

# **Catholic Relief Services**

## **Beoog Biiga II Program Baseline Report**

***Updated***

***November 2017***

**Authors:**

Jean Paul Pétraud  
Elnaz Safarha  
Laurence Dessein



***Submitted to:***

Neda Sobhani  
Food for Education Chief of Party  
Catholic Relief Services – Burkina Faso  
Zone industrielle Gounghin, 01 BP 469  
Ouagadougou Burkina Faso

***Submitted by:***

Jean Paul Pétraud  
IMPAQ International, LLC  
10420 Little Patuxent Parkway  
Suite 300  
Columbia, MD 21044  
[www.impagint.com](http://www.impagint.com)

# FY 2014 McGovern Dole Beoog Biiga II Program in Burkina Faso Baseline Report

Program: McGovern-Dole International Food for Education and Child Nutrition

Agreement Number: FFE-686-2014/047-00

Funding Year: Fiscal Year 2014

Project Duration:

Implemented by: Catholic Relief Services

Evaluation Authored by: IMPAQ International, LLC

Jean Paul Pétraud

*DISCLAIMER: This evaluation was conducted by an independent third party firm. The views expressed in this report do not necessarily reflect the views of the United States Department of Agriculture or the United States Government.*

*Accessibility Note: An accessible version of this report can be made available upon request by emailing [FAS.monitoring.evaluation@usda.gov](mailto:FAS.monitoring.evaluation@usda.gov).*

## TABLE OF CONTENTS

---

<b>ACKNOWLEDGEMENT [REDACTED]</b> .....	<b>I</b>
<b>TABLE OF CONTENTS</b> .....	<b>II</b>
<b>TABLE OF EXHIBITS</b> .....	<b>IV</b>
<b>ACRONYM LIST</b> .....	<b>VI</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>1</b>
<b>I. PROGRAM BACKGROUND</b> .....	<b>3</b>
1.1 Beoog Biiga I .....	3
1.2 Beoog Biiga II .....	5
<b>II. RESEARCH QUESTIONS AND KEY INDICATORS</b> .....	<b>7</b>
<b>III. EMPIRICAL STRATEGY</b> .....	<b>10</b>
3.1 Mixed Evaluation Strategy .....	10
3.2 Impact Evaluation .....	12
3.2.1 Sample Selection .....	12
3.2.2 Random Assignment .....	14
3.2.3 Baseline Data and Endline Data .....	15
3.2.4 Estimation Strategy .....	15
3.2.5 Sample Size and Minimum Detectable Effects.....	16
3.3 Performance Evaluation .....	18
3.3.1 Sample Selection .....	18
3.3.2 Sample Size, Minimum Detectable Effects, and Analysis Plan.....	19
3.3.3 Quantitative and Qualitative Data .....	20
3.3.4 Work Plan .....	20
<b>IV. INSTRUMENTS</b> .....	<b>22</b>
4.1 ASER-Reading Test .....	22
4.1.1 Overview .....	22
4.1.2 Calibration .....	23
4.2 Questionnaires.....	25
4.2.1 Overview .....	25
4.2.2 Calibration .....	28
4.3 Other Data .....	28
4.3.1 Attendance Data.....	28

4.3.2 Attentiveness Data .....	29
4.3.3 Qualitative Data .....	29
<b>V. PERFORMANCE EVALUATION BASELINE LEVELS .....</b>	<b>31</b>
<b>5.1 Student Outcomes.....</b>	<b>31</b>
5.1.1 Household Environment.....	31
5.1.2 Food Security.....	32
5.1.3 Attendance and Reading Test .....	33
<b>5.2 Other Student Environment Outcomes .....</b>	<b>37</b>
5.2.1 Parents and School.....	37
5.2.2 Parent-Teacher Association Representatives.....	38
<b>5.3 Teacher Outcomes .....</b>	<b>40</b>
5.3.1 Teacher Characteristics .....	40
5.3.2 Classroom Practices.....	42
5.3.3 Knowledge of Nutrition .....	48
5.3.4 Attendance .....	49
<b>5.4 Districts Managers.....</b>	<b>49</b>
<b>VI. RCT VARIABLES: BASELINE LEVELS AND BALANCE TEST.....</b>	<b>51</b>
6.1 Second Grade Student Characteristics and Food Security.....	51
6.2 Second Grade Students Attendance and Literacy.....	53
6.3 Second Grade Students Other Learning Outcomes.....	54
6.3.1 Hygiene.....	54
6.3.2 Nutrition Knowledge .....	54
<b>VII. CONCLUSION AND LESSONS LEARNED .....</b>	<b>56</b>
<b>VIII. REFERENCE LIST .....</b>	<b>59</b>
<b>APPENDICES .....</b>	<b>60</b>
<b>APPENDIX A. INDICATOR VARIABLES: BASELINE LEVELS .....</b>	<b>61</b>
<b>APPENDIX B. ADDITIONAL DETAILED DATA BREAKDOWN .....</b>	<b>69</b>
<b>APPENDIX C. ASER-READING TEST [REDACTED].....</b>	<b>74</b>
<b>APPENDIX D. ASER-READING TEST INSTRUCTIONS [REDACTED].....</b>	<b>75</b>
<b>APPENDIX F. QUESTIONNAIRES [REDACTED].....</b>	<b>77</b>
<b>APPENDIX E. TREATMENT AND CONTROL SCHOOLS [REDACTED].....</b>	<b>100</b>

## TABLE OF EXHIBITS

---

Exhibit 1. BBI and BBII Interventions: Bam and Sanmatenga Provinces.....	4
Exhibit 2. Key Project Performance Indicators and Baseline Levels .....	8
Exhibit 3. RCT and Performance Evaluation Design.....	11
Exhibit 4. Sample Schools .....	14
Exhibit 5. Minimum Detectable Effects .....	16
Exhibit 6. Performance Evaluation Sample – 22 Control Schools in the RCT.....	18
Exhibit 7. Minimum Detectable Effects .....	19
Exhibit 8. Work Plan.....	21
Exhibit 9: ASER-Reading Test Structure .....	25
Exhibit 10. Student Characteristics .....	27
Exhibit 11. Household Size .....	32
Exhibit 12. Household Characteristics .....	32
Exhibit 13. Educational Attainment of Mothers .....	32
Exhibit 14. Student Food Security.....	33
Exhibit 15. Student Attendance .....	34
Exhibit 16. Students Reading at Second Grade Level .....	34
Exhibit 17. ASER-Reading Test Results, All Students (Grade 2) .....	35
Exhibit 18. ASER-Reading Test Results, Disaggregated by Sex (Grade 2) .....	35
Exhibit 19. Student Reading Groups .....	36
Exhibit 20. Student Knowledge of Nutrition .....	36
Exhibit 21. Student Hygiene Practices .....	37
Exhibit 22. Parent Knowledge of Nutrition .....	37
Exhibit 23. Parents and School.....	37
Exhibit 24. Parents’ Opinions of Girls’ Schooling .....	38
Exhibit 25. Parents’ Opinions of Why Girls’ Schooling is Good.....	38
Exhibit 26. Parent - Teacher Associations Activities .....	39
Exhibit 27 PTA representative involvement.....	39
Exhibit 28. Educational Attainment of Teachers.....	40
Exhibit 29. Teacher Training.....	41
Exhibit 30. Years of Teaching Experience .....	41
Exhibit 31. Classroom Size.....	42
Exhibit 32. Teacher Reported Utilization of Classroom Activities .....	43
Exhibit 33. Teachers Using New Techniques or Tools.....	44
Exhibit 34. Teachers Spending at Least 45 Minutes per Day on Literacy Instruction.....	44
Exhibit 35. Average Time Spent Per Week on Literacy Related Activities .....	45
Exhibit 36. How Often Does the Teacher Usually Have Students Read by Themselves? .....	46
Exhibit 37. How Often Does the Teacher Usually Have Students Read to Others? .....	46
Exhibit 38. How Often Does the Teacher Usually Have Students Listen to Others Read Aloud? .....	47
Exhibit 39. How Often Does the Teacher Usually Have Students Write? .....	47
Exhibit 40. How Often Does the Teacher Usually Have Students Do Word Play? .....	48
Exhibit 41. Student Reported Classroom Activities .....	48
Exhibit 42. Teachers’ Knowledge of Nutrition .....	49
Exhibit 43. Proportion of Teachers Who Were Present at least 90 Percent of School Days .....	49
Exhibit 44. Learning Styles Mentioned by District Managers .....	50
Exhibit 45. Techniques to Teach Literacy Mentioned by District Managers .....	50

Exhibit 46. Second Grade Student Characteristics.....	51
Exhibit 47. Second Grade Students Food Security.....	52
Exhibit 48. Second Grade Student Attendance.....	53
Exhibit 49. Second Grade Literacy Activities.....	53
Exhibit 50. Second Grade Student Hygiene Outcomes.....	54
Exhibit 51. Second Grade Student Knowledge of Vitamin A.....	55
Exhibit 52. Minimum Acceptable Diet Reported by Parents .....	69
Exhibit 53. Second Grade Students Reading at Level B or Above.....	69
Exhibit 54. ASER-Reading Test Results for RCT Sample (Grade 2) .....	70
Exhibit 55. ASER-Reading Test Results for RCT Sample, Disaggregated by Sex (Grade 2) .....	70
Exhibit 56. Proportion of Children that Washed Themselves in the Morning in Control Schools.....	71
Exhibit 57. Proportion of Children that Washed Themselves in the Morning in Treatment Schools.....	71
Exhibit 58. Proportion of Children that Passed the Minimum Acceptable Diet in Control Schools .....	72
Exhibit 59. Proportion of Children that Passed the Minimum Acceptable Diet in Treatment Schools .....	73

## ACRONYM LIST

---

AGIR	Action Gouvernance Integration-Renforcement
ASER	Annual Status of Education Report
BBI	Beoog Biiga I
BBII	Beoog Biiga II
BEPC	Brevet d'Etudes du Premier Cycle (English translation)
CCEB	Chef de Circonscription d'Education de Base (District Manager)
COGES	Committee de Gestion (School Management Committees)
CRS	Catholic Relief Services
DECEB	Direction des Examens et Concours de l'Education de Base (Department of Exams and Testing for Basic Education)
DGESS	Direction Générale des Études et des Statistiques Sectorielles (General Department for Research and Statistics)
DGRIEF	Direction Generale de la Recherche, des Innovations Educatives et de la Formation (General Department for Research, Educational Innovations and Training)
EGRA	Early Grade Reading Assessment
FAO	Food and Agriculture Organization of the United Nations
FAS	Foreign Agriculture Service
GIS	Geographic Information System
MDE	Minimum Detectable Effect
MENA	Ministere de l'Education Nationale et de l'Alphabetisation (Ministry of Basic Education and Literacy)
PTA	Parent Teacher Association
RCT	Randomized Control Trial
SILC	Savings and Internal Lending Communities
USDA	United States Department of Agriculture

## EXECUTIVE SUMMARY

---

This report presents the evaluation plan and baseline levels of key indicators for the Beoog Biiga II (BBII) program implemented by Catholic Relief Services (CRS) in Burkina Faso. The intervention is funded by a United States Department of Agriculture (USDA) McGovern-Dole International Food for Education and Child Nutrition Program award and is a follow up to Beoog Biiga I (BBI), implemented from 2011- 2014. BBII will provide school meals and take home rations, trainings to teachers, administrators and PTA members, school health and nutrition interventions, literacy activities, and girls mentoring in primary schools in the provinces of Bam and Sanmatenga. Additional activities include establishing Savings and Internal Lending Groups (SILC) for parents, and opening community libraries. The program activities will begin in fall 2015 and continue until August 2018.

In March 2015, CRS contracted IMPAQ International, LLC (IMPAQ) to conduct an evaluation of the BBII program in accordance to the evaluation plan CRS submitted to USDA. This evaluation consists of both an impact and a performance evaluation. The impact evaluation will measure the causal effect of the mentoring program on girls' literacy levels. The performance evaluation will measure changes in health, hygiene and dietary practices, as well as attendance and enrollment. IMPAQ met with CRS BBII staff in Ouagadougou to gather information about the details of the program implementation and in order to determine sampling and randomization strategies. Taking into account the imperatives for a robust evaluation strategy and the evaluation budget constraints, IMPAQ defined a strategy where the sample frames of the performance and the impact evaluations overlap and where the mentoring program is only implemented in a randomized group of BBII schools.

The IMPAQ team designed four quantitative questionnaires, one qualitative questionnaire and an ASER-Reading test calibrated for the Burkina Faso educational context. The questionnaires were designed to ask questions about health, nutrition, literacy and community involvement in schools to five types of respondents: students in grades 2 to 6, students' mothers, teachers, members of parent-teacher associations (PTA) and School District Managers. The sample frame included 44 schools scattered across the Bam and Sanmatenga provinces. This document reports on data collected with 920 students, 343 mothers, 140 teachers and 50 members of PTAs and 13 School District Managers.

The IMPAQ team has gathered a considerable amount of information, with over 1,000 variables across all surveys. This report highlights some of the key variables of interest for the evaluation and provides a full list of data, organized according to the indicators listed by the CRS evaluation plan, in the appendix. The data provide some interesting insight into the students' environment and their academic performance, while pointing to the need for programs such as BBII to raise the literacy levels and nutrition awareness of children. Highlights from the study include:



- Among second-graders, only 17 percent of girls and 12 percent of boys can read at grade level (complex sounds in French);
- 96 percent of children come to school after having had a meal and are not hungry;
- 86 percent of the children who did not eat a meal before going to school state that this was for the lack of food;
- 32 percent of students know about vitamin A, but only 13 percent can cite one of its benefits;
- 9 percent of students know about iron, but only 3 percent can cite one of its benefits;
- The treatment and control groups of the RCT samples are balanced on most observables. This means that, for all we can see, the random assignment led to a comparison group of students in schools that are not receiving the mentoring program that is the same on average as students in the schools that are receiving the mentoring program.

## I. PROGRAM BACKGROUND

The Beog Biiga II program is one of many school-meal projects that Catholic Relief Services (CRS) has implemented in Burkina Faso since CRS began working in Burkina Faso in 1960. From 1962 to 1988, CRS' school feeding program reached all primary schools in the country. Since then, CRS has been gradually transferring the school-feeding program to the Ministry of Basic Education and Literacy (MENA) of Burkina Faso. Today, the government's school feeding program is nationwide; local communities contribute more and more to schools' canteens as well. Beog Biiga II is focused on the Bam and Sanmatenga provinces in which CRS supports school canteens for only part of the year; CRS works directly with MENA and the community to cover the rest of the year.



Children enjoying lunch provided by the Beog Biiga program. Photo: Kim Pozniak/CRS.

### 1.1 Beog Biiga I

Catholic Relief Services (CRS), in partnership with the government of Burkina Faso and local organizations, implemented the Beog Biiga I (Tomorrow's Child) (BBI) program from 2011 to 2014 in response to recurrent food crises, high levels of malnutrition, and low and inequitable levels of education in the Central North region. The program was funded by the International Child Nutrition and Food for Education program, which is managed by the Foreign Agriculture Service (FAS) of the United States Department of Agriculture (USDA).

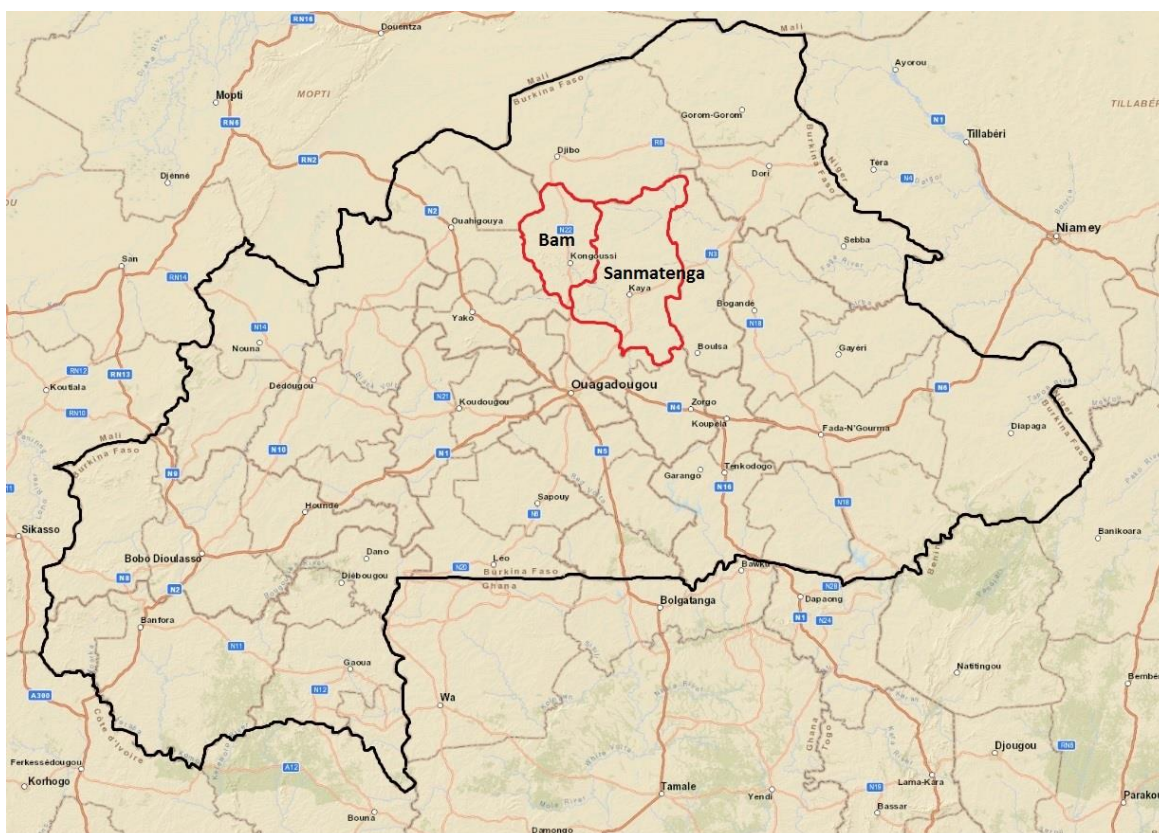
The three-year, multi-sectoral program sought to improve food security through education, health, and capacity building activities. It was implemented in more than 750 primary schools and preschools in the provinces of Bam and Sanmatenga (Exhibit 1). The BBI program's activities included providing daily lunches, take-home rations for girls, deworming, and micronutrient supplements, as well as conducting health, hygiene, and nutrition trainings for teachers and cooks, and trainings for parent-teacher associations (PTA). CRS selected 50 of the participating schools to receive additional services including mentoring for girls, additional trainings for teachers, and also created savings and microfinance



Students using hand washing stations provided to all Beog Biiga schools. Photo: SEMFILM

groups called Saving and Internal Lending Communities (SILC) for parents.<sup>1</sup> The authors refer to these 50 schools as “pilot schools” in the map in Exhibit 4 and in the sampling design sections below.

**Exhibit 1. BBI and BBII Interventions: Bam and Sanmatenga Provinces**



<sup>1</sup> Holistic programming approach that offers households a strategy to protect assets, smooth cash flow, and increase income. SILC members contribute to a fund from which group members may internally borrow at a predetermined interest rate and term.

Vanmeenen, G. (2010). *Savings and Internal Lending Communities (SILC) - Voices from Africa*. Nairobi: Catholic Relief Services.

## 1.2 Beoog Biiga II

Beoog Biiga II (BBII) is an extension and an expansion of BBI and is being implemented between October 2014 and August 2018. The \$21.6 million program is funded by a USDA McGovern-Dole International Food for Education and Child Nutrition Program award. The program's value includes 8,910 metric tons of commodities (soy-fortified bulgur, lentils, cornmeal, and vegetable oil), which CRS will import during the last three years of the program for school meals and take home rations. CRS is leading the project's implementation in partnership with MENA, Organisation Catholique pour le Developpement et la Solidarite, and Friends of African Village Libraries. CRS is also working in collaboration with the Burkina Faso Ministry of Health and the Ministry of Social Action.

CRS is continuing and expanding upon BBI's successful interventions. BBII's interventions include providing school meals to all students and take home rations for girls in select schools, training teachers, administrators and PTA members, school health and nutrition activities, and SILC. BBII will build upon BBI's initiatives towards sustaining the school meals program by continuing the program's transition to a joint community-government effort. CRS is also continuing to implement a number of health and nutrition interventions aimed at improving students' health, attendance, and attentiveness.

BBII is focusing on improving students' learning outcomes. BBII is working with government agents and non-governmental partners to improve literacy instruction and students' achievements through innovative teacher and administrator trainings. The program seeks to increase students' access to reading opportunities outside of regular class time, improve teachers' and schools' accountability, as well as improve school supplies and materials. BBII will also create community libraries and provide additional literacy support to students through mentoring, reading groups, and summer reading camps, which will motivate students and give them the skills they need to achieve higher literacy.

Furthermore, BBII will continue to improve community support for education through awareness-raising and capacity-building activities. These activities will focus on building the economic capacities of parents through SILC, as well as building the capacity of PTAs, school management committees, and local government to assure the sustainability of the program's activities and local ownership.



A community mentor assisting a student with her homework. Photo: Regina Kane/CRS.

Over the last three years of the project, BBII will expand the mentoring program to girls in 150 new schools, with a progression of 50 schools per year. In this program, BBII supports



volunteers, women who are role models in their communities, to mentor girls in the program schools. Mentor activities include spending time with the girls to support their homework, ensure they maintain good attendance, and encourage the family's support for their education through home visits and meetings with parents.

Beoog Biiga II will reach approximately 409,000 direct beneficiaries over the life of the project. CRS will work in all of the primary and preschools of Bam and Sanmatenga Provinces. The BBII program aims to reach the following beneficiaries:

- 292,000 primary and preschool students
- 3,630 teachers
- 90 preschool caregivers
- 9,900 PTAs, community management committees, and cooks
- 51,100 SILC group members
- 2,250 community mentors
- 20,000 community members (using libraries)
- 36 local government officials
- 30,000 parents
- 81 school district administrators (27 district level managers, 27 inspectors and 27 pedagogic advisors)

## II. RESEARCH QUESTIONS AND KEY INDICATORS

---

This evaluation will assess the extent to which BBII achieves its objectives. BBII focuses on literacy, health, and nutrition, and targets various beneficiaries, including teachers, students, and PTA members, with an emphasis on improving outcomes for girls. The research questions are driven by the program objectives, as well as by guidelines from the McGovern-Dole International Food for Education and Child Nutrition Program.

For simplicity and easy reference, Exhibit 2 presents the indicators as they are presented in the BBII Evaluation Plan submitted to the FAS by CRS. Exhibit 2 also indicates which evaluation methodology, impact or performance evaluation, addresses the target indicators. The two methodologies that IMPAQ will use to evaluate BBII are described in detail in section 3, Empirical Strategy. The table also relates each BBII indicator to the McGovern-Dole standard indicators.

In Exhibit 2, IMPAQ shows the data source that will be used to inform each indicator. IMPAQ will collect information about some crucial indicators from different actors with different perspectives. For example, IMPAQ will ask teachers about the techniques that they use to teach literacy skills and how often they use them, and will also ask students. Both types of information contain a bias: The teacher might give the answers that they think will make them look good and the students might not remember or be able to identify what happens in their classrooms. Yet together this information paints a picture of what happens in the classroom and how teachers and students both perceive the classroom activities. Another example involves the food security of BBII children. IMPAQ will ask children if they feel hungry and will also ask their mothers an extensive set of questions to assess the nutritional vulnerability of the household.

Exhibit 2 also displays the baseline levels of each key indicator. This information comes from a sample of 22 randomly selected schools, as described in the sample section of the report. IMPAQ used the survey data to calculate each proportion. To calculate the total number, IMPAQ applied the proportion to the total population of beneficiaries in the Bam and Sanmatenga provinces. This included 165,518 students (86,242 boys/79,276 girls), 3,265 teachers, 81 School District Managers, and 30,745 SILC group members. This method is described in CRS' performance monitoring plan, and was approved by USDA. The findings and other interesting outcomes are discussed further in sections 5 and 6.

## Exhibit 2. Key Project Performance Indicators and Baseline Levels

McGovern-Dole Indicators <sup>2</sup>	Key Project Performance Indicators	Data Source	Evaluation method	Proportions	Total Number	95% Confidence Interval
Indicator 24	1. 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 (boys and girls).	ASER reading test	Impact evaluation	Boys: 12%	10349	2% - 38%
				Girls: 17%	13,477	5% - 37%
	2. Number of teachers who devote at least an average of 45 minutes a day to literacy instruction.	Teacher questionnaire	Performance evaluation	55%	1,795	44% - 65%
	3. Number of teachers in target schools who attend and teach school at least 90% of scheduled school days per year.	District Managers	Performance evaluation	85%	2,775	NA <sup>3</sup>
Indicator 3	4. Number of school administrators and officials in targeted schools who demonstrate use of at least one new technique, skill, or tools as a result of USDA assistance. <sup>4</sup>	District Managers	Performance evaluation	61%	49	NA <sup>5</sup>
	5. Number of preschool teachers (bissongo caregivers) in target preschools who demonstrate use of at least one new teaching technique, skill, or tool.	Will be collected with pre and post testing at the time of training provided by BBII	N/A	N/A	N/A	N/A
Indicator 6	6. Number of teachers/educators/teaching assistants in target schools who demonstrate use of new and quality teaching techniques or tools as a result of USDA assistance. <sup>6</sup>	Teacher questionnaire Student questionnaire	Performance evaluation	51%	1,675	42% - 63%
	7. Number of students in target schools who are identified as attentive <sup>7</sup> during class/instruction.	Teacher questionnaire	Performance evaluation	43%	71,173	33% - 53%

<sup>2</sup> United States Department of Agriculture. (2014). *Food for Progress and McGovern-Dole Indicators and Definitions*. Washington DC: Food Assistance Division, Office of Capacity Building, Foreign Agricultural Service.

<sup>3</sup> The small sample size of District Managers does not permit statistical inferences and a confidence interval is not meaningful. Please see section 4.3.3 for details.

<sup>4</sup> In the BBII project, this indicator focuses on school administrators who are school district level staff

<sup>5</sup> Same as # 3.

<sup>6</sup> Teachers who declared using at least 5 of the 7 activities.

<sup>7</sup> Passing attentiveness score defined as a score of (according to teachers) 7 or more on a scale of 1 to 10.

McGovern-Dole Indicators	Key Project Performance Indicators	Data Source	Evaluation method	Proportions	Total Number	95% Confidence Interval
	8. Number of students in target schools who indicate that they are hungry or very hungry during the school days.	Student questionnaire Parent questionnaire	Impact and performance evaluation	4%	6,621	2% - 7%
Indicator 1	9. Number of students regularly (80%) attending USDA supported classrooms/schools (boys and girls).	School registries	Impact and performance evaluation	Boys: 100%	86,242	N/A
				Girls: 100%	79,276	N/A
	10. Number of group members that spend money from SILC on education costs.	Parent questionnaire PTA questionnaire	Performance evaluation	48%	14,758	26% - 70%
	11. Number of girl students who cite mentors as one of the top 3 reasons for their success or improvement in school.	Student questionnaire	Impact evaluation <sup>1</sup>	0.00%	0	N/A
Indicator 23	12. Percent of school-age children receiving a minimum acceptable diet (boys and girls).	Student and parent questionnaire	Performance evaluation	Boys: 56 %	48,296	48% - 64%
				Girls: 63%	49944	55% - 70%
	13. Percent of school children that wash their hands at critical moments <sup>8</sup> .	Student questionnaire	Impact and performance evaluation	78%	129,129,104	73% - 83%
	14. Number of students in target schools who achieve a passing score <sup>9</sup> on a test of good health and hygiene practices.	Student questionnaire	Impact and performance evaluation	5%	8,275	3% - 8%
	15. Number of students in target schools who achieve a passing score <sup>10</sup> on a test on food nutrition and dietary practices.	Student questionnaire	Impact and performance evaluation	1%	1,655	0% - 2%
	16. Number of months of community and/or government support for school canteens.	PTA questionnaire Parent questionnaire	Performance evaluation	5 <sup>11</sup> months	N/A	4% - 7%

<sup>8</sup> Children who reported washing their hands both as before eating and after using the latrine.

<sup>9</sup> Passing defined as students that could specify at least four situations that they should wash their hands before or after.

<sup>10</sup> Passing defined as students that could name at least one food with iron, one food with vitamin A, one benefit of iron, and one benefit of vitamin A.

<sup>11</sup> Based on responses from 14 PTA representatives. This figure is doubtful. Please see section 5.2.2 for details.



### III. EMPIRICAL STRATEGY

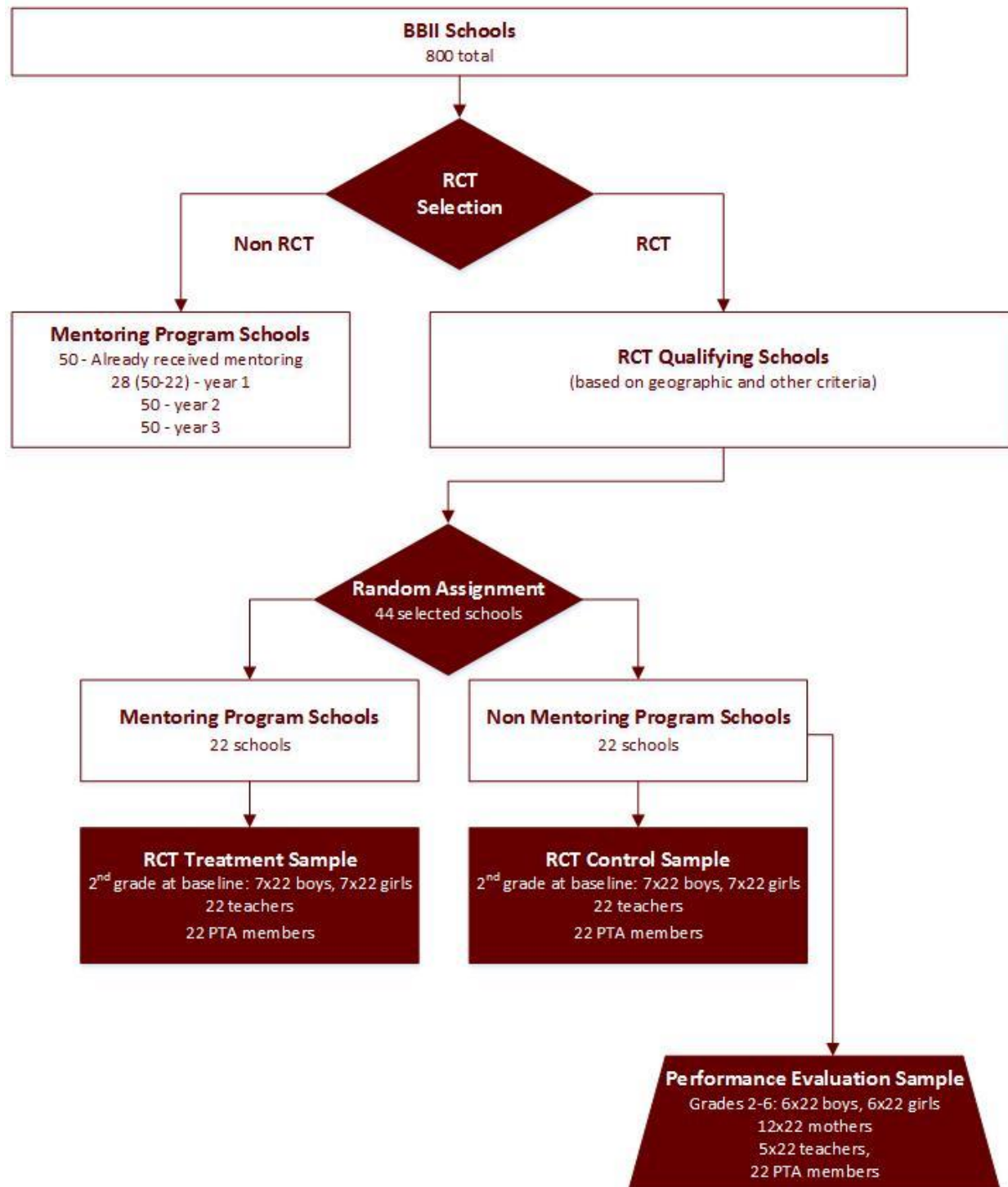
---

#### 3.1 Mixed Evaluation Strategy

The BBII evaluation consists of two main components. One component of the evaluation seeks to identify the causal impact of the mentoring program, which is only implemented in a subset of BBII program schools. In this impact evaluation, IMPAQ will estimate how the mentoring program affects students' reading skills. This will be done by comparing students' outcomes between the randomly assigned control and treatment groups at the end of the project in 2018, when IMPAQ will collect endline data. IMPAQ will also estimate the impact of mentoring on food security, hygiene practices and nutrition knowledge outcomes.

The second component seeks to measure how BBII reaches its program goals. In this performance evaluation, IMPAQ will measure and track the program beneficiaries' literacy, health and nutrition understanding, attitudes, and behaviors at three points of time: at the beginning of the program (baseline data), midway through the program cycle (midterm data), and the closing of the program (endline data). The flow chart in Exhibit 3 summarizes the sampling strategies for the two components and shows how they are related. Note that figures in Exhibit 3 represent the sample design and not necessarily the actual sample size in the data. The next section describes both the sampling strategies and the methodologies that IMPAQ is implementing for both types of evaluations.

**Exhibit 3. RCT and Performance Evaluation Design<sup>12</sup>**



<sup>12</sup> Sample size figures in this graph refer to the ex-ante sampling design. They may differ from actual (ex-post) number of respondents whose information is presented in the data section.

## 3.2 Impact Evaluation

The impact evaluation will measure the causal effect of the mentoring program on student outcomes. The primary measure of interest is the estimate of the impact of the program on the proportion of girls who achieve the grade appropriate reading level. Although mentoring is offered to girls only, the program may have spillover effects on boys in the program schools. For example, boys' reading skills may improve as they emulate the better study habits their female peers learn through the mentoring program; or, alternatively, boys' reading skills may suffer because more attention is given to girls as a result of the mentoring program. Because of the possible spillover effects, the evaluation will also estimate the program's impact on boys' reading skills.

The measure of the reading skills proportion is based on a benchmark provided by the results of an ASER test that IMPAQ designed and calibrated. Section 4.1 describes the instrument and the process of its calibration in detail. IMPAQ and CRS are also collecting attendance data. These data will enable IMPAQ to estimate the impact of the mentoring program on student attendance. In order to get insights into RCT school communities and potential spillover effects of the mentoring program, IMPAQ will interview a member of the PTA and the students' teachers in each school.

IMPAQ is conducting the impact evaluation by using a clustered randomized control trial (RCT) methodology. An RCT consists of randomly choosing a treatment and a control group in a target population. The randomization ensures that the treatment and control groups are similar on average. IMPAQ, in collaboration with CRS, randomized assignment into control and treatment group at the school level: some schools in the evaluation receive the mentoring program (the treatment group) and some do not (the control group).

### 3.2.1 Sample Selection

The population of primary interest is girls who receive three years of the mentoring program while in primary school. Therefore, IMPAQ selected girls who were in the second grade at the end of the school year preceding the beginning of the program (spring 2015). These girls will be in fifth grade in the third year of the project. IMPAQ selected boys in the same grades as the girls. IMPAQ also collected data on second-grade teachers in spring 2015 and will collect data on fifth-grade teachers in spring 2018.

The mentoring program was designed only to take place in 200 out of the 800 Food for Education (BBII) primary schools in the Sanmatenga and Bam provinces. Among these 200 schools, the top of the flow chart in Exhibit 3 shows that 50 BBII pilot schools were already using the mentoring program and were therefore excluded from the sample of schools for the BBII mentoring program. The additional 150 schools will be phased in over the first three years of the project, at a rate of 50 schools each year.

IMPAQ worked in close collaboration with the BBII team and made a selection of schools best suited for the RCT study. Exhibit 3 shows this selection process. The team was careful that the

selection did not bias the sample but, on the contrary, that it strengthened the evaluation methodology by ensuring homogeneity of the type of schools while avoiding contamination (spillover) from the treatment group to the control group. The schools that IMPAQ excluded from the RCT are not included in the impact evaluation; however, they will still receive the mentoring program.

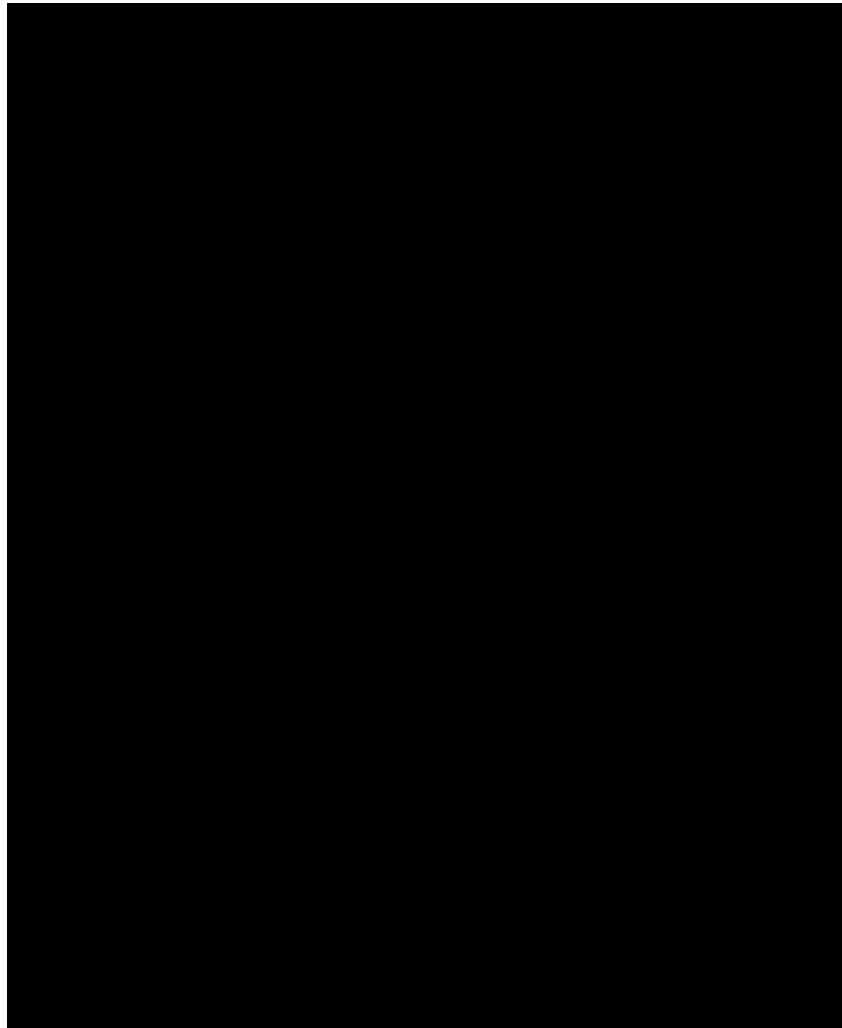
IMPAQ used the following criteria to select schools for the RCT study.

- I. Criteria to ensure school homogeneity:
  - Schools teaching exclusively in French. This is to ensure that the RCT schools are similar; Madrasas and Franco-Arab schools are excluded from the impact evaluation.
  - Schools that include a separate class for each of all of the six grade levels.
  - Rural schools. Schools in urban or more densely populated areas have different characteristics and were excluded.
  - Schools that are a reasonable distance (at least 5 km) from BBI pilot schools. The map in Exhibit 4 shows the dispersion of BBI (in red) and RCT schools.
- II. Criteria to prevent contamination:
  - Schools that have not been exposed to the mentoring program. In addition to the 50 pilot schools that already received the program, IMPAQ excluded some schools that had been visited by CRS staff and presented with the mentoring activity, as well as schools that are close to BBI pilot schools.
  - Schools that are at a reasonable distance (more than 5 km) from each other. This is so that the control schools may not be inspired by their neighbor and start their own mentoring program independently. The map in Exhibit 4 shows the dispersion of BBII RCT control and treatment schools: schools are far away from each other and they are scattered across both BBII provinces.
- III. Program feasibility criteria: School's population should be appropriate for the program. IMPAQ excluded very small and very large schools, as well as schools that were not suitable in terms of curriculum.

IMPAQ determined which schools matched the criteria with information provided by BBII program staff and with GIS data. These criteria resulted in a sample of 44 schools for the RCT: 22 in the treatment group and 22 in the control group. Determining the number of clusters (schools) in the control and treatment group is a critical step of the impact evaluation methodology. It maximizes the statistical power to detect a given effect (see section 3.1.5 for details), constrained by the scope and particularities of the program implementation itself and by the evaluation budget. In this case, 44 schools is a sufficient number of clusters to enable IMPAQ to detect the program's effects; this number also is the most that IMPAQ could select given the homogeneity and contamination criteria.

The BBII staff will select the remaining schools that will receive the mentoring program: 28 in the first year, and 50 in each subsequent year. BBII’s implementation staff is mindful not to contaminate the RCT sample. IMPAQ has coached the staff to respect the geographic contamination criteria above, especially as staff select the 28 remaining schools in the program’s first year.

#### **Exhibit 4. Sample Schools**



#### **3.2.2 Random Assignment**

Exhibit 3 illustrates that IMPAQ randomly split the sample of 44 RCT schools into two groups. IMPAQ assigned 22 schools to the control group that will not receive the mentoring program in the 3-year period of the evaluation. The random assignment and the condition that the control schools not benefit from the program are justified ethically by the fact that CRS does not have the funding to implement the program in all of the qualifying schools.

In spring 2015, after IMPAQ selected the schools, but before IMPAQ informed school staff and communities of their selection, IMPAQ randomly selected a sample of seven girls and seven boys from the second grade class<sup>13</sup> to participate in a baseline survey. Thanks to a unique student ID given to each student by CRS, IMPAQ will track each baseline second grader and give these same students an endline survey and a reading test in spring 2018.

### 3.2.3 Baseline Data and Endline Data

The baseline survey for the impact evaluation collected data on sampled students' reading skills and comprehension, their school attendance, their diet, their understanding of the roles of vitamin A and iron, as well as their hygiene practices. IMPAQ also collected data on second-grade teachers and a member of the PTA for each school. IMPAQ describes the survey instruments in more detail in section 4.2. These baseline data are important for the impact evaluation because they enable IMPAQ to accomplish the following:

- Check the integrity of random assignment by comparing the characteristics of members of the treatment and control groups;
- Control for differences between members of the treatment and control groups that may occur because of random sampling or survey non-response; and,
- Increase the precision of the impact estimates by explaining some of the variation in the outcome measures.

IMPAQ will administer an endline survey that will replicate the baseline questionnaire and will provide the data for the impact evaluation. The primary outcome will be the proportion of children reaching 5<sup>th</sup>-grade-level reading skills as measured by the ASER test. IMPAQ will use the baseline data to control for potential differences between the control and the treatment groups and to increase the statistical efficiency of the estimate. Data from the PTA and teachers surveys will also provide useful additional control variables as described in the estimation strategy below.

### 3.2.4 Estimation Strategy

Technically, IMPAQ will estimate the program's impact using the following regression model:

$$Y_{ij} = \alpha_1 + \alpha_2 M_{ij} + \beta_1 T_{ij} + \beta_2 T_j M_{ij} + \gamma X_{ij} + u_{ij} \quad (1)$$

Where the elements are:

- $Y_{ij}$ , is the child's score  $\{i = 1, \dots, n\}$  in school  $\{j = 1, 2, 3, \dots, 22\}$
- $T_{ij}$ , is the treatment indicator, which equals 1 if the child is from a school that was assigned to the treatment and 0 otherwise;

---

<sup>13</sup> There is only one class for each grade in each school.

- $M_{ij}$ , is the gender indicator, which equals 1 if the child is a boy and 0 otherwise;
- $X_{ij}$ , are child level control variables;
- $u_{ij}$ , is the error term.

The parameters of interest in this model are:

- $\beta_1$ , the regression-adjusted Average Treatment Effect of the intervention for girls;
- $\beta_1 + \beta_2$ , the regression-adjusted Average Treatment Effect of the intervention for boys;
- $\alpha_1$ , the conditional mean outcome for girls in control schools;
- $\alpha_1 + \alpha_2$ , the conditional mean outcome for boys in control schools.

IMPAQ will use estimators with clustered standard errors at the school level to account for differences in the variance of outcomes between schools.

### 3.2.5 Sample Size and Minimum Detectable Effects

To calculate the minimum detectable effects (MDEs), IMPAQ will apply<sup>14</sup> the following formula for cluster randomized trials:

$$MDE \sim M_{J-K} \sqrt{\frac{\rho}{P(1-P)J} + \frac{(1-\rho)}{P(1-P)nJ}} \quad (2)$$

Where:

- $\rho$ : unconditional intra-cluster (school) correlation
- J: total number of schools (22)
- n: number of sampled student per school (7)
- P: proportion of schools in treatment (0.5)
- $M_{J-K} = 2.8$  for 80 percent power at the 0.05 level of significance for a two-sided test

#### Exhibit 5. Minimum Detectable Effects

	Girls: $p_1 = 0.15$	Boys: $p_1 = 0.15$	All: $p_1 = 0.15$
MDE	0.22	0.22	0.22

<sup>14</sup> This formula pertains to continuous outcomes but is a good approximation of MDE for proportions. In the table for Exhibit 2, IMPAQ used the exact formula for proportions, programmed in the -clustersampsi- function in Stata.

Exhibit 5 shows Minimum Detectable Effects for the proportion of students reaching grade appropriate reading scores. IMPAQ assumed the intra-cluster correlation  $\rho$  to be 0.3, a reasonable value, based on literature describing similar studies.<sup>15</sup> To ensure a standard level of statistical power, IMPAQ assumed a level of 80 percent for power and a 0.05 significance for a two-sided test. The sample size for these calculations is 7 students for each group (7 girls and 7 boys) for 22 control and 22 treatment schools.

Exhibit 5 shows that the sample design will allow IMPAQ to detect an increase of 22 percentage points, based on the results of the ASER test at baseline. Assuming that the same proportion of fifth graders will reach grade level reading skills as they did in second grade, absent the mentoring program (that is only 15 percent of girls in the control group will have reached the grade appropriate threshold), IMPAQ will be able to detect a statistically significant effect of the mentoring program if 37 percent reach the grade appropriate level in the treatment group, a policy relevant jump in achievement.

In reality, the MDE levels in Exhibit 5 are conservative estimates because they do not take into account two characteristics of equation (1), which increase the precision of the estimates that IMPAQ will get. IMPAQ will use covariates (information about students and their household, information about their teacher and school) and will jointly estimate the parameters for boys and girls (pool both samples together) instead of estimating each separately. This will enable IMPAQ to detect significant effects smaller than a 16 percent change.

A concern when tracking respondents for data collection at a later point in time is attrition, which may be caused by students dropping out of school or moving out of the area. Comparing attrition between control and treatment groups constitutes in itself a measure of the program's success in retaining students. If that information is available, IMPAQ will attempt to find out the causes of attrition. For example: Did the household move out of the area because of a shock? Are the students now working?

However, differentials in attrition might introduce a selection bias in the impact estimate. Attrition may also substantially lower the sample size and limit IMPAQ's ability to capture the minimum detectable effect according to plan. Because students are tracked and identified with CRS-assigned identification numbers and because CRS will be collecting attendance data along the project, IMPAQ will know the level of attrition ahead of endline data collection. IMPAQ will then be in a position to re-evaluate the evaluation strategy, including re-sampling by randomly drawing additional students in both groups.

---

<sup>15</sup>Duflo, E., Glennerster, R. & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In T. Schultz & J. Strauss, Eds., *Handbook of development economics*. Vol. 4. Amsterdam and New York: North Holland.

Hedges, L. & Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials. *Education. Educational Evaluation and Policy Analysis*. 29, 60-87.



### 3.3 Performance Evaluation

The performance evaluation will measure the changes in outcomes related to the BBII core educational and community interventions, aside from the mentoring program. It will use data collected directly by CRS at school sites, as well as quantitative and qualitative data collected by IMPAQ. Student outcomes include reading habits, attendance, enrollment, as well as health, hygiene and dietary practices. Teacher outcomes pertain to content knowledge, teaching practices, and attendance. The performance evaluation will also look at outcomes about the environment of the child with a survey of the mother of the child and a survey of a member of the PTA for each school. IMPAQ will collect data on meals at school and at home, parents' involvement with the school, attitudes towards girls' education, access to reading material, and parents' understanding of nutrition and hygiene.

#### 3.3.1 Sample Selection

The performance evaluation sample frame is the schools in the control group in the impact evaluation. The sample excludes schools that receive the mentoring program so that the changes in outcomes that are detected by the performance evaluation are not attributed to the mentoring program, but instead, to the other BBII components evaluated by the performance evaluation. Exhibit 6 shows the performance evaluation sample.

**Exhibit 6. Performance Evaluation Sample – 22 Control Schools in the RCT**

Each School	2 <sup>nd</sup> grade	3 <sup>rd</sup> grade	4 <sup>th</sup> grade	5 <sup>th</sup> grade	6 <sup>th</sup> grade	Per School	Total All Schools	Total with all RCT sample
Girls	2	1	1	1	1	6	132	396
Boys	2	1	1	1	1	6	132	396
Mothers	4	2	2	2	2	12	264	264
Teachers	1	1	1	1	1	5	110	132
PTA	1					1	22	44

Furthermore, within each school, the study will target a group of students that is quite different from that targeted in the impact evaluation. Instead of looking only at students in one grade level (second grade at baseline), IMPAQ will assess students in all grade levels, except in the first grade.<sup>16</sup> As shown in Exhibit 6, the sample will include four of the second graders from the impact evaluation, in addition to one boy and one girl from each higher grade. The mothers<sup>17</sup> of each of these added children, and of two second-grade girls and two second grade boys in the RCT sample will take a survey. The teachers of each grade, as well as the PTA representative, will also

<sup>16</sup> Originally, IMPAQ planned to include first graders, but field-testing of the survey showed that they were emotionally and cognitively not able to answer a survey.

<sup>17</sup> Mothers of two boys and two girls in second grade compensate for the absence of first graders (and their mothers) in the sample.

take a survey. In total, as shown in Exhibit 6, the sample design for the performance evaluation comprises 132 girls and 132 boys, 264 mothers, 110 teachers (second to sixth grade), and 22 PTA representatives.

IMPAQ has devised a sampling strategy for the performance evaluation that is different from the strategy for the impact evaluation because IMPAQ will draw a new sample (of the same size) at midterm and endline instead of tracking the same students over time. Because the performance evaluation pertains more to the school's culture and general outcomes such as nutrition and hygiene, IMPAQ is not concerned with tracking individual students for the endline. For example, the sixth graders at baseline will no longer attend the school at endline, and second graders at endline did not attend the school at baseline.

### 3.3.2 Sample Size, Minimum Detectable Effects, and Analysis Plan

IMPAQ's analysis will compare the means of the sample at baseline, with those at midterm and endline. Because of considerable budget and time constraints, IMPAQ and CRS needed to prioritize which methodology and which outcomes to focus on and devised an approach that would still yield interesting results. IMPAQ and CRS made the following decisions, which depart from the original BBII evaluation plan<sup>18</sup>:

1. Prioritize the impact evaluation, which, as described in section 3.1.4, is given the appropriate sample size to detect a policy relevant minimum effect and will collect a considerable amount of data on students (not just their reading test scores) and their learning environments (questionnaires with teachers and PTA's).
2. Reduce data collection cost by overlapping the impact and the performance evaluations and limiting the performance evaluation sample to the control schools of the RCT.
3. Because food security is such a foundational outcome for BBII, the team decided to add interviews with mothers for each child in the performance evaluation sample.

**Exhibit 7. Minimum Detectable Effects**

	<b>Food security</b> $p_1 = 0.96$	<b>Vitamin A</b> $p_1 = 0.32$	<b>Iron</b> $p_1 = 0.09$	<b>Wash body</b> $p_1 = 0.67$
MDE	N/A	0.17	0.12	0.15

Exhibit 7 assumes a sample size of 132 students at baseline, comparing the means of two samples of the same size at baseline and endline. The exhibit shows that IMPAQ can detect a statistically significant effect between 12 and 17 percentage points for the nutrition and hygiene outcomes, which are policy relevant sizes. For example, the study will detect a change in the proportion of students who know that there is iron in foods from 9 percent to 21 percent. As for the food

<sup>18</sup> Although the evaluation differed from the approved evaluation plan, the rigor of the design was maintained. CRS resubmitted the evaluation plan to USDA in September, 2016 to reflect changes made on the baseline.

security outcomes, whereas the policy goal is that no child is hungry (100 percent not hungry) a change from 96 (key performance indicator # 8 in Exhibit 1) to 100 percent is in practice not statistically discernable.

Unlike the impact evaluation, attrition is not a concern for the performance evaluation sample because the sample at endline is drawn again and students are not tracked. However, CRS is tracking students by their assigned identification numbers and will be able to monitor attendance throughout BBII. This will provide IMPAQ with important data for the performance evaluation.

### **3.3.3 Quantitative and Qualitative Data**

In the baseline survey for the performance evaluation, IMPAQ collected data on students' understanding of the roles of vitamin A and iron, their hygiene practices, and diets. IMPAQ also asked teachers and mothers questions related to these three topics. The data from the baseline surveys will enable IMPAQ to draw inferences in its analysis from the correlation of answers between the three groups. In addition, these data make up the baseline that IMPAQ will use for the evaluation, which consists of comparing the levels of key outcomes at baseline with their levels at midterm and endline. IMPAQ will collect the same information from a similar sample at the midterm and the endline of the evaluation.

The student, teacher, and PTA questionnaires are the same as those used by IMPAQ in the impact evaluation. IMPAQ designed an additional questionnaire for the students' mothers. All survey instruments are described in detail in section 4.2.

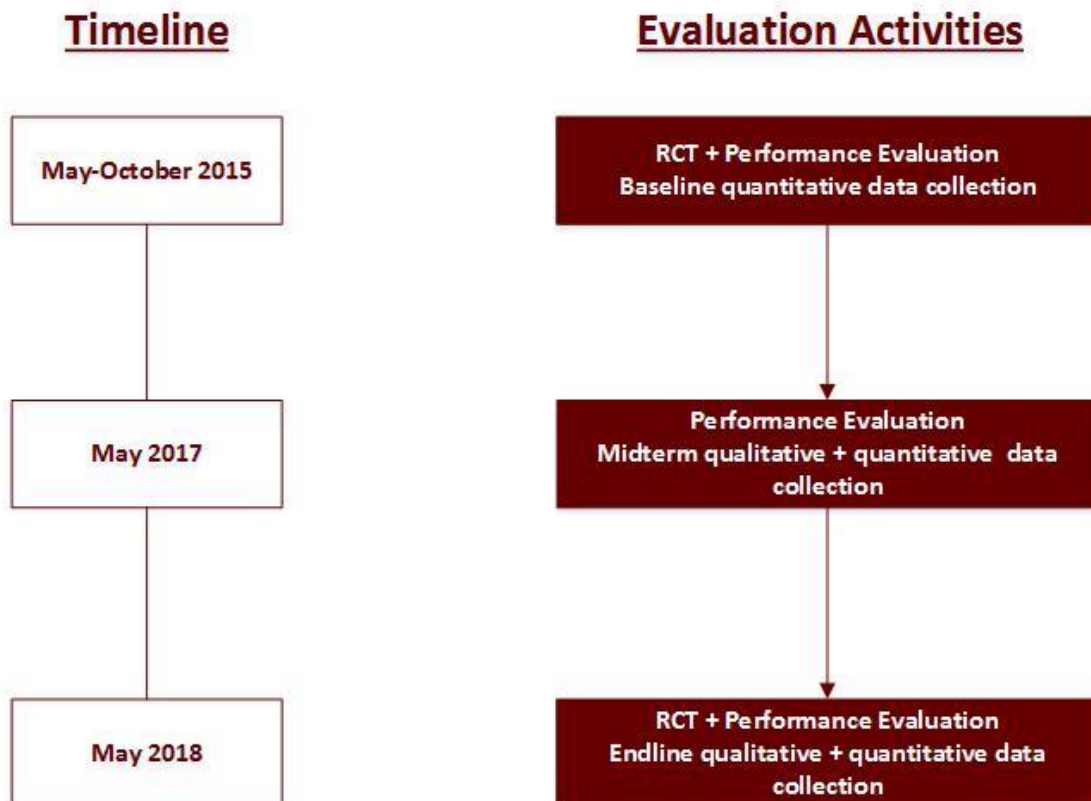
Qualitative methods will complement this design to enable IMPAQ to better understand existing motivations among the randomly assigned participants, their expectations for the program, their family's support of girls' education, and any other relevant contextual information. IMPAQ will determine the specific content and strategy for the qualitative data collection using the information from and the questions that arise after baseline and midterm data collection. A qualitative data collection team will accompany the survey team at midline and endline in a sample of five schools. More details on the qualitative information collected are provided in the Instrument Section. At baseline, the Action Gouvernance Integration-Renforcement (AGIR) team conducted qualitative interviews with the District Managers. Results from these interviews are described in details in section 5.8.

### **3.3.4 Work Plan**

Exhibit 8 summarizes the performance and RCT quantitative data collection activities. IMPAQ collected the baseline data in May, June and October of 2015. IMPAQ will collect additional data for the performance evaluation in 2017 at the project's midterm. This evaluation will measure the progress of BBII in achieving the desired outcomes and will inform potential additional questions to ask at endline, both for the performance and the impact evaluations. The midterm evaluation will also provide details about the implementation of BBII. At endline, IMPAQ will collect data for both components again. IMPAQ will use these data to compare the measures of

the indicators of the performance evaluation with the levels achieved at midterm and at baseline. IMPAQ will also collect data from both the control and the treatment groups of the impact evaluation, which will enable IMPAQ to measure the impact of mentoring program by comparing the levels in both groups. Additionally, IMPAQ will collect qualitative data from key informants to gather insights on the mechanisms leading to the observed results.

#### Exhibit 8. Work Plan



## IV. INSTRUMENTS

### 4.1 ASER-Reading Test

#### 4.1.1 Overview

IMPAQ developed and used an adaptation of the ASER-Reading test (and the test's administration instructions) to measure reading levels at baseline. IMPAQ will use the same tool to measure the program's effects on students' literacy outcomes at the midterm and endline. The Annual Status of Education Report (ASER) Center pioneered in 2005 a nationwide survey, composed of a reading and math test, to measure the achievements of children in primary school in rural India in reading and math. Since their inception, adapted versions of the tests have been implemented in Pakistan, Kenya, Mali, Senegal, Tanzania, Uganda, Ghana, and now Burkina Faso (Pratham 2015; USAID, 2012).

The ASER tests are designed as criterion-referenced assessments to categorize children on an ordinal scale indexing mastery in each of the basic skills in reading (letter identifications, word decoding, etc.) and math (number recognitions, subtraction, etc.). For example, the ASER-Reading test can classify children at the 'nothing', 'letter', 'word', 'sentence', or 'story' levels based on defined performance criteria or cut-off scores that allow testers to classify children as masters or non-masters of any given level. For example, a child's inability to correctly identify 4 out of 5 letters will classify the child at the 'nothing' level. Researchers administer the test orally and individually in about 10 minutes. Empirical research that has been conducted to evaluate the reliability and validity of the ASER testing tools shows that the tools have substantial reliability of decisions across repeated measurements, satisfactory inter-rater reliability, and suitable evidence for concurrent and convergent-discriminant validity (Vagh, 2009).



IMPAQ/AGIR team preparing to perform ASER reading assessment. Photo: IMPAQ

During the inception phase of the impact evaluation, IMPAQ convened a team of education and literacy experts to discuss early grade reading assessments and identify the most appropriate assessment to use in this evaluation. The team considered the two best-known oral early-grade reading assessments: ASER-Reading and EGRA. Following the team's in-depth analysis, as well as extensive further research on the topic, IMPAQ selected the ASER test for the following reasons:

- The ASER-Reading test is more appropriate to measure *grade-level* reading skills, and thus, more appropriate to answer the evaluation's research question as related to

McGovern Dole indicator 24 (Exhibit 2): *“What are student’s reading levels? What percent of students (male and female) can read and understand grade level text?”* In fact, the ASER-Reading test measures students’ reading performance against a fixed set of learning standards (such as national curriculum standards for each grade) while EGRA is a reading fluency assessment measured by words read per minute that is not specific to a learning standard (Simons, 2012).

- The ASER-Reading test is likely more appropriate for the French language (the instructional language in Burkina Faso), than the EGRA test, as evidence suggests that the EGRA method has complexities such as sight words or short words that are more appropriate for the English language (Abadzi, 2011).
- The ASER-Reading test is a less complex assessment tool to design and implement, and since the test is technically more appropriate for answering this evaluation’s research questions as compared to the EGRA test, ASER, thus, is a more cost-effective test for the purposes of this evaluation.

#### **4.1.2 Calibration**

In collaboration with CRS staff, IMPAQ conducted an adaptation workshop and a pretest to ensure that the IMPAQ-developed test and test administration instructions were culturally appropriate and consistent with Burkina Faso’s learning standards for each grade level in primary school.

The one-day adaptation workshop convened a group of local reading, curriculum, and assessment experts from MENA to assess the appropriateness of the test and its administration instructions with respect to the following factors: (1) the language; (2) the grade level; and (3) the research questions. The CRS Educational team as well as four coordinators from the AGIR data collection team participated in the workshop in addition to a number of high level local experts:

- [REDACTED] Curriculum expert of MENA’s editorial team for language and communication curriculum;
- Three specialists from the education research department of MENA, the General Department for Research, Educational Innovations and Training (DGRIFE):
  - [REDACTED] Primary School Inspector, Director of curriculum for primary education
  - [REDACTED] Primary School Inspector, Curriculum Division, Member of the editorial team for language and communication curriculum,
  - [REDACTED] Advisor for School and University Administration, Curriculum Division, Member of the editorial team for language and communication curriculum,
- Two specialists from the Department of Exams and Testing for Basic Education (DECEB):

- [REDACTED] Pedagogic Advisor, division for competitive testing
- [REDACTED] Primary School Inspector, division for competitive testing
- And two specialists from the General Department for Research and Statistics (DGESS):
  - [REDACTED] [REDACTED], Director of monitoring and learning, Division of monitoring, learning and evaluation,
  - [REDACTED] Pedagogic Advisor, member of evaluation team.

IMPAQ conducted the pretest at a school outside of Ouagadougou that has similar characteristics to the rural schools in the evaluation sample. Four trained enumerators supervised by two IMPAQ staff tested about 25 students. IMPAQ used the results from the pretest to further improve the ASER-Reading test and the test administration instructions.

The final version of the test included 11 levels (A-K), which roughly correspond to the reading standards for each grade level. Exhibit 9 presents the structure of the ASER-Reading test, including the test's levels and corresponding grades and reading skills.



[REDACTED]  
[REDACTED] Photo:  
IMPAQ



### Exhibit 9: ASER-Reading Test Structure

Level	Corresponding Grade	Reading Skills
Level 0	None	None
Level A	Grade 1 (CP1) – Lower level	Identify letters
Level B	Grade 1 (CP1) – Upper level	Read simple sounds
Level C	Grade 2 (CP2) – Lower level	Read complex sounds
Level D	Grade 2 (CP2) – Upper level	Decode simple words (1-2 syllables)
Level E	Grade 3 (CE1) – Lower level	Decode complex words (2-3 syllables)
Level F	Grade 3 (CE1) – Upper level	Read simple sentences
Level G	Grade 4 (CE2) – Lower level	Read complex sentences
Level H	Grade 4 (CE2) – Upper level	Read simple stories
Level I	Grade 5 (CM1) – Lower level	Answer reading comprehension questions on simple stories
Level J	Grade 5 (CM1) – Upper level	Read complex stories
Level K	Grade 6 (CM2)	Answer reading comprehension questions on complex stories

Appendix C presents the ASER-Reading test. Appendix D presents the test administration instructions.

## 4.2 Questionnaires

### 4.2.1 Overview

IMPAQ collected the baseline data using four questionnaires: (1) one for students, (2) one for parents, (3) one for teachers, and (4) one for PTA leaders. All four questionnaires collected basic demographic information, such as age and gender. IMPAQ developed the questionnaires in collaboration with BBII staff to insure the quality and internal consistency of the data. Working with the BBII staff, IMPAQ made sure the questionnaires' content and the methodology addressed the evaluation indicators and were coherent within the BBII context. IMPAQ's team for this process consisted of content area experts in nutrition and education, as well as survey and evaluation specialists. The full English version of the questionnaires is in Appendix D.

IMPAQ experts informed all specialized sections of the questionnaires including, for example, how to ask children about hunger. IMPAQ decided that a better way to assess children's food security was not to ask them directly if they were hungry, but rather to ask them if they would have eaten more during each meal that they had the day before the interview. The food security section of the mothers' questionnaire complements the information collected directly from children. A key indicator for children's nutrition is the food diversity index<sup>19</sup> recommended by FAS and developed by the Food and Agriculture Organization of the United Nations (FAO). IMPAQ asked mothers and children whether they ate food in any of 15 out of the 16 food groups defined by the FAO. We left out the last FAO food group about spices because it is not salient to children.

---

<sup>19</sup> Food and Agriculture Organization (2010). *Guidelines for Measuring Household and Individual Dietary Diversity*. Rome, Italy: United Nations.



Students in all grades completed the same student questionnaire. They were asked about their literacy activities at school, as well as their diets and knowledge of nutrition and hygiene. Students answered questions related to the different foods that they ate the previous day and whether they felt full after eating. Their knowledge and understanding of the value of vitamin A and iron was also tested, and they were asked about their hand-washing practices as well.

The parent questionnaire is targeted to mothers, who are assumed to be most involved in decisions regarding their children's nutrition. The questionnaire includes the same questions as the student questionnaire regarding their knowledge of nutrition and hygiene. Mothers also answered questions related to their interactions with the school in the last year, foods they regularly serve their children, and whether there were times that their households did not have enough food. Lastly, mothers were also asked about their involvement in the PTA and if they are part of a SILC. If in an SILC, they were asked to what extent they used these savings to cover school expenses.

One PTA leader from each school also completed a questionnaire with questions related to how often the PTA held meetings in the past year, the functioning of the school canteen, and the extent of community support for the school. As parents, they were also asked some questions about the socio-economic status of their households and about the frequency and the content of their interactions with their child's teacher. Assuming that many PTA members are also part of SILC groups, the respondent was also asked about participation in SILC and the extent to which the savings are used to cover school expenses. Both PTA representatives and mothers were asked about their opinions on girls going to school, and specific reasons as to why they think girls' schooling is good or bad.

All teachers completed the same teacher questionnaire, regardless of the grade of their students. They were asked about their literacy activities and other teaching methods they use in the classroom, including whether they taught hygiene practices during class. Teachers also answered questions about their involvement with parents over the past year and their knowledge of vitamin A and iron.

In two waves of data collection, in May and October 2015, the IMPAQ/AGIR team collected survey data from 920 students, 343 mothers, 140 teachers and 50 PTA representatives. Exhibit 10 shows the breakdown of the student sample in the 22 control group schools. Comparing with exhibit 6 and the desired sample size, the overall goals were achieved: approximately 44 students in each grade 3 to 5 were sampled. Sixth grade students were harder to reach at the end of the school year in May 2015 (due to exams), and students from the same cohort were no longer available when IMPAQ visited the sites again in October 2015. This resulted in a somewhat smaller sample of 34 students for that grade.

There is an overlap between the performance evaluation sample and the RCT sample of second graders. According to the sampling design, out of the seven boys and seven girls in second grade sampled for the RCT, two boys and two girls are selected for the performance evaluation and

their mothers are also interviewed. The team actually interviewed more mothers than was originally planned. IMPAQ kept these data, which resulted in a larger sample of second-graders for the performance evaluation.

The RCT sample at baseline is made up of 381 second-grade students from the 22 control schools (same sample frame as the performance evaluation) and 382 second-grade students from the 22 treatment schools. This is well above the target of 7 girls and 7 boys from each of the 44 schools (616). This gives an additional level of confidence that IMPAQ will be able to track a sufficient sample to detect the minimum detectable effects described in Exhibit 5.

#### **Exhibit 10. Student Characteristics**

<b>Grade</b>		<b>Observations</b>
Second Grade	All students	193
	Girls	102
	Boys	91
Third Grade	All students	38
	Girls	21
	Boys	17
Fourth Grade	All students	39
	Girls	21
	Boys	18
Fifth Grade	All students	41
	Girls	21
	Boys	20
Sixth Grade	All students	32
	Girls	15
	Boys	17

### 4.2.2 Calibration

After IMPAQ developed the first versions of the questionnaires and they were translated into French, the questionnaires were reviewed by IMPAQ's Burkinabé data collection partners, AGIR, to ensure that the formulation of the questions made sense to data collectors and were appropriate for the local context. This led to some adjustments to the survey with respect to, for example, the option choices of foods containing iron and vitamin A. It was important to include choices that reflected the types of foods that were available during the season in which data were collected and in the Sanmatenga and Bam provinces. Also important was identifying the foods with the vocabulary used in these provinces. Other questions that were adjusted to the local context were the variables regarding types of latrines and the evening meal.

IMPAQ then field-tested the questionnaires in a pilot data collection in primary schools in the area of Ouagadougou. IMPAQ selected three public schools that served students of socio-economic backgrounds similar to those in the sample. The purpose of the pilot was two-fold. First, it tested the data collection protocol and gave the survey team the opportunity to get better acquainted with the surveys' content and the target population. Second, it permitted IMPAQ to further adjust the language of the questions for better flow and to ensure the internal consistency of the responses. In addition to minor adjustments, the pilot revealed that it was unrealistic to collect data from first graders because they are not cognitively and emotionally able to answer a survey. Understanding that the survey methodology was not appropriate for this age group, IMPAQ decided to adjust the performance evaluation sampling strategy and, instead of including first-graders, to double the sample of second graders (and their mothers). This over-sampled the lower tail of the student population to compensate for the absence of first-graders.

## 4.3 Other Data

### 4.3.1 Attendance Data

All attendance data is collected from school administrative records. Student attendance is recorded by teachers in school registries. At baseline, the IMPAQ data collection team accessed these registries when available and recorded the number of absences of each student in our sample for three significant months across the school year. These data will be collected each year by BBII staff. Collecting these attendance data for different times through the academic year, in November, January and April, reveals variations in attendance and is representative of all seasons and the crop cycle<sup>20</sup>. The data reported in this report comes from information recorded on school registries by teachers themselves, who note when a student is absent. Many teachers did not record these data and, when they did, the IMPAQ team could not ascertain how rigorously. Even with the best intent, it is possible that teachers omitted to record missing children. Because only

---

<sup>20</sup> Harvest is typically in November and April is typically a period of low agricultural activity.

absences are recorded, this mechanically leads our estimator of student attendance to be biased towards reporting higher attendance.

Collecting teacher attendance data is more sensitive than student attendance as it is a direct indicator of teachers' performance. Teachers, and even administrators, sometimes resist sharing this data. The IMPAQ team at baseline collected data shared by school directors<sup>21</sup> and District Managers.

#### **4.3.2 Attentiveness Data**

Together with CRS, IMPAQ decided how to best measure student attentiveness, which is the seventh key indicator in Exhibit 2. IMPAQ and CRS noted that it is very hard to judge objectively of the degree to which a given child is attentive. Some children are very restless when they are attentive, while others, on the contrary, are very still. Conversely, a very still student might be sleepy and an active one might only be distracted. Hence, even classroom observations may not be a good measure of attentiveness. Moreover, classroom observations are tricky because they must be completed in exactly the same circumstances in each classroom in order to get quality, comparable data. IMPAQ and CRS, therefore, decided that the best measure of a child's attentiveness was the teacher's own assessment of his or her students' attentiveness. This measure, then, is currently provided by the teacher—the person who is best able to judge the relative degree of the students' attentiveness. However, given the difficulty and subjectivity of measuring attentiveness, IMPAQ suggests this indicator be eliminated.

#### **4.3.3 Qualitative Data**

Qualitative data will be collected at midline and endline. Such data will be collected from key beneficiaries of BBII, including students and parents, as well as key agents who are the channel of the intervention, such as, PTA members, District Managers and teachers. Interviews with these stakeholders will probe attitudes towards the program and assess the efficacy of practical aspects of program implementation and mechanisms for its successes or shortcomings. Schools in the performance evaluation sample (RCT control group) will be randomly selected for qualitative data collection. In these schools, students, mothers, PTA members and teachers will be interviewed in addition to being surveyed.

Qualitative data will be collected at five school sites at baseline and endline. The data will enable IMPAQ to better understand progress made towards objectives and indicators. Such data will give IMPAQ insight into the mechanisms leading to the results obtained in the quantitative data analysis, point to components that might be particularly efficient, shed light to help interpret results, and determine the extent to which teachers use improved teaching techniques. The classroom observations and key informant interviews with teachers will help to understand which new teaching techniques they use and find most useful, and what changes they have observed in their classrooms. Key informant interviews with teachers will also probe what

---

<sup>21</sup> Most often, one of the teacher doubles as director.

improvements they have seen in their students' abilities (such as reading, writing, health nutrition) and what supplies provided through the project they use and find most useful. Student observations will confirm to what extent they have adopted good hand-washing practices.

Data collected from District Managers will also be qualitative, in the form of a key informant interview, rather than a quantitative survey. Because there are only 27 school districts in Bam and Sanmatenga, IMPAQ and CRS cannot achieve any statistical significance from the quantitative data collected from District Managers. Also, because of their position, IMPAQ and CRS determined that some questions about the District Managers' content knowledge were better asked in the format of a key informant interview rather than a survey. These qualitative data will inform indicator 4 about administrators' (District Managers) understanding of literacy teaching methods and their involvement in their district's schools, including the frequency of literacy trainings and follow up visits.

## V. PERFORMANCE EVALUATION BASELINE LEVELS

---

The performance evaluation will evaluate the larger effects of BBII by comparing outcomes across the life span of the program. This section highlights the means for some key outcomes for the sample of the performance evaluation: students grade 2 to 6 in selected school which will not receive the mentoring program. The section is organized around the key target of the intervention: the child. It starts with the child and their household, continues with their educational environment and their teachers and concludes with the District Managers. A table in Appendix A reports the means for a more extensive list of questions, organized according to the program's key indicators in Exhibit 2 for easy reference.

Data for this section come from surveys conducted in 2015 with 343 students in grades 2 to 6, as well as their mothers, 101 teachers, 25 PTA representatives and 13 District Managers.

### 5.1 Student Outcomes

As with the impact evaluation, the performance evaluation will look at children's perceptions of hunger, as well as their understanding of nutrition and hygiene, which addresses indicators 8, 12, 13, and 14. The questionnaire that IMPAQ administers to the students' mothers will provide detailed information on the meals served at home. This information will enable IMPAQ to determine the children's diet diversity index. Student attendance is collected from school registries.



IMPAQ/AGIR enumerator interviewing a student in Zincko Classique School. Photo: IMPAQ

#### 5.1.1 Household Environment

Exhibits 11, 12 and 13 show the household environment of the children in the sample. These are important because they are the conditions in which children achieve the measured outcomes, and they may be limiting factors. Absence of electricity (and maybe light) in almost two thirds of households in our sample can limit children's capacity to study and develop reading skills. Lack of access to a latrine for half of the children in the sample is likely a limiting factor when it comes to hygiene practices. Judging from the education level of their mothers, most girls in the sample are the first generation of girls to attend any formal school. There is one mature person (over 16 year old) for each child below 16 on average, suggesting that children benefit from adult care and supervision.

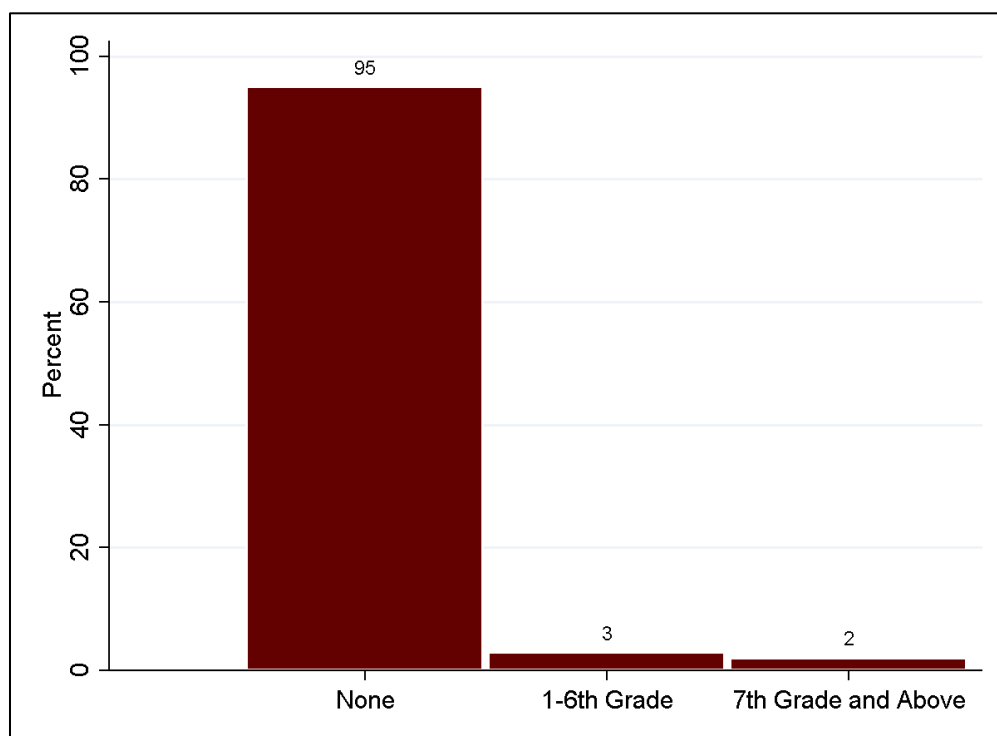
### Exhibit 11. Household Size

Indicator	People	Observations
Average number of people under 16 years old including the respondent in the household	7	343
Average number of people over 16 years old including the respondent in the household	6	343

### Exhibit 12. Household Characteristics

Indicator	Percentage	Observations
Households not using any source of electricity	41%	343
Households that have a latrine at home	49%	343

### Exhibit 13. Educational Attainment of Mothers



#### 5.1.2 Food Security

Exhibit 14 shows that 96 percent of children came to school after having had a meal feel satiated and not hungry. Likewise, the majority of children (94 percent) who eat a meal at home for lunch were also not hungry.

Furthermore, over 40 percent do not receive an acceptable diet in terms of its diversity. A key indicator for children's nutrition is the FAO food diversity index<sup>22</sup> recommended by FAS. IMPAQ removed spices from the list of 16 food groups in the index, because it is not relevant for child nutrition, and defined a threshold for an acceptable diet to be a minimum of five of the 15 different types of foods per day. Boys report less food diversity than girls, but it is possible that they might not report accurately because the response provided by mothers (details in Exhibit 50 in Appendix B), about the food they prepare at home, is closer to the average reported by girls. Here also, the data in the Appendix A (under Indicator 12) provides deeper insight. Although 98 percent of children eat cereals (such as pasta, rice, noodles), very few eat vitamin A rich vegetables such as carrots, but 43 percent eat vitamin A rich fruits. Girls report eating more meat than boys (26 percent and 20 percent, respectively). Yet only 12 percent of mothers report preparing meat.

#### Exhibit 14. Student Food Security

Indicator	Percentage	Observations
Children that felt full after the meal s/he ate before going to school	96%	293
Children that felt full after eating lunch at home	94%	282
Boys receiving minimum acceptable diet <sup>1</sup>	56%	156
Girls receiving minimum acceptable diet <sup>1</sup>	63%	168
Households receiving minimum acceptable diet <sup>2</sup>	35%	343

<sup>1</sup>Minimum acceptable diet is defined as having a minimum of five different foods per day as reported by children

<sup>2</sup>Minimum acceptable diet is defined as having a minimum of five different foods per day as reported by parents

#### 5.1.3 Attendance and Reading Test

Attendance data was collected for each child<sup>23</sup> in the sample from the school registries. IMPAQ and CRS selected three representative months, November, January and April, across the agricultural crop cycle and the academic school year. Many teachers do not record these data and, when they did, the IMPAQ team could not ascertain how rigorously. Even with the best intent, it is possible that teachers omitted to record missing children. In this context, the average attendance rate of 100 percent in Exhibit 15 is likely an overestimation of how regularly students attend classes. It follows that figures reported for project performance indicator 9 in Exhibit 1 is likely inflated as well.

<sup>22</sup> Food and Agriculture Organization (2010). *Guidelines for Measuring Household and Individual Dietary Diversity*. Rome, Italy: United Nations.

<sup>23</sup> Note that the sample size for attendance does not exactly match the sample size of the performance evaluation. This is because incomplete records were dropped from the performance evaluation data, and IMPAQ could not match the student and the attendance data. IMPAQ kept all the representative attendance data for Exhibit 15.



### Exhibit 15. Student Attendance

Indicator	Percentage	Observations
All students regularly attending class days	100%	510
Boys regularly attending class days	100%	219
Girls regularly attending class days	100%	204

Exhibits 16-18 show the distribution of ASER test results. The threshold for an acceptable reading level for a second grader was determined according to the Burkina curriculum guidelines and the calibration workshop IMPAQ and CRS held in April 2015 (please refer to Exhibit 9 for the map of test levels to each elementary school grade in Burkina). The outcome of the calibration is that level C in the ASER test is the minimum acceptable level for a second grader. It corresponds to the demonstrated ability to read complex sounds.

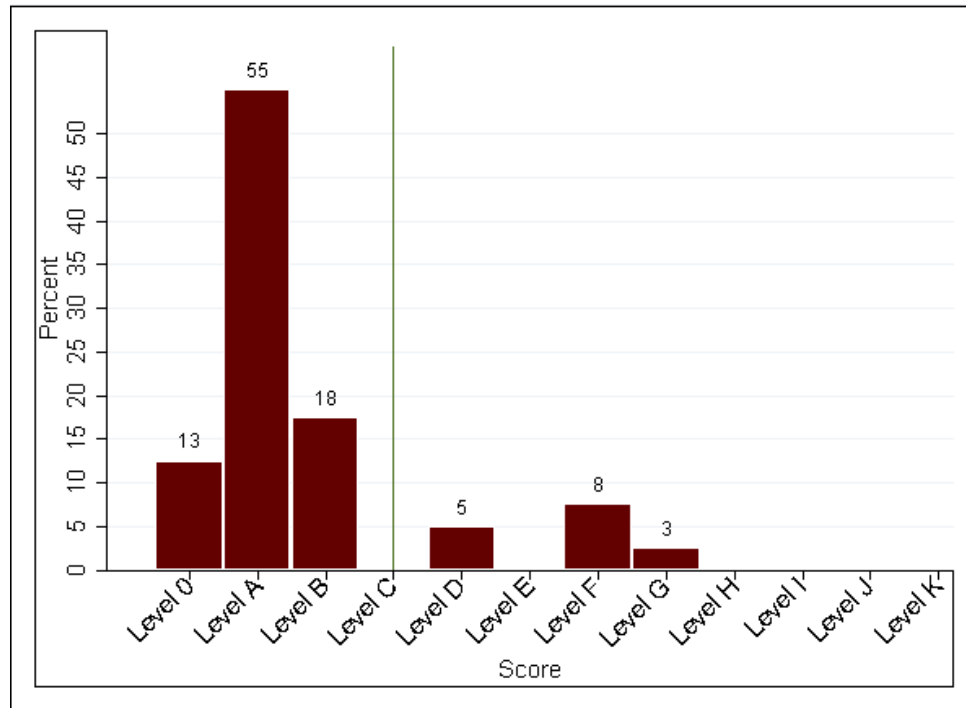
The data shows that most second graders do not achieve grade level expectations in reading, and there is no significant difference between boys and girls. Only 15 percent of the sample can read complex sounds. The distribution of second-graders by the level achieved in the test, in Exhibit 18, shows that most students who achieve level C also manage to read simple words (16 percent of students reach level D).

There are differences between boys and girls in the distribution of achievement levels (Exhibit 18). Many more boys (25 percent) than girls (4 percent) cannot read at all (level zero). If we choose level B as the reading achievement threshold (please see Exhibit 53 in Appendix B) more girls than boys reach that level and the difference is significant. Many more girls can read at least simple sounds than do boys.

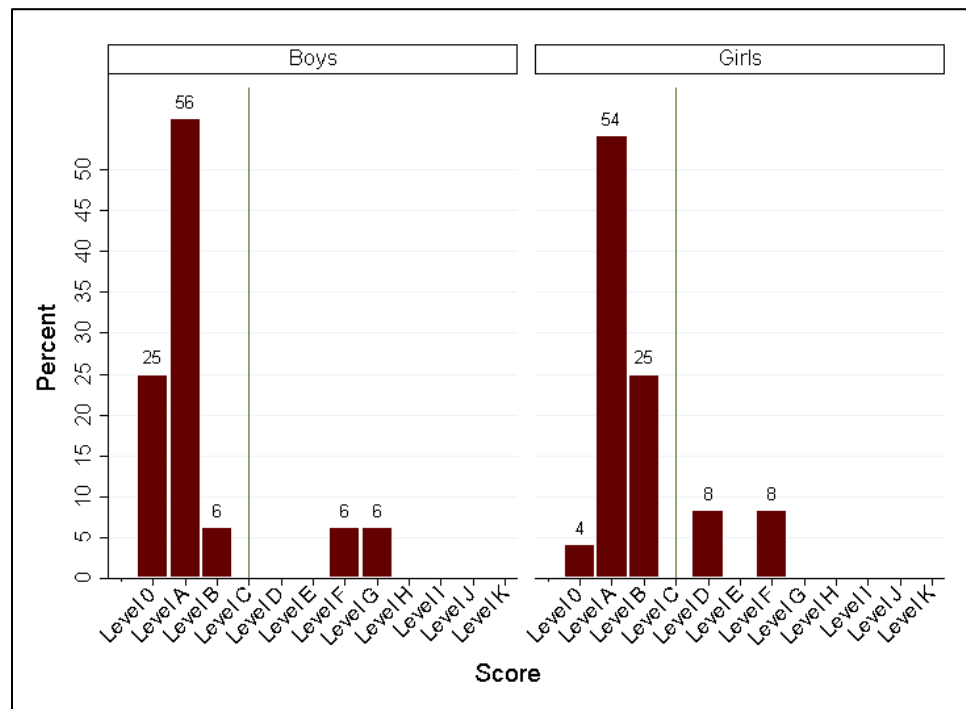
### Exhibit 16. Students Reading at Second Grade Level (ASER Level C: Reading Complex Sounds)

Indicator	Percentage	Observations
Second grade students demonstrating reading ability at grade level or above	15%	40
Male students demonstrating reading ability at grade level or above	12%	16
Female students demonstrating reading ability at grade level or above	17%	24

**Exhibit 17. ASER-Reading Test Results, All Students (Grade 2)**



**Exhibit 18. ASER-Reading Test Results, Disaggregated by Sex (Grade 2)**



As shown below in Exhibit 19, most of the students in this sample, 60 percent belonged to a reading group. Most families though (98 percent) are not aware of the location of a nearby library and have not visited it.

### Exhibit 19. Student Reading Groups

Indicator	Percentage	Observations
Students in a reading group	60%	325
Reading groups that consist of the entire class	19	195
Students that visit library with their parents	2%	343

Exhibit 20 shows that students' knowledge of nutrition is low. Only 32 percent and 9 percent of students know that foods contain vitamin A and iron, respectively. Of those who said they had heard of vitamin A and iron, less than half of these students can actually cite a food that contains them, or cite a benefit of either vitamin A or iron. For example, of the students who had heard of vitamin A, just 41 percent could cite one of its benefits.

In order to aggregate these data and determine a "passing score" on nutrition, IMPAQ counted the number of students who could name at least one food containing vitamin A, one food containing iron, one benefit of vitamin A and one benefit of iron. Based on this criteria, only less than 1 percent of students have basic knowledge of health and nutrition, across all grade levels.

### Exhibit 20. Student Knowledge of Nutrition

Indicator	Percentage	Observations
Students that have heard of vitamin A	32%	324
Students that can cite one benefit of vitamin A <sup>1</sup>	41%	103
Students that can cite a food containing vitamin A <sup>1</sup>	23%	92
Students that have heard of iron	9%	324
Students that can cite one benefit of iron <sup>1</sup>	32%	28
Students that can cite a food containing iron <sup>1</sup>	29%	28
Students that achieve a passing score on a test of food nutrition and dietary practices	1%	324

<sup>1</sup>Smaller sample size comprises only students who declared they had heard of vitamin A/iron.

Exhibit 21 shows that students' have solid basic hygiene practices with 94 percent reporting that they washed their hands on the day they were asked. A smaller proportion of students, 66 percent, reported that they wash their whole body daily (self-reported), but this may be more representative of limited access to water (see Appendix A) than lack of awareness of hygiene. IMPAQ defined a threshold to determine if students are knowledgeable of hygiene. Students are considered knowledgeable if they can identify at least four situations in which they should wash their hands. Although many students reported good hygiene practices, very few (5 percent) passed this test of knowledge of hygiene.

## Exhibit 21. Student Hygiene Practices

Indicator	Percentage	Observations
Students that washed their hands today	94%	324
Students that washed their body today	66%	324
Students that achieve a passing score on a test of good health and hygiene practices	5%	324

## 5.2 Other Student Environment Outcomes

### 5.2.1 Parents and School

Exhibits 22, 23, and 24 display outcomes related to information collected in the surveys conducted with mothers. These measures were taken because students are largely influenced by their parents, and mothers make decisions regarding their children's meals.

## Exhibit 22. Parent Knowledge of Nutrition

Indicator	Percentage	Observations
Parents that have heard of vitamin A	35%	343
Parents that can cite one benefit of vitamin A <sup>1</sup>	60%	120
Parents that can cite a food containing vitamin A <sup>1</sup>	59%	120
Parents that have heard of iron	8%	343
Parents that can cite one benefit of iron <sup>1</sup>	78%	23
Parents that can cite a food containing iron <sup>1</sup>	28%	29

<sup>1</sup>Smaller sample size comprises only parents who had heard of vitamin A/iron.

Exhibit 22 shows that parents lack knowledge of nutrition. Parents' knowledge of nutrition are at levels similar to that of students: Only 8 percent of parents have heard of iron and 35 percent have heard of vitamin A. Contrary to students though, most parents who have heard of iron or vitamin A are able to cite their benefits.

## Exhibit 23. Parents and School

Indicator	Percentage	Observations
Parents that are members of the PTA	13%	343
Parents that are members of a savings group	9%	220
Parents that have used their savings for school expenses <sup>1</sup>	48%	21
Parents indicating the savings helped a lot <sup>2</sup>	44%	9
Parents indicating the savings helped some <sup>2</sup>	11%	9
Parents indicating the savings helped a little <sup>2</sup>	44%	9

<sup>1</sup>Smaller sample size comprises only parents who are members of a savings group

<sup>2</sup>Smaller sample size comprises only parents who used their savings for school expenses

Exhibit 23 shows that the majority of parents are not involved in the school, with only 13 percent being active in the PTA. Very few (9 percent) parents are involved in a savings group (SILC). On the other hand, showing the potential educational benefits of the saving groups, 48 percent of parents who are in a SILC say they used the funds to cover educational expenses and 44 percent of these find it helpful.

#### Exhibit 24. Parents' Opinions of Girls' Schooling

Indicator	Percentage	Observations
Parents indicating girls' schooling is good	79%	343
Parents indicating girls' schooling is bad	2%	343
Don't know	20%	343

#### Exhibit 25. Parents' Opinions of Why Girls' Schooling is Good

Indicator	Percentage	Observations
Proportion of parents who say educating girls improve living standards	71%	270
Proportion of parents who say educating girls improve their health	18%	270
Proportion of parents who says educating girls improve the health of their future children	19%	270
Proportion of parents who say girls should be able to fulfill themselves	38%	270
Proportion of parents that says educating girls allows them to find better work	36%	270

Parents view girls going to school positively, as shown in exhibit 24, with 79 percent indicating it is good and only 2 percent indicating it is bad. Confirming this result, 71 percent of the parents who have a positive view of girls' education say that they believe it would improve their living standards, and 36 percent believe that school will help girls get a better job (Exhibit 25).



IMPAQ/AGIR enumerator interviewing a mother in Simildougou. Photo: IMPAQ

### 5.2.2 Parent-Teacher Association Representatives

Community involvement in the school, starting with key Parent Teacher Association representatives, is also an indication of the quality of the learning environment. Exhibit 26 shows that an average of 3 General Assemblies were held annually and were all attended by the PTA member who was interviewed.

Although only around half (52 percent) of PTA representatives interviewed (Exhibit 27) help directly with preparing school meals, most are involved in other ways (for example, 92 percent help with cleaning the school.)

#### Exhibit 26. Parent - Teacher Associations Activities

Indicator	Mean	Observations
Number of general assemblies held past school year	3	21
Number of PTA Council meetings held past school year	4	21

#### Exhibit 27 PTA representative involvement

Indicator	Percentage	Observations
Proportion of PTA representatives who had a meal with students while in school?	4%	25
Proportion of PTA representatives who attended a general assembly	100%	25
Proportion of PTA representatives who participate in a school community projects such as cleaning	92%	25
Proportion of PTA representatives who help the school as cook or store keeper	52%	25
Proportion of PTA representatives who help watch over a reading group	0	25
Proportion of PTA representatives who visit the classroom	76%	25
Proportion of PTA representatives who attend a performance put on by the Children in Class	44%	25
Proportion of PTA representatives who help in other ways	8%	25

The PTA representatives in the sample were asked about how often the canteen operated in their school. More specifically, they told how many months per year the canteen operated, in order to inform Indicator 16 in Exhibit 2. Only fourteen of the twenty-five PTA representatives in the performance evaluation sample answered the question. The average number of months they reported the canteen operated (regardless of who provided the meals, including the government or the community) is 5.3. IMPAQ fears that this is a misleading representation of canteen operations because it does not take into account weeks when the canteen works only for a few days. This is also not consistent with other sources that suggest that the government was able to cover two months of canteen at most.

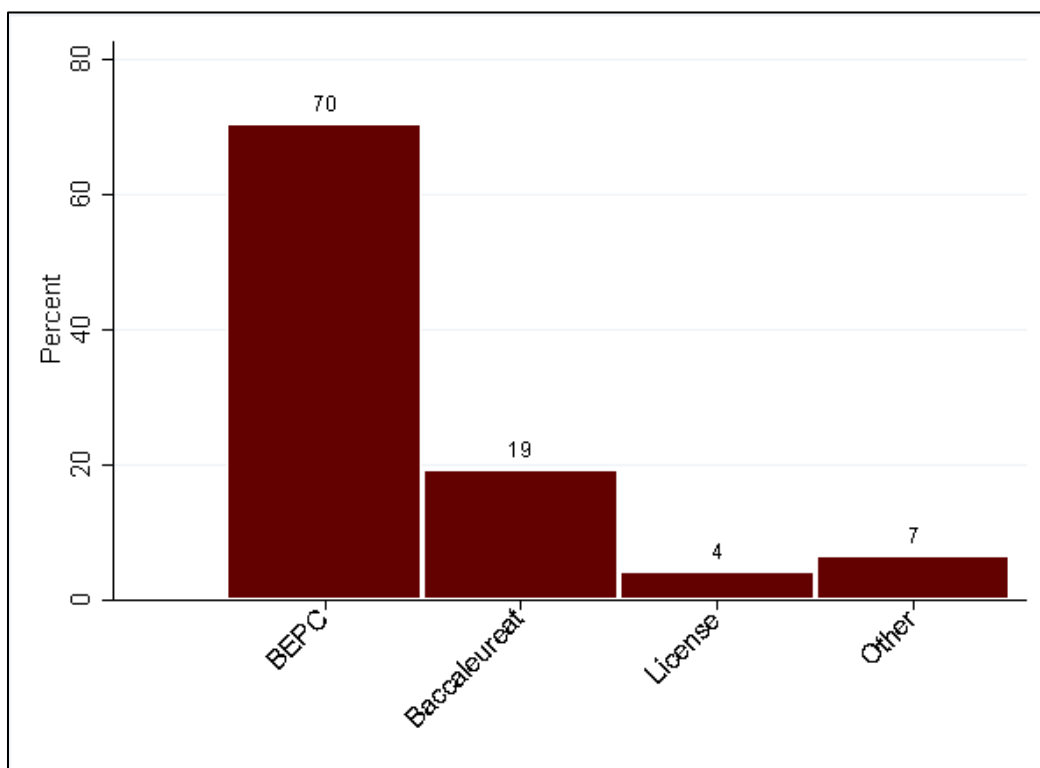
## 5.3 Teacher Outcomes

The teacher outcomes concern teachers' practices and how their classroom practices reflect the support and trainings that they receive from BBII. The first categories of outcomes relate to nutrition, health, and hygiene. These categories are measured in the teacher survey, with questions about their content knowledge and their classroom practices and they inform indicator 6. Additionally, IMPAQ collected data about outcomes related to the literacy component of BBII, as in indicator 2. IMPAQ collected this information in the teacher questionnaire, in which teachers report their literacy instruction practices. Furthermore, the student questionnaire also asks students about the literacy instruction they receive in the classroom. IMPAQ will cross reference this information with the information provided by teachers. The questionnaire also asks teachers about their teaching practices in general, which addresses indicator 6.

### 5.3.1 Teacher Characteristics

Exhibit 28 shows that the majority of teachers only have a Brevet d'Etudes du Premier Cycle (BEPC), which corresponds to a 10<sup>th</sup> grade level of education. In other words, the majority of teachers did not graduate from high school.

**Exhibit 28. Educational Attainment of Teachers**

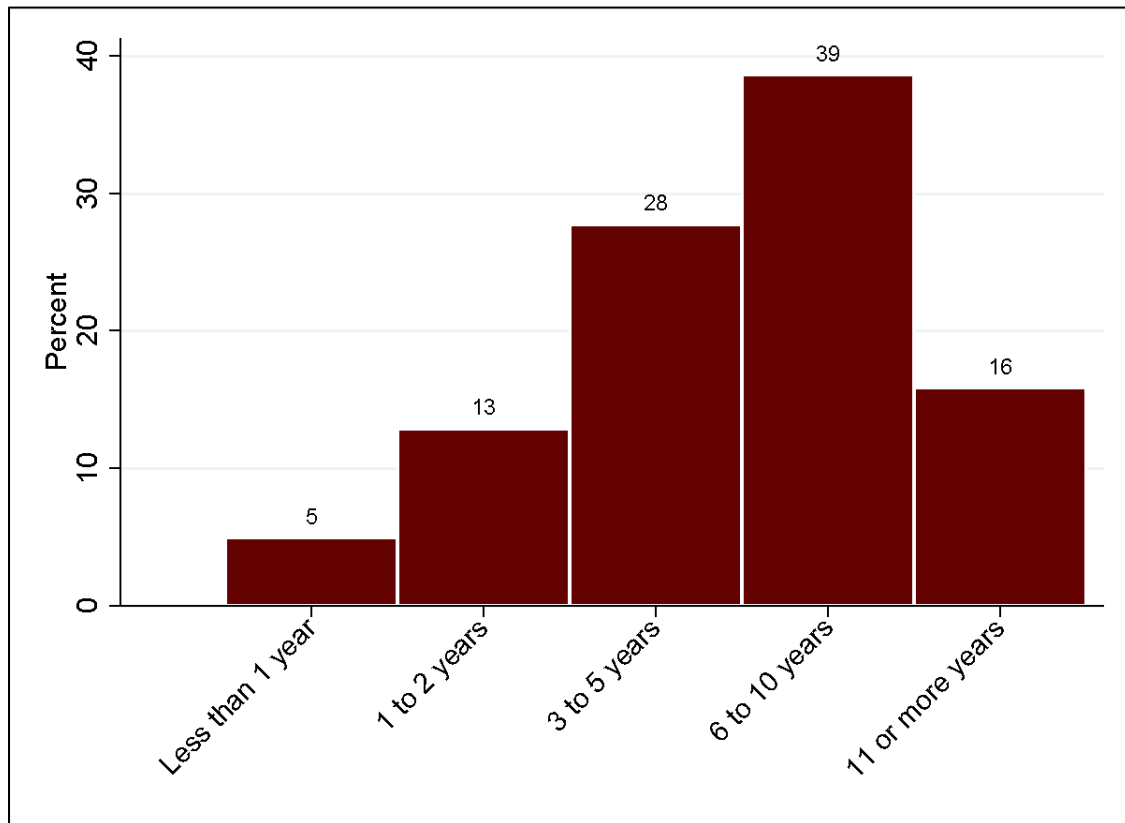


While almost all (96 percent) have received some teacher training, only 16 percent have received training related to teaching reading skills (Exhibit 29). Meanwhile, their average class size is 49 pupils (Exhibit 31). These numbers point to the policy challenge regarding increasing students' performances: teachers have low qualifications and receive limited support, while facing large class sizes. Yet, despite these challenges, most (52 percent) teachers have more than six years of experience (Exhibit 30).

**Exhibit 29. Teacher Training**

Indicator	Percentage	Observations
Teachers that have not received teacher training	4%	101
Teachers that have not received literacy training	84%	89

**Exhibit 30. Years of Teaching Experience**





### Exhibit 31. Classroom Size

Indicator	Students
Average number of students per classroom, all grades	50
Average number of students in CP2 classrooms	55
Average number of students in CE1 classrooms	51
Average number of students in CE2 classrooms	46
Average number of students in CM1 classrooms	47
Average number of students in CM2 classrooms	46

#### 5.3.2 Classroom Practices

##### 5.3.2.1 General Teaching Techniques and Classroom Activities

Exhibit 32 shows the diversity of activities a teacher does in a day or a week. Teachers who use a variety of activity types, including switching between activities that are teacher centered, student centered and group activities, are more likely to be effective at teaching children with a diversity of learning styles. Varying activities is also more likely to keep students engaged as a whole. More than two third of teachers (not necessarily the same for each activity) indicate having students check each other's work, have the whole class check the work of a student (most likely on a black board), have students of different skill level work together, have students write solutions on a slate as share with the class. The last activity, somewhat traditional, is the most popular. This shows a diversity of approaches and suggests that most teachers use at least two or three types of activities daily. Yet some activities, such as students assessing their own work, or students paired by level, are still relatively unpopular, even when the time frame is expanded to a whole week.

The sixth McGovern Dole key program indicator in Exhibit 2 is about teachers who demonstrate the use of new and quality teaching techniques. IMPAQ defined indicator 6 as “teachers who use at least 5 of the 7 activities” listed in Exhibit 30 on a daily basis. Exhibit 33 shows that 53 percent of teachers demonstrate the use of new techniques or tools.

Exhibit 32 shows the diversity of activities a teacher does in a day or a week. Teachers who use a variety of activity types, including switching between activities that are teacher-centered, with others that are student-centered or group activities, are more likely to be effective at teaching children with a diversity of learning styles. Varying activities is also more likely to keep students engaged as a whole. More than two



IMPAQ/AGIR enumerator interviewing a CP2 teacher in Yargo Yarce. Photo: IMPAQ

thirds of teachers indicate that in a given day, they have students check each other's work, have the whole class check the work of a student (most likely on a black board), have students of different skill level work together, have students write solutions on a slate as share with the class. This last activity, somewhat traditional, is the most popular. Yet some activities, such as students assessing their own work, or students paired by level, are still relatively unpopular, even when the time frame is expanded to a whole week.

### Exhibit 32. Teacher Reported Utilization of Classroom Activities

Utilization	Today	This Week
Individual student checks his or her own work and gives himself/herself a mark/comments		
Already used	23%	27%
Planning to use	13%	20%
Have not and not planning to use	64%	53%
Students check each other's work		
Already used	46%	67%
Planning to use	22%	17%
Have not and not planning to use	33%	16%
The whole class checks the work of a student		
Already used	43%	62%
Planning to use	25%	21%
Have not and not planning to use	33%	17%
Students write solutions on a slate and show to the teacher and class		
Already used	64%	77%
Planning to use	19%	18%
Have not and not planning to use	17%	5%
Students of different skill levels are paired together		
Already used	50%	69%
Planning to use	29%	24%
Have not and not planning to use	22%	7%
Students of the same skill level are paired together		
Already used	27%	41%
Planning to use	22%	16%
Have not and not planning to use	51%	44%
Teacher asks group (3 or more) of students to work together on a project and later provides feedback on group performance		
Already used	39%	48%
Planning to use	21%	19%
Have not and not planning to use	41%	34%

### Exhibit 33. Teachers Using New Techniques or Tools

Indicator	Percentage	Observations
Teachers that demonstrate use of new techniques or tools	53%	101

#### 5.3.2.2 Reading Instruction

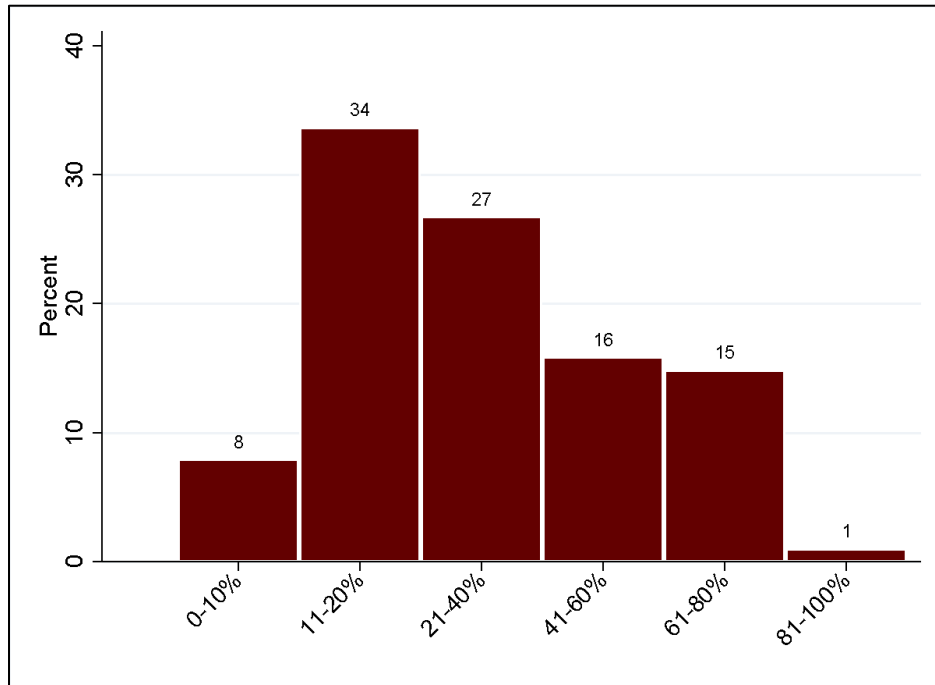
Turning to teaching reading, and literacy skills in general, Exhibits 34 and 35 provide different information on teacher practices. The indicator for Exhibit 34 was built from the questionnaire questions that ask teachers to give examples of activities that they use to teach reading. If a teacher mentioned a key reading activity (phonemic, phonics, fluency, vocabulary, identification of key words or key sounds...), they were then asked how many minutes they do such an activity on average, per day. They were asked this four times (unless they did not have an answer – only 24 teachers mentioned up to four activities) and the total time was added. This indicator reflects both teachers' understanding and practices of literacy instruction. It shows that 55 percent of teachers in the sample cited key reading activities on which they spent, in total, at least 45 minutes per day (10 teachers could not cite any and were not counted here.)

### Exhibit 34. Teachers Spending at Least 45 Minutes per Day on Literacy Instruction

Indicator	Percentage	Observations
Proportion of teachers who devote at least 45 minutes a day on literacy instruction	55%	91

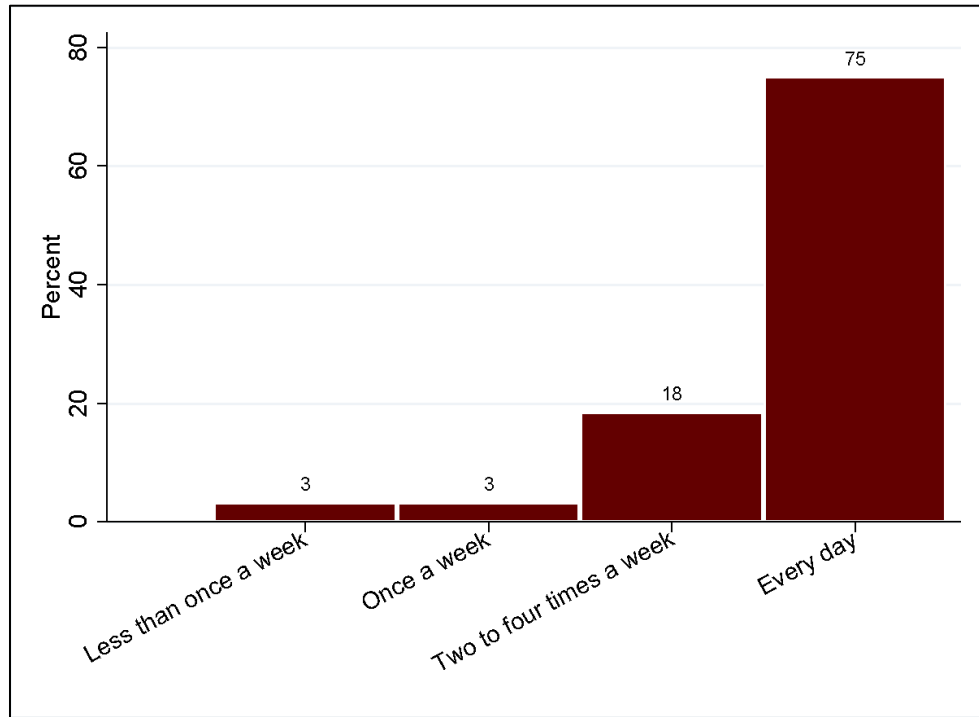
The variable for Exhibit 35, by contrast, is a question asking teachers directly how much time they spend teaching reading every week. This is a broader question and it does not require teachers to cite specific techniques. This is why it shows teachers engaging students in reading related activities more often than Exhibit 34.

**Exhibit 35. Average Time Spent Per Week on Literacy Related Activities**

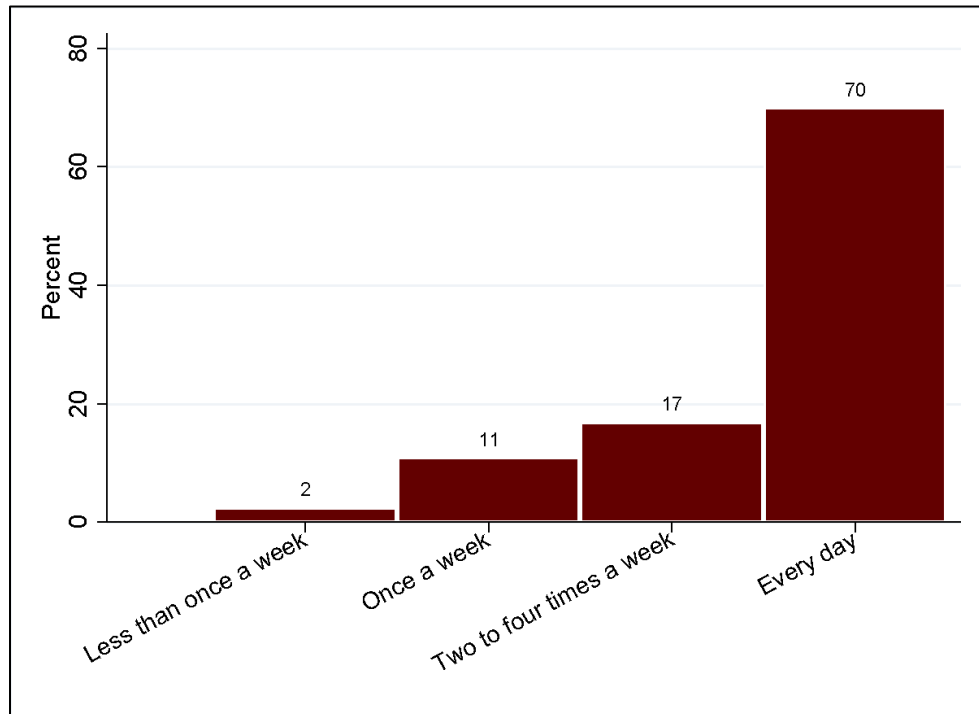


Exhibits 36, 37 and 38 show how often teachers ask their students to read. The last column of the graphs, for activities every day, is telling: 75 percent of teacher report that students read by themselves every day, 70 percent report that students read to others, and 68 percent that students read aloud. On the other hand, less than half report that students write or do word play every day (Exhibits 39 and 40). These are important literacy activities although they maybe are more taxing on the teacher because they require more preparation or more feedback. Lastly, Exhibit 41 shows literacy activities as reported by students. The student reported information supports the information provided by teachers.

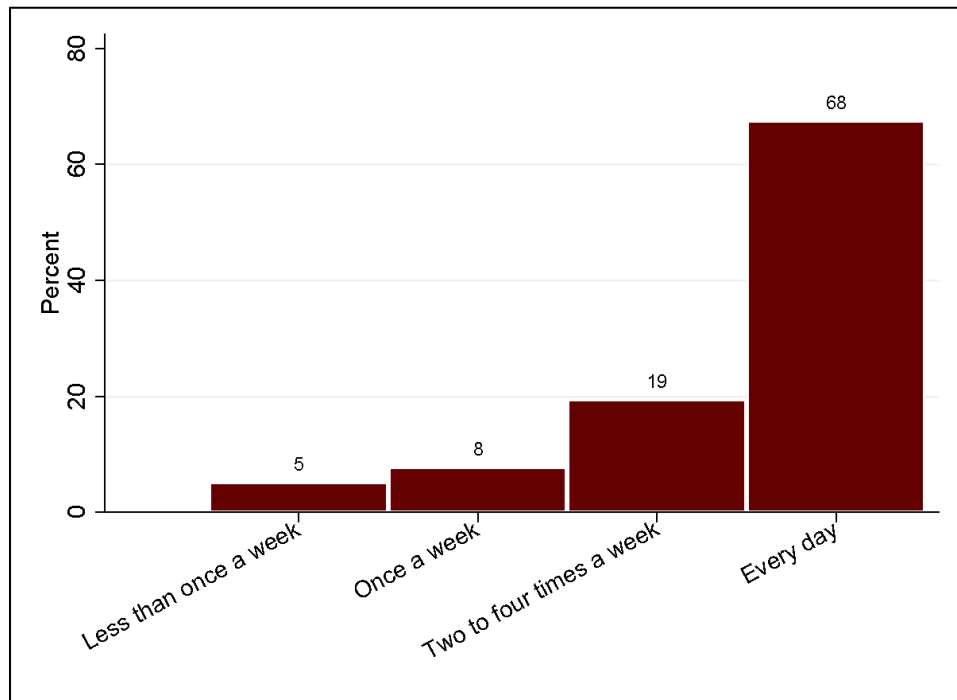
**Exhibit 36. How Often Does the Teacher Usually Have Students Read by Themselves?**



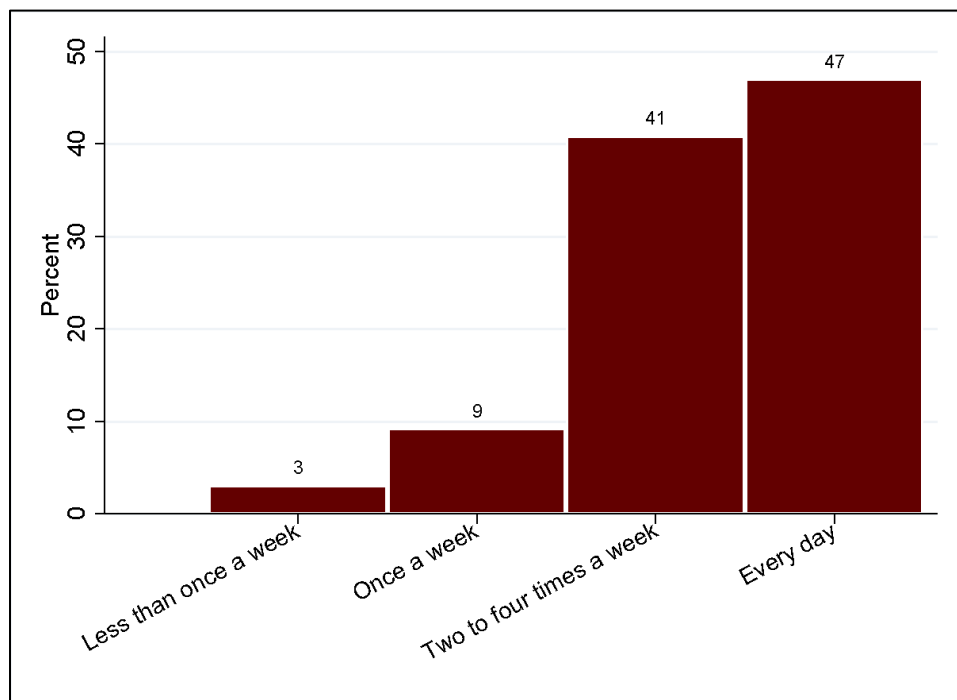
**Exhibit 37. How Often Does the Teacher Usually Have Students Read to Others?**



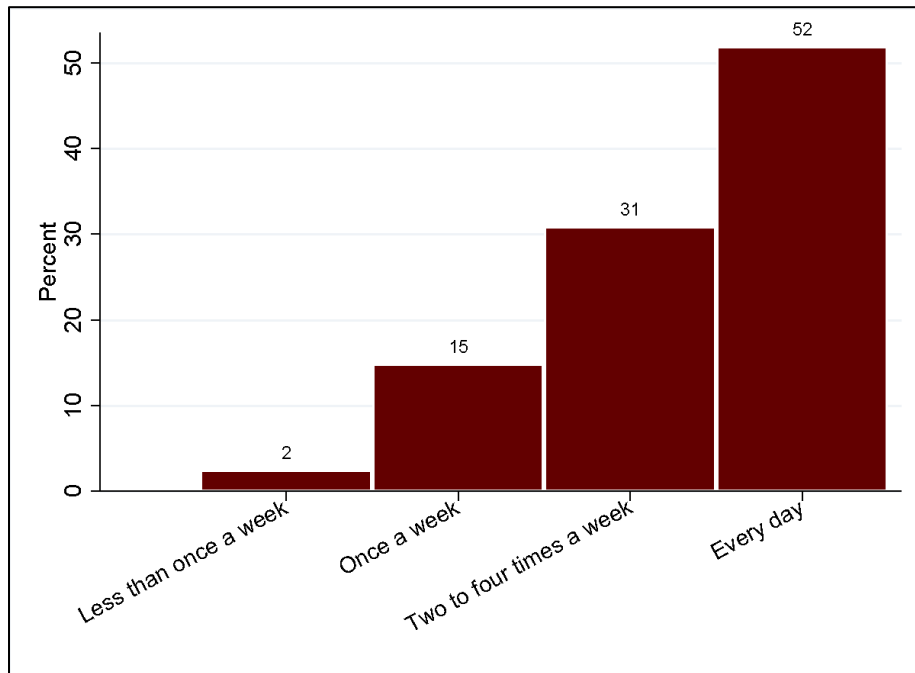
**Exhibit 38. How Often Does the Teacher Usually Have Students Listen to Others Read Aloud?**



**Exhibit 39. How Often Does the Teacher Usually Have Students Write?**



**Exhibit 40. How Often Does the Teacher Usually Have Students Do Word Play?**



**Exhibit 41. Student Reported Classroom Activities**

Indicator	Percentage	Observations
Spent time reading alone in class yesterday	63%	325
Spent time reading to someone else in class yesterday	62%	325
Spent time writing in class yesterday	71%	325
Spent time doing word work in class yesterday	43%	325
Spent time listening to reading in class yesterday	72%	325

### 5.3.3 Knowledge of Nutrition

Teachers are well aware of the benefits of vitamin A and iron and of which foods contain them. Among the 96 teachers who say they have heard of vitamin A, 78 percent can cite at least one benefit of it, and 71 percent can cite a food that contains it. The figures are similar with iron. This confirms that most teachers do not lack the knowledge necessary to increase students (and community) nutrition awareness, but may need support on how to teach it effectively.

#### Exhibit 42. Teachers' Knowledge of Nutrition

Indicator	Percentage	Observations
Teachers that have heard of vitamin A	95%	101
Teachers that can cite one benefit of vitamin A	78%	96
Teachers that can cite a food containing vitamin A	71%	77
Teachers that have heard of iron	98%	101
Teachers that can cite one benefit of iron	69%	87
Teachers that can cite a food containing iron	75%	69

#### 5.3.4 Attendance

CRS and the AGIR team collected official teacher attendance information from the District Managers for 20 of the 22 performance evaluation schools. The outcome of interest is teachers who are absent more than 10 percent of workdays, or 2 school days per month. IMPAQ gathered attendance data for a sample of 120 teachers of second to six grade classes in the 20 schools in February, March and April 2015. Exhibit 43 shows that attendance rate varies and that, on average over the three months, 28 percent of teachers missed more than 2 days per month. Thus, 85 percent of teachers attend and teach school at least 90 percent of scheduled school days per year (Indicator 3).

#### Exhibit 43. Proportion of Teachers Who Were Present at least 90 Percent of School Days

February 2015	March 2015	April 2015
76%	92%	86%

#### 5.4 Districts Managers

The AGIR supervisor conducted qualitative interviews with 13 District Managers or Chef de Circonscription d'Education de Base (CCEB), in order to determine their level of confidence with training teachers in the methods of classroom management and literacy instruction. The team determined that the best way to get at this information was through a key informant interview instead of directly asking sensitive questions.

Only one District Manager interviewed was a woman and on average, they had 7 years of professional experience. Regarding learning styles, which is an important pedagogical factor in increasing students' learning, Exhibit 44 shows that although all District Managers mention "watching", which refers to a traditional, teacher centered way of teaching, very few mention "sensing and feeling", which is particularly important for younger children. Strikingly, "thinking" is mentioned only by 8 of the 13 Managers. The interviews also probed into the District Managers' recommendations for teaching reading. Exhibit 45 shows that, although all District Managers mentioned phonics, only 8 of the 13 mention phonemics (hearing, identifying and



manipulating phonemes), which is considered to be the basis for learning phonics (hearing, identifying and manipulating larger units of sound).

**Exhibit 44. Learning Styles Mentioned by District Managers**

Type of style	Number of District Managers who mentioned	Total number of observations
Sensing and feeling	5	13
Watching	13	13
Thinking	8	13
Doing	11	13

**Exhibit 45. Techniques to Teach Literacy Mentioned by District Managers**

Technique	Number of District Managers who mentioned	Total number of observations
Phonemic	8	13
Phonics	13	13
Fluency	13	13
Vocabulary	8	13
Comprehension	11	13

The interviews revealed that most District Managers are engaged with teachers in their districts, both through training workshops and direct classroom observations. During the academic year, 8 District Managers had held at least one training session about reading skills. IMPAQ defined indicator 4 (school administrators demonstrating use of new techniques or tools) as District Managers transferring skills to others through trainings. Thus, the baseline level for indicator 4 is 61 percent.

During the academic year, 9 District Managers had observed classrooms. In classroom observations, District Managers report that 80 percent of the teachers they observe teach literacy, through at least one of the “daily five”, that is writing, listening to others reading, reading to others, reading alone and playing with words. In general feedback about teaching literacy, District Managers expressed the need for more support for teachers, with teaching materials as well as workshops to learn and share with others. Another theme that emerged is the question of the motivation and the mobility of teachers.

## VI. RCT VARIABLES: BASELINE LEVELS AND BALANCE TEST

The impact evaluation is focused on the expected educational effects of the mentoring program, specifically with respect to participants' reading skills (indicator 1 in Exhibit 2) and attendance. IMPAQ will track students and measure their reading skill attainment at endline. Because student literacy outcomes are important to the BBII program, the impact evaluation will also estimate the program's impact on student participation in reading groups outside of school.

The student questionnaire also collects information on other learning outcomes that might be positively affected by the mentoring program, specifically whether students understand the value of iron and vitamin A and hygiene practices. The mentor's home visits might also lead parents to be more cognizant of their children's diets; IMPAQ will look at possible impacts on the child's perceived food security. Together, these address indicators 8, and 12 to 15 in Exhibit 2.

The tables in this section show balance checks for key variables pertaining to the impact evaluation of the mentoring program. Recall that the sample of this evaluation at baseline is second grade students in 44 schools. Balance checks are important because they show that students in the control and treatment groups are similar on average, so control group students can be used as counterfactual, and represent the students who received the intervention, *had they not received it*. The criteria of interest to judge of the balance test is the p-value of the statistical test of the difference of the means. A p-value above 0.10 indicates that there is no statistically significant difference between the control and treatment groups at baseline.

The tests indicate that the treatment group is no different from the control group with regards to most outcomes, particularly the literacy outcomes. However, IMPAQ found some statistically significant differences in one hygiene and in the dietary diversity outcomes. Inspecting the data closer points to some possible variations in water access for the hygiene outcome and the reliability of child reported dietary data for the dietary diversity outcome. Both can be controlled for in the analysis.

### 6.1 Second Grade Student Characteristics and Food Security

The sample is made up of 761 second-graders (surpassing the target of 616 in the sample design), with more girls in the control schools than in the treatment schools (Exhibit 46). This in itself is not an issue. Our goal is to have a representative sample of boys and girls in both the control and the treatment groups and results will be disaggregated by sex.

**Exhibit 46. Second Grade Student Characteristics**

Indicator	Control group	Treatment group	P-value of t-test
-----------	---------------	-----------------	-------------------

	Mean	Observations	Mean	Observations	
Proportion of girls	0.53	380	0.49	381	0.01

As with the performance evaluation sample, Exhibit 47 shows that most students (on average 86 percent) eat before going to school and there is a slight significant difference between control and treatment groups (89 percent and 82 percent respectively) at the marginal level (10 percent). Almost all who eat are satiated. There are more girls who eat at the school canteen in the treatment group than in the control group, but there is no particular reason why this should have an effect on the outcomes. It can be controlled for in robustness checks in the impact analysis at endline. There is a significant difference between control and treatment groups for the minimum acceptable diet variable, as reported by the students themselves. However, a closer look at the data in Exhibits 58 and 59 in Appendix B shows a lot of heterogeneity in the data from village to village, with no clear pattern emerging from a comparison of the control and treatment villages. Furthermore, the performance evaluation data shows that dietary reports from children are quite different from that of their mothers'. We cannot perform such a comparison with the RCT data because we did not interview mothers.

**Exhibit 47. Second Grade Students Food Security**

Indicator	Control group		Treatment group		P-value of t-test between the means
	Mean	Observations	Mean	Observations	
Proportion of children that pass the minimum acceptable diet	0.56	378	0.68	381	0.01
Girls	0.61	199	0.72	184	0.04
Boys	0.50	178	0.64	196	0.05
Proportion of children that ate at home before coming to school	0.89	378	0.82	381	0.07
Girls	0.91	199	0.83	184	0.05
Boys	0.86	178	0.82	196	0.30
Proportion of children that ate at the canteen while at school	0.21	378	0.35	381	0.10
Girls	0.21	199	0.38	184	0.07
Boys	0.21	178	0.33	196	0.22
Proportion of students that could have eaten more before going to school	0.04	336	0.05	314	0.54
Girls	0.04	182	0.04	153	0.87
Boys	0.03	153	0.06	160	0.34
Proportion of children that could have eaten more after lunch in the canteen	0.09	85	0.10	136	0.98

Indicator	Control group		Treatment group		P-value of t-test between the means
	Mean	Observations	Mean	Observations	
Girls	0.09	45	0.13	71	0.62
Boys	0.10	40	0.06	64	0.48

## 6.2 Second Grade Students Attendance and Literacy

Exhibits 48 and 49 show there are no significant difference between the control and the treatment groups in attendance and literacy outcomes. Note that we do not find a significant difference between boys and girls who can read at least simple sounds (ability to read at grade level at least), as opposed to what we found in the (smaller) performance evaluation sample.

**Exhibit 48. Second Grade Student Attendance**

Indicator	Control group		Treatment group		P-value of t-test between the means
	Mean	Observations	Mean	Observations	
Total number of times a student was absent over 3 months (November, January, April)	0.09	334	0.06	379	0.44

**Exhibit 49. Second Grade Literacy Activities**

Indicator	Control group		Treatment group		P-value of t-test between the means
	Mean	Observations	Mean	Observations	
Proportion of students who can read simple sounds at least (ASER level B)	0.25	104	0.34	98	0.31
Girls	0.28	61	0.28	43	1.00
Boys	0.21	43	0.38	55	0.07
If the child is in a reading group	0.53	378	0.52	381	0.80
Girls	0.56	199	0.53	184	0.60
Boys	0.50	178	0.51	196	0.89
Reading group consists of all children in the classroom	0.12	202	0.12	198	0.95
Girls	0.14	112	0.10	97	0.40
Boys	0.09	89	0.13	100	0.47
Reading group consists of some of the students in the class	0.88	202	0.88	198	0.95
Girls	0.86	112	0.90	97	0.40

Indicator	Control group		Treatment group		P-value of t-test between the means
	Mean	Observations	Mean	Observations	
Boys	0.91	89	0.87	100	0.47

## 6.3 Second Grade Students Other Learning Outcomes

Children in the treatment group are more likely to have washed themselves in the morning (Exhibit 50). Boys in the treatment group are more likely to have washed their body in the morning.

### 6.3.1 Hygiene

Students are not different in the control and treatment groups, and, as with the performance evaluation sample, most of students (90 percent) have washed their hands at least once the day of the interview. There is however, a significant difference between the proportion of children who washed their body in the control and the treatment groups, for both boys and girls. This difference could be caused by varying conditions of access to water, or simply with issues of quality of the children-reported data. Exhibits 56 and 57 in the Appendix B show data that vary a lot across villages. Particularly, some villages seem to be outliers and show quite small proportions of children who wash their hands, as in [REDACTED], for example. In any event, these significant differences will be controlled for in the regression analysis of the impact of the mentoring program.

**Exhibit 50. Second Grade Student Hygiene Outcomes**

Indicator	Control group		Treatment group		P-value of t-test Between the means
	Mean	Observations	Mean	Observations	
Proportion of children that washed themselves in the morning	0.60	378	0.72	381	0.01
Girls	0.66	199	0.80	184	0.01
Boys	0.54	178	0.64	196	0.07
Proportion of children that washed their hands that day	0.90	378	0.91	381	0.49
Girls	0.89	199	0.91	184	0.44
Boys	0.91	178	0.91	196	0.91

### 6.3.2 Nutrition Knowledge

Exhibit 51 shows that slightly less boys in the treatment group had heard of vitamin A than boys in the control group. It also shows that more girls in the treatment group were able to name a

food containing vitamin A. Exhibit 51 also shows that there are no significance differences between control and treatment groups regarding their nutrition knowledge.

**Exhibit 51. Second Grade Student Knowledge of Vitamin A**

Indicator	Control group		Treatment group		Difference in means
	Mean	Observations	Mean	Observations	p-value of t-test
Proportion of students that have not heard of vitamin A before or were not able to specify anything about vitamin A	0.93	378	0.90	381	0.22
Girls	0.93	199	0.90	184	0.30
Boys	0.93	178	0.89	196	0.30
Proportion of children that were not able to cite a benefit of vitamin A	0.61	62	0.60	96	0.93
Girls	0.59	32	0.60	43	0.93
Boys	0.63	30	0.60	53	0.83
Proportion of children that were not able to cite a food containing vitamin A	0.92	62	0.88	96	0.47
Girls	0.97	32	0.86	43	0.12
Boys	0.87	30	0.89	53	0.82
Proportion of children that have not heard about iron or were not able to cite anything about iron	0.99	378	0.99	381	0.38
Girls	0.99	199	0.99	184	0.51
Boys	0.99	178	0.99	196	0.61
Proportion of children that were not able to cite a benefit of iron	0.82	11	0.77	13	0.77
Girls	0.83	6	0.86	7	0.91
Boys	0.80	5	0.67	6	0.65
Proportion of children that were not able to cite a food containing iron	0.91	11	0.77	13	0.35
Girls	1.00	6	0.71	7	0.18
Boys	0.80	5	0.83	6	0.91

## VII. CONCLUSION AND LESSONS LEARNED

---

IMPAQ designed a methodology and instruments to conduct a performance evaluation of the BBII intervention and an impact evaluation of its mentoring component. For this evaluation, a representative sample of 44 BBII schools was selected, and 22 schools were randomly assigned to receive the mentoring program. The remaining 22 schools will not receive the mentoring program but will receive all other BBII components, as will the other 22 treatment schools. The control schools constitute the sample frame for the performance evaluation. The impact evaluation will measure the causal effect of the mentoring program on girls' literacy levels. The performance evaluation will measure changes in health, hygiene and dietary practices, as well as attendance and enrollment.

The IMPAQ team designed four quantitative questionnaires, one qualitative questionnaire and an ASER-Reading test calibrated for the Burkina Faso educational context. The questionnaires were designed to ask questions about health, nutrition, literacy and community involvement in schools to five types of respondents: students grades 2 to 6, students' mothers, teachers, members of the PTA and District Managers. This document reports on data collected from 44 schools with 920 students, 343 mothers, 140 teachers and 50 members of PTAs and 13 District Managers.

The IMPAQ team gathered a considerable amount of information, with over 1000 variables across all surveys. This report showcases some of the key variables of interest for the evaluation and provides a full display of the data, organized according to the indicators listed by the CRS evaluation plan, in the appendix. The data are providing many interesting insights into the students' environment and their academic performances. The data point to the need for programs such as BBII to raise literacy levels and nutrition awareness of children. A good news is that most children have food before going to school, but many, 16 percent, go to school with an empty stomach. The highlights from the data are:

- Among second-graders, only 15 percent of girls and boys can read at grade level;
- 96 percent of children come to school after having had a meal and are not hungry;
- 86 percent of the children who did not eat a meal before going to school state that this was for the lack of food;
- 32 percent of students know about vitamin A, but only 13 percent can cite one of its benefits;
- 9 percent of students know about iron as a food item, but only 3 percent can cite one of its benefits;
- The treatment and control groups of the RCT samples are balanced on most observables. This means that, for all we can see, the random assignment led to a comparison group of students in schools that are not receiving the mentoring program that is the same on average as students in the schools that are receiving the mentoring program.



The lessons learned from the baseline data collection and analysis concern the BBII communities and programs, and the main points are outlined here. Additionally, there were some methodological lessons about the data collection that can further inform FAS's Mc Govern Dole indicators.

This report and feedback from the data collection process illustrate some challenges encountered when collecting information on children's nutritional intake. In order to address Mc Govern Dole Indicator 23, IMPAQ sought to collect information about the FAO Food Diversity Index recommended by FAS. The Indicator Handbook produced by USDA (draft, 07/15/2014) recommends that these data be collected through surveys with children, at the school, rather than at the household. To measure the index, the



Children at break in [REDACTED] Photo: IMPAQ

Handbook defines a diverse food intake for children as a daily intake of least four food groups out of a list of seven. Because this list does not disentangle iron from vitamin A intake, IMPAQ and CRS agreed to use a full FAO list of 15 foods instead (all except for spices). Given the data collected, IMPAQ's assessment is that data from children is unreliable because many children do not pay attention to what they eat and gender differences are more likely to stem from how much they pay attention rather than actual differences in the meals they take. This report shows large variations which cannot be explained by heterogeneity of children because it does not correspond to nutrition data collected from mothers directly. Although the average number of different food types reported by boys is the same as girls, there are wide variations in what they report (Appendix A, Indicator 12), and the distribution is different, so that many more girls than boys report at least five good groups (Exhibit 14).

This finding suggests there is no simple methodological solution, and in-depth measures of children dietary intakes require specialized enumerators with training in nutrition, which is beyond the scope of this evaluation. Key informant interviews with mothers around nutritional practices (including, where/how they procure foods, how they store and cook it, seasonal variations) would help to better understand locally available food items and how they fit in defined food groups, identify local food terms, discuss issues of food availability, get information on ingredients used in local dishes. The FAO guide recommends that the questionnaire ask respondents to:



“Please describe the foods (meals and snacks) that you ate or drank yesterday during the day and night, whether at home or outside the home. Start with the first food or drink of the morning.” The person conducting the survey writes down all foods and drinks mentioned, asks for the list of ingredients if a mixed dish is mentioned, and then probes for meals and snacks not mentioned. Then to person conducting the survey fills in a chart of the food groups based on what the respondent said and probes on any food groups not mentioned and asks if a food item from this group was consumed.

Although this method should provide rich data, this is a rather extensive process and the quality of the data would depend on the nutritional knowledge of the enumerator. For the evaluation of BBII, the lesson is to not ask nutrition questions to the children in the sample, but rather to their mothers. For the sake of comparison between baseline and midline/endline data, the exact same question should be asked from mothers again. However, additional questions, probing with a more extensive list of local foods, could be added for deeper analysis.

Another methodological lesson learned concerns the canteen information. The information collected from PTA representatives does not appear to represent the reality of canteen meals available to children. IMPAQ and CRS will discuss potentially better approaches.



Children waiting to be interviewed in [REDACTED] Photo: IMPAQ

## VIII. REFERENCE LIST

---

- Abadzi, H. (2011). *Reading fluency measurements in EFA FTI partner countries: Outcomes and improvement prospects* (GPE Working Paper Series on Learning, No. 1). Washington, DC: World Bank.
- Catholic Relief Services. (2015). IMPAQ International, LLC Consulting Agreement and Statement of Work. Ouagadougou, Burkina Faso.
- Food and Agriculture Organization (2010). *Guidelines for Measuring Household and Individual Dietary Diversity*. Rome, Italy: United Nations.
- Pratham. (2015). Pratham Around the World. Retrieved from <http://www.prathamusa.org/programs/pratham-around-world>.
- Simons, K.A. (2012). Discussion Document #1: Multi-Country Assessments of Learning. Washington, DC: Center for Universal Education at the Brookings Institution.
- United States Department of Agriculture. (2014). *Food for progress and McGovern-Dole indicators and definitions*. Washington DC: Food Assistance Division, Office of Capacity Building, Foreign Agricultural Service.
- USAID. (2012). *2011 USAID education strategy reference materials*. Washington, DC.
- Vagh, S.B. (2009). *Evaluating the reliability and validity of the ASER testing tools*. New Delhi: ASER Centre.
- Vanmeenen, G. (2010). *Savings and Internal Lending Communities (SILC) - Voices from Africa*. Nairobi: Catholic Relief Services.

## **APPENDICES**

---

- A. Indicator Variables: Baseline Levels**
- B. Additional detailed data breakdown**
- C. ASER-Reading Test Instructions**
- D. Questionnaires**
- E. Treatment and Control Schools**

## APPENDIX A. INDICATOR VARIABLES: BASELINE LEVELS

Outcome Variable	Mean	Observations
Indicator 1: 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		
Average reading level of second graders	2.72 <sup>24</sup>	40
Proportion of boys that can read and understand the meaning of grade level text	0.12	16
Proportion of girls that can read and understand the meaning of grade level text	0.17	24
Indicator 2: Number of teachers who devote at least an average of 45 minutes a day to literacy instruction		
Average number of minutes teachers spent on phonemics	5	3
Average number of minutes teachers spent on phonics	30	9
Average number of minutes teachers spent on fluency	30.61	59
Average number of minutes teachers spent on vocabulary	17.92	26
Average number of minutes teachers spent on comprehension	21.82	33
Average number of minutes teachers spent on identification of key words in a sentence	22.41	17
Average number of minutes teachers spent on identification of key sounds	24.21	19
Average number of minutes teachers spent on ability to repeat words, sounds and syllables	27.63	41
Average number of minutes teachers spent on reading punctuation	27.29	7
Proportion of that teachers didn't know more than one techniques	0.81	101
Average number of minutes students read by themselves today	28.08	80
Average number of minutes students usually read by themselves during class	27.61	92
Average number of minutes students read to others today	18.68	74
Average number of minutes students usually read to others during class	21.90	83
Average number of minutes students listened to others read today	28.44	75
Average number of minutes students usually listen to others read during class	23.71	77
Average number of minutes students wrote today	26.07	70
Average number of minutes students usually write during class	24.41	99
Average number of minutes students did word play today	11.05	62
Average number of minutes students usually do word play during class	11.10	81
Indicator 3: Number of teachers in target schools who attend and teach school at least 90% of scheduled school days per year		
Proportion of teachers in target schools who attend and teach school at least 90% of scheduled school days per year	.85	95

<sup>24</sup> This value corresponds to between reading levels B (level 2) and C (level 3).

Outcome Variable	Mean	Observations
Indicator 4: Number of school administrators in target schools who demonstrate use of new techniques or tools as a result of USDA assistance		
Proportion of school administrators in targeted districts who demonstrate use of new techniques or tools as a result of USDA assistance	.61	13
Indicator 5: Number of preschool teachers (bissongo caregivers) in target preschools who demonstrate use of at least one new teaching technique, skill or tool		
Proportion of preschool teachers (bissongo caregivers) in target preschools who demonstrate use of at least one new teaching technique, skill, or tool <sup>25</sup>	N/A	N/A
Indicator 6: Number of teachers/educators/teaching assistants in target schools who demonstrate use of new and quality teaching techniques or tools as a result of USDA assistance		
Proportion of teachers who have received literacy training in the last 12 months	0.16	89
Proportion of teachers who have had students check their own work this week	0.27	101
Proportion of teachers who are planning to have students check their own work this week	0.20	101
Proportion of teachers who have not and are not planning to have students check their own work this week	0.53	101
Proportion of teachers who have had students check each other's work this week	0.67	101
Proportion of teachers who are planning to have students check each other's work this week	0.17	101
Proportion of teachers who have not and are not planning to have students check each other's work this week	0.16	101
Proportion of teachers who have had the whole class check the work of a student this week	0.62	101
Proportion of teachers who are planning to have the whole class check the work of a student this week	0.21	101
Proportion of teachers who have not and are not planning to have the whole class check the work of a student this week	0.17	101
Proportion of teachers who have checked students' work this week	0.77	101
Proportion of teachers who are planning to check students' work this week	0.18	101
Proportion of teachers who have not and are not planning to check students' work this week	0.05	101
Proportion of teachers who have paired students of different skill levels to work together this week	0.69	101
Proportion of teachers who are planning to pair students of different skill levels to work together this week	0.24	101
Proportion of teachers who have not and are not planning to pair students of different skill levels to work together this week	0.07	101
Proportion of teachers who have paired students of the same skill levels to work together this week	0.41	101
Proportion of teachers who are planning to pair students of the same skill levels to work together this week	0.16	101
Proportion of teachers who have not and are not planning to pair students of the same skill levels to work together this week	0.44	101

<sup>25</sup> Will be collected with pre and post testing at the time of training provided by BBII

Outcome Variable	Mean	Observations
Proportion of teachers who have had students work together in groups this week	0.48	101
Proportion of teachers who are planning to have students work together in groups this week	0.19	101
Proportion of teachers who have not and are not planning have students work together in groups this week	0.34	101
Proportion of teachers using the phonemics technique	0.03	101
Proportion of teachers using the phonics technique	0.09	101
Proportion of teachers using the fluency technique	0.58	101
Proportion of teachers using the vocabulary technique	0.26	101
Proportion of teachers using the comprehension technique	0.33	101
Proportion of teachers using the identification of key words in a sentence technique	0.17	101
Proportion of teachers using the identification of key sounds technique	0.19	101
Proportion of teachers using the ability to repeat words, sounds and syllables technique	0.41	101
Proportion of teachers using the reading punctuation technique	0.07	101
<b>Indicator 7: Number of students in target schools who are identified as attentive during class/instruction</b>		
Average students' attentiveness on a scale of 1 (the lowest) to 10 (the highest), as reported by teachers	5.91	101
Proportion of students identified as 7 or above on the attentiveness scale, as reported by teachers	0.43	101
<b>Indicator 8: Number of students in target schools who indicate that they are hungry or very hungry during the school day</b>		
Proportion of children that felt full after the meal s/he ate before going to school	0.96	293
Proportion of students that could have eaten more before going to school	0.04	293
Proportion of students that cite lack of food as the reason for not having a meal before going to school	0.86	14
Proportion of students that mentioned there was no food that they liked before school	0.14	14
Proportion of students indicating another reason why they did not eat more before school	0.00	14
Proportion of children that felt full after eating lunch at home	0.94	282
Proportion of students that could have eaten more for lunch at home	0.06	282
Proportion of students that cite lack of food as the reason why they did not eat a meal for lunch at home	0.89	19
Proportion of students that mentioned there was no food that they liked for lunch at home	0.00	19
Proportion of students indicating another reason why they did not eat more for lunch at home	0.11	19
Proportion of children that felt full after eating lunch at the school canteen	0.94	83
Proportion of students that could have eaten more for lunch at the school canteen	0.09	83
Proportion of students that cite lack of food as the reason why they did not eat a meal for lunch at the school canteen	1.00	9
Proportion of students that mentioned there was no food that they liked for lunch at the school canteen	0.00	9
Proportion of students indicating another reason why they did not eat more for lunch at the school canteen	0.00	9
Proportion of children that felt full before going to bed	0.94	324
Proportion of children that could have eaten more before going to bed	0.06	324

Outcome Variable	Mean	Observations
Proportion of children that did not eat a meal before going to bed for the lack of food	0.86	21
Proportion of children that mentioned there was no food they liked before going to bed	0.00	21
Proportion of students indicating another reason why they did not eat more before going to bed	0.14	21
Proportion of parents that indicated 1 or more of their children ate less because there was no food	0.34	343
Proportion of parents that indicated their children did not eat because there was no food	0.21	343
Proportion of parents that indicated 1 or more of their children went a full day without food	0.08	343
<b>Indicator 9: Number of students regularly (80%) attending USDA supported classrooms/schools</b>		
Proportion of students attending regularly (80%)	1	510
Proportion of boys attending regularly (80%)	1	219
Proportion of girls attending regularly (80%)	1	204
<b>Indicator 10: Number of SILC group members that spend money from SILC on education costs</b>		
Proportion of parents that are members of SILC	0.09	220
Proportion of parents that have used a portion of their savings accumulated in SILC for school expenses	0.48	21
<b>Indicator 11: Number of girl students who cite mentors as one of the top 3 reasons for their success or improvement in school</b>		
Proportion of girl students who cite mentors as one of the top 3 reasons for their success or improvement in school.	0	0
<b>Indicator 12: Percent of school-age children receiving a minimum acceptable diet</b>		
Proportion of boys that ate cereals (e.g. pasta, rice,noodles, etc)	0.98	156
Proportion of girls that ate cereals (e.g. pasta, rice,noodles, etc)	0.98	168
Proportion of boys that ate white roots like white potato	0.04	156
Proportion of girls that ate white roots like white potato	0.06	168
Proportion of boys that ate vitamin A rich vegetables (e.g. carrot, etc)	0.03	156
Proportion of girls that ate vitamin A rich vegetables (e.g. carrot, etc)	0.05	168
Proportion of boys that ate dark green leafy vegetables like spinach	0.74	156
Proportion of girls that ate dark green leafy vegetables like spinach	0.77	168
Proportion of boys that ate other vegetables (e.g. eggplant, tomato)	0.55	156
Proportion of girls that ate other vegetables (e.g. eggplant, tomato)	0.59	168
Proportion of boys that ate vitamin A rich fruit like mango & papaya	0.43	156
Proportion of girls that ate vitamin A rich fruit like mango & papaya	0.43	168
Proportion of boys that ate other fruit (e.g. watermelon, coconut)	0.33	156
Proportion of girls that ate other fruit (e.g. watermelon, coconut)	0.26	168
Proportion of boys that ate internal organs (e.g. liver, heart)	0.04	156
Proportion of girls that ate internal organs (e.g. liver, heart)	0.02	168
Proportion of boys that ate flesh meat (e.g. chicken, pork, beef)	0.20	156
Proportion of girls that ate flesh meat (e.g. chicken, pork, beef)	0.26	168

Outcome Variable	Mean	Observations
Proportion of boys that ate any bird's eggs	0.03	156
Proportion of girls that ate any bird's eggs	0.02	168
Proportion of boys that ate seafood (e.g. shrimp, fish)	0.38	156
Proportion of girls that ate seafood (e.g. shrimp, fish)	0.34	168
Proportion of boys that ate nuts and seeds (e.g. lentils, beans)	0.58	156
Proportion of girls that ate nuts and seeds (e.g. lentils, beans)	0.67	168
Proportion of boys that ate dairy products (e.g. milk, cheese)	0.06	156
Proportion of girls that ate dairy products (e.g. milk, cheese)	0.10	168
Proportion of boys that ate oils and fats like butter	0.39	156
Proportion of girls that ate oils and fats like butter	0.52	168
Proportion of boys ate any sweets (e.g. sugar, honey)	0.27	156
Proportion of girls ate any sweets (e.g. sugar, honey)	0.38	168
On the scale of 1 to 15, FAO dietary diversity index for boys	5.05	156
On the scale of 1 to 15, FAO dietary diversity index for girls	5.43	168

**Indicator 13: Percent of school children that wash their hands at critical moments**

Proportion of school children that wash their hands at critical moments	0.78	307
Proportion of children that washed their hands before eating	0.94	307
Proportion of children that did not wash their hands before eating	0.01	307
Proportion of children that did not wash their hands because they did not eat that day	0.05	307
Proportion of children that washed their hands before preparing food	0.37	307
Proportion of children that did not wash their hands before preparing food	0.09	307
Proportion of children that did not wash their hands because they did not prepare food that day	0.54	307
Proportion of children that washed their hands before giving food to others	0.25	307
Proportion of children that did not wash their hands before giving food to others	0.10	307
Proportion of children that did not wash their hands because they did not give food to anyone that day	0.64	307
Proportion of children that washed their hands when they were dirty	0.53	307
Proportion of children that did not wash their hands when they were dirty	0.24	307
Proportion of children that did not wash their hands because they were not dirty that day	0.22	307
Proportion of children that washed their hands after touching dirt	0.45	307
Proportion of children that did not wash their hands after touching dirt	0.24	307
Proportion of children that did not wash their hands because they did not touch dirt that day	0.31	307
Proportion of children that washed their hands after using the latrine	0.46	307
Proportion of children that did not wash their hands after using the latrine	0.18	307
Proportion of children that did not wash their hands because they did not use the latrine that day	0.36	307



Outcome Variable	Mean	Observations
Indicator 14: Number of students in target schools who achieve a passing score on a test of good health and hygiene practices		
Proportion of students that believe they should wash their hands before eating	0.98	324
Proportion of students that believe they should wash their hands before preparing food	0.05	324
Proportion of students that believe they should wash their hands before feeding others	0.02	324
Proportion of students that believe they should wash their hands if they are dirty	0.13	324
Proportion of students that believe they should wash their hands after touching dirt	0.07	324
Proportion of students that believe they should wash their hands after using the latrine	0.22	324
Indicator 15: Number of students in target schools who achieve a passing score on a test of food nutrition and dietary practices		
Proportion of children that have heard of vitamin A before	0.32	324
Proportion of children that say vitamin A provides good vision	0.01	107
Proportion of children that say vitamin A protects the body against diseases	0.33	107
Proportion of children that say vitamin A helps kids to grow up healthy	0.14	107
Proportion of children that say vitamin A keeps the skin healthy	0.05	107
Proportion of children that specify something else vitamin A does	0.25	107
Proportion of children that were not able to specify what vitamin A does	0.48	107
Proportion of children that say liver has vitamin A	0.02	106
Proportion of children that say small whole fish has vitamin A	0.05	106
Proportion of children that say dairy products have vitamin A	0.11	106
Proportion of children that say egg yolks have vitamin A	0.04	106
Proportion of children that say carrot has vitamin A	0.03	106
Proportion of children that say sweet potato has vitamin A	0.01	106
Proportion of children that say squash has vitamin A	0.02	106
Proportion of children that say dark green leafy vegetables have vitamin A	0.08	106
Proportion of children that say red palm oil has vitamin A	0.01	106
Proportion of children that say papaya has vitamin A	0.04	106
Proportion of children that say mango has vitamin A	0.08	106
Proportion of children that say Nere four (leaves of a local tree) has vitamin A	0.00	106
Proportion of children that say green pepper has vitamin A	0.01	106
Proportion of children that say melon has vitamin A	0.01	106
Proportion of children that specify another food having vitamin A	0.25	106
Proportion of children that were not able to specify a food containing vitamin A	0.55	106
Proportion of children that knew there is iron in the food they eat	0.09	324
Proportion of children that say iron provides good vision	0.10	29
Proportion of children that say iron protects the body against diseases	0.14	29
Proportion of children that say iron helps kids to grow up healthy	0.07	29

Outcome Variable	Mean	Observations
Proportion of children that say helps the muscle function	0.03	29
Proportion of children that say iron helps the brain function	0.03	29
Proportion of children that say iron regulates the body temperature	0.03	29
Proportion of children that say iron prevents fatigue	0.00	4
Proportion of children that say iron is the oxygen carrier in the body	0.03	29
Proportion of children that say iron maintains healthy skin	0.00	4
Proportion of children that say iron helps hemoglobin formation	0.10	29
Proportion of children that specify something else iron does	0.14	29
Proportion of children that were not able to specify what iron does	0.62	29
Proportion of children that say fish has iron	0.07	29
Proportion of children that say the poultry family has iron	0.07	29
Proportion of children that say red meat has iron	0.14	29
Proportion of children that organ meat has iron	0.00	4
Proportion of children that say beans have iron	0.07	29
Proportion of children that say dry peas have iron	0.03	29
Proportion of children that say lentils has iron	0.00	4
Proportion of children that say cowpeas has iron	0.00	4
Proportion of children that say dark green leafy vegetables have iron	0.00	29
Proportion of children that say eggs have iron	0.03	29
Proportion of children that specify another food having iron	0.14	29
Proportion of children that were not able to specify a food containing iron	0.69	29
<b>Indicator 16: Number of months of community and/or government support for school canteens</b>		
Number of months of community and or government support for school canteen	5.28	14
Proportion of parents that contribute wood every day that the canteen operates	0.43	14
Proportion of parents that contribute wood most of the days that the canteen operates	0.43	14
Proportion of parents that rarely contribute wood of the days that the canteen operates	0.14	14
Proportion of parents that never contribute wood of the days that the canteen operates	0	14
Proportion of parents that contribute utensils every day that the canteen operates	0.64	14
Proportion of parents that contribute utensils most of the days that the canteen operates	0.07	14
Proportion of parents that rarely contribute utensils every day that the canteen operates	0	14
Proportion of parents that never contribute utensils every day that the canteen operates	0.29	14
Proportion of parents that contribute to the storage of food every day the canteen operates	0	4
Proportion of parents that contribute to the storage of food most of the days the canteen operates	0.29	14
Proportion of parents that rarely contribute to the storage of food when the canteen operates	0.14	14
Proportion of parents that never contribute to the storage of food when the canteen operates	0.29	14

Outcome Variable	Mean	Observations
Proportion of communities that pay women to cook for the canteen	0.29	14
Proportion of students that contribute food to the canteen every day it operates	0.5	14
Proportion of students that contribute food to the canteen most of the days it operates	0.36	14
Proportion of students that rarely contribute food to the canteen when it operates	0	14
Proportion of students that never contribute food to the canteen when it operates	.14	14

## APPENDIX B. ADDITIONAL DETAILED DATA BREAKDOWN

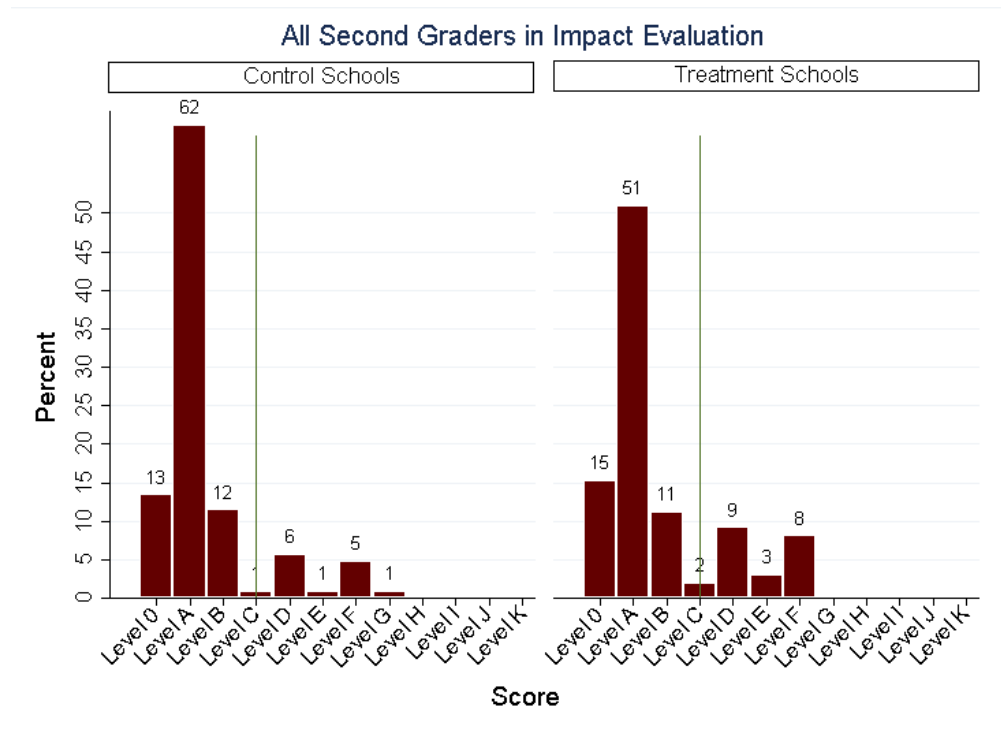
**Exhibit 52. Minimum Acceptable Diet Reported by Parents**

Outcome Variable	Mean	Observations
Household receiving a minimum acceptable diet		
The proportion of parents that offered their child cereals (e.g. pasta, rice, noodles, etc.)	0.75	343
The proportion of parents that offered their child white roots like white potato	0.03	343
The proportion of parents that offered their child vitamin A rich vegetables (e.g. carrot, etc.)	0.01	343
The proportion of parents that offered their child dark green leafy vegetables like spinach	0.58	343
The proportion of parents that offered their child other vegetables (e.g. eggplant, tomato)	0.38	343
The proportion of parents that offered their child vitamin A rich fruit like mango & papaya	0.24	343
The proportion of parents that offered their child other fruit (e.g. watermelon, coconut)	0.13	343
The proportion of parents that offered their child internal organs (e.g. liver, heart)	0.02	343
The proportion of parents that offered their child flesh meat (e.g. chicken, pork, beef)	0.12	343
The proportion of parents that offered their child any bird's eggs	0.02	343
The proportion of parents that offered their child seafood (e.g. shrimp, fish)	0.30	343
The proportion of parents that offered their child nuts and seeds (e.g. lentils, beans)	0.41	343
The proportion of parents that offered their child dairy products (e.g. milk, cheese)	0.04	343
The proportion of parents that offered their child oils and fats like butter	0.27	343
The proportion of parents that offered their child any sweets (e.g. sugar, honey)	0.15	343
On the scale of 1 to 15, FAO dietary diversity index reported by Parents	3.45	343

**Exhibit 53. Second Grade Students Reading at Level B or Above  
(Reading Simple Sounds – Grade Expectation for First Graders)**

Indicator	Percentage	Observations
Male students demonstrating reading ability at grade level or above	19%	16
Female students demonstrating reading ability at grade level or above	42%	24

**Exhibit 54. ASER-Reading Test Results for RCT Sample (Grade 2)**



**Exhibit 55. ASER-Reading Test Results for RCT Sample, Disaggregated by Sex (Grade 2)**

