, 555 for comparison, and 607 for LB+SF.

Age. Exhibit 5 breaks down Grade 2 students' age by mean, median, and range. The average age was between 8.2 and 8.4 across the three school types, with no differences between each of the treatment arms and comparison group. The average age shows consistently that in each group students were the correct age for their grade (Grade 2). However, students' ages range from 5 to 16 years, consistent across

⁴⁰ The first cohort of the LB intervention started the implementation with Grade 1 in 2017-2018 school year, and continued with Grade 2 in 2018-2019 school year. We considered two full years of exposure to the program if students started in Grade 1 at baseline and were in Grade 2 at follow up in the same school. Thus, we excluded 70 students that have been repeating Grade 2 since 2017 and 15 students enrolled in our target school in 2018-2019 school year.

⁴¹ We did not include students who were still in school but were repeating Grade 1, because they missed the second year of exposure through Grade 2 as LB intervention planned.

the treatment and comparison arms. Late entrance to primary school,⁴² high grade repetition in the early grades,⁴³ and students being unaware of their real ages could account for the wide age range.

Exhibit 5. Students' Age by Treatment Group

Age	SF	Comparison	SF+LB	Overall
Mean Age	8.2	8.3	8.4	8.3
Median Age	8	8	8	8
Range of Ages	6-14	5-16	5-15	5-16

Source: Student survey; authors' calculations. N = 349 for SF, 365 for comparison, and 400 for LB+SF. Note: Many students did not know their age; therefore, Ns are smaller for this question.

Grade repetition and pre-school participation. Overall, 29 percent of the sample reported that they repeated a grade. Although, the distribution was balanced between SF+LB and comparison groups, fewer students in SF (23 percent) than in comparison (33 percent) reported repeating a grade, statistically significant at the 1 percent level. We also asked students if they had ever attended pre-school or attended an early childhood development (ECD) program. Approximately, half of the sample (52 percent), said yes, with no significant differences between each of the treatment arms and comparison.

Exhibit 6. Grade Repetition and Preschool Participation by Treatment Group

Variable	SF	N	Comparison	N	SF+LB	N	Overall
Ever repeated a grade	23%***	493	33%	548	29%	604	29%
Attended ECD/preschool	56%	494	47%	551	52%	596	52%

Source: Student survey; Authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group.

4.2.2 Household Environment

In this section, we compared students' key household characteristics, including household size, main language spoken at home, socioeconomic status, availability of books at home, and students' home literacy environment.

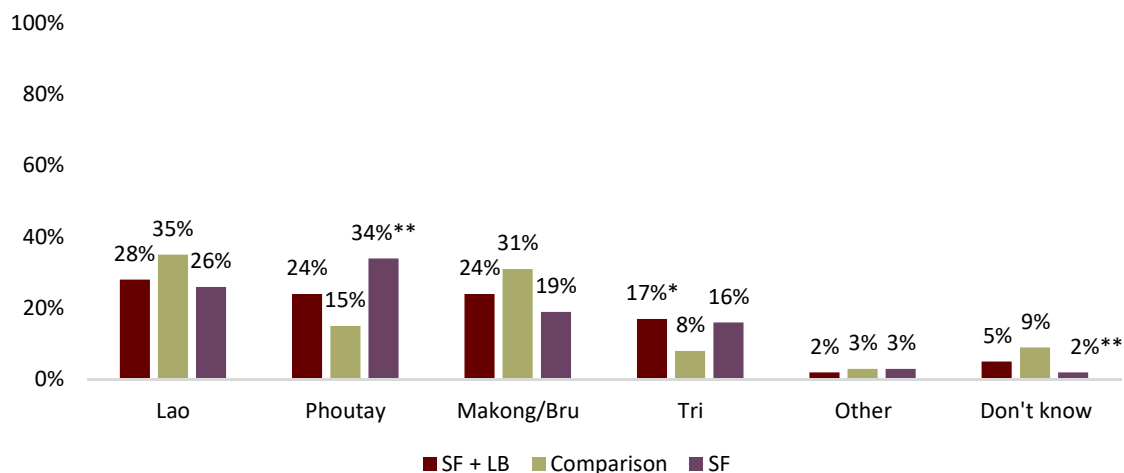
Household size. The average household size was 6.5 people, consistent between each of the treatment arms and the comparison group, with similar ranges from 2 to 17 people.

Language spoken at home. We asked children about their primary language spoken at home, without reading the options to them. If the child reported a language other than those listed, the enumerators chose "other" as the response. Exhibit 7 shows the breakdown of main languages spoken by our sample. We found some differences in the reported language spoken by treatment and comparison group; however, not all the differences were statistically significant, except for Phoutay. We noted a statistically significant difference in the proportion of Phoutay speakers between the SF treatment and comparison group ($p < 0.05$). Among multilingual students, 18 percent of the follow-up sample reported speaking at least two languages, and 95 percent of multilingual speakers speak Lao as one of their languages. There was no significant variation between each of the treatment arms and the comparison group.

⁴² FHI 360. 2016. Education and Policy and Data Center. Retrieved from: https://www.epdc.org/sites/default/files/documents/Laos_coreusaid.pdf

⁴³ World Bank. 2016. Reducing Early Grade Drop Out and Low Learning Achievement in Lao PDR: Root Causes and Possible Interventions. Washington, D.C.

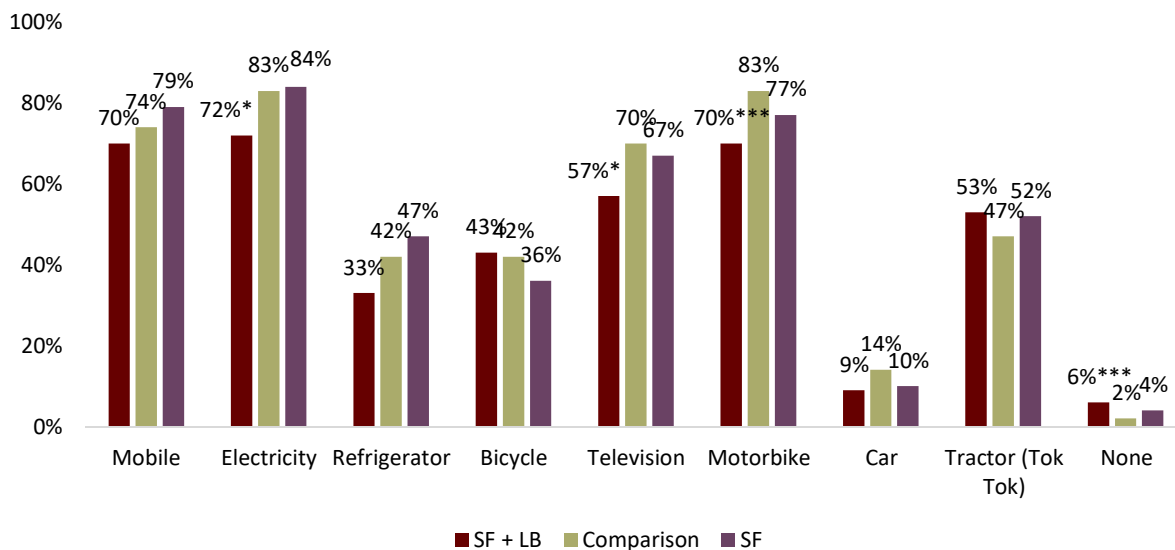
Exhibit 7. Main Language Spoken at Home by Treatment Group



Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 607$ for LB+SF, 555 for comparison, and 495 for SF.

Socioeconomic status. We asked students about their household's socio-economic status as measured by a household's possession of any eight durable goods, including electricity, refrigerator, bicycle, motorbike, *tok tok* (tractor), television, mobile phone, and car. We also created an index for students' socioeconomic status by adding their responses on whether their household possessed any of eight goods or services. On average, students reported that their households possessed 4.4 of these goods, with no difference between each of the treatment arms and comparison group. As Exhibit 8 shows, however, there were some significant differences with specific items. Fewer students in the SF+LB group (70 percent) reported that they owned a motorbike compared to students in the comparison group (83 percent), ($p < 0.01$). Additionally, students in the SF+LB group were more likely to own none of the eight items (6 percent compared to 2 percent, $p < 0.01$).

Exhibit 8. SES by Treatment Group

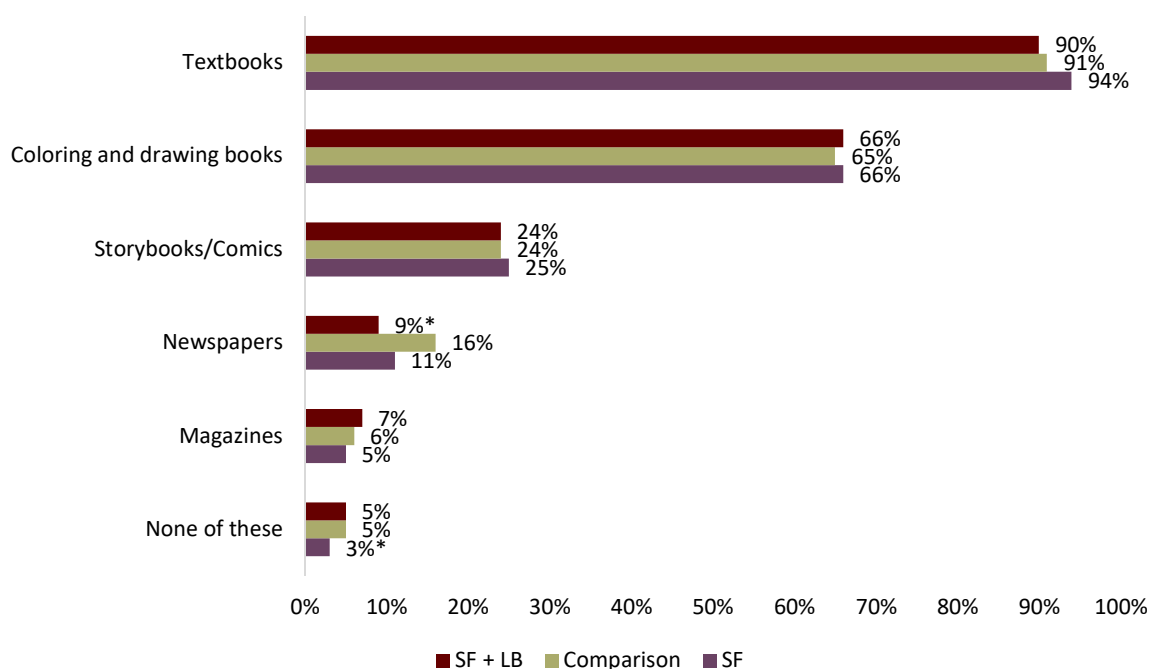


Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 607$ for SF+LB, 555 for comparison, and 495 for SF.

Availability of reading materials at home. Students exposed to literacy activities at home have better opportunities for literacy acquisition.⁴⁴ Numerous studies point to the role of the home literacy environment in influencing early reading skills—in particular, children’s exposure to print materials at home and opportunities to engage in reading with other household members.⁴⁵ Our survey therefore asked questions about the availability of print materials at home and home literacy practices.

We asked students about the types of books available to them at home including textbooks, coloring/drawing books, storybooks/comics, newspapers, and magazines. Exhibit 9 shows that the sample is balanced in terms of the availability of books at home between each of the treatment arms and comparison group. Consistently, students in each group most frequently reported having textbooks at home, followed by coloring books.

Exhibit 9. Availability of Books at Home by Treatment Group



Source: Student survey; authors’ calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 607$ for LB+SF, 555 for comparison, and 495 for SF.

Home literacy practices. We asked if, during the last week, anyone in their household encouraged students to study, read to them, told them a story, or asked them questions about the stories. As well, we inquired whether students saw anyone reading at home. For each of these four activities, we created an index by assigning one point to each of these activities, but doubling the weight if students reported that someone read to her/him.

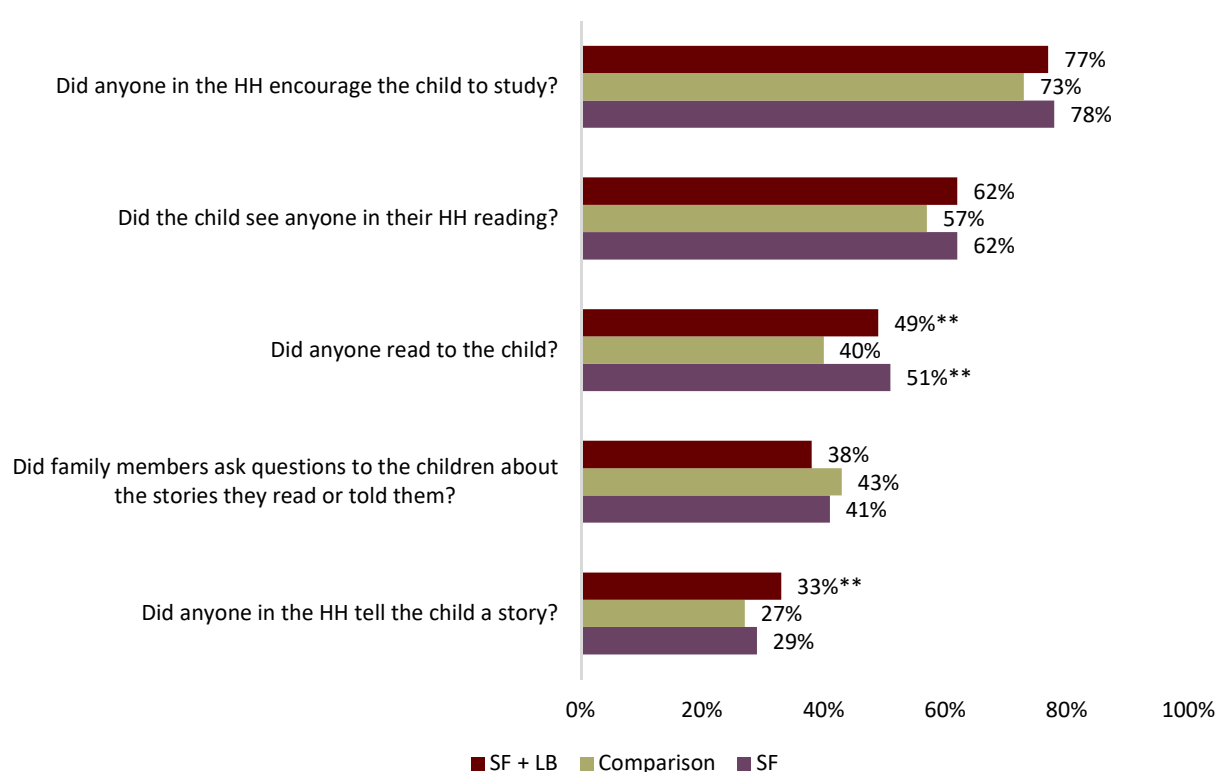
⁴⁴ Kim, Y. S. 2009. The relationship between home literacy practices and developmental trajectories of emergent literacy and conventional literacy skills for Korean children. *Reading and Writing*, 22(1), 57-84.

⁴⁵ Hess, R. D. & Holloway, S. D. 1984. Family and school as educational institutions. *Review of Child Development Research*, 7, 179-222. Dowd, A.J., Pisani, L. & Borisava, I. 2016. “Evaluating Early Learning from Age 3 to Grade 3” in *Understanding What Works in Oral Reading Assessments*. Montreal: UNESCO Institute for Statistics (UIS).

On average, students in each of the SF+LB group and the SF group reported slightly higher literacy interactions (2.7 on average in each arm) compared to students in comparison schools (2.5 practices); The differences were statistically significant at the 5 and 1 percent level, respectively.

Exhibit 10 shows a breakdown of the components in the literacy index by treatment group. Overall, households were more likely to participate in passive literacy interactions with children such as encouraging studying (76 percent) or reading themselves in the presence of children (60 percent) than the more active ones, such as reading directly to children (47 percent) or helping children study (30 percent). Households in the SF+LB group (49 percent) and in the SF group (51 percent) were significantly more likely to read to the children than the comparison group (40 percent), each significant at the 5 percent level. In addition, households in the SF+LB group (33 percent) were six percentage points more likely than the comparison group (27 percent) to tell a story to children ($p < 0.05$).

Exhibit 10. Home Literacy Practices by Treatment Group



Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 607$ for LB+SF, 555 for comparison, and 495 for SF.

4.3 Attrition at Follow Up

Attrition is one of the major methodological problems of longitudinal studies. In the impact evaluation, we have two main types of attrition – (1) *attrition from the school* such as students who dropped out of school or moved from the area), and (2) *attrition from the evaluation sample*, including students who are repeating Grade 1 and/or were absent on the survey day. While these are distinct reasons for attrition, for the purpose of the impact evaluation analysis and for simplicity, we classified all of them as “missing” students since we only have baseline information for them. Thus, throughout this report, we use attrition in a more general way to identify all these types of students.

For the impact evaluation, we experienced a high attrition rate (overall 52 percent) at the follow up. As shown in Exhibit 11, we could only track 44 percent of students in comparison schools and 51 percent in SF schools. The proportion of students tracked was higher in Atsaphone (60 percent) compared to Sepon (41 percent). Despite the high attrition rate in all the school types, there are no statistically significant differences in the level of attrition between each treatment arm and the comparison group.

Exhibit 11. Sample Distribution by Original Students

District	Percent of Original Students Tracked from Baseline to Follow Up		
	SF	SF+LB	Comparison
Atsaphone	65%	57%	55%
Sepon	41%	44%	37%
Overall	51%	50%	44%
	48%		

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group.

One of the main concerns with attrition in a quasi-experimental setting is differential attrition, i.e. different rates of attrition between the treatment arms and the comparison group. We performed two checks to alleviate this concern. First, we checked for the reasons causing attrition and whether they differ by treatment arm and comparison group. Second, we performed additional balance tests for the main background characteristics between tracked and replacement students.

Check 1: Reasons for Attrition. As noted in Overview of Lao Context, a number of factors could cause attrition in an education study: grade repetition; absenteeism; dropout; relocation to other schools. Exhibit 12 presents the main reasons for attrition in our sample and tests for differences between each of the treatment and comparison groups. The primary reasons for sample attrition align with the literature, as presented in Section 1. Introduction on the challenges to schooling. Most commonly, principals and teachers reported grade repetition (24 percent overall), followed by absenteeism on the day of the survey (14 percent overall) as the main reasons. Except for dropouts, all other reasons for attrition are similar between each of the treatment arms and the comparison group. For dropouts, the comparison group had a greater proportion (11 percent) compared to either of the two treatments (6 percent in the SF intervention and 4 percent in the SF+LB intervention, statistically significant at the 10 and 1 percent level, respectively). To examine if differential dropout between the treatment arms and the comparison group poses a bias to the interpretation of our findings, we estimated the regression results separately for three sub-samples: (1) a sample that includes only the tracked students; (2) a sample that includes only tracked and missing students; and (3) a sample that includes tracked, missing, and replacement students. If differential dropout is a concern, then the three samples would give significantly different results. However, as shown in Section 5. Results, we find that for all three samples the results are almost identical, thus indicating that differential attrition, while present, was not large enough to affect the main results.

Exhibit 12. Reasons for Attrition by Treatment Arm

Reason	SF	N	Comparison	N	SF+LB	N	Overall
Repetition	21%	127	26%	216	24%	178	24%
Absent	16%	94	12%	103	15%	109	14%
Dropped out	6%*	34	11%	90	4%***	33	7%
Unknown	5%	28	6%	48	4%	33	6%
Moved	1%	8	1%	6	1%	8	1%

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. Note: The columns do not add up to 100 percent because those who were tracked are omitted. Tracked students made up 51 percent for SF, 44 percent for Comparison, 52 percent for SF+LB, and 48 percent of the overall sample.

Check 2: Balance between tracked and replacement students. We tested for balance in background characteristics for tracked vs. replacement students. These background characteristics include students'

gender, grade repetition, participation in a preschool or ECD program, and main language spoken at home. We also looked at students' socioeconomic status index.

We looked at equivalences on these characteristics between tracked vs. replacement students to ensure students who made it to Grade 2 at the follow up are similar to the replacement students, which is important to ensure that the treatment and comparison groups at follow up remain “comparable”. Exhibit 13 presents the balance tests for tracked students vs. replacements. We found that the socioeconomic index, grade repetition, attendance at an ECD/preschool, and main language have significant differences across tracked vs. replacement students. This implies that tracked students are better off compared to replacement students because of a higher socioeconomic index and better proficiency in Lao (since a greater proportion of students noted themselves as Lao and Phoutay speakers). Consequently, in the results presented in Section 5. Results, we include these variables as covariates to control for these differences.

Further, in Appendix C. Additional Exhibits we also present balance checks between tracked and missing students (Exhibit 29) and missing and replacement students (Exhibit 30). We find that missing students were more likely to be poorer than tracked students, indicating that poorer students were more likely to drop out than those who were tracked (Exhibit 29). The replacement students were also better off than missing students (Exhibit 30). This is not unsurprising because replacement students were selected from students who were present in school on the day of the survey. Thus, they would be students who are more likely to come to school and who may be better off. Consequently, given the differences between tracked and missing students *and* missing and replacement students, we control for an indicator for the main language being Lao, the socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program.

Exhibit 13. Balance Tests between Tracked and Replacement Students at the Follow Up

Background Characteristics	Tracked	Replacement	Difference
Gender	48.5%	45.3%	3.2%
Age	8.2	8.6	0.4***
Ever repeated a grade	26.6%	32.2%	5.6%*
Attended ECD/pre-school	55.3%	44.6%	10.7%***
Socioeconomic Index	4.53	4.46	0.427***
Main language – Lao	31.3%	26.6%	4.7%*
Main language – Phoutay	27%	19.2%	7.8%***
Main language – Makong/Bru	23.2%	27.6%	4.4%
Main language – Tri	10.7%	18.2%	-7.5%***

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 1,040$ for tracked and 616 for replacement. For age, $N = 711$ for tracked students and 403 for replacements. Note: we only include languages with five or more speakers in the sample. We excluded those who selected “other” or “don't know”.

SECTION 5. RESULTS

In this section we describe our main findings from the impact evaluation of the SF and SF+LB interventions. Our key outcomes of interest are reading with comprehension and letter recognition, and we intended to capture the effect of two years of exposure to the intervention (SF or SF+LB) on Grade 2 students in SF/SF+LB schools (treatment groups) at follow up compared to Grade 2 students in schools with no interventions (comparison group).

As described in the methodology section, we used a DID method to estimate the two-year effect of the SF intervention and the combined two-year effect of SF+LB on improvements in literacy outcomes. Because of attrition related concerns from baseline to follow up, outlined in Section 4.3, we controlled for any variables in the data that could influence the literacy outcomes such as gender, an indicator for the main language being Lao, socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program.

We used the following DID specification for each treatment arm:

$$Y_{ist} = \alpha + \beta Post + \gamma Treatment + \Delta Treatment * Post + X_{ist} + \mu_c + \varepsilon_{ist}$$

where:

- Y_{ist} is the literacy outcome (e.g. reading with comprehension and letter recognition) of student i in school s at time t ⁴⁶
- $Post$ is an indicator variable, equal to 0 if the outcome is from baseline and 1 if at follow up
- $Treatment$ is an indicator variable equal to 0 for comparison group and equal to 1 for the treatment group in consideration (either SF or SF+LB)
- The coefficient on the interaction term $Treatment * Post$ (Δ) represents the DID estimate, i.e. the change in outcome from baseline to follow up in the treatment group relative to the change in the same outcome for the comparison group
- X_{ist} includes a list of variables, unrelated to the program, that IMPAQ used to control for their effect on literacy outcomes to isolate the effect of the intervention. These factors include gender, an indicator for the main language being Lao, the socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program, of student i in school s at time t , which could potentially be related to the literacy outcome of the student^{47,48}
- μ_c represents school cluster-level fixed effects to control for time-invariant school geographic and other characteristic that vary across schools and can affect outcomes
- ε_{ist} is the error term

Students in the same school might be similar to each other on observable and unobservable characteristics. Therefore, the literacy outcomes of students within a particular school, may be related to

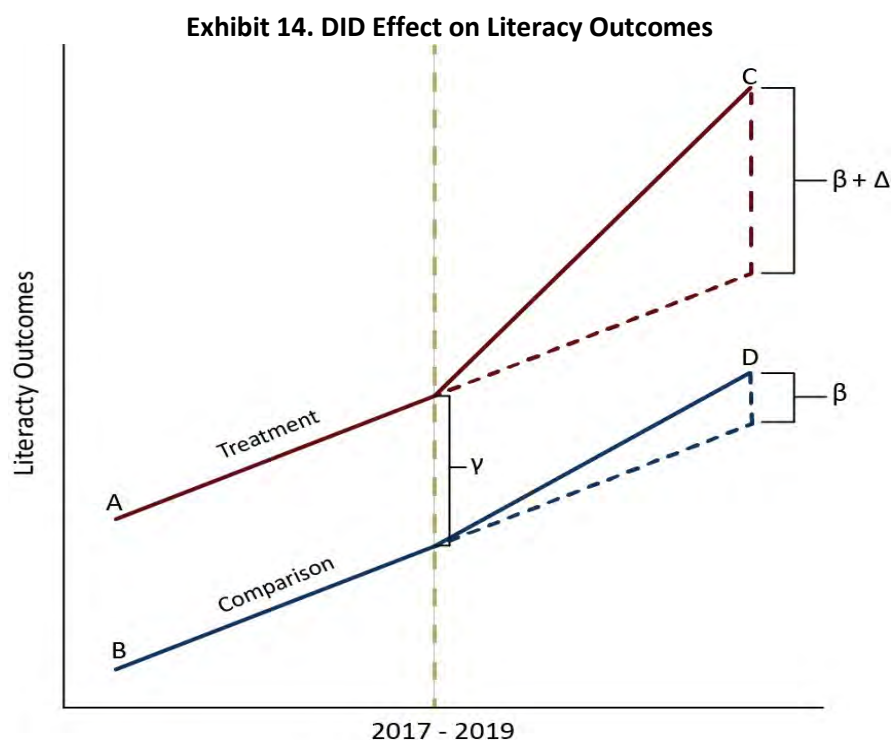
⁴⁶ Both dependent variables of interest are binary and we use a linear probability model.

⁴⁷ We believe the following variables, i.e. an indicator for the main language being Lao, the socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program, to be time varying. This is because, many children at baseline/follow up were unsure of the values for these variables since they were so young. Hence, we take these variables to be time varying. The only time invariant variable is gender.

⁴⁸ We include these specific covariates because, as outlined in Section 4, these characteristics were significantly different between tracked and replacement students, tracked and missing students, and missing and replacement students.

each other (i.e. the correlation between literacy outcomes of children in the same school may be high).⁴⁹ To account for this correlation of error terms across students in the same school, we clustered the standard errors at the school level.

Below we visually show the impact of the SF and SF+LB interventions to facilitate easier comprehension of the intuition of the approach. As shown, β is the time trend as represented by comparison schools' change in literacy outcomes. γ is the average difference in literacy outcomes between treatment and comparison schools at baseline. Our main outcome of interest is Δ , which we measure as the difference in the average outcome in treatment schools before and after treatment **minus** the difference in average outcome in control schools before and after treatment, controlling for time effects.



Source: IMPAQ

As a result of attrition at follow up, we had three types of students: tracked, missing, and replacements. To ensure that selective attrition is not driving our results, we present DID results for three sub-samples.

1. *Balanced Sample*. This sample includes only the students surveyed at the baseline and at follow up (i.e. tracked students).
2. *Unbalanced Sample*. This sample includes students surveyed at baseline and follow up (i.e. sample in [1]), as well as students who we surveyed only at baseline (missing students); and
3. *Full Sample*. The full sample contains students surveyed at baseline and follow up; students surveyed only at baseline but not at follow up; and students surveyed at follow up but not at baseline (i.e. replacements).

⁴⁹ "The clustering problem is caused by the presence of a common unobserved random shock at the group level that will lead to correlation between all observations within each group" - Christian B Hansen. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, 140(2):670–694, 2007.

We present the results from the specified regression above for the SF+LB intervention in Section 5.1 School Feeding and Literacy Boost Interventions and for the SF intervention alone in Section 5.2 School Feeding Intervention.

We also present various robustness checks to check the sensitivity of the regression results to inclusion of other covariates, as well as different statistical specification. In Appendix C. Additional Exhibits, we present three additional robustness checks. First, we conducted robustness checks by including age⁵⁰ and prior exposure to Room to Read intervention⁵¹ in addition to other covariates.⁵² Second, we examined a different specification for the DID method in the balanced sample — the first difference estimator.⁵³ Finally, we estimated an alternative regression specification by simply controlling for the baseline value of the outcome variable rather than using the DID approach.⁵⁴

5.1 School Feeding and Literacy Boost Interventions

We have structured each exhibit in this section as follows: the highlighted row (*Treatment*Post*) presents the key program effects on letter recognition or reading with comprehension. We also present coefficients on the differences in outcomes between treatment and comparison schools at baseline (*Treatment*) and the time trend (*Post*). The first column for each sample type presents treatment effects without controlling for any covariates. The second column for each sample type illustrates the treatment effects after controlling for the following covariates: gender, an indicator for the main language being Lao, the socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program.⁵⁵

5.1.1 Letter Recognition

We describe the effect of SF+LB on students' letter recognition for the three samples below. Exhibit 15 on the next page summarizes the impact results on letter recognition.

Balanced Sample: Exposure to the SF+LB interventions increased Grade 2 students' proficiency in letter recognition by 42.9 percentage points in the balanced sample. This effect is large and statistically significant ($p < 0.01$). At baseline, only 5 percent of students could recognize letters. The time effect is also

⁵⁰ Among all the students at baseline and follow up, 53 percent did not know their age. We left age out of the main list of covariates to maintain the power because sample size could be reduced considerably. However, we checked the sensitivity of the results by controlling for age.

⁵¹ In our impact evaluation sample, 26 schools were previously in the Room to Read education program under LEAPS I in Sepon and Atsaphone districts. All of these 26 schools received a school library, and three also received a full package of literacy activities, which included teacher training, school library, and in-class learning aids.

⁵² See footnote 44.

⁵³ The first difference estimator is the equivalent of taking differences for the same observations over the two time periods. It compares changes over time for the same individual. Therefore, to implement the first difference estimator, it is critical to have data for two time periods for the same individual (Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT Press, 2010).

⁵⁴ Recent literature has shown that simply controlling for the baseline value of the outcome variable gives you more statistical power than using the DID approach. Further, this approach also helps in examining the coefficients on time invariant covariates, which is useful from a programmatic standpoint (Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2008.)

⁵⁵ Because having repeated a grade is potentially endogenous, we also estimated the DID model without including this covariate and the results are similar.

significant, which means letter recognition increased by 44 percentage points after introduction of the program.

Unbalanced Sample: For the unbalanced sample, the results match closely the balanced sample in terms of magnitude and precision. Students in SF+LB schools are more likely to identify letters at the end of Grade 2 by 40.7 percentage points ($p < 0.01$). Adding covariates does not change these estimates.

Full Sample: Finally, for the full sample that includes all three sets of children, the results, again, are similar to the other two samples. Exposure to the SF+LB interventions increased students' letter recognition skills at the end of Grade 2 by 40.7 percentage points ($p < 0.01$).

The results show strong evidence of SF+LB impact on letter recognition, especially given the similarity of results across all three samples, indicating that differential attrition did not drive our results. If selective attrition did pose a problem, then we should have seen different results in the balanced and unbalanced samples. Strong praise of reading camps in the qualitative data collected for our performance evaluation provides some context for these findings. Almost all qualitative respondents linked improved letter recognition to the lessons taught to students during these camps, which are scheduled to occur during the summer months and during the academic year in all LB schools. As well, for regular classroom instruction, teachers and principals remarked that they have acquired better knowledge of how to teach the Lao language in smaller components, either word by word or by vowels. Possibly, because students learn letter recognition in their first couple of years in primary schools, time might drive the results. However, the DID method captures the time effect by looking at the “post” coefficient to isolate the program's effect.

Exhibit 15. Impact of SF+LB on Letter Recognition

Variables	Balanced Sample		Unbalanced Sample		Full Sample	
	Coefficient (SE)		Coefficient (SE)		Coefficient (SE)	
	I	II	III	IV	V	VI
Treatment*Post (DID)	0.429 (0.045)***	0.430 (0.045)***	0.407 (0.051)***	0.410 (0.049)***	0.407 (0.046)***	0.404 (0.045)***
Treatment	0.011 (0.032)	0.020 (0.029)	0.007 (0.022)	0.010 (0.021)	0.000 (0.024)	0.003 (0.023)
Post	0.440 (0.045)***	0.429 (0.041)***	0.479 (0.048)***	0.451 (0.045)***	0.486 (0.043)***	0.463 (0.041)***
Controls	No	Yes	No	Yes	No	Yes
N	1,472	1,437	2,318	2,226	2,743	2,646

Source: Student survey; authors' calculations; p -value < 0.1 ** p -value < 0.05 *** p -value < 0.01 ; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster-level fixed effects.

Sub-group Analysis. Most of the results from the overall sample described above continued to hold after running the regressions by gender (Exhibit 16) and by main language spoken at home (Lao vs. non-Lao speakers, Exhibit 16).

IMPAQ found a large and statistically significant impact for both girls and boys ($p < 0.01$), but the difference in their impacts are not statistically significant. Girls, on average, saw a 39.6 to 41 percentage point increase in letter recognition across different samples, while for boys the increase was similar and ranged from 39 percentage points to 44 percentage points. At baseline, only 3 percent of boys and 8 percent of girls in the sample could recognize letters. These results imply boys have witnessed a larger increase relative to baseline since they had a lower baseline value to begin with. Possibly, the slight preference for

girls to remain at home either to help with housework or childcare, as revealed in the qualitative data for the performance evaluation, accounts for boys remaining in school and absorbing more of the lessons.

With respect to the Lao vs. non-Lao sample in Exhibit 17, both samples witnessed positive and statistically meaningful impacts on letter recognition ($p < 0.01$). The magnitude of the effect is larger for the non-Lao sample (45.2 percentage points for non-Lao in the full sample vs. 35 percentage points for Lao in the full sample). Treatment effects for all sub-groups are stable across all three sample types. As explained in Section 1. Introduction, the literature explicitly refers to gender disparity and language-group disadvantages as known education issues at the national level. These results could also reflect gains because of inclusive education, which aims to train school staff and the community to make schools a more welcoming environment for all children, including children who do not speak Lao at home.

Exhibit 16. Impact of SF+LB on Letter Recognition by Gender

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Female Sample			
Treatment*Post (DID)	0.409 (0.050)***	0.396 (0.050)***	0.408 (0.049)***
Treatment	0.018 (0.038)	0.007 (0.028)	0.002 (0.028)
Post	0.438 (0.042)***	0.461 (0.044)***	0.457 (0.043)***
N	717	1,047	1,237
Male Sample			
Treatment*Post (DID)	0.440 (0.062)***	0.416 (0.063)***	0.396 (0.055)***
Treatment	0.023 (0.037)	0.012 (0.021)	0.003 (0.025)
Post	0.427 (0.053)**	0.444 (0.056)***	0.468 (0.049)***
N	722	1,181	1,411

Source: Student survey; authors' calculations; p -value < 0.1 ** p -value < 0.05 *** p -value < 0.01 ; Standard errors shown in parentheses are clustered at the school level. Female and male sample regressions include controls for an indicator for the main language being Lao, socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster level fixed effects.

Exhibit 17. Impact of SF+LB on Letter Recognition by Main Language (Lao vs. non-Lao)

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Lao Sample			
Treatment*Post (DID)	0.343 (0.082)***	0.324 (0.076)***	0.350 (0.075)***
Treatment	0.315 (0.096)***	0.271 (0.091)***	0.190 (0.082)**
Post	0.524 (0.056)***	0.539 (0.055)***	0.554 (0.059)***
N	317	355	481

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Non-Lao Sample			
Treatment*Post (DID)	0.487 (0.054)***	0.470 (0.055)***	0.452 (0.047)***
Treatment	-0.010 (0.028)	-0.005 (0.019)	-0.011 (0.021)
Post	0.381 (0.049)***	0.405 (0.051)***	0.425 (0.043)***
N	1,122	1,873	2,167

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 * p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Lao and non-Lao sample regressions include controls for gender, the socioeconomic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster fixed effects.*

5.1.2 Reading Comprehension

In this section, we discuss the effect of SF+LB on students' reading with comprehension skills. See Exhibit 18 on the next page for a summary of the impact results.

Balanced Sample: As shown in Exhibit 18, Grade 2 students in SF+LB schools are more likely to read with comprehension by 4.3 percentage points. However, this effect is not statistically significant. The time effect (i.e. the coefficient on the "post" variable) is statistically significant, showing students can read with comprehension at their grade level by 9.2 percentage points over time.

Unbalanced Sample: The results for the unbalanced sample aligned closely with the balanced sample. Exposure to the SF+LB interventions increased the reading comprehension proficiency of Grade 2 students by 4.4 percentage points, but, again, not statistically significant. Adding covariates does not change these estimates, as well.

Full Sample: The effect of the SF+LB intervention on Grade 2 students' reading with comprehension skills in the full sample also shows that despite improvements by 4.1 percentage points, such gains are not statistically meaningful.

Consistent with letter recognition results, selective attrition is a much smaller concern, because the results for balanced, unbalanced, and full samples varied little. The impact findings are consistent with performance results on reading with comprehension. They also match the literature and the responses from LEAPS II staff in our qualitative findings, which suggest reading with comprehension usually emerges in Grades 3 and 4, and the majority of students achieve full fluency and comprehension in Grade 5.⁵⁶

⁵⁶ The World Bank. May 2016. Reducing early grade dropout and low learning achievement in Lao PDR. Retrieved from: <http://documents.worldbank.org/curated/en/161641483590757065/pdf/111627-REVISED-PUBLIC-Lao-PDR-Root-Causes-of-Early-School-Leaving-f.pdf>

Exhibit 18. Impact of SF+LB on Reading with Comprehension

Variables	Reading Comprehension					
	Balanced Sample Coefficient (SE)		Unbalanced Sample Coefficient (SE)		Full Sample Coefficient (SE)	
	I	II	III	IV	V	VI
Treatment*Post (DID)	0.043 (0.039)	0.041 (0.039)	0.044 (0.038)	0.043 (0.038)	0.041 (0.035)	0.039 (0.035)
Treatment	-0.001 (0.018)	-0.000 (0.019)	-0.002 (0.011)	-0.003 (0.011)	-0.011 (0.012)	-0.012 (0.013)
Post	0.092 (0.031)***	0.091 (0.033)***	0.090 (0.030)***	0.086 (0.031)***	0.094 (0.026)***	0.094 (0.025)***
Controls	NO	YES	NO	YES	NO	YES
N	1,472	1,437	2,317	2,226	2,742	2,646

* Source: Student survey; Authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

Sub-group Analysis. Most of the results from the overall sample described above continue to hold. While we found no statistically significant impact for both girls and boys, the magnitude of the coefficient is higher for girls than boys, indicating a potentially larger program effect for girls (4.5 percentage points for girls vs. 3.2 percentage points for boys in the full sample). Statements from principals, teachers, and parents in our qualitative data indicated that girls prefer to read more than boys.

Exhibit 19. Impact of SF+LB on Reading with Comprehension by Gender

Variables	Reading Comprehension		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Female Sample			
Treatment*Post (DID)	0.047 (0.050)	0.049 (0.047)	0.045 (0.044)
Treatment	-0.012 (0.025)	-0.013 (0.015)	-0.019 (0.018)
Post	0.113 (0.035)***	0.108 (0.033)***	0.104 (0.029)***
N	717	1,047	1,237
Male Sample			
Treatment*Post (DID)	0.028 (0.041)	0.028 (0.041)	0.032 (0.036)
Treatment	0.009 (0.019)	0.005 (0.010)	-0.005 (0.013)
Post	0.077 (0.035)**	0.074 (0.033)**	0.086 (0.027)***
N	722	1,181	1,411

*Source: Student survey; Authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Female and male sample regressions include controls for an indicator for the main language being Lao, a socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster level fixed effects.

With respect to the Lao vs. non-Lao sample, the non-Lao sample witnessed a marginally significant impact on reading comprehension by 6.1 percent, which is consistent across all three sample types. Although the

results contrast the literature, they could, as with letter recognition, reflect the effect of the inclusive education component of LEAPS II and the extra attention paid to non-Lao speakers in reading camps. One CLV interviewed in Atsaphone remarked that the camps serve as an additional class to complement school instruction on the Lao language. Meanwhile, the Lao sample had very low sample sizes and did not show any statistically significant impact of the SF+LB intervention.

Exhibit 20. Impact of SF+LB on Reading with Comprehension by Main Language Spoken at Home

Variables	Reading Comprehension		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Lao Sample			
Treatment*Post (DID)	0.022 (0.061)	-0.052 (0.088)	-0.054 (0.079)
Treatment	-0.041 (0.046)	0.099 (0.061)	0.066 (0.062)
Post	0.141 (0.044)***	0.174 (0.057)***	0.158 (0.058)***
N	382	355	481
Non-Lao Sample			
Treatment*Post (DID)	0.061 (0.035)*	0.062 (0.034)	0.061 (0.033)*
Treatment	-0.013 (0.013)	-0.009 (0.007)	-0.018 (0.010)
Post	0.068 (0.024)***	0.067 (0.024)***	0.078 (0.022)***
N	1,122	1,873	2,167

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 * p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Lao and non-Lao sample regressions include controls for gender, the socioeconomic index, an indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster-level fixed effects.*

5.2 School Feeding Intervention

The structure of each exhibit in this section follows the same format described in 5.1 School Feeding and Literacy Boost Interventions.

5.2.1 Letter Recognition

As presented in Exhibit 21, IMPAQ examined the effect of SF intervention alone on students' letter recognition. The highlighted row represents the key program effects, and we describe such impacts in greater detail below for the three samples.

Balanced Sample: Exposure to the SF intervention did not lead to any statistically significant effect. Further, the magnitude of the effect for the balanced sample is not distinguishable from zero. The time effect is significant: letter recognition of Grade-2 students increased by 44 percentage points after the introduction of the program. Adding covariates does not change these findings.

Unbalanced Sample: The results are almost identical to the balanced sample in magnitude and precision. Again, the effect is not distinguishable from zero. Adding covariates does not change these estimates.

Full Sample: Finally, for the full sample that includes all three sets of children, the results are again similar to the balanced and unbalanced samples. There is no statistically significant effect of the SF intervention on letter recognition and the effect is not distinguishable from zero ($p < 0.7898$).

Two potential reasons could account for the differences in the impact of the SF and SF+LB interventions on letter recognition. First, SF might only be effective at the extensive margin (i.e. increasing school attendance and attentiveness), and only when combined with LB activities does it become effective at the intensive margin (i.e. literacy outcomes). Second, SF might not be the most important binding constraint for children. Most of the positive commentary regarding the perceived impact of LEAPS II in our qualitative interviews considered the blend of all LEAPS II interventions, and not school meals alone. These findings perhaps would support the finding that SF on its own might not be as relevant for enhanced literacy outcomes.

Sub-group Analysis. Most of the results from the overall sample continue to hold for the sub-group analysis by gender (Exhibit 22) and by main language spoken at home (Exhibit 23).

Exhibit 21. Impact of SF Intervention on Letter Recognition

Variables	Letter Recognition					
	Balanced Sample Coefficient (SE)		Unbalanced Sample Coefficient (SE)		Full Sample Coefficient (SE)	
	I	II	III	IV	V	VI
Treatment*Post (DID)	0.033 (0.062)	0.027 (0.062)	0.034 (0.067)	0.030 (0.065)	-0.014 (0.060)	-0.011 (0.058)
Treatment	-0.030 (0.043)	-0.023 (0.038)	-0.025 (0.032)	-0.013 (0.027)	-0.026 (0.033)	-0.010 (0.029)
Post	0.440*** (0.043)	0.425*** (0.041)	0.475*** (0.047)	0.439 (0.045)	0.484*** (0.044)	0.451*** (0.042)
Controls	No	Yes	No	Yes	No	Yes
N	1,344	1,325	2,113	2,054	2,490	2,426

* Source: Student survey; authors' calculations; p -value < 0.1 ** p -value < 0.05 *** p -value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

Exhibit 22. Impact of SF Intervention on Letter Recognition by Gender

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Female Sample			
Treatment*Post (DID)	-0.033 (0.073)	-0.051 (0.073)	-0.007 (0.068)
Treatment	0.011 (0.050)	0.001 (0.043)	0.008 (0.043)
Post	0.438 (0.042)***	0.449 (0.046)***	0.442 (0.043)***
N	637	930	1,082
Male Sample			
Treatment*Post (DID)	0.075 (0.076)	0.070 (0.076)	-0.013 (0.066)
Treatment	-0.039 (0.045)	-0.022 (0.024)	-0.022 (0.027)

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Post	0.421 (0.053)**	0.436 (0.056)**	0.459 (0.050)***
N	689	1,125	1,345

*Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Female and male sample regressions include controls for an indicator for the main language being Lao, a socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program.

Exhibit 23. Impact of SF Intervention on Letter Recognition by Language

Variables	Letter Recognition		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Lao Sample			
Treatment*Post (DID)	0.047 (0.154)	0.001 (0.117)	-0.006 (0.110)
Treatment	-0.002 (0.141)	0.062 (0.093)	0.006 (0.089)
Post	0.533 (0.056)***	0.548 (0.053)***	0.568 (0.059)***
N	307	347	441
Non-Lao Sample			
Treatment*Post (DID)	0.044 (0.073)	0.055 (0.076)	-0.000 (0.062)
Treatment	-0.027 (0.040)	-0.018 (0.026)	-0.009 (0.028)
Post	0.386 (0.049)***	0.404 (0.050)***	0.422 (0.042)***
N	1,019	1,708	1,986

*Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Lao and non-Lao sample regressions include controls for gender, a socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster level fixed effects.

5.2.2 Reading Comprehension

This section summarizes the impact results of the SF intervention on reading comprehension. The results are shown in Exhibit 24 after the analyses.

Balanced Sample: Exposure to the SF intervention increased the reading comprehension score of Grade 2 students by 2.6 percentage points in the balanced sample. This effect is not statistically significant. Adding covariates does not change these findings.

Unbalanced Sample: The results are almost identical to the balanced sample in magnitude and precision. Exposure to the SF intervention increased the reading comprehension score of Grade 2 students by 2.9 percentage points. However, this finding is not statistically significant. Adding covariates does not change these estimates, as well.

Full Sample: Finally, the full sample results again match closely the balanced and unbalanced samples. Exposure to the SF intervention increased the reading comprehension score of Grade 2 students by 1.9 to 2.1 percentage points, across different sub-samples, but not in any statistically meaningful way.

The similarity of results for all three samples means that selective attrition poses a lesser concern in this case as well.

Exhibit 24. Impact of SF Intervention on Reading Comprehension

Variables	Reading Comprehension					
	Balanced Sample Coefficient (SE)		Unbalanced Sample Coefficient (SE)		Full Sample Coefficient (SE)	
	I	II	III	IV	V	VI
Treatment*Post (DID)	0.026 (0.042)	0.024 (0.043)	0.029 (0.042)	0.029 (0.043)	0.019 (0.035)	0.021 (0.035)
Treatment	-0.003 (0.015)	-0.004 (0.017)	-0.005 (0.009)	-0.007 (0.011)	-0.002 (0.010)	-0.003 (0.011)
Post	0.092 (0.031)***	0.096 (0.033)***	0.089 (0.030)***	0.089 (0.031)***	0.093 (0.026)***	0.094 (0.026)***
Controls	NO	YES	NO	YES	NO	YES
N	1,344	1,325	2,113	2,054	2,490	2,426

*Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

Sub-group Analysis. We also performed two sub-group analyses: by gender (Exhibit 25) and by main language spoken at home (Exhibit 26). Most of the results from the overall sample described above continue to hold.

Exhibit 25. Impact of SF Intervention on Reading Comprehension by Gender

Variables	Reading Comprehension		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Female Sample			
Treatment*Post (DID)	0.046 (0.052)	0.054 (0.051)	0.036 (0.042)
Treatment	-0.006 (0.021)	-0.012 (0.015)	-0.001 (0.014)
Post	0.119 (0.035)***	0.112 (0.033)***	0.103 (0.028)***
N	637	930	1,082
Male Sample			
Treatment*Post (DID)	0.002 (0.042)	0.005 (0.042)	0.007 (0.037)
Treatment	-0.004 (0.015)	-0.002 (0.009)	-0.006 (0.012)
Post	0.078 (0.035)**	0.074 (0.033)**	0.089 (0.027)***
N	689	1,125	1,345

*Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Female and male sample regressions include controls for an indicator for the main

language being Lao, a socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster level fixed effects.

Exhibit 26. Impact of SF Intervention on Reading Comprehension by Language

Variables	Reading Comprehension		
	Balanced Sample Coefficient (SE)	Unbalanced Sample Coefficient (SE)	Full Sample Coefficient (SE)
	I	II	III
Lao Sample			
Treatment*Post (DID)	-0.105 (0.077)	-0.093 (0.074)	-0.040 (0.065)
Treatment	0.061 (0.055)	0.048 (0.043)	0.028 (0.044)
Post	0.174 (0.059)***	0.169 (0.058)	0.151 (0.056)***
N	307	347	441
Non-Lao Sample			
Treatment*Post (DID)	0.052 (0.041)	0.060 (0.042)	0.035 (0.034)
Treatment	-0.010 (0.015)	-0.012 (0.010)	-0.007 (0.011)
Post	0.074 (0.024)***	0.067 (0.023)***	0.080 (0.021)***
N	1,019	1,708	1,986

*Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. Lao and non-Lao sample regressions include controls for gender, a socio-economic index, indicator for ever having repeated a grade, and an indicator for having attended an ECD program. All regressions include cluster-level fixed effects.

5.3 Comparison of the Treatment Arms

In this section, we aim to determine to what extent changes in literacy outcomes are attributable to SF and LB activities separately. From a programmatic standpoint, understanding which specific activities are driving the effects on literacy is critical to help expand activities which are doing well, and dive deeper into those which do not indicate any impact. Because the two interventions, SF and SF+LB, are disjointed, up to now we have estimated their treatment effects separately. However, to test whether the two treatment arms differ significantly and thereby uncover the impact of each of these interventions separately, we estimate them in the same equation as outlined below:

$$Y_{ist} = \alpha + \beta \text{Post} + \gamma_1 T_{\text{SF}} + \gamma_2 T_{\text{SF+LB}} + \Delta_1 T_{\text{SF}} * \text{Post} + \Delta_2 T_{\text{SF+LB}} * \text{Post} + X_{ist} + \mu_c + \varepsilon_{ist}$$

where:

- T_{SF} is an indicator variable with $T_{\text{SF}} = 0$ for comparison group and $T_{\text{SF}} = 1$ if the treatment group is SF
- $T_{\text{SF+LB}}$ is an indicator variable with $T_{\text{SF+LB}} = 0$ for comparison group and $T_{\text{SF+LB}} = 1$ if the treatment group is SF+LB
- $\Delta_2 - \Delta_1$ is the *additional* impact of the LB intervention
- All other variables are defined as before.

Exhibits 27 and 28 presents the results on letter recognition and reading comprehension, respectively, by estimating the above equation. The shaded row presents the *additional* impact of the LB intervention. That is, comparing the literacy outcomes at the students' level provides an unbiased estimate of the incremental two-year effect of the LB intervention, which will allow us to understand if the effect on letter recognition observed previously for SF+LB arose from SF only, LB only, or both.

For letter recognition (Exhibit 27), we found a large and statistically significant additional impact of the LB intervention. The LB intervention led to an additional impact of 39 to 42 percentage points across various samples ($p < 0.01$). Thus, the additional impact of the LB activities drives most of the impact of the SF+LB intervention on letter recognition, which we estimated in 5.1.2 Reading Comprehension as 42.7 percentage points. This means the LB activities account for more than 95 percent of the effect. This aligns with our qualitative data that noted the high praise from DESB, teachers, principals, and PAs regarding the various LB components (e.g. teacher trainings and reading camps). For reading comprehension (Exhibit 27), the LB intervention does not contribute any significant additional impact.

The results from this section point to the fact that SF alone might not lead to any meaningful impact on literacy outcomes, such as reading comprehension and letter recognition. In other words, consistent with the literature, although SF may increase students' attentiveness and attendance rate, e, SF may require an accompanying intervention to improve students' classroom performance.⁵⁷ Depending on the context, various factors could cause low literacy rates among school-aged children, which a SF intervention alone cannot address, such as limited experience of teachers. As shown in Exhibit 27 below, the impact on letter recognition seems to mainly come from LB activities.

Exhibit 27. Comparing Treatment Arms Letter Recognition

Letter Recognition						
Variables	Balanced Sample Coefficient (SE)		Unbalanced Sample Coefficient (SE)		Full Sample Coefficient (SE)	
	I	II	III	IV	V	VI
SF+LB*Post	0.428 (0.044)***	0.428 (0.045)***	0.407 (0.050)***	0.410 (0.048)***	0.404 (0.046)***	0.403 (0.045)***
SF*Post	0.032 (0.061)	0.025 (0.061)	0.027 (0.066)	0.024 (0.065)	-0.019 (0.059)	-0.015 (0.058)
Post	0.441 (0.040)***	0.432 (0.041)***	0.480 (0.047)	0.452 (0.044)	0.488 (0.043)	0.461 (0.041)
$\Delta_2 - \Delta_1$	0.39***	0.40***	0.38***	0.39***	0.42***	0.42***
N	2,082	2,040	3,225	3,110	3,841	3,720

*Source: Student survey; authors' calculations; p -value < 0.1 ** p -value < 0.05 *** p -value < 0.01 ; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

⁵⁷ Kremer, M., & Vermeersch, C. (2004). School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation

Exhibit 28. Comparing Treatment Arms - Reading Comprehension

Reading Comprehension						
Variables	Balanced Sample Coefficient (SE)		Unbalanced Sample Coefficient (SE)		Full Sample Coefficient (SE)	
	I	II	III	IV	V	VI
SF+LB*Post	0.041 (0.039)	0.038 (0.039)	0.043 (0.038)	0.041 (0.038)	0.039 (0.035)	0.038 (0.035)
SF*Post	0.023 (0.042)	0.022 (0.008)	0.024 (0.041)	0.025 (0.042)	0.017 (0.035)	0.018 (0.035)
Post	0.094 (0.031)***	0.098 (0.032)***	0.092 (0.030)***	0.092 (0.030)	0.095 (0.026)***	0.096 (0.026)***
$\Delta_2 - \Delta_1$	0.018	0.016	0.019	0.016	0.022	0.020
Controls	No	Yes	No	Yes	No	Yes
N	2,082	2,040	3,224	3,110	3,840	3,720

*Source: Student survey; authors' calculations; p -value < 0.1 ** p -value < 0.05 *** p -value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster-level fixed effects.

SECTION 6. LIMITATIONS

We note below a few limitations of the study that may be relevant to the final results:

Reliance on self-reported data. The quantitative approach relies entirely on self-reported data from children in Grade 2, who may not be emotionally and cognitively able to answer survey questions effectively. To minimize this limitation and improve data reliability, IMPAQ adapted the questions to the Lao context through cognitive interviews before baseline data collection (February 2017). In collaboration with CRS, the evaluation team conducted the cognitive interviews in school areas where students' main language at home was Lao and non-Lao to identify – in the local context and in different languages – what works, what does not work, and why. During the cognitive testing, the interviewers discussed the meaning of each item with Grade 1 and Grade 2 students to assess the clarity of the question and the appropriateness of the proposed categories. At the end, in consultation with our data collection partner, we adjusted phrasing to ensure children would understand the questions and feel comfortable answering. IMPAQ used the same instrument for the impact evaluation.

Tracking students at follow up. During baseline data collection (September, 2017), we reached a sample of 2,184 students in 45 SF schools, 59 SF+LB schools, and 55 comparison. In March and April 2019, we visited the same schools to follow up with students surveyed at baseline. With baseline students repeating their grade, dropping out of school, being absent, relocating to another school, and/or other reasons, we experienced a high attrition rate (overall 52 percent) from the baseline sample, which ranged from 44 percent in comparison schools to 51 percent in SF schools. We checked if this differential attrition is biased and performed extra analyses to minimize this concern. First, we ran a p-test on the level of attrition between each treatment arm and the comparison group, and the results were not statistically significant. We also checked the differences in background characteristics between tracked students vs. replacements (i.e. students surveyed only at follow up). For characteristics that were statistically different between these groups, we included them as control variables in our analysis to check the sensitivity of our estimates.

Absence of electronic class lists. The absence of electronic class lists left no option other than random sampling from students present at school on the day of data collection, rather than from the full classroom of enrolled students. While this approach ensures sampling consistency across schools (SF, SF+LB, and comparison schools) and achieves a random sample of students, the approach has two constraints. First, the possibility of systematic absences risks sampling bias for the original sample (tracked students) by selecting only present students. Such a bias could arise from excluding information on children who are more likely to be absent from school, including those from vulnerable socioeconomic backgrounds who may live far from schools and face difficulties commuting on muddy roads in rainy seasons; or children who may be absent due to health issues. Second, this challenge also exists while sampling replacement students from those students who were present on the day of data collection at follow up. However, we examine differences in characteristics of tracked and replacement students in Section 4. Evaluation Sample to understand if such selection is an issue. We include any observable characteristics that were significantly different between these two groups as controls in our main regression specifications.

Restricted sample. The proposed impact evaluation methodology samples from only Cohort 1 of the three LB cohorts that SCI will target over the life of the LEAPS II project. Two years may not be enough to see the effect of the literacy intervention, especially in communities with a higher proportion of ethnic students who only become exposed to Lao for the first time in school. However, focusing the impact evaluation on Cohort 1 does have advantages. Early assessment of attribution for observed changes in literacy to LB and/or SF after the full “dose” (i.e. the planned two years of support) will result in actionable

information at midterm that CRS and partners can use to inform mid-course corrections to improve and potentially accelerate literacy gains during the last half of the project.

Assumptions of the DID Model. Our impact evaluation design is a quasi-experimental DID methodology that rests on the assumption that in the absence of the intervention, Grade 2 students in the treatment and comparison groups would have experienced similar outcomes (“parallel trends”). This assumption is not directly testable with the current data because we would need several rounds of pre-program data on literacy outcomes, which is not available under the current SOW and budget. Moreover, if any external policy changed the probability of participation in the program differentially for schools, then the quasi-experimental methodology may not depict an accurate picture of the program impact on literacy. However, IMPAQ did check with CRS and SCI for any potential programs that may have been implemented from baseline to follow up. To the best of our knowledge, no other programs overlapped with LEAPS II implementation in Sepon and Atsaphone, and consequently we believe that these results do represent the true impacts of the SF and SF+LB interventions.

Spillover effects. The present study focuses only on the impact of the interventions on direct program participants. However, accounting for unintended effects on siblings of the program participants could potentially imply larger program effects on letter recognition. That is, it is possible that siblings within the same household as program beneficiaries may also benefit from the improved outcomes of their beneficiary siblings. For example, beneficiary students might help their non-beneficiary siblings with letter recognition.

SECTION 7. CONCLUSIONS

We conclude by providing a brief summary of the key regression analysis results and the robustness of these findings. As well, based on our results, detailed in Section 4. Evaluation Sample and Section 5. Results and the lessons learned from our experience in the field, IMPAQ provides our recommendations to improve the effectiveness of LEAPS II in the next half of program implementation.

7.1 Summary of Key Outcomes

Below we provide a snapshot of the key impact evaluation outcomes for the SF and SF+LB interventions on reading comprehension and letter recognition, controlling for any observed differences between each of the treatment arms and comparison group and across the three sub-samples (balanced, unbalanced, and the full sample). Section 5. Results provides more details.

Letter Recognition

- In the past two years, exposure to the SF+LB interventions increased students' proficiency in identifying letters by approximately 41 percentage points compared to the comparison group students, statistically significant at the 1 percent. However, the impact of the SF intervention alone on students' letter proficiency was minimal (3 percentage points) and not statistically meaningful.
- We also performed two sub-group analyses: girls vs. boys and Lao vs. non-Lao speakers. Impact results for the SF+LB intervention by gender are large, statistically significant, and consistent with the results for the overall sample ($p < 0.01$). Both girls and boys were more likely to identify letters by approximately 40 percentage points compared to their peers in the comparison group. The variation of the SF+LB impact on boys was a bit higher than girls across different sample types, which ranged from 40 to 44 percentage points. With respect to main language spoken at home, both samples experienced positive and statistically meaningful impacts compared to Lao and non-Lao speakers in comparison schools. However, the magnitude of the effect on letter recognition was larger for non-Lao speakers (approximately a 47 percentage-point increase) than Lao speakers (approximately a 32 percentage-point increase) compared to comparison group students ($p < 0.01$) in SF+LB schools. The sub-group analyses in SF did not show any meaningful impact on letter proficiency.

Reading Comprehension

- We did not find any evidence that either of the treatment arms (SF and SF+LB) had any statistically meaningful impact on reading comprehension of Grade 2 students. The magnitude of the effect was also trivial for SF (approximately a 4 percentage-point increase) and SF+LB (a 2 percentage-point increase) on reading comprehension compared to comparison group students. Results for all three samples (balanced, unbalanced, and the full sample) were consistent across the SF and SF+LB interventions.
- While we did not find a statistically meaningful impact of the program on reading comprehension of either girls or boys, the magnitude of the coefficient was higher for girls than boys across SF and SF+LB. The effects of SF and SF+LB were higher for non-Lao speakers compared to Lao speakers, however, the impact was not statistically significant for either of the language groups.

Comparing the Treatment Arms (SF vs. SF+LB)

- We also determined the effect of incremental impacts of LB by comparing the SF and SF+LB with each other. We noted a large and statistically significant impact of LB on letter recognition, but the results show no significant additional impact of the LB intervention on reading comprehension. The LB intervention leads to an additional impact on letter recognition of 39 to 42 percentage points across various samples ($p < 0.01$). Since the overall impact of SF+LB on letter recognition was approximately 41 percentage points, this implies that LB drives most of the impact of the SF+LB intervention on letter recognition.

7.2 Recommendations

We present the following recommendations to CRS based on lessons learned in the field and findings after analyzing the collected data. We divide this section into three parts:

- **Quick wins** are cost-effective recommendations for relatively simple problems, which can be implemented immediately and yield maximum impact.
- **Medium-effort** items require a modified evaluation budget and SOW, but are not necessarily as time or cost intensive as long-term strategy items. Medium-effort recommendations are important to consider for bridge or future projects.
- **Long-term strategy** recommendations may require more time, effort, and cost to implement, and therefore would be most useful to consider for future studies beyond LEAPS II.

7.2.1 Quick Wins

- **Reinforce the importance of education for *all*, especially for students from different ethnicities, in LB teacher training modules.** As mentioned in Overview of Lao Context, gender disparity and ethnic group disadvantages are known education issues at the national level in Lao PDR. Responses from participants in our qualitative interviews indicated that students who do not speak Lao as the primary language at home, which is the official language of instruction, may feel discouraged to continue their schooling, which results in dropouts. However, in the impact evaluation, the subgroup analysis by language shows a statistically significant higher effect of SF+LB on letter recognition for non-Lao speakers than for Lao speakers. Even for reading comprehension, the non-Lao sample witnessed a marginally significant impact. Thus updating the LB training materials to emphasize on training school staff and the community to make schools a more welcoming environment for *all* children (i.e. the objective of inclusive education component), but with the focus on marginalized students, non-Lao speakers, and girls, could result in even greater impact for those who usually fall behind.

7.2.2 Medium-Effort

- **Study the mechanism responsible for no effects of the program on reading comprehension, and unpack the main driving force behind the impact of the program on letter recognition.** CRS could consider a study to better understand the mechanism that hinders improvements in reading comprehension by speaking to teachers or teacher trainers and incorporate the results in implementation to the extent possible. In addition, since the impact evaluation estimates suggest the SF+LB interventions improved letter recognition but not reading comprehension, LEAPS II could consider unpacking specific LB activities to understand which LB activities have led to these changes. IMPAQ could study the effect of the various LB activities on additional outcome

variables, such as teacher trainings, reading camps, and/or school supplies. IMPAQ could collect additional quantitative information from students and teachers on these outcomes. Further, this kind of study would help future programming by pointing out which aspects of LB activities are least impactful, and thus should be modified for future iterations.

- **Consider alternative outcome indicators to study the effect of SF on students' schooling.** School meals did not seem to affect letter recognition and reading comprehension directly. Further, all impacts of the SF+LB intervention on letter recognition came from LB activities. These results reflect prevailing literature that indicate direct linkages between school meals and cognitive outcomes, including literacy gains, are rather complex; rather, SF may have more indirect effects on student cognition through increased school participation (attendance and attentiveness, two of the main outputs of LEAPS II according to the Theory of Change) and improved nutrition intake.⁵⁸ As well, SF programs are more likely to improve cognition in the presence of complementary inputs, such as teacher ability, instructional materials, and school infrastructure.⁵⁹ Qualitative findings and CRS' M&E data confirm that LEAPS II has progressed quite well on some of these mediating factors in the causal chain between SF and literacy, such as higher attendance. IMPAQ could conduct further investigation on the causal effect of SF on some of these intermediary outcomes, such as attendance, to better understand improvements on the extensive margin as a result of school meals.

7.2.3 Long-term Strategy

- **Modifying the longitudinal nature of the study.** The current impact evaluation was originally designed as a longitudinal quasi-experimental study (following the same students over time). However, studies have shown that longitudinal studies suffer from high attrition. In this study too, we could only track 48 percent of students from baseline to follow-up. Consequently, we recommend a repeated cross-sectional design in future studies (sampling a fresh group of students at each point in time). This will have two main advantages: (1) the study will not suffer from high attrition because a new group of students will be sampled each time and (2) the study will be less expensive since tracking the same student can be quite costly.

⁵⁸ Buttenheim, A., Alderman, H., & Friedman, J. 2011. Impact evaluation of school feeding programmes in Lao People's Democratic Republic. *Journal of Development Effectiveness*. 3 (4), 520-542. Adelman, S., et al., 2008. The impact of alternative food for education programs on learning achievement and cognitive development in northern Uganda. Washington, DC: The World Bank.

⁵⁹ Chakraborty, T. & Jayaraman, R. 2016. School Feeding and Learning Achievement: Evidence from India's Midday Meal Program. IZA Discussion Papers, No. 10086, Institute for the Study of Labor (IZA), Bonn, Germany

REFERENCES

- Catholic Relief Services. 2016. Learning and Engaging All in Primary School (LEAPS): Final Evaluation Report.
- Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences. Routledge. ISBN 1-134-74270-3.
- Dowd, A.J., Pisani, L. & Borisava, I. 2016. “Evaluating Early Learning from Age 3 to Grade 3” in Understanding What Works in Oral Reading Assessments. Montreal: UNESCO Institute for Statistics (UIS).
- FHi 360. 2016. Education and Policy and Data Center. Retrieved from: https://www.epdc.org/sites/default/files/documents/Laos_coreusaid.pdf
- Hess, R. D. & Holloway, S. D. 1984. Family and school as educational institutions. Review of Child Development Research, 7, 179-222.
- IMPAQ International. 2017. McGovern-Dole International Food For Education and Child Nutrition Program: Beoog Biiga II Midline Performance Evaluation Report.
- Itthida, Gnangnouvong. 2015. “Progressive Promotion”: Dropout Issue in Lao Primary Education and Misplaced Policy. Procedia – Social and Behavioral Sciences. Vol 197. P.2355-2361.
- Kim, Y. S. 2009. The relationship between home literacy practices and developmental trajectories of emergent literacy and conventional literacy skills for Korean children. Reading and Writing, 22(1), 57-84.
- Ministry of Education and Sports. 2013. Education Sector Development Plan (2011-2015), Review and Update: Final Report.
- Ministry of Education and Sports. 2015. Situational Analysis: Student Learning Outcomes in Primary Education in Lao PDR.
- Onphanhdala, P., and Suruga, T. 2008. Obstacles to Basic Education Development in Rural Lao PDR: Delayed Enrollment, Grade Repetition, and Dropout. Graduate School of International Cooperation Studies, Kobe University.
- RTI International. 2012. Early Grade Reading Assessment 2012: Lao PDR. Washington, D.C.: US Agency for International Development.
- World Bank. 2016. Reducing Early Grade Dropout and Low Learning Achievement in Lao PDR: Root Causes and Possible Interventions. Washington, D.C.: World Bank Group
- World Food Programme. 2018. Decentralized Evaluation of the USDA McGovern-Dole FY14 Endline Evaluation in Lao PDR [FY 14-16].

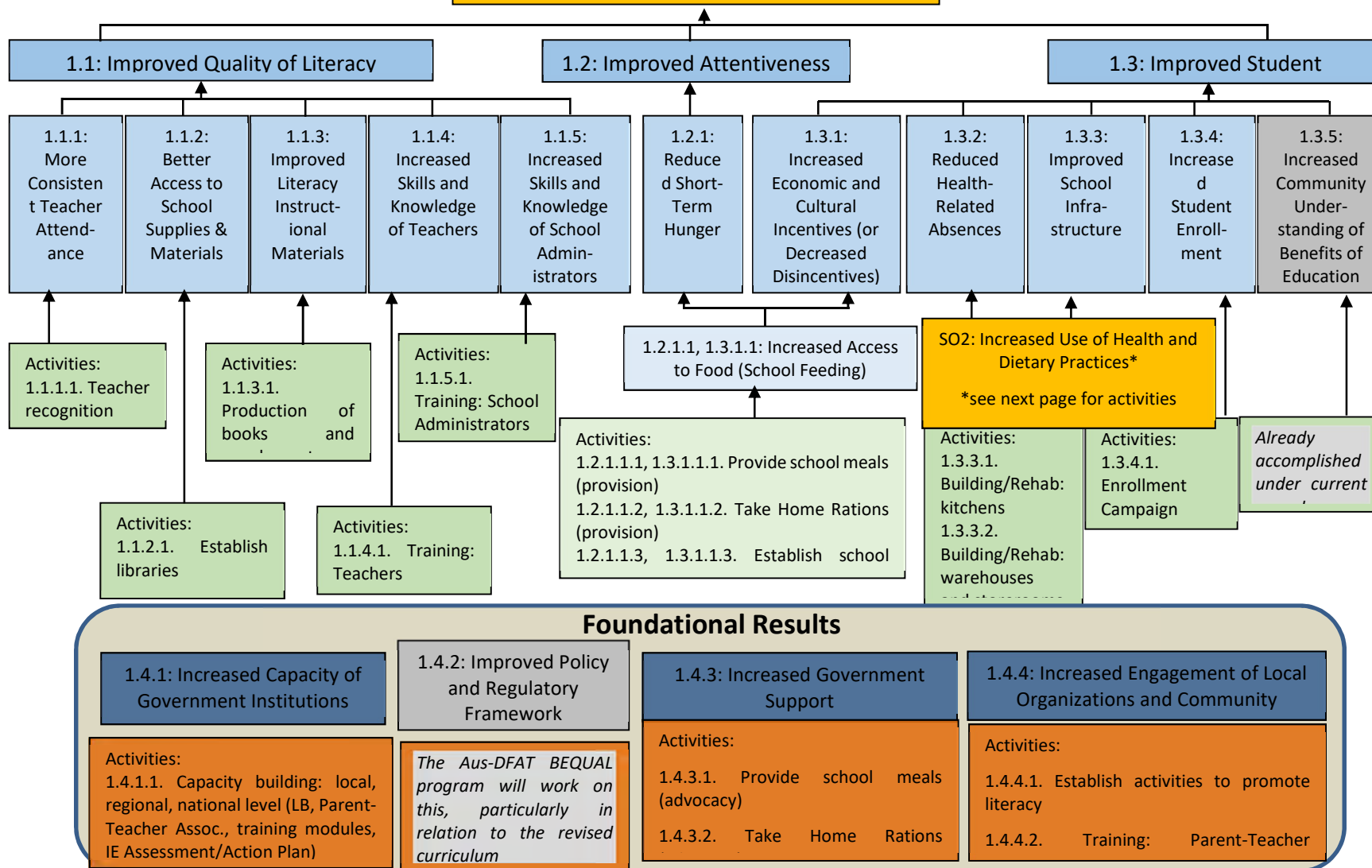
APPENDICES

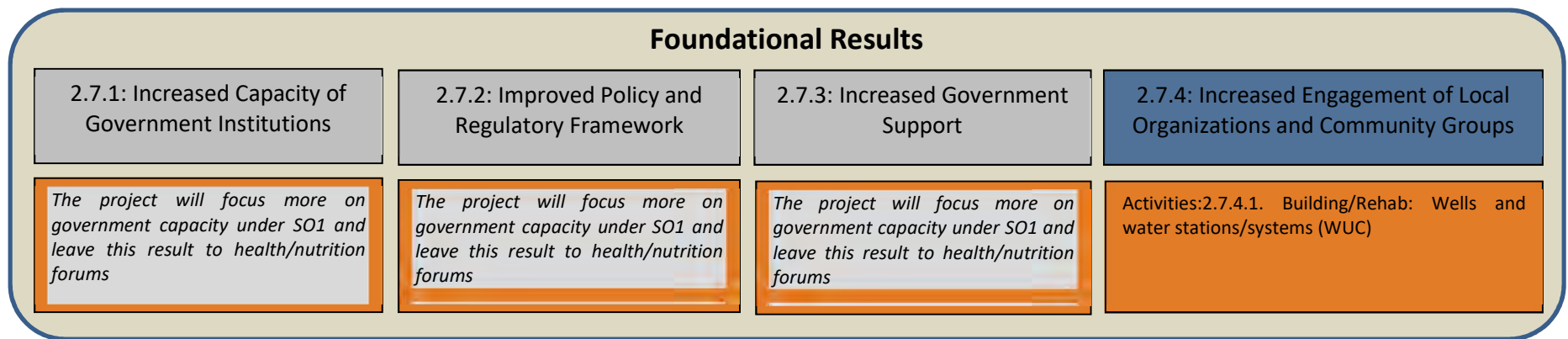
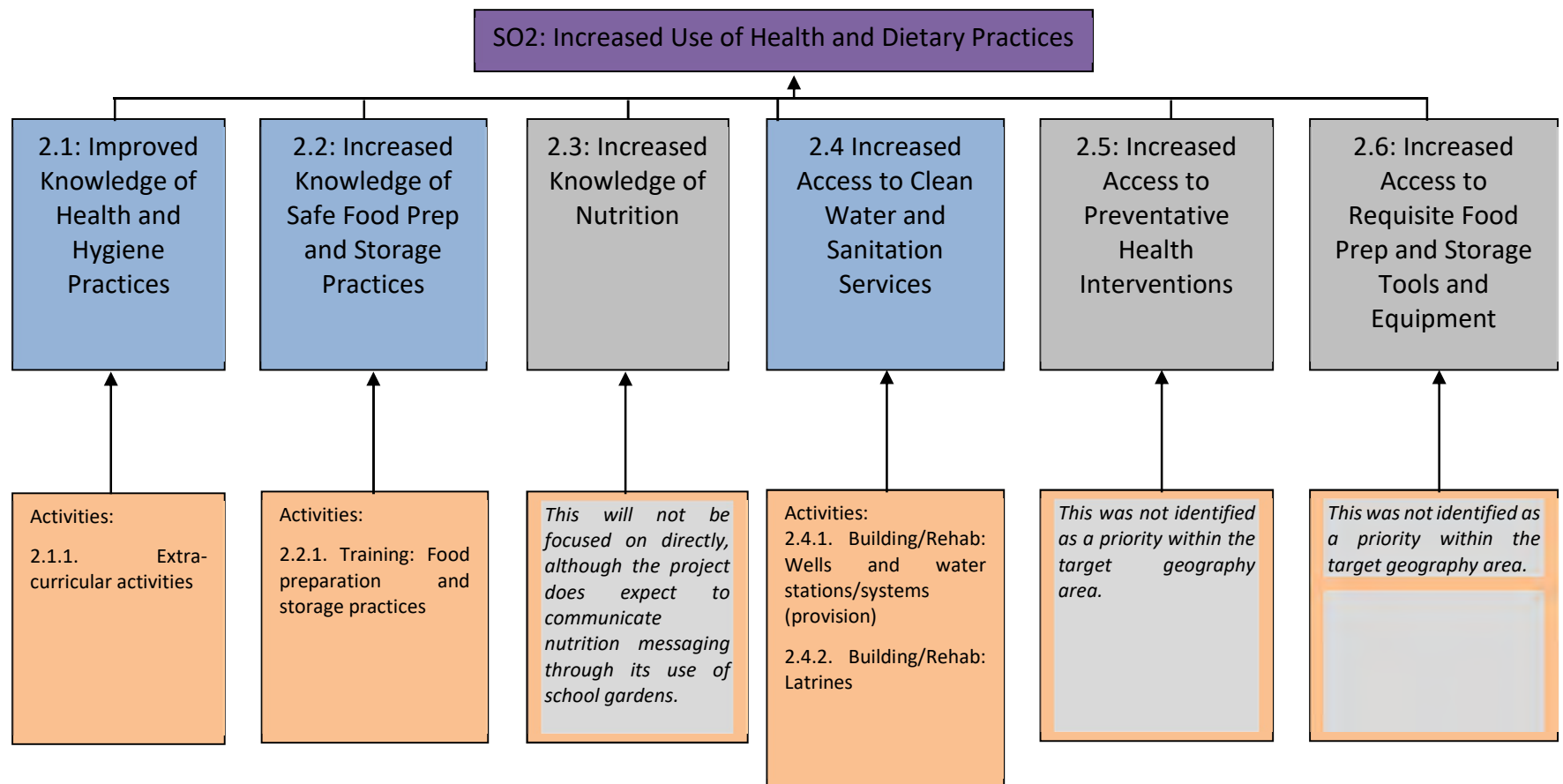
- A. LEAPS II Results Framework**
- B. School Lists by group**
- C. Additional Exhibits**
- D. Inter-rater Reliability**
- E. Survey Instrument**
- F. Evaluation SOW and Terms of Reference**

APPENDIX A. LEAPS II RESULTS FRAMEWORK

Exhibit 29. LEAPS II Conceptual Framework

SO1: Improved Literacy of School-Age





APPENDIX B. SCHOOL LISTS BY GROUP

REDACTED

APPENDIX C. ADDITIONAL EXHIBITS

C.1 Additional Balance Checks

Exhibit 30. Balance Tests between Tracked and Missing Students at Baseline

Background Characteristics	Tracked	Missing	Difference
Gender	49%	42%	-7%***
Age	6.6	6.9	-0.3**
Ever repeated a grade	25%	32%	-7%***
Attended ECD/pre-school	64%	44%	-20%***
Socioeconomic Index	4.6	3.5	-0.826***
Main language – Lao	11%	4%	-6%***
Main language – Phoutay	26%	15%	-11%***
Main language – Makong/Bru	12%	12%	0%
Main language – Tri	0%	1%	<1%

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 1,040$ for tracked and 1,143 for replacement. For age, $N = 390$ for tracked students and 303 for missing. Note: we only include languages with five or more speakers in the sample. We excluded those who selected "other" or "don't know".

Exhibit 31. Balance Tests between Missing at Baseline and Replacement Students at the Follow Up

Background Characteristics	Missing	Replacement	Difference
Gender	42%	45%	3%
Age	6.9	8.6	1.7***
Ever repeated a grade	32%	32%	0%
Attended ECD/pre-school	44%	45%	1%
Socioeconomic Index	3.5	4.1	0.6***
Main language – Lao	4%	27%	22%***
Main language – Phoutay	15%	19%	4%
Main language – Makong/Bru	12%	28%	15%***
Main language – Tri	0%	18%	18%***

Source: Student survey; authors' calculations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ in relation to the comparison group. $N = 1,143$ for missing and 616 for replacement. For age, $N = 303$ for missing and 403 for replacements. Note: we only include languages with five or more speakers in the sample. We excluded those who selected "other" or "don't know".

C.2 Robustness Checks

In this section we present three robustness checks for our main specifications, and throughout we make reference to certain exhibits found in Section 5. Results.

Check I: Controlling for Additional Covariates

We checked the sensitivity of the regression results to inclusion of other covariates, including age and prior exposure to Room to Read intervention in LEAPS I.

As shown in Exhibits 30 and 31, the age difference was statistically significant between various sample types (tracked vs missing students, and missing vs replacement). However, we did not include age in the main regression specifications because among all sampled students at baseline and follow up, 53 percent did not know their age. Thus, we left age out of the main list of covariates to maintain the power because sample size could be reduced considerably. However, we checked the sensitivity of the results by controlling for age.

Furthermore, in our impact evaluation sample, 26 schools were previously in the Room to Read education program under LEAPS I in Atsaphone and Sepon districts. All 26 schools received a school library, and three also received a full package of literacy activities, which included teacher training, school library, and in-class learning aids. To ensure that participation in those programs did not influence the impact results in LEAPS II, we included them as an additional covariate in LB+SF specifications and checked for the robustness of the results.

SF and LB interventions

Letter Recognition. For all three sample types, adding age as a covariate in estimating the effect of SF+LB on letter recognition reduced the magnitude of the treatment effect. We expected this finding because the sample size drops considerably once we include age since 53 percent of respondents did not know their age. However, the results remain positive and statistically significant. More specifically, for the balanced sample, after including age the coefficient on *Treatment*Post* is 0.347, which is marginally statistically different from the coefficient in Column (II) in Exhibit 15 (p-value = 0.0465); 2) for the unbalanced sample after including age, the coefficient on *Treatment*Post* is 0.307, statistically different from the coefficient in Column (IV) in Exhibit 16 (p-value = 0.0058); 3) for the full sample after including age the coefficient on *Treatment*Post* is 0.305, statistically different from the coefficient in Column (IV) in Exhibit 16 (p-value = 0.0025). However, in all three samples, the treatment effect continued to remain positive and statistically significant (p<0.01).

We also ran the specification after including exposure to Room to Read as a covariate. For all three sample types, adding Room to Read did not meaningfully change the magnitude or significance of the coefficients. More specifically: 1) for the balanced sample, after including exposure to Room to Read, the coefficient on *Treatment*Post* was 0.429; 2) for the unbalanced sample, the coefficient on *Treatment*Post* was 0.410, and 3) for the full sample the coefficient on *Treatment*Post* is 0.405.

Reading Comprehension. The results were not sensitive to inclusion of age in any of the three sub-samples, even with reduced sample sizes. More specifically: 1) for the balanced sample after including age, the coefficient on *Treatment*Post* is 0.019, and not statistically different from the coefficient in Column (II) in Exhibit 18 (p-value = 0.3479); 2) for the unbalanced sample after including age, the coefficient on *Treatment*Post* is 0.020, not statistically different from the coefficient in Column (IV) in Exhibit 18 (p-value = 0.1732); 3) for the full sample after including age, the coefficient on *Treatment*Post* is 0.021, not statistically different from the coefficient in Column (VI) in Exhibit 18 (p-value = 0.2759).

Our results are also not sensitive to the inclusion of exposure to Room to Read variable in any of the three sub-samples. More specifically: 1) for the balanced sample, after including exposure to room-to-read, the coefficient on *Treatment*Post* was 0.041; 2) for the unbalanced sample, the coefficient on *Treatment*Post* was 0.043, and 3) for the full sample the coefficient on *Treatment*Post* is 0.041.

SF intervention

Letter Recognition. Including age as a covariate to estimate the effect of SF on students' letter recognition for all three sample A, all coefficients continue to remain statistically indistinguishable from zero. More specifically: 1) the balanced sample after including age the coefficient on *Treatment*Post* is -0.077, marginally statistically different from the coefficient in Column (II) in Exhibit 21 (p-value = 0.0295); 2) the unbalanced sample after including age the coefficient on *Treatment*Post* is -0.066, statistically different from the coefficient in Column (IV) in Exhibit 21 (p-value = 0.0202); 3) for the full sample after including age the coefficient on *Treatment*Post* is -0.096, statistically different from the coefficient in Column (VI) in Exhibit 21 (p-value = 0.0198).

Reading Comprehension. Further, we performed robustness checks by adding age as a covariate in SF specifications for reading with comprehension. For all three sub-samples, even with reduced sample sizes, the results are similar to those reported in Exhibit 24 in 5.2.2 Reading Comprehension. More specifically for the balanced sample, after including age, the coefficient on *Treatment*Post*, is 0.002, and not statistically different from the coefficient in Column (II) in Exhibit 24 (p-value = 0.3223); for the unbalanced sample after including age the coefficient on *Treatment*Post*, is 0.010, and not statistically different from the coefficient in Column (IV) in Exhibit 25 (p-value = 0.3358); and 3) the full sample after including age, the coefficient on *Treatment*Post*, is -0.003, and not statistically different from the coefficient in Column (VI) in Exhibit 25 (p-value = 0.1502).

Check II: The First Difference Estimator

The first difference estimator is the equivalent of taking differences for the same observations over the two time periods. It compares changes over time for the same individual. Therefore, to implement the first difference estimator, it is critical to have data for two time periods for the *same* individual.

We estimate the first difference specification only for the balanced sample because the first difference estimator will give the same result as our main specification in the absence of attrition (i.e. balanced panel). However, if there is attrition, the first difference estimator will not be accurate since it will only use a subset of observations that are balanced. Thus, the fixed effects estimator used in the main body of the report has the advantage of being easily implementable for unbalanced samples. Hence, we present this estimator only as a robustness check for the balanced sample. The equation we estimate is:

$$\Delta Y_{is} = \alpha + \gamma \Delta T_s + \lambda \Delta X_i + \Delta \varepsilon_{is}$$

where

- ΔY_{is} is the change in the literacy outcome (e.g. reading comprehension and letter recognition) for student *i* in school *s* at time *t* from baseline to midterm.
- ΔX_i is the change in time varying covariates for student *i* in school *s* at time *t* from baseline to midterm.
- ΔT_s reflects the change in the treatment variable from baseline to midterm (i.e. at baseline *T* is zero for all).
- $\Delta \varepsilon_{is}$ is the change in the random shock from baseline to midterm for student *i* in school *s* at time *t*.

We present the results for the first difference estimator for the SF+LB intervention in Exhibit 31 and for the SF intervention in Exhibit 32. We find that for SF+LB (Exhibit 31), the first difference estimates are very similar to those presented in Exhibit 14 and Exhibit 17 for letter recognition and reading comprehension, respectively, for the balanced sample.

Similarly, for the SF intervention (Exhibit 32) the results for letter recognition and reading comprehension are almost identical to those presented for the balanced sample in Exhibits 20 and 23, respectively.

Exhibit 32. First Difference Estimator: SF+LB Intervention

	Letter Recognition		Reading Comprehension	
Treatment	0.468 (0.058)***	0.464 (0.576)***	0.046 (0.031)	0.037 (0.031)
Controls	NO	YES	NO	YES
N	736	701	736	701

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster-level fixed effects.

Exhibit 33. First Difference Estimator: SF Intervention

	Letter Recognition		Reading Comprehension	
Treatment	0.002 (0.063)	0.005 (0.064)	0.020 (0.037)	0.020 (0.037)
Controls	NO	YES	NO	YES
N	672	653	672	653

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster-level fixed effects.

Check III: Lagged Dependent Variable

We also checked the robustness of our findings to estimating an alternative specification to the DID approach in Section 5. Results. We follow recent literature that has shown that simply controlling for the baseline value of the outcome variable gives you more statistical power than using the DID approach.⁶⁰ This also gave us the added benefit of simultaneously examining time-invariant heterogeneity. Once again, this approach is only useful for the balanced sample as to implement this approach one needs to have baseline and follow-up information from the *same individual*. Consequently, we only present results for the balanced sample. We estimated the following equation:

$$y_{is2} = \alpha + \eta y_{is1} + \gamma T_s + \lambda X_{ist} + \varepsilon_{ist}$$

where

- y_{is2} is the value of the outcome at follow up
- y_{is1} is the value of the outcome at baseline
- T_s is an indicator variable which takes the value 1 for the treatment group and 0 for the control group at follow up and takes the value 0 at baseline for both treatment and control groups

We present the impact of the SF+LB intervention and SF interventions on letter recognition in Exhibits 35 and 36, respectively. We do not present results for reading comprehension because the baseline values for *all* students in the balanced sample were zero for reading comprehension. Hence, this specification will not be meaningful for reading comprehension. We find that the impacts for letter recognition for both the SF+LB and SF interventions are very similar to those from Exhibit 14 and Exhibit 20 in Section 5. Results, indicating that the results in Section 5. Results are very robust.

Exhibit 34. Lagged Dependent Variable: SF+LB on Letter Recognition

Variables	Coefficient (SE)	
	No Controls	Controls
Treatment	0.462 (0.063)***	0.478 (0.060)***
Baseline Value of Dependent Variable	0.0293 (0.052)***	0.238 (0.057)***
Gender	n/a	0.037 (0.029)
Main Language Lao	n/a	0.067 (0.040)*

⁶⁰ Teerenstra, Steven, Maud Graff, Esther de Hoop, and George F. Borm. 2012. "A simple sample size formula for analysis of covariance in cluster randomized trials." *Statistics in Medicine* 31 (20). doi: <https://doi.org/10.1002/sim.5352>.

Variables	Coefficient (SE)	
	No Controls	Controls
Socioeconomic Index	n/a	0.022 (0.009)**
Grade Repetition	n/a	0.032 (0.034)
ECD	n/a	0.059 (0.044)
N	736	728

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

Exhibit 35. Lagged Dependent Variable: SF on Letter Recognition

Variables	Coefficient (SE)	
	No Controls	Controls
Treatment*Post (DID)	-0.007 (0.071)	-0.004 (0.063)
Baseline Value of Dependent Variable	0.378 (0.056)***	0.320 (0.060)***
Gender	n/a	0.008 (0.035)
Main Language Lao	n/a	0.095 (0.048)*
Socioeconomic Index	n/a	0.025 (0.012)**
Grade Repetition	n/a	-0.010 (0.043)
ECD	n/a	0.071 (0.050)
N	672	664

Source: Student survey; authors' calculations; p-value < 0.1 ** p-value < 0.05 *** p-value < 0.01; Standard errors shown in parentheses are clustered at the school level. All regressions include cluster level fixed effects.

APPENDIX D. INTER-RATER RELIABILITY

Reading Assessment

To measure the reliability and level of homogeneity of enumerators' scores on children's literacy skills, 16 percent of the overall second grade sample (258 out of 1,635) were assessed by two different enumerators simultaneously. In other words, two enumerators surveyed one student at the same time: One enumerator was asking questions from the child, while the other enumerator was observing their survey and capturing the child's responses on his/her own tablet. The evaluation team used long one-way Analysis of Variance (ANOVA) techniques to determine whether the mean of a dependent variable is the same in two or more unrelated and independent groups, were used to calculate the intra-class correlation within pairs of assessors for a measure of inter-rater reliability. Adapted from Fleiss et al. (1973), we interpreted the intra-class correlations as it follows:

- Less than .40 – Poor
- Between .40 and .75 – Good or fair
- Greater than .75 – Excellent

Exhibit 28 shows the percent of agreement between the raters, as well as inter-rater reliability ratings for the project evaluation sample. Overall, the inter-rater reliability (IRR) across the project evaluation sample was excellent for every one of the literacy skill measures, showing high internal validity of the scores.

Exhibit 36. IRR by Literacy Skill Subtests for Impact Sample

Literacy Skill Sub-test	IRR	Rating
Letter Knowledge	98%	Excellent
Expressive Vocabulary	100%	Excellent
Word Pair	99%	Excellent
Matching	100%	Excellent
Most Commonly Used Words	100%	Excellent
Reader	99%	Excellent
Reader Comprehension	100%	Excellent
Listener Comprehension	100%	Excellent

Source: Student survey, IMPAQ calculation. N = 258 Grade 2 students

APPENDIX E. SURVEY INSTRUMENTS



DATA COLLECTION FOR USDA FOOD FOR EDUCATION (LEAPS II) IN LAOS

Student Survey

Start Time _____ Date _____

INTRODUCTION

districts	Enter the name of the district -----		
school name	Enter the school name -----		
motherconsent	Has the teacher given consent for her child to participate in this survey? 0. No consent → thank them and terminate the survey and select the next child on your list. 1. Teacher Consented → "timeofsuevey"	I__I	
 If teacher says No, thank them, and terminate the survey and proceed to the next child on your list.			
timeofsurvey	Is the survey administered in the... 1. Morning (before 12 pm) 2. Noon (between 12pm and 1pm) 3. Afternoon (after 1pm)	I__I	select only one option

Dear student:

Hi, my name is ____, and I am here asking some questions from children like you to understand more about the LEAPS project also known as the Primary School Lunch Program for Small Children. Nothing you say here will be repeated to your parents or teacher will be kept a secret. There aren't any right or wrong answers. I want you to answer honestly and as best as you can. Do you have any questions for me? You can interrupt me to ask a question at any time. Also, if you don't know the answer to a question or don't want to answer it, just let me know and we can skip it. Are you ready to begin?

assent	Do you accept if I ask you some questions? 0. No → thank him/her, terminate the survey and proceed to the next child on your list. 1. Yes → continue with the background section.	I _ I	
 If child says No, thank him/her, terminate the survey and proceed to the next child on your list.			

Student Code:

Studentcode Please get the student code from the team leader. It is very important to use the correct student code, so please enter the code twice. If you are unsure, please check again with the team leader			
Track	Is this student: 0. A new student 1. Selected from tracking list	I _ I	
Stcode1	Please enter the student code CAREFULLY		Record student code
Stcode2	Please enter the student code CAREFULLY again		Record student code
reliab	Is this an individual assessment or a pair assessment? 0. Individual → “nickname” 1. Pair assessment → “reliabtype”	I _ I	
reliabtype	Talking enumerator or observing enumerator? 1. Talking 2. Watching	I _ I	

Background information

nickname	What is your nickname?
fname	What is your first name?
mothername	What is your mother's name? -----

fathername	What is your father's name? -----		
gender	1. Male 2. Female	I__I	*Ask only if necessary
age	How old are you?		*RECORD AGE >=5 & <17 If the child doesn't know her/his age enter 999
newsch	When you started at this school, which grade were you in? 0. Preschool 1. Grade 1 2. Grade 2 3. Grade 3 4. Grade 4 5. Grade 5 999. Don't know	I__I	*Select only one option
grade	Which grade/class are you in? 1. Grade 1 2. Grade 2 3. Grade 3 4. Grade 4 5. Grade 5	I__I	*Select only one option
everrpt	Did you repeat any grades? 0. No 1. Yes 999. Don't know	I__I	*Select only one option
Everrpt_b	Which grade had you repeated? 1. Grade 1 2. Grade 2 3. Grade 3 4. Grade 4 5. Grade 5	I__I	*Select all that apply
Everrpt_c1	How many times did you repeat Grade 1?	...	Enter the frequency

Everrpt_c2	How many times did you repeat Grade 2?	...	Enter the frequency
Everrpt_c3	How many times did you repeat Grade 3?	...	Enter the frequency
Everrpt_c4	How many times did you repeat Grade 4?	...	Enter the frequency
Everrpt_c5	How many times did you repeat Grade 5?	...	Enter the frequency
ecdattended	Did you attend ECD/preschool? 0. No 1. Yes 999. Don't know		*Select only one option
mainlang	What language do you often speak at home? 1. Lao 2. Phoutay 3. Makong/Bru 4. Tri 5. Taoy 6. Katang 7. Thai 8. Other 999 Don'tKnow/No response	__	*Select only one option
otherlang	At home, do you speak any other languages? 1. Lao 2. Phoutay 3. Makong/Bru 4. Tri 5. Taoy 6. Katang	__ __ __ __ __ __	*Select all that apply

	7. Thai 8. Other (Specify) 9. None 999 Don't know/No response	☐	
ses	Does your home have the following: 1. Mobile 2. Electricity 3. Refrigerator 4. Bicycle 5. TV 6. Motorbike 7. Car 8. Tractor [TOK TOK]) 9. None 999. Don't know	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	*Please read all the options to the child and select all that apply
book	At home do you have : 1. Textbooks 2. Magazines 3. Newspapers 4. Storybooks/COMICS 5. Coloring an drawing books 6. None of these 999. Don't know	☐ ☐ ☐ ☐ ☐	*Please read all the options to the child and select all that apply

HEALTH

Thank you! Now, I would like to ask you a few questions about your health.

health1	Have you been sick anytime during the last week? 0. No → "health4" 1. Yes → "health1a" 999. Don't know	☐	*Select only one
----------------	--	--------------------------	------------------

health1a	What kind of sickness did you have? 1. Sore stomach 2. Fever 3. Headache 4. Tiredness 5. Coughing 6. Other 999. Don't Know	_ _ _ _ _ _	*If the child cannot name the sickness, ask them to describe the symptoms for you and select the right answer. *Give examples for a sore stomach such as food poisoning or diarrhoea *Select all that apply
health2	During last week, did you miss school because you were sick? 0. No → "health4" 1. Yes → "health3" 999. Don't know	_	*Explain that this means they stayed at home *Select only one
health3	How many days did you miss school because you were sick during the last week? 1. 1 2. 2 3. 3 4. 4 5. 5 → fs3 999. Don't Know	_	*Select only one
Health4	During the last week did you miss school for any other reasons? 0. No 1. Yes 999. Don't know	_	*Explain to student if necessary: When the school was open, not because the school was closed *Select only one

FOOD SECURITY

Thank you! Now, I would like you to think about all the meals you had today.

fs3	Did you eat something for breakfast today? 0. No → "fs5" 1. Yes → "fs4a" 999. Don't know	_	*Select only one
fs4a	Did you feel full after eating breakfast? Or could you have eaten more? 0. I felt full → "fs5"	_	*Select only one

	1. I could have eaten more → “fs4b”		
fs4b	Why didn't you eat more in the morning? 1. There wasn't any more food 2. There was nothing I liked 3. Other	I__I	*Select only one
The next set of questions will be asked if “timeofsurvey” is equal to 2 or 3, else go to “fsx3”.			
fs5	Has the school lunch meal already been served today? 0. No → “fs6” 1. Yes → “fsX1” 999. Don't know	I__I	*Select only one
fs6	Did you already eat something for lunch today? 0. No → “fsX3” 1. Yes → “fsX3” 999. Don't know	I__I	*Select only one
fsX1	Did you already eat the school meal? 0. No 1. Yes 999. Don't know	I__I	*Select only one
fsX2	Did you like the taste of the school meal? 0. Not at all 1. A little bit 2. Somewhat 3. Yes, very much	I__I	*Select only one
fx2b	Why didn't you eat the school meal? 1. I was sick 2. I was playing 3. There was not enough food 4. I didn't like the taste 5. Other 999. Don't know	I__I	*Don't read the options to the child *Select only one
fsX3	How hungry do you feel right now? 0. Not at all hungry 1. A little hungry 2. Somewhat hungry 3. Very hungry	I__I	*Select only one

	999. Don't know		
--	-----------------	--	--

Household Environment

Very good! We have a few more questions about your home.

nhhold	How many people are there in your household, including yourself?		*Record the number > 0 & < 21 *Mark 999 if no response/don't know
seeread	During the last week, did you see anyone in your house reading? 0. No → "helpstudy" 1. Yes → "nseeread" 999. Don't know	I__I	*Select only one
nseeread	Who did you see reading last week? 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 999. Don't know	I__I	(enter the number of household members the child counted) *If the child counted more than 6 select 6
helpstudy	During the last week, did anyone in your house encourage you to study? 0. No → "story2u" 1. Yes → "nhelpstudy" 999. Don't know	I__I	Select only one option
nhelpstudy	Who encouraged you to study last week? 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 999. Don't know	I__I	(enter the number of household members the child counted) *If the child counted more than 6 select 6

story2u	During the last week, did anyone in your house tell you a story? 0. No → "read2u" 1. Yes → "nstory2u" 999. Don't know	I__I	Select only one option
nstory2u	Who told you a story in the past week? 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 999. Don't know	I__I	(enter the number of household members the child counted) *If the child counted more than 6 select 6
Read2u	During the last week, did anyone in your house read to you? 0. No → "hhenviro3" 1. Yes 999. Don't know	I__I	*Select only one option
nread2u	Who read to you last week? 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 999. Don't know	I__I	(enter the number of household members the child counted) *If the child counted more than 6 select 6
Hhenviro3	Are your parents or family members asking you questions about the stories they tell you or read to you? 0. No 1. Yes 999. Don't know	I__I	Select only one option Only ask if the answer to read2u OR story2u is yes
readout	During the last week, did you read outside of school? 0. No 1. Yes 999. Don't know	I__I	select only one option

SCHOOL ENVIRONMENT AND PARTICIPATION

We are almost done! Now, I would like to ask you some questions about your school.

enviro1	Do you like coming to school? 1. No 2. Yes 999. Don't know	I__I	*Select only one option
enviro2	What do you like about your school? 1. Reading 2. Writing 3. Participating in class activities/games 4. Dancing 5. Teacher reading 6. Being with my friends 7. Working in groups 8. Answering questions 9. Presenting ideas in class 10. Food is provided 11. Physical education/sports at school 12. Learning new things 13. other 14. Nothing 999. Don't know	I__I I__I I__I I__I I__I I__I I__I I__I I__I I__I I__I I__I	*Don't give examples or read the list to the respondent *Select all that apply
Skip enviro 6, enviro4, and enviro7 if health3 = 5 (missed school for 5 days)			
enviro6	How often in the last week did the teacher tell a story or read a poem to the classroom which was not in the textbook? 1. Every day 2. A few times during the week; 3. Once during the week; 4. Never → "enviro7" 999. Don't know	I__I	*Read the list to the respondent, but don't read 'don't know' *Select only one
enviro4	How often in the last week did the teacher ask you about the story s/he told or the poem s/he read during class? 1. Often 2. Sometimes 3. Rarely 4. Never 999. Don't know	I__I	*Read the list to the respondent, but don't read "don't know" *Select only one

enviro7	How often in the last week did you play a game in the classroom around the alphabet or a reading activity? 1. Every day 2. A few times during the week; 3. Once during the week; 4. Never 999. Don't know	I__I	*Read the list to the respondent, but don't read 'don't know' *Select only one
enviro10	Does your school have story books other than textbook for you to take home and borrow? 0. No → expressvocab1 if the child is in second grade; otherwise thanks the child and terminate the survey. 1. Yes → enviro10a 999. Don't know	I__I	*Select only one
Enviro10a	How often in the last week did you borrow story books other than textbook from school to take home to read? 1. Every day 2. A few times during the week; 3. Once during the week; 4. Never 999. Don't know	I__I	*Read the list to the respondent, but don't read 'don't know' *Select only one
 If child is not in grade 2 thank him/her and terminate the survey. If s/he is in grade two continue with expressive vocabulary			

LITERACY BOOST ASSESSMENT:

Expressive vocabulary

Now let's try a word game. Imagine you are going to the market and name some foods that you can eat. Try to name as many things as you can think of.

Record the number of items the child lists until the child has listed 10 items. You can tally on the score sheet as the child enumerates the objects.

When the child pauses for 5 seconds or more, PROMPT ONCE by saying, Can you think of any others?

When the child cannot think of more items, move on to the next question and say:

Now, I want to know what animals you are familiar with. Tell me the names of some animals that you know. Try to name as many animals as you can think of and I will keep count again.

When the child pauses for 5 seconds or more, PROMPT ONCE by saying, Can you think of any others?

expressvocab1	Can you tell me the names of things you can eat in Lao? (Specify the number of items child says they can eat in Lao 0-10) 0. 0 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 7. 7 8. 8 9. 9 10. 10	I__I	*Select only one option
Expressvocab2	Can you tell me the names of animals in Lao? (Specify the number of animals a child counted in Lao 0-10) 0. 0 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 7. 7 8. 8 9. 9 10. 10	I__I	*Select only one option

Letter Sounds

Now we will play a listening game. This one is about the sounds in words. The word “cat” starts with /c/ (*Say the sound, not the letter name*). /c/ is the first sound in cat. Now listen to the words I say and tell me which one starts with the same sound, the sound /c/ (*Say the sound, not the letter name*) star, ball, or cup?

If the child gives an incorrect response, say: cup starts with /c/ just like cat.

Wordpair1	I will read three words to you with the sound (“C” Child matches the letter “C” and the correct word) 0. Not able to match/Don’t know 1. Able to match 999. Did not understand the question	I__I	ຫມາ, ແມວ, ເດືອນ
------------------	---	------	-----------------

Wordpair2	I will read three words to you with the sound "ຕ" -child matches the letter "ຕ" with the correct word 0. Not able to match/Don't know 1. Able to match 999. Did not understand the question	I__I	ໄມ້, ກ້ວຍ, ຕົ້ນ
Wordpair3	I will read three words to you with the sound "ມ" - Child was able to matches the letter "ມ" with the correct word 0. Not able to match/Don't know 1. Able to match 999. Did not understand the question	I__I	ໄກ່, ຫນູ, ມ້າ

Understanding Letters

1. Give the child the list of letters and say to the child:
2. Say: *Let's look at some letters. Can you start here* (point to first letter) *and tell me what these letters are moving in this direction?* (indicate left to right direction) *Do you understand? Ok, you can begin.*
3. Correct letters are:
 - the letter name in the home language or language of instruction
 - any sound that is acceptable for in the home or instructional language
 - a response which says "It begins like..." giving a word for which the letter is the initial letter
5. If the child read the letters out of order, then remember to bring his/her attention to the ones they might have skipped.
6. Make sure you marked all of the incorrect letters
7. Move to the Most Used Words section.

What to do if a student is struggling:

- If the student is struggling, and hesitates at any letter for five seconds, ask follow-up questions: *Do you know its name? What sound does it make? Do you know a word that starts with this letter?*
- If the student still hesitates for five seconds, ask: *Can you tell me any of these letters?*
- If the student still hesitates for five seconds, then stop and thank him/her for trying his/her best.
- Mark letters not identified or not attempted as incorrect.
- Move to the Most Used Words section.

ຈ	ນ	ຮ	ຊ	ຜ
ມ	ຂ	ພ	ຟ	ງ
ດ	ອ	ຫ	ທ	ລ
ປ	ຍ	ກ	ວ	ຖ
ຄ	ຮ	ບ	ສ	ຜ
ຕ	ຢ	ຫວ	ຫງ	ຫຍ
ໄລ	ຫຼ	ໄນ		

Most Used Words

1. Give the pupil the laminated copy of the "Most Used Words" list.

2. Say: *I would like you to read some words to me. They are words from your textbook. Please point to and say each of these words starting here (point to first word) and moving across each line like this (indicate left to right direction). Do you understand? Ok, you can begin.*
3. Remember that pronunciations of words in local dialects are acceptable.
4. If the child read the words out of order then remember to bring his/her attention to the ones they might have skipped.
5. Make sure you marked all of the incorrect words.
6. Move to the Matching Section.

What to do if a student is struggling:

- If the student is struggling, and hesitates at any letter for five seconds ask the child, *Are there any words on the list that you know? Tell me or say the words you know.* Repeat the request to encourage the child to continue.
- If the student still hesitates for five seconds, then stop and thank him/her for trying his/her best.
- Mark words not identified or not attempted as incorrect.

Move to the Decoding section.

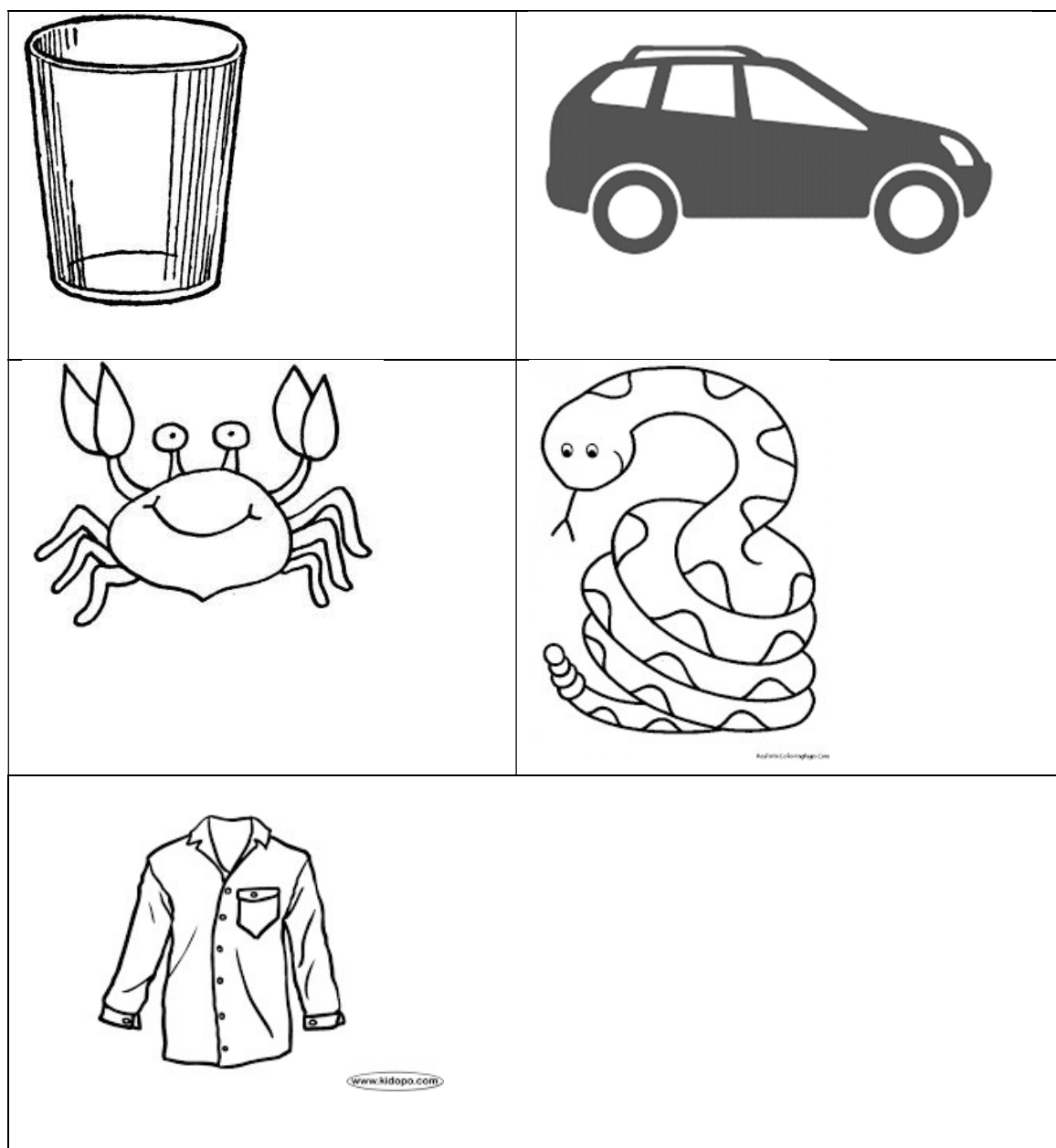
ລົດ	ນາ	ງູ	ຈານ
ມາ	ກອບ	ຢູງ	ກະບູ
ໄຟ	ປຶ້ມ	ຕະຫຼາດ	ອ້າຍ
ແຂງແຮງ	ເສື້ອ	ອະນຸຍາດ	ໂສ້ງ
ອະນາໄມ	ແມ່	ຕັ້ງ	ເດືອນ

Matching 1

Point to the words and ask to the child which of these pictures the words mean. If the child cannot match the first three words then mark everything incorrect and move to the next test.

Practice: Car → Point at the word for "car". Then point at the picture of the car. Ask if the child understands.

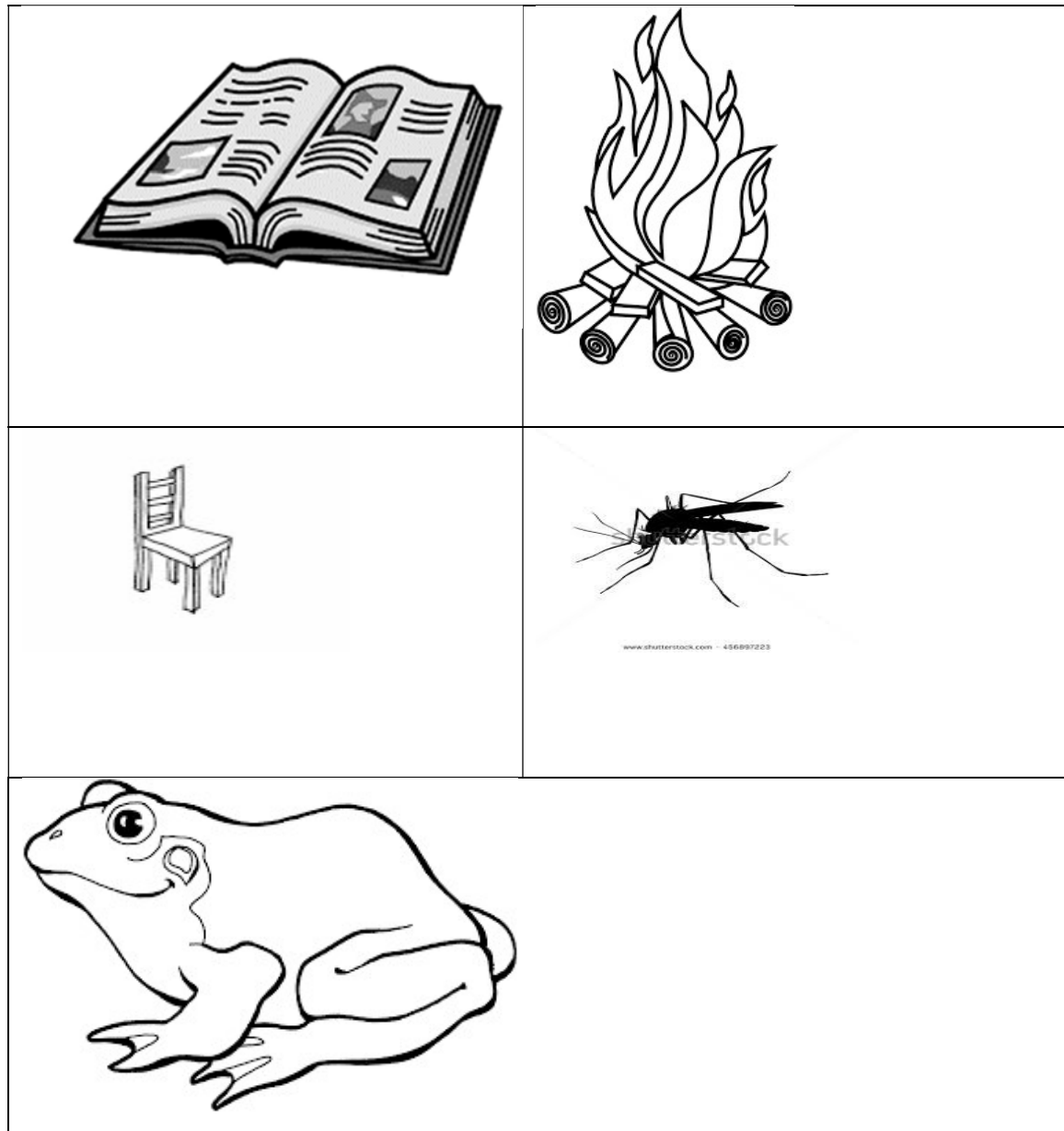
shirt, snake, glass, crab



Matching 2

Point to the words and ask to the child which of these pictures the words mean. If the child cannot match the first three words then mark everything incorrect and move to the next test.

frog, mosquito, chair, fire, book



COMPREHENSION PASSAGES AND QUESTIONS

The mother chicken teaches her chick.

One day when the sky was transparent and clear, the sunlight was shining down on the ground and the cool wind sometimes blew. A flock of little chicks running under their mother's tail happily went to find

food. The mother chicken ordered her children “Don’t you go far away from me because a hawk will come and take you to eat. You won’t see your mother again!” One of the chicks went to peck food by itself and became separated from the group and was alone. The chicken felt afraid and tried to call for its mother. At the same time a hawk approached flying in the sky which made the chick even more scared. Later when the mother found the lost chick she took them all to safety.

reader	Is child a reader or a non-reader? 0. A non-reader read less accurate than 5 per 30 seconds) 1. A reader (read correctly 5 per 30 seconds)	I__I	Select only one option
---------------	--	------	------------------------

Comprehension Questions

Comp1	What happened in the story? 1. mother chicken and baby chickens go out to find food 2. mother chicken tells baby chickens to stay close to her 3. one baby wanders away from the flock and is scared of the hawk 4. in the end, the mother chicken finds the baby and brings him home safely 5. None	I__I I__I I__I I__I	mark every main point mentioned by the child
Comp2	How many baby chickens were lost? (one chicken) 0. False 1. True	I__I	Don’t read the answer to them
Comp3	Why did the baby chicken get lost? (a chicken finds food alone) 0.	I__I	Don’t read the answer to them
Comp4	Why was the lost chicken frightened? (The baby chicken saw a hawk flying in the sky/separated from her mother) 0. False 1. True	I__I	Don’t read the answer to them
Comp5	Did the mother chicken find the lost chicken? (Yes) 0. False 1. True	I__I	Don’t read the answer to them
Comp6	What could happen if the hawk catches the baby chicken? (the chicken will be hawk’s food/chicken has died/ can’t find mother chicken)	I__I	Don’t read the answer to them

	0. False 1. True		
Comp7	Where did baby chicken find food? (Somewhere around rice field) 0. False 1. True	I__I	Don't read the answer to them
Comp8	Why should baby chicken not look for food by themselves? 1. Student could explain their answer with information from the story 2. Student count NOT explain their answer with information from the story	I__I	Don't read the answer to them
Thank you very much for answering my questions.			

End Time, Date _____

APPENDIX F. IMPACT EVALUATION SOW AND TOR

Learning and Engaging All in Primary School (LEAPS) Phase II

USDA – McGovern-Dole International Food for Education and Child Nutrition Program

2016 – 2021 Laos PDR

Project Midterm Evaluation:

Scope of Work and Terms of Reference

October 2018

The **purpose** of this terms of reference (TOR) is to describe the objectives and minimum methodological requirements for midterm performance and impact evaluation of the Learning and Engaging All in Primary Schools Project Phase II (LEAPS II). Also included is a description of the scope of work (SOW) for an experienced, independent third-party consultant or firm to conduct the project's baseline, midterm and final evaluations. CRS reserves the right to terminate its relationship with the third-party evaluator if the baseline performance or impact study and/or mid-term performance or impact evaluations are not satisfactory. CRS will re-engage in a competitive recruitment process if the first evaluator does not meet expectations.

BACKGROUND

Catholic Relief Services (CRS) is in the second, 5-year phase (2016 – 2021) of the USDA-funded Learning and Engaging All in Primary School project as part of the McGovern-Dole (MGD) International Food for Education (FFE) and Child Nutrition Program. LEAPS II is valued at \$27 million and will be implemented by a consortium of CRS, government partners, sub-recipient Save the Children International (SCI), and technical partner University of Oregon (UO), with the overarching objective of improving the literacy skills of approximately 86,049 new and continuing students in 350 schools across 7 educationally-disadvantaged districts of Laos' Savannakhet Province. To support improved literacy, LEAPS II will ensure schools have access to resources and community support to 1) improve quality of instruction with an intensive model of training and accompaniment of school staff, 2) increase enrollment of children in schools through greater inclusion of children with disabilities (CWD) and other traditionally-excluded groups, and 3) maintain strong attendance by ensuring a healthy and hunger-free learning environment for all students.

CRS will build on gains achieved in LEAPS I (2012 – 2016) which, with support from USDA/MGD-FFE, reached over 36,000 students from 310 communities in Savannakhet Province, increased boys' and girls' attendance by 38% and enhanced reading skills through approaches that empower government, parents, teachers, and community leaders. LEAPS II will continue effective interventions from LEAPS I, while also incorporating new strategies built on lessons learned and identified needs to: 1) expand project coverage to approximately 3,400 additional students in 45 newly targeted schools in educationally-disadvantaged Nong district; 2) intensify and scale efforts to develop pedagogical skills of teachers and administrators in literacy instruction through the Literacy Boost (LB) program; 3) deepen focus on building a supportive community environment for language and learning; 4) strengthen capacities of government, school, and community actors in inclusive education (IE) practices and techniques; and 5) improve access to water and sanitation in target schools. CRS has designed LEAPS II with a focus on sustainability and will work with government and community stakeholders to phase-over key interventions, strengthening key capacities, commitments, and resources to support continued success of students in Savannakhet following FFE's exit.

The LEAPS II results framework detailing the project's strategic objectives, intermediate and foundational results is available in [Annex 1](#).

PURPOSE AND SCOPE OF THE MIDTERM EVALUATION

Purpose and Scope: The purpose of the midterm evaluation is to critically and objectively review and take stock of the project's implementing experience and the implementing environment; assess whether targeted beneficiaries are receiving services as expected, assess whether the project is on track in meeting its stated goals and objectives, assess the impact of the Literacy Boost and school feeding interventions on literacy outcomes, review the project-level results frameworks and assumptions, document initial

lessons learned, and discuss mid-course corrections that may be necessary to effectively and efficiently meet the stated goals and objectives (*USDA/FAS M&E Policy, 2013*).

Midterm Evaluation Questions

The midterm evaluation will assess the relevance, effectiveness, efficiency, sustainability, and impact of the key project interventions. Key evaluation questions for the MTE are detailed below:

Relevance

- Do stakeholders feel that their voices are heard and their needs considered throughout the project? Are stakeholders aware of how to share information with CRS?
- To what extent is school feeding a relevant incentive for parents to send their children to school?
- How well do teacher and administrator trainings organized by the LEAPS project support teachers to address the issues they face in their schools and communities? Are the topics being offered relevant to their needs?
- To what extent are the trainings organized by the LEAPS project in health/hygiene, safe food prep/storage practices, and clean water/sanitation services addressing pre-existing knowledge gaps as identified by CRS?
- Did the project staff properly identify school/community priority needs? To what extent has the program design and intervention met the needs of the participants, including students, teachers, school administrators, and parents?
- To what extent are project activities aligned with Government of Laos priorities and policies? If they contrast, why?
- According to parents and students, how appropriate/acceptable are the food items, the timing of the meal, and the food preparation?

Effectiveness

- Is there evidence that the training of teachers led to improved teaching practices? To what degree are objectives related to improved quality of instruction likely to be achieved by the end of the project?
- Is there evidence that capacity strengthening activities contributed to the implementation of IE policies and programs?
- Is there evidence that WASH interventions contributed to changes in the use of health and dietary practices? Is there evidence that the objective of increased use of health and dietary practices likely to be achieved by the end of the project? Is there evidence that schools demonstrating improved knowledge of health and hygiene practices? Is there evidence that the objective of increased use of health and dietary practices likely to be achieved by the end of the project?
- Is there evidence that cooks and storekeepers displaying improved knowledge of safe food preparation and storage practices? To what degree is the objective of increased use of health and dietary practices likely to be achieved by the end of the project?
- Is there evidence that schools achieving increased access to clean water and sanitation services? To what degree is the objective of increased use of health and dietary practices likely to be achieved by the end of the project?
- To what extent is school feeding an effective incentive for parents to send their children to school? What has been the resulting effect on attendance?
- Is there evidence that school feeding improve student attentiveness? To what degree is the objective of improved attentiveness likely to be achieved by the end of the project?
- In what ways is the project increasing the capacities of MoES at various levels, including teachers, administrators, and VEDC/communities?

- To what extent is learning occurring between the various partners? What is the value-added of the government-NGO-academic partnership model? Are the learning mechanisms effective in facilitating learning between all?

Efficiency

- How efficient are the management and staffing structures for project implementation, monitoring, learning, and timely decision-making?
- Was the monitoring system designed efficiently to meet the needs and requirements of the project?

Sustainability

- What types of support do communities, including parents and VEDC members, provide to support schools, including school meals? Based on current practice, which type(s) of support are most commonly provided by communities?
- What mechanisms are in place for the local stakeholders (including VEDC) to continue school feeding after the duration of the project, and with high standards for health/dietary practices? How well did the VEDCs and cooks incorporate locally-available foods into school meals? What mechanisms are and/or need to be in place for MOES staff to support and scale-up literacy education in LB schools upon completion of project support?
- What, or who, incentivizes VEDCs and water user committees (WUCs) to sustain/maintain school latrines, water points, and handwashing facilities?

Impact

- What have been the consequences of the program, in terms of both intended and unintended results and changes? Did these results vary for different groups, communities, schools, or households (e.g. boys vs. girls, ethnic vs. Lao communities, etc.) within the targeted areas? If so, why?
- What changes in literacy and drop-out rates have been observed as a result of the Literacy Boost and/or school feeding interventions? To what extent are these changes attributable to Literacy Boost program and/or school feeding?
- Did school feeding and educational quality interventions have the intended impact on attendance, attentiveness, quality of instruction, and literacy of school-age children?

EVALUATIONS APPROACH AND METHODOLOGY

LEAPS II will use both quantitative and qualitative methods to conduct the midterm performance and impact evaluations. This mixed-methods approach is replicated from baseline and will again be replicated for the project's final performance evaluation in project year 5. Surveys associated with the baseline study established baseline values for key performance indicators. Table 1 summarizes both the survey-based indicators assessed at baseline and will again be assessed through midterm surveys.

Table 1. Indicators Measured at Baseline, Midterm and Final Performance Evaluations^{61*}

Performance Indicator	Baseline	Endline Target
Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand grade level text	▪ 3%	▪ 26%
Percent of students who, by the end of two grades of primary schooling, demonstrate proficiency in identifying letters.	▪ 47%	▪ 55%

⁶¹ Baseline values and endline targets were reviewed and updated following actual baseline data collection, in March 2017.

Percent of students who are attentive in the classroom.	▪ 84%	▪ 92%
Percent of students reporting that they are 'somewhat' or 'very' hungry during their afternoon class.	▪ 7%	▪ 5%
Percent of students in target schools reporting health-related absences.	▪ 33%	▪ 20%

Performance Evaluation Methodology

The **quantitative** portion of the midterm performance evaluation will involve: 1) a student survey and classroom observation drawing from children in Grades 1 – 5 across target schools in the 7 districts of Savannakhet province where LEAPS II intervenes; and 2) a Literacy Boost assessment of grade 2 students from the 180 schools in 6 districts of Savannakhet that will be targeted for LB interventions over the life of the LEAPS II project. The student survey will include classroom observation of student attentiveness combined with some teacher observations. The student survey and LB assessment will both follow a probability sampling design to detect targeted changes in required performance indicators between baseline and endline with standard statistical precision using the FANTA sampling guidance suggested for USAID/FFP Title II development projects.⁶²

The **sample design** for the proposed student survey will use a two-stage probability proportional to size (PPS) cluster sampling based on an **estimated sample size** of 1,525 student interviews drawn from the total 40,872 targeted Grade 1 – 5 students. Stage-one sampling will involve PPS selection of 61 schools from among the 350 schools targeted by LEAPS II. End-stage sampling will involve a systematic random selection of 25 Grade 1 – 5 students from each sampled school.⁶³

In addition, the classroom observations will be conducted in the same schools that were visited at baseline to measure student attentiveness, excluding the mentioned schools in Nong.

The **sample design** for the proposed Literacy Boost assessment will use a two-stage probability proportional to size cluster sample based on an **estimated sample size** of 490 student literacy assessments drawn from the estimated total 4,470 targeted Grade 2 students in LB schools. Stage-one sampling will involve PPS selection of 49 schools from among the 180 schools targeted for LB interventions. End-stage sampling will involve a systematic random selection of Grade 2 students from sampled schools.

Sample size calculations for both the student survey and literacy assessment were built off of a list of performance indicators (see **Table 1**) using expected baseline proportions for each indicator and the desired end of project targets for each, which varied by indicator. Other parameters included an alpha of 0.05 (95% confidence), a desired power of 80%, and a contingency factor of 10% for non-response. The sample size for the student survey uses an expected design effect of 2 and, based on SCI's experience conducting the LB assessment in Laos, an expected design effect of 3.15 for the LB assessment sample.

Literacy Boost Assessment Tool Development: The Literacy Boost assessment tool is used to collect data for the baseline, midterm and final performance evaluations as well as for the impact evaluation of literacy activities that is proposed for the LEAPS II project (see the section on [Impact Evaluation](#)). This is a modified

⁶² Magnani, Robert. 1999. Sampling Guide. Washington, D.C.: FHI 360/FANTA.

⁶³ Note that when sampling was done for baseline data collection in February 2017, school selection in Nong had not yet been finalized. Four schools from that sample **[Redacted]** will not be part of the midterm or final performance evaluation.

early grade reading assessment (EGRA) developed by SCI in order to assess the percent of students in Grade 2 who are able to read and understand grade level text. The LB assessment measures key mechanics of reading, similar to the EGRA, but with more focus on measuring authentic reading and reading comprehension. Since LEAPS II targets marginalized communities, where student literacy skills are at the lower end of the national distribution, the LB assessment tests a more basic version of the skills included in an EGRA. The assessment has been translated into Lao and has undergone two rounds of field-testing and validation in Laos. As part of that field-testing, 4 reading passages were tested to establish relative equivalency among the passages in case different passages are needed for future LBRA assessments. The tool was also revised to include a test of expressive vocabulary, a revised list of most common words to match national curriculum, and changes to the words used for assessing phonemic awareness. In addition to the literacy test, the assessment includes questions on potential factors contributing to literacy improvement, including questions on home literacy environment, student access to breakfast, etc. As a part of preparation for the baseline study, SCI's Literacy Advisor and Basic Education and Literacy Research Senior Specialist trained the independent evaluation team on the LB assessment tool and LB data analysis. SCI's advisors also worked with the evaluation team to refine the LB assessment tool for the [impact evaluation baseline](#) conducted with students entering grade 1 in September 2017.

At midterm, the **qualitative portions** of the performance evaluation will use purposive sampling of **key informants** and groups from at least 7 LEAPS II schools selected from among those schools selected during stage-one sampling for the student survey. **Purposive sampling** criteria to be used should be best-worst case sampling. For students and parents, data for attentiveness and attendance should be used to rate schools as best-worst case. For VEDC members and cooks, rates of cooking should be used to rate schools as best-worst case. For teachers, DESB, Pedagogical Advisors (PAs) and Community Literacy Volunteers (CLVs), previous LBRA performance should be used to rate schools as best-worst case. Focus group discussions and semi-structured key informant interviews will be the principle qualitative methods used to collect in-depth information. See **Table 2** for proposed qualitative methods, key informants, and qualitative sample sizes for the baseline, midterm and final performance evaluations.

Table 2: Summary of Key Informants and Proposed Qualitative Sample

Method	Key Informant	Estimated number of FGDs and KIIs
Focus Discussion	Group Students	6 FGDs - 3 Male and 3 Female FGD, 6-8 participants per FGD
Focus Discussion	Group Parents	6 FGDs - 3 Male and 3 Female FGD, 6-8 participants per FGD
Focus Discussion	Group VEDC members	7 FGDs - 1 VEDC per district in 7 project districts, 6-8 participants per FGD
Focus Discussion	Group Teachers and Principals	7 FGDs - 1 Teacher/Principal FGD per district in 7 project districts, 3-8 participants per FGD
Focus Discussion	Group Cooks	7 FGDs - 1 Cook FGD per district in 7 project districts, 6-8 participants per FGD
Key Interview	Informant Various	1 KII with a PESS representative, 7 KIIs with DESB (1 from each district), 6 KIIs with Pedagogical Advisors (3 from Cohort 1 and 3 from Cohort 2), 6 KIIs with Community Literacy Volunteers (3 from Cohort 1 and 3 from Cohort 2)

Key Interview	Informant	CRS staff	Interviews to be conducted with LEAPS Chief of Party, Project Manager, MEAL Manager and Officer, Commodities Manager, and Administrative and Finance Manager
Key Interview	Informant	SCI staff	Interviews to be conducted with relevant project staff
Key Interview	Informant	UO staff	Interviews to be conducted with relevant project staff
Key Interview	Informant	WFP staff	Interviews to be conducted with relevant representative
Key Interview	Informant	USDA	Interview with USDA/FFE representative

CRS and partners will **use the midterm evaluation findings and recommendations** from both the performance evaluation and [impact evaluation](#) to support evidence-based decision making about modifications and mid-course corrections that may be necessary to ensure that LEAPS II effectively and efficiently meets its strategic objectives and expected results by end-of-project. Any necessary changes identified through the MTE will be reflected in updates to the project’s detailed implementation plan and, as appropriate, in revisions to the project MEAL system.

Table 3: Timeline for the Conduct of Key Midterm Evaluation Activities

Month	Activity
October 2018	Updated Evaluation TOR shared with USDA, including description of midterm performance evaluation and impact evaluation *NOTE: The same third-party contracted for the baseline study will conduct the MTE. CRS reserves the right to terminate its relationship with the third-party evaluator if the baseline study is not satisfactory.
Dec 2018 - Feb 2019	Refine evaluation methodology and data collection tools
March-May 2019	Conduct data collection and analysis
June 2019	Stakeholder consultation, including a presentation of preliminary findings with government partners, to share and discuss initial findings, and to formulate lessons and recommendations
July 2019	Finalize Midterm Evaluation report
July 2019	Submit final report to USDA
July 2019	Develop action plan to address findings and recommendations

Limitations of the Study Design: The plan for the project’s non-literacy related performance evaluation will allow results to be generalized to all schools targeted by LEAPS II. However, the performance achievement of literacy indicators will only be generalizable to the 180 target schools in 6 districts of Savannakhet that are reached through the Literacy Boost intervention.

Key Audience and Stakeholders (midterm):To ensure high quality information that reflects priority perspectives during the midterm evaluation, CRS will engage all project stakeholders including: implementing and technical partners (SCI and University of Oregon); local and national government

partners (MOES, MOH); project beneficiaries (e.g., students, teachers, school administrators, Village Education Development Committees (VEDC), cooks, community literacy volunteers, etc.); other school feeding actors in Laos (World Food Programme); and USDA. Regional and global CRS representatives will also be engaged as a key audience for the planned LEAPS II midterm performance and impact evaluations.

IMPACT EVALUATION

In addition to the midterm performance evaluation described above, LEAPS II will incorporate an impact evaluation focused on whether any observed improvements in the literacy of school-aged children can be attributed to the project's Literacy Boost and/or school feeding interventions.

An impact evaluation using quantitative and qualitative methods will investigate the attribution of Literacy Boost and/or school feeding interventions to the literacy outcomes observed in LEAPS II. A quasi-experimental design using propensity score matching and difference in differences is proposed for the quantitative portions of the LEAPS II impact evaluation of the overall (SF+LB) LEAPS II. The impact evaluation targets the two districts of Atsaphone and Sepon where SCI began its first cohort for LB activities in year 2. The Literacy Boost approach includes a package of interventions implemented over a two or three-year period. Targeting LB's cohort 1, which begins in project year 2, allows LEAPS II to assess impact of literacy activities through a baseline literacy assessment of incoming grade 1 students completed in September 2017 and an impact evaluation that will be undertaken as part of the planned midterm evaluation process in year 3 (March 2019)—i.e. the final year of LB intervention for Cohort 1. See [Annex III](#) for a more detailed description of key project components in LEAPS II that are targeted in the impact assessment.

The impact evaluation includes two treatment arms and one well-defined comparison arm. Treatment one consists of clusters of Cohort 1 schools that receive both school feeding and Literacy Boost interventions. Treatment two includes clusters of Cohort 1 schools that only receive school feeding. The comparison group consists of clusters of schools that receive neither school feeding nor Literacy Boost support through LEAPS II.

The two districts targeted for the impact evaluation includes schools that began receiving school feeding support during LEAPS phase one and that will continue to receive school feeding in phase two. These school-feeding schools have been grouped into clusters of 3 to 4 schools based on their geographic proximity. After school feeding schools were clustered, participation in treatment arm one (school feeding + Literacy Boost) or two (school feeding only) were randomly assigned across the existing school feeding schools in the two cohort 1 districts. Clusters of non-school feeding schools in the two districts were then matched to the treatment one and treatment two clusters based on a propensity score. Characteristics used for propensity score matching included: total students, total teachers, teacher-student ratio, average number of students per school, average number of teachers per school, percent Lao ethnicity, total schools, percent of complete schools (i.e. schools that contain all grades 1 – 5).

Table 4. Number of Clusters, Schools and Students per Cohort 1 District and Treatment Group for the LEAPS II Impact Evaluation of Literacy

	Atsaphone				Sepone				Total			
	Clusters	Schools	Avg. students per school	Total # of Students	Clusters	Schools	Avg. students per school	Total # of Students	Clusters	Schools	Avg. students per school	Total # of Students
Treatment 1: Literacy Boost + School Feeding	7	29	12	345	7	30	13	398	14	59	13	743
Treatment 2: School Feeding Only	5	19	13	248	7	26	14	353	12	45	13	601
Control: No Treatment	9	30	10	298	10	25	22	542	19	55	15	840
Total	21	78	11	891	24	81	16	1,293	39	156	14	2,184

The sampling frame for the impact evaluation is the cohort of students beginning grade 1 in September 2017 (baseline) in the schools identified above. Students were randomly selected from the clusters of treatment and comparison schools described above.

A two-stage cluster **sample design** was used at impact baseline (September 2017) and impact evaluation (year 3, March 2019). The sampling frame for the impact evaluation is a cohort of students beginning grade 1 in September 2017 in clusters of schools in the two districts targeted for LB cohort 1 intervention. Stage-one sampling randomly selected from the clusters of treatment and comparison schools described above. Within sampled clusters, end-stage sampling was drawn from a random selection of cohort 1 Grade 1 students in September 2017. The same group of cohort 1 students will be sampled again as they complete grade 2 in March 2019 (year 3) Optimal Design was used to calculate an initial **expected sample size** of 2,912 Grade 2 students (see [Annex IV](#)) for the impact evaluation. The proposed sample size was calculated using the expected baseline value for the MGD indicator *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* and the desired detectable difference in LB cohort 1 by project year 3. Sampling parameters for the impact evaluation include an alpha of 0.05 (95% confidence), a desired power of 80% and a minimum detectable effect size of 0.32. The sample size calculation also assumes a 28% rate of attrition for cohort 1 students between first and second grade.

To reach this expected sample size in the baseline impact evaluation (September 2017), external evaluators planned to sample 14 students from Grade 1 (7 boys and 7 girls) within each LEAPS II school, and 28 students from Grade 1 (14 boys and 14 girls) within each comparison school. However, at the time of data collection, given the lower than anticipated enrollment rates – especially in comparison schools – the evaluation team surveyed virtually all present Grade 1 students in most schools.

Moreover, the list of comparison schools had to be further changed because of logistical challenges in the field. The external evaluators performed additional power computations to investigate different scenarios reflecting different school and cluster lists. After careful review, the evaluation team concluded that the study was still powered to detect moderate effects on literacy across different scenarios (about 0.32 standard deviations effect in literacy). In the end, the baseline impact evaluation sample included 104 treatment schools and 55 comparison schools for a total of 2,184 students.

Selecting from only present students in the absence of electronic lists, as well as surveying all present

students because of low enrollment rates, ensured a large enough sample for the evaluation and sampling consistency across schools. However, this approach can introduce sampling bias which could arise from excluding information on children who are more likely to be absent from school, like those from vulnerable socio-economic backgrounds that may live farther from schools and have difficulties commuting to schools on muddy roads in rainy seasons; or children who may be absent due to health issues.

The evaluation team will use a regression model to analyze and test the differential impacts of LB and school feeding on literacy in single and multi-grade classrooms, rural and urban schools, and demographic characteristics.

The same **Literacy Boost Assessment Tool** used in the LEAPS II baseline study and performance evaluations will be used for the impact evaluation. See further description of this tool and its development in [the section on performance evaluation methodology](#).

CRS has engaged IMPAQ, an **independent evaluation firm**, to conduct the impact and performance evaluations of LEAPS II. CRS reserves the right to terminate its relationship with the third-party evaluator if the midterm performance or impact study is not satisfactory. CRS will re-engage in a competitive recruitment process if IMPAQ does not meet expectations. A description of [the evaluation team qualifications](#) is available below.

Limitations of Impact Evaluation Design: The LEAPS II evaluation plan proposes to conduct an impact evaluation at midterm, rather than at endline. This methodology for measuring impact and assessing attribution at midterm offers both **strengths and weaknesses**. The proposed evaluation plan limits the exploration of attribution of program impact to only literacy outcomes. Another potential limitation of the proposed impact evaluation methodology is that it samples from only cohort 1 of the three Literacy Boost cohorts that SCI will target over the life of the LEAPS II project. However, focusing the impact evaluation on cohort 1 has advantages. Early assessment of attribution for observed changes in literacy to LB and/or school feeding after the full “dose” (i.e. planned 2 years of support) of the Literacy Boost intervention is provided to cohort 1 will result in actionable information at midterm that CRS and partners can use to inform mid-course corrections to improve and potentially accelerate literacy gains during the last half of the project.

Table 5: Timeline for the Conduct of Impact Evaluation Activities

Month	Activity
October 2016	(Completed) Develop Baseline Study and Evaluation TOR that includes details of selection of a single evaluation firm that will conduct baseline study, performance evaluations and impact evaluation.
November 2016	(Completed) Submit Comprehensive Baseline, Performance and Impact Evaluation TOR to USDA for approval
November 2016	(Completed) Advertise for evaluation firm
December 2016	(Completed) Finalize external evaluation firm contract
Dec 2016 - Feb 2017	(Completed) Refine and validate Literacy Boost assessment tool
June – August 2017	(Completed) Refine and finalize impact evaluation design
Sept - Oct 2017	(Completed) Conduct data collection for impact evaluation baseline
November 2017	(Completed) Finalize impact assessment baseline report
December 2017	(Completed) Submit final impact assessment baseline report to USDA

Table 5: Timeline for the Conduct of Impact Evaluation Activities

Month	Activity
October 2018	Updated Evaluation TOR shared with USDA, including description of midterm performance evaluation and final impact evaluation
Dec 2018 - Feb 2019	Refine evaluation methodology and data collection tools
March-May 2019	Conduct data collection and analysis
July 2019	Stakeholder consultation, including a presentation of preliminary findings with government partners, to share and discuss initial findings, and to formulate lessons and recommendations
June 2019	Finalize Impact Evaluation report
July 2019	Submit final impact evaluation report to USDA

Dissemination Strategy

Information arising from LEAPS II evaluations will be shared with stakeholders such as beneficiaries, local authorities, government agencies, donors and sectoral peers. CRS will ensure that results are shared widely in appropriate formats and at various venues. This includes both annual events such as the MEAL stakeholder meeting, government partnership meetings, and the circulation of key program findings following major evaluations. The midterm performance evaluation and final impact evaluation includes dissemination events where findings are presented to key government figures and community members.

Report Writing

Two reports and two short briefs will be required as part of the midterm evaluation process. These include

1. Midterm Performance Evaluation Report (maximum 50 pages)
2. Final Impact Evaluation Report (maximum 40 pages)
3. Two stand-alone briefs for each of the reports cited above (2-3 pages each)

The page limits for these studies do not include cover page, executive summary, table of contents, list of acronyms and annexes.

Write-up of midterm performance evaluation and final impact evaluation findings will be ensured by the evaluation team. The LEAPS II team, CRS East and South Asia (ESA) Regional Office and Headquarters, and SCI will be involved in providing review and feedback on the reports before it is finalized by the evaluators.

The following is a suggested outline for the reports:

1. **Cover page, Table of Contents, List of Acronyms**
2. **Executive Summary** should be a clear and concise stand-alone document that states the most salient findings, conclusions, and recommendations and gives readers the essential contents of the report in two or three pages. The Executive Summary helps readers to build a mental framework for organizing and understanding the detailed information within the report.
3. **Introduction** should include purpose, audience, and synopsis of task.
4. **Overview of the Programmatic Context and Implementation** should provide a brief overview of the programmatic context Laos and LEAPS II including the theory of change and expected results.
5. **Methodology** should clearly describe the methodologies used to collect and analyze data

6. **Findings** should respond to the objectives defined for the LEAPS II baseline or evaluation and the achievement against outcomes.
7. **Conclusions** should provide additional analysis of the data and results responding to evaluation questions, drawing on programmatic context and industry best practices.
8. **Lessons Learned**
9. **Recommendations** should provide recommendations for LEAPS II implementation. Recommendations must be relevant to program and context and include concrete and realistic steps for CRS and its LEAPS II implementing partners.
10. **Limitations** should describe any constraints of the baseline or evaluation processes.
11. **Annexes** should document the study, scope of work, schedule and tables and be succinct, pertinent and readable.
 - a. A table presenting the baseline and subsequent midterm and final values for all performance indicators (including those derived from project monitoring data), and any other indicators relevant to the key findings with levels of significance of differences noted.
 - b. References, including bibliographical documentation, meetings, interviews, and focus group discussions.
 - c. List of stakeholder group with number, type, and date of interactions.
 - d. Evaluation SOW and TOR.

In addition to the full reports described above, a 2-3 page stand-alone brief should be produced for each report describing the evaluation design, key findings and other relevant considerations. These two briefs will serve to inform any interested stakeholders of the evaluations and should be written in language easy to understand by non-evaluators and with appropriate graphics and tables.

Ethical Guidance

The evaluation team, including any enumerators or contractors that the evaluation team may hire, must ensure that the baseline study adheres to ethical guidelines as outlined in the American Evaluation Association's (AEA) Guiding Principles for Evaluators. A summary of these guidelines is provided below:

1. **Informed Consent:** All participants are expected to provide informed consent following standard and pre-agreed consent protocols. As in the baseline performance evaluation, and as per the approved IRB protocols⁶⁴ for the LBRA, teachers should provide written consent for the children to be surveyed and students must provide verbal assent. For qualitative surveys, participants are required to give verbal consent.
2. **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries.
3. **Competence:** Evaluators provide competent performance to stakeholders.
4. **Integrity/Honesty:** Evaluators display honesty and integrity in their own behavior and attempt to ensure the honesty and integrity of the entire evaluation process.
5. **Respect for People:** Evaluators respect the security, dignity and self-worth of respondents, program participants, clients, and other evaluation stakeholders. It is expected that the evaluator

⁶⁴ Chesapeake Institutional Review Board, February 28, 2017 Pro00020840)

will obtain the informed consent of participants to ensure that they can decide in a conscious, deliberate way whether they want to participate.

6. Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.
7. Evaluators shall provide copies of the evaluation reports that are free of personally identifiable information (PII) and proprietary information.

■

A link to a more detailed description of AEA's Guiding Principles for Evaluators can be found at: <http://comm.eval.org/eval/Go.aspx?c=ViewDocument&DocumentKey=ba879c95-f810-4c6b-bf50-524da31144c1>

The evaluation team is responsible for ensuring all local (Lao) and international ethical review and approval processes are followed for the studies outlined in this ToR.

SCOPE OF WORK

CRS engaged IMPAQ, an experienced, independent third-party to conduct the baseline, midterm and final performance and impact evaluations of LEAPS II. CRS reserves the right to terminate its relationship with the third-party evaluator if the baseline performance or impact study and/or mid-term performance or impact evaluations are not satisfactory. CRS will re-engage in a competitive recruitment process if the first evaluator does not meet expectations.

As specified in US government regulations (see [7 CFR Part 1499.13](#) and [7 CFR Part 1599.13](#)), the third-party consultant or firm will:

- Be financially and legally separate from CSR and LEAPS II partner organizations;
- Have staff with demonstrated knowledge, analytical capability, language skills and experience in conducting evaluations of development programs involving agriculture, education, and nutrition;
- Have knowledge of the region and ability to work within a multicultural environment;
- Use acceptable analytical frameworks such as comparison with non-project areas, surveys, involvement of stakeholders in the evaluation, and statistical analyses;
- Use local consultants, as appropriate, to conduct portions of the evaluation;
- Provide a detailed outline of the evaluation, major tasks, and specific schedules prior to initiating the evaluation; and,
- Provide a copy of evaluation reports that are free of personally identifiable information (PII) and proprietary information.

Independence of the evaluation function from program design and management is a core principle of USDA evaluation. Independence helps to ensure both credible and objective evaluations. USDA supported evaluations should be conducted by people who are not involved in the design and implementation of the project and the evaluation process must be free from political influence and organizational pressure.⁶⁵

⁶⁵ USDA/FAS. *Monitoring and Evaluation Policy*. Washington DC:USDA/FAS (5/6/2013).

Evaluation Team Qualifications

The evaluation team will comprise one international program evaluator (Team Leader), and three or more local or international consultants or members of a consulting firm selected for their technical expertise. One consultant should be an expert in the design of impact evaluations. The evaluation team should also include an expert(s) in survey design, management and multivariate analyses of quantitative social science data. In addition, a team member(s) is needed with experience implementing an EGRA tool, as well as analysis of EGRA data.

The aggregate technical expertise of the evaluation team members should encompass the sectors: childhood literacy, primary education, and WASH and nutrition in the context of school feeding programs. The team should have Lao language ability and fluent English speakers. Inclusion of team members with previous experience working in Lao PDR is preferred.

All evaluation team members should meet the following qualifications and experience:

- No previous involvement in the LEAPS II program design or implementation;
- Proficient in spoken and written English;
- Master's or PhD degree in a relevant field;
- 10+ years of experience in applying his/her field of expertise to program design and oversight or 8+ years' experience with program baselines and/or evaluation in his/her field in developing countries (Laos context preferred);
- Demonstrated ability to achieve results and meet deadlines in a demanding environment;
- Preferred skills: Knowledge of IRB requirements and ethical considerations when working in schools.

All evaluation team members and all baseline and evaluation data collectors are responsible for adhering to all terms and conditions stipulated in the contract with CRS, including but not limited to CRS' Code of Conduct and Policy on Protection from Abuse and Exploitation. All enumerators and evaluation team members working in schools must also complete SCI's child protection training.

The Team Leader and any other international consultants on the evaluation team are responsible for and must provide proof of emergency evacuation insurance.

Evaluation Team Leader Responsibilities and Qualifications

The evaluation team leader will lead the team to accomplish the studies and deliverables outlined in this TOR and SOW, ensuring that all team members fulfill assigned tasks.

The Team Leader will:

- Serve as the primary point of contact with the LEAPS II management to assure adequate logistics and good adherence to local protocols;
- Lead the timely development of appropriate study plans for the baseline performance evaluation, mid-term performance evaluation, final performance evaluation, and the impact evaluation.
- Assure rational sampling of targeted schools, communities, and other key informant;
- Ensure adequate triangulation and validation of findings;
- Oversee data collection training for enumerators, including time for field testing quantitative and qualitative tools;
- Monitor and assure the quality of data collection and analyses;
- Oversee the data analyses and integration of qualitative and quantitative findings;

- Lead a presentation of preliminary findings to CRS, USDA, SCI, and other key stakeholders to be confirmed by the LEAPS II team;
- Assure timely submission of draft and final reports that flow logically and clearly separate findings, conclusions, and recommendations, and in which all conclusions and recommendations are based on evidence presented in the report;
- Assure timely submission of the deliverables to the appropriate recipients or entities; and
- Be available to respond to USDA questions on baseline and evaluation reports.

■ The Team Leader should have the following **qualifications and experience**:

- Master's degree required, PhD in a relevant field preferred;
- Minimum of fifteen years of experience in international development evaluation, preferably of education programs;
- Experience leading similar evaluations;
- Familiarity working with USDA and/or USAID and their project and performance management frameworks;
- Strong management, communication and administrative skills;
- Demonstrated exceptional report writing skills that emphasize logical flow and objective analyses;
- Demonstrated ability to achieve results and meet deadlines in a demanding environment;
- Previous work experience in Laos and/or proficiency in Laos language and/or indigenous Laos languages preferred.

Evaluation Team Deliverables

All reports and deliverables should be completed in English, be free of typos or grammatical errors, and be a polished document ready for publication. This means the document contains no factual errors or inaccuracies and citations are properly used.

- CRS-approved study plans (including consultant responsibilities for identifying, interviewing, contracting, training and overseeing enumerators) for: 1) baseline performance evaluation, 2) baseline impact evaluation, 3) midterm performance evaluation, 4) final impact evaluation, 5) final performance evaluation;
- Local (Lao) and/or international IRB or other ethical approval documentation, as required;
- Meet with CRS & SCI for training on Literacy Boost tools and methodology;
- Sampling plan including file with actual sample size calculation for all quantitative surveys and assessments associated with performance and impact evaluations;
- Soft copies of all quantitative and qualitative data collection tools for performance and impact evaluations;
- Train enumerators and supervisors;
- Data collection guide for enumerators & supervisors;
- Completed copies of survey questionnaires, consent forms, and qualitative notes;
- Soft copies of both raw and final, clean quantitative datasets with syntax, in a readable format, with accompanying codebook/data dictionary;
- Soft copy of qualitative matrices;

- Survey implementation report: summary of issues encountered and solutions during the fielding of questionnaire, data entry and data analysis. The purpose is to: 1) document the issues that might affect the analysis (i.e. if one of the module has failed then the analysis will show this; and 2) highlight lessons learned to improve the implementation of mid- or end-line survey;
- Draft reports for: 1) Baseline performance evaluation [completed], 2) baseline impact evaluation [completed], 3) midterm performance evaluation, 4) final impact evaluation, and 5) final performance evaluation following the outline and pages limits defined on in the [Report Writing](#) section of this TOR and SOW;
- Final reports must not contain any propriety or personally identifiable information (PII). PII is any information that directly or indirectly identifies an individual. This information can be used on its own or with other information to identify, contact or locate a single person, or to identify an individual in a specific situation. This may include, for example, a name, national ID number, address, birthplace, etc. PII includes both direct and indirect identifiers that, when taken together, could allow for identification of an individual (such as a village name, gender, age, name, and/ or facial image).
- Final reports should not allow for the identification of individual schools or communities. Any list of schools or communities provided should be included in the report as an annex, so that it can be easily removed for external sharing. CRS-approved final reports for: 1) Baseline performance evaluation [completed], 2) baseline impact evaluation [completed], 3) midterm performance evaluation 4) final impact evaluation, and 5) final performance evaluation following the outline and pages limits defined on in the [Report Writing](#) section of this TOR and SOW;
- Each full report [starting with the midterm reports] should be accompanied by a 2-3-page stand-alone brief describing the evaluation design, key findings and other relevant considerations. It will serve to inform any interested stakeholders of the evaluations and should be written in language easy to understand by non-evaluators along with appropriate graphics and tables.
- Delivery of a short webinar, with time for Q&A to the global CRS FFE and MEAL audience after the completion of the final impact evaluation;
- Participation in dissemination events and provision of soft copies of presentations developed and delivered during dissemination events; and
- Written responses to any questions raised by the donor on any of the final reports that require feedback from the Evaluation Team.⁶⁶

Evaluation Coordination and Management

CRS will contract an independent third-party consultant firm to collect evaluation data that are reliable, accurate, valid and timely. CRS will support the independent consultant through review of the performance evaluation and impact evaluation plans, survey and data collection instruments, sampling methods, and the development of a data analysis plan based on the project indicators. In-country

⁶⁶ Timing of donor feedback on evaluation reports cannot be estimated. The evaluation team is suggested to plan the equivalent of 0.5 day of the Team Leader's time within the 6 months following the submission of each of the 5 required evaluation reports (2.5 days total) to allow for time required to respond to donor queries on final evaluation reports.

coordination of the baseline study and evaluations will be ensured by the CRS Laos Head of Programs, LEAPS II Chief of Party, or MEAL Manager.

Table 6: Evaluation Roles and Responsibilities

Catholic Relief Services

LEAPS MEAL Manager: Draft and revise baseline and evaluation TORs based on USDA comments; Review draft baseline, midterm and final evaluation methodology and tools; Review draft midterm and final performance evaluation reports and impact evaluation reports; store all evaluation-related data provided by the external evaluator.

CRS Laos Head of Programs & LEAPS Chief of Party: Review draft baseline and evaluation TOR (year 1); Recruit and contract external evaluation consultant/firm for baseline, midterm and final evaluations; Provide external evaluator with key project documents, performance monitoring data, and logistical support to contact program participants; Review draft baseline, midterm and final evaluation methodology and tools; Review draft midterm & final performance evaluation reports and impact evaluation reports; Approve final deliverables that will be submitted to USDA; Share evaluation findings and reports, as appropriate, with key stakeholders, Coordinate management response related to USDA's and evaluator's questions and recommendations; Implement any post-evaluation activities.

Global MEAL Team Technical Advisor and Senior Technical Advisor for Education: Review baseline and evaluation TORs; support selection of third-party evaluation consultant; Review draft baseline, midterm & final evaluation methodology and tools; Review impact evaluation methodology and tools; Review draft baseline, midterm and final performance evaluation reports and impact evaluation reports; Ensure technical approval of evaluation-related deliverables.

Sub-Recipient

SCI Literacy Boost Assessment Advisor: Revise Literacy Boost assessment tools with evaluation firm, as necessary; Train external evaluation team on LB assessment and analysis; Remote support to external consultant to train LB enumerators; Remote support to external consultant to conduct LB assessment and data collection; Review draft baseline, midterm and final performance evaluation reports and impact evaluation reports; Technical approval of literacy-related evaluation deliverables.

External Evaluator (independent third-party)

Independent third party: Draft and finalize tools and methodology; Conduct an interview with USDA representatives prior to finalization of tools and field work, at baseline, mid-term, and final evaluations; Train enumerators and field test tools; Manage data collection and ensure data quality; liaise with LB assessment advisor observing a random selection of literacy assessments in line with both performance evaluation and impact evaluation design; Analyze data; Facilitate reflection event for participatory analysis of preliminary results; Co-facilitate evaluation dissemination events; Prepare draft and final reports

Government Partners

MOES: Contribute to interpretation of baseline, midterm and final evaluation findings, accompany enumerators in the field, Participate in dissemination workshops

Key Evaluation Resources

The evaluation team will have access to the following key documents:

- Donor-approved proposal;
- Donor-approved evaluation plan;
- Donor-approved performance monitoring plan (PMP);
- Updated indicator performance tracking table (IPTT);
- Previous evaluations from phase I of the project;

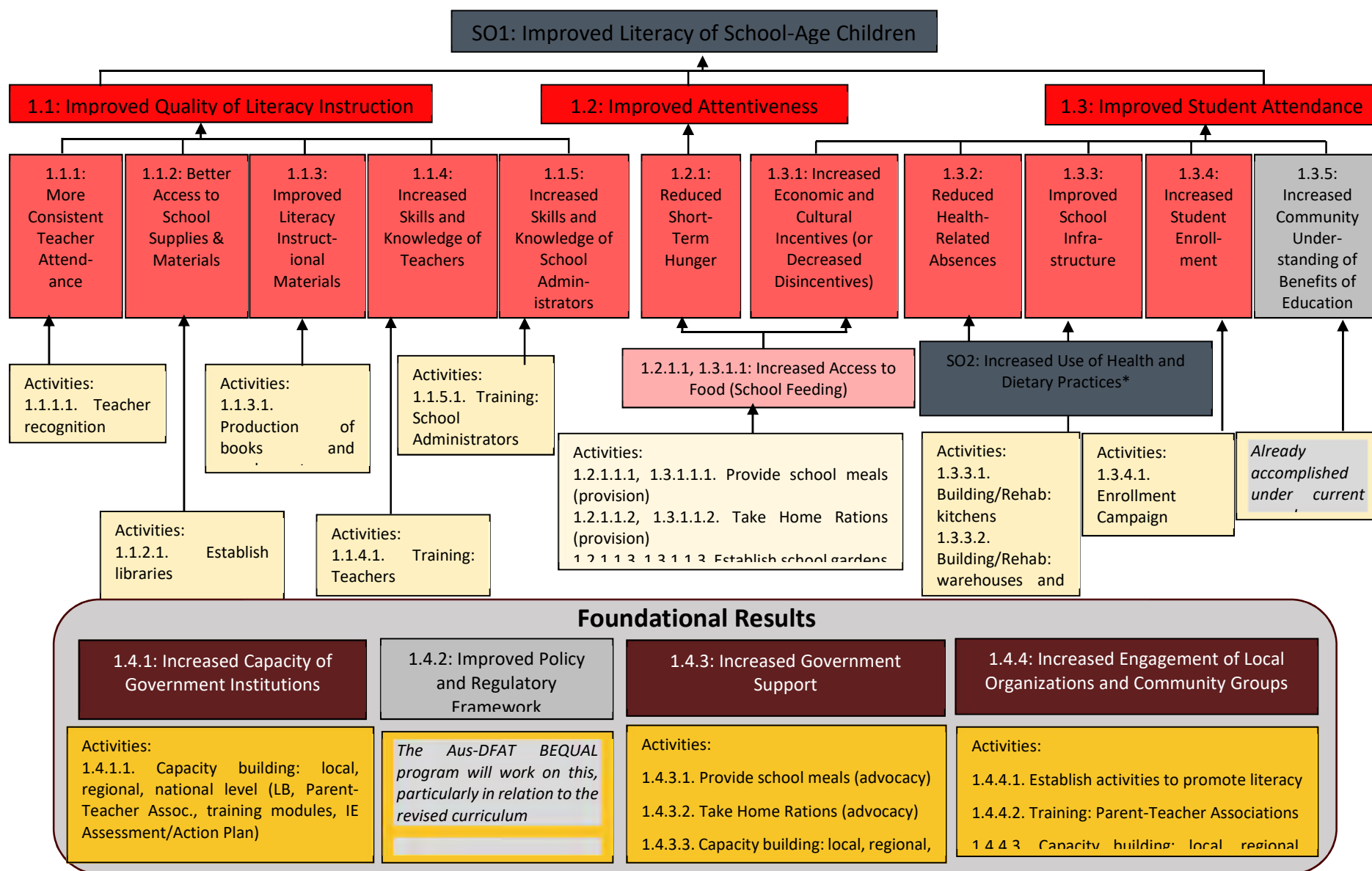
- Baseline performance and impact evaluations for LEAPS II;
- Baseline data collection tools;
- Financial documents;
- Semi-annual project status reports;
- Relevant population data for calculation of survey sample sizes;
- Definition files used for digital data collecting tools used by the LEAPS team;
- Donor-approved MTE TOR

CRS will facilitate introductions to government stakeholders, but the evaluation team is responsible for engaging the government to allow for their input and feedback throughout the study processes.

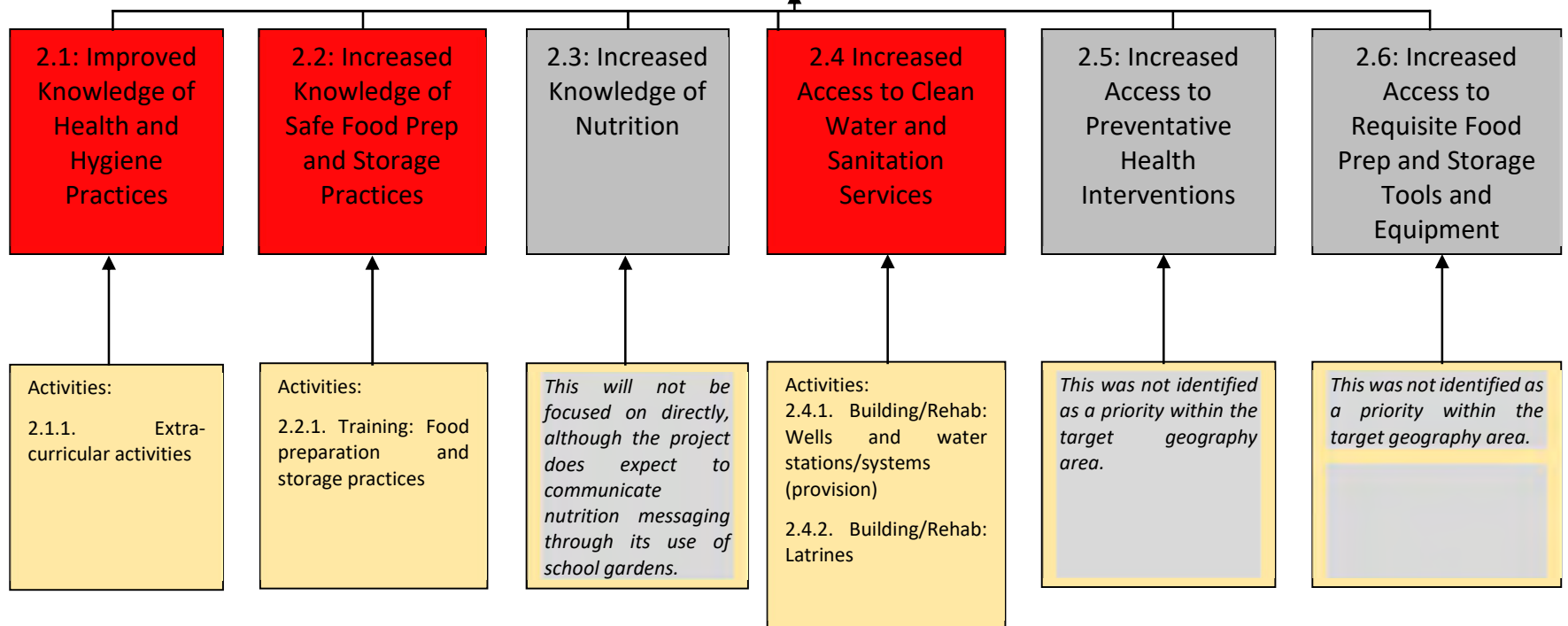
Ownership of the Evaluation Data, Results and Report

All data collected for the studies outlined in these TOR shall remain the property of USDA and CRS as its custodian. Any work product resulting from the baseline study, midterm or final performance/impact evaluations must cite the participating NGOs and USDA. The developed Literacy Boost Assessment tool shall be the property of SCI and is not to be used by or shared with parties outside of this TOR and Scope of Work.

Project-Level Results Framework (Annex I in the TOR)



SO2: Increased Use of Health and Dietary Practices



2.7.1: Increased Capacity of Government Institutions

The project will focus more on government capacity under SO1 and leave this result to health/nutrition forums

2.7.2: Improved Policy and Regulatory Framework

The project will focus more on government capacity under SO1 and leave this result to health/nutrition forums

2.7.3: Increased Government Support

The project will focus more on government capacity under SO1 and leave this result to health/nutrition forums

2.7.4: Increased Engagement of Local Organizations and Community Groups

Activities: 2.7.4.1. Building/Rehab: Wells and water stations/systems (WUC)

LEAPS II Performance Indicators (Annex II in the TOR)

Results Stream	Intermediate Result	Expected Result
MGD SO 1: Improved literacy of school-age children		
• (MGD #24) 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 • (Custom) Percent of students who, by the end of two grades of primary schooling, can demonstrate proficiency in identifying letters • (MGD #25) Number of individuals benefiting directly from USDA-funded interventions • (MGD #26) Number of individuals benefiting indirectly from USDA funded interventions		
1.1 Improved quality of literacy instruction (MGD #5) Number of teachers/educators/ teaching assistants in target schools who demonstrate use of new and quality teaching techniques or tools	<i>1.1.1 More consistent teacher attendance</i>	(Custom) Percent of teachers in attendance on day of school visit at USDA supported schools
	<i>1.1.2 Better access to school supplies and materials</i>	(MGD #2) Number of textbooks and other teaching and learning materials provided as a result of USDA assistance
	<i>1.1.3 Improved literacy instructional materials</i>	(Custom) Number of schools with improved literacy instructional materials as a result of USDA assistance.
	<i>1.1.4 Increased skills and knowledge of teachers</i>	(MGD #6) Number of teachers/educators/teaching assistants trained or certified as a result of USDA assistance
	<i>1.1.5 Increased skills and knowledge of school administrators</i>	(MGD #3) Number of school administrators in targeted schools who demonstrate use of new techniques or tools (MGD #4) Number of school administrators and officials trained or certified as a result of USDA assistance
1.2 Improved attentiveness (Custom) Percent of students who are attentive in the classroom	<i>1.2.1 Reduced short-term hunger</i>	(Custom) Percent of students reporting that they are 'somewhat' or 'very hungry' during their afternoon class
	<i>1.2.1.1 Increased access to food (school feeding)</i>	(MGD #15) Number of daily school meals (lunch) provided to school-age children as a result of USDA assistance
		(MGD #14) Number of individuals receiving take-home rations as a result of USDA assistance

Results Stream	Intermediate Result	Expected Result
		(MGD #16) Number of school-aged children receiving daily school meals (lunch) as a result of USDA assistance
		(MGD #13) Number of take-home rations provided as a result of USDA assistance
1.3 Improved student attendance (MGD #1) Number of students (male/female) regularly (80%) attending USDA supported classrooms/schools	<i>1.3.1 Increased economic and cultural incentives (or decreased disincentives)</i>	(MGD #17) Number of social assistance beneficiaries participating in productive safety nets as a result of USDA assistance
	<i>1.3.2 Reduced health related absences</i>	(Custom) Percent of students in target schools with health-related absences
	<i>1.3.3 Improved school infrastructure</i>	(MGD #7) Number of educational facilities (i.e. school buildings, classrooms, and latrines)
	<i>1.3.4 Increased student enrollment</i>	(MGD #8) Number of students enrolled in schools receiving USDA assistance
1.4 Foundational results	<i>1.4.1 Increased capacity of government institutions</i>	(Custom) Percent of officials trained through education related ToTs who go on to conduct trainings
	<i>1.4.3 Increased government support</i>	(MGD #11) Value of public and private sector investments leveraged as a result of USDA assistance
	<i>1.4.4 Increased engagement of local organization and community groups</i>	(MGD #9) Number of Parent Teacher Associations or similar “school” governance structures supported as a result of USDA assistance.
		(MGD #10) Number of public-private partnerships formed as a result of USDA assistance

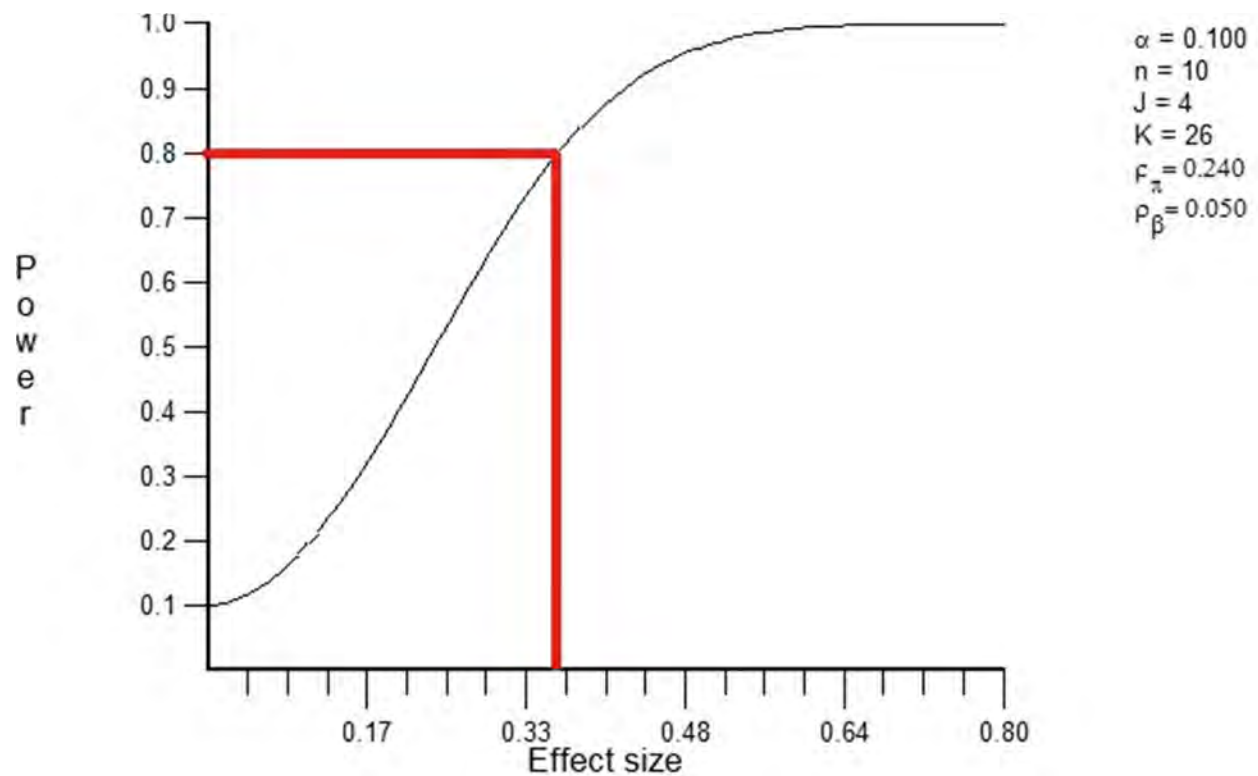
Results Stream	Intermediate Result	Expected Result
		(Custom) Percent of VEDCs that are 'high functioning'
MGD SO 2: Increased Use of Health and Dietary Practices		
(Custom) Percent of schools that provide access to safe and nutritious food		
2.1 Improved Knowledge of Health and Hygiene Practices		(Custom) Percent of schools targeted for WASH interventions that achieve at least a "1-star" rating
2.2 Increased Knowledge of Safe Food Prep and Storage Practices		(MGD #18) Number of people trained in child health and nutrition as a result of USDA assistance
		(Custom) Percent of food preparers at target schools who demonstrate safe food preparation and storage
2.4 Increased Access to Clean Water and Sanitation Services		(MGD #19) Number of schools using an improved water source
		(MGD #20) Number of schools with improved sanitation facilities
Foundational Result 2.7	<i>2.7.4: Increased Engagement of Local Organizations and Community Groups</i>	(Custom) Percent of water user committees collecting user fees for maintenance of water points provided through USDA assistance

Annex III Summary of Key Project Components for LEAPS II Impact Evaluation⁶⁷

REDACTED

⁶⁷ Figures are reflective of the state of the project in September 2016

Annex IV Sampling calculations for impact evaluation completed using Optimal Design.



Annex V Potential Local Data Collection Agencies

REDACTED