EDUCATION CANNOT WAIT: EDUCATION FIRST PROJECT

AN ENDLINE REPORT FOR EARLY GRADE MATHEMATICS ASSESSMENT (EGMA) IN PALABEK REFUGEE SETTLEMENT, LAMWO DISTRICT

NOVEMBER, 2019

LUIGI GIUSSANI INSTITUTE OF HIGHER EDUCATION

EDUCATION CANNOT WAIT
a global fund for education in emergencies

AVSI
People for development

ACKNOWLEDGEMENTS

Special consideration goes to Save the Children and other Education Cannot Wait consortium partners for supporting this assessment.

We are also sincerely grateful to RTI for developing the EGMA tool that was adapted for this assessment. Furthermore, we immensely thank the learners, teachers and headteachers of the participating schools for their support and collaboration in this assessment.

Finally, we acknowledge the contribution of the following LGIHE and AVSI staff: John Mary Vianney Mitana (Ph.D), Mauro Giacomazzi, Monica Fontana Abad (Ph.D), Betty Okot (Ph.D), Sara Cohen, Alfred Agaba, Martin Ariapa, Sarah Akello, Brian Kisembo, Godfrey Rukundo, Amos Okiring, Gillian Atuheire, Edimond Serwanga and Diana Nakasujja, among others. This accomplishment would not have been possible without you.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

APPCO	African Partners for Child Poverty
CRS	Catholic Relief Services
DID	Difference in Differences
DIDE	Difference in Differences Estimate
ECW	Education Cannot Wait
EGM	Early Grade Mathematics
EGMA	Early Grade Mathematics Assessment
EGR	Early Grade Reading
EGRA	Early Grade Reading Assessment
FCA	Finn Church Aid
HI	Humanity and Inclusion
LGIHE	Luigi Giussani Institute of Higher Education
NRC	Norwegian Refugee Council
QED	Quasi Experimental Design
RTI	Research Triangle Institute
TICC	Teachers in Crisis Context
UNESCO	United Nations Educational, Scientific and Cultural Organization
WIU	Windle International Uganda

EXECUTIVE SUMMARY

Introduction

AVSI Foundation and Luigi Giussani Institute of Higher Education (LGIHE) have been piloting Early Grade Mathematics (EGM) in Aywee Nursery and Primary School in Palabek refugee settlement, Lamwo district. The EGM intervention was based on Singapore Math as an instructional method derived from the national curriculum of Singapore. The other school – Awich Nursery and Primary School was used as a control to facilitate the evaluation design. In order to assess the effectiveness of the EGM intervention, an endline was conducted in the stated schools.

Objectives

The EGMA was conducted to specifically (i) assess the learners' early grade mathematics skills in the classes of P1, P2 and P3; and (ii) ascertain the level of improvement in learners' numeracy abilities as a result of the EGM intervention.

Methods used

A quasi experimental design, where measurements were done in the 2 schools both before and after exposure to the intervention, was adopted to assess the effectiveness of the EGM intervention. During the baseline, a total of 327 learners from P.1 to P.3 of the 2 schools were assessed (treatment=175 learners; control=152 learners). Thereafter, an intervention was implemented with 12 teachers of the treatment school. After that, an endline was conducted with a total of 315 learners whereby in the treatment school a total of 172 learners (P1=57, P2=58, P3=57) were assessed while the rest 143 (P1=50, P2=47, P3=46) were assessed in the control school.

The EGMA tool used was adapted from the Research Triangle Institute (RTI) for purposes of validity, reliability and comparability. This tool included the following subtask areas: number identification, number discrimination, missing number, addition, subtraction and word problems. The assessments were conducted in pairs by trained Assessors - a teacher of a respective school and an AVSI/LGIHE staff. The endline was conducted from 2nd to 8th November, 2019.

This being a paper-and-pencil assessment, learners' scores were captured into an Epidata (Version 3.1) system and exported to STATA (Version 13.0) for cleaning and analysis. The analysis was done in terms of proportions, percentages and means. Furthermore, in order to investigate the casual effects of the EGM intervention on the treatment school over the control school, difference in differences (DID) analysis was conducted.

Summary of Findings

i. Subtask 1: Number Identification
   - The project improved the number identification knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items within 60 seconds (DIDE=5%).
   - The project improved the number identification knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners within 60 seconds (DIDE=2.0 items).

ii. Subtask 2: Number Discrimination
   - The project improved the number discrimination knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=20%).
   - The project improved the number discrimination knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.2 items).

iii. Subtask 3: Missing Number
   - The project improved the missing number knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=21%).
   - The project improved the missing number knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.3 items).

iv. Subtask 4: Addition
   - The project improved the addition knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=21%).
   - The project improved the addition knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.5 items).

v. Subtask 5: Subtraction
   - The project improved the subtraction knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=23%).
   - The project improved the subtraction knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.7 items).

vi. Subtask 6: Word problems
   - The project improved the word problem knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=26%).
   - The project improved the word problem knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.6 items).

Conclusion

The results reveal improvements in the treatment school compared to the control over the period of 4 months of EGM intervention. Specifically, significant improvements were observed in the subtask areas of: number discrimination, missing numbers, addition, subtraction and word problems. A debrief with teachers reveals these improvements to be attributed to the EGM intervention that was implemented in the treatment school immediately after the baseline assessment.

Based on the findings of this pilot, it is important to increase on the length of the intervention in order to allow teachers to fully comprehend and effectively implement the knowledge and skills acquired. To maximize the impact of the Singapore Math Approach, there is need for an extensive training of the teachers on materials development to enable them create authentic and appropriate teaching materials which foster teaching and learning of numeracy.

CHAPTER ONE: INTRODUCTION

1.1. Background to the Assessment

Education Cannot Wait (ECW) is a one-year project being implemented by AVSI Foundation as a lead implementer and Luigi Giussani Institute of Higher Education (LGIHE), as a co-implementer. The ECW project is being implemented in a consortium of 13 implementing partners, including: APPCO, AVSI, CRS, FCA, HI, Plan International, NRC, Save the Children, Street Child, UNESCO, WIU and ZOA. However, the activities of this project which LGIHE is directly involved in include; Adaptation and development of training materials, and training of teachers using the adapted Teachers in Crisis Context (TiCC) materials.

AVSI and LGIHE have been piloting Early Grade Mathematics (EGM) in one of the schools (Aywee Nursery and Primary School) in Palabek refugee settlement, Lamwo district. The second school (Awich Nursery and Primary School) was used as a control to facilitate the evaluation design. LGIHE's interventions are expected to contribute to the achievement of the following outcomes:

- Outcome 1: Improved Equitable Access to Inclusive Relevant Learning Opportunities
- Outcome 2: Improved Delivery of Quality Education and Training.

1.2. Description of the EGM Intervention

Singapore Math is an instructional method derived from the national curriculum of Singapore. The backbone of this pedagogy is the goal of empowering learners with the ability to both acquire and master a smaller range of mathematical concepts in greater depth. This is accomplished through a phased approach to learning that heavily focuses on problem-solving skills.

i. Phase 1, the concrete phase, is hands-on: learners explore mathematical concepts using concrete objects such as dice, paper clips, and blocks.
ii. In phase 2 (the pictorial phase), learners are introduced to pictorial representations of the same mathematical concepts and are challenged to produce their own pictorial representations.
iii. Finally, in phase 3 (the abstract phase) learners extend the knowledge they have gained to solving abstract mathematical problems using numbers and symbols.

As an instructional methodology, the Singapore Math approach is a solid fit with the realities of education in Palabek and other emergency settings. This approach was implemented in a series of trainings and follow-up sessions as follows:

✶ 1st Training

The first training was conducted in a period of 3 days. This training was intended to respond to the observations made during the pre-intervention assessment and daily debriefs with teachers. It covered the following:

- Concrete, Pictorial and Abstract (CPA) methodology (Singapore Math)
- Material creation

- Community of practice (COP): This is a platform where teachers share ideas on how to improve classroom practices based on the knowledge and skills acquired from the EGM intervention. COP meetings were held weekly. During their meetings, teachers sought to answer the following questions: (i) how have the skills and knowledge acquired as a result of the trainings from AVSI/LGIHE influenced the way you teach EGM to your pupils? (ii) what has worked well for you? How do the pupils feel about these approaches? (iii) what does not work well for you? Why? (iv) what form of additional support do you need to effectively apply the skills and knowledge acquired, to the teaching and learning of EGM?
- Classroom management

✶ 1st Follow-up Training

The follow-up was conducted in a period of 5 days: It covered the following:

- Classroom observations
- Training on:
    - How to plan classroom activities
    - Classroom management
    - Material creation
    - Strategies of teaching Singapore Math
    - Assessment for learning in Mathematics
- Community of practice
    - Role of teamwork as an avenue for their academic and social growth
    - Encouraging them to keep consulting among themselves and also involve the LGIHE Field Officer

✶ 2nd Follow-up Training

The follow-up was conducted in a period of 5 days: It covered the following:

- Training on:
    - Scheming
    - Lesson planning
    - Materials creation
    - Application of Singapore Math

✶ 3rd Follow-up Training

The follow-up was conducted in a period of 4 days: It covered the following:

- Classroom observations
- Community of practice - reflection on the fundamental social nature of human learning.

Based on this, AVSI and LGIHE organized to conduct an Endline for EGMA in the 2 schools (Aywee Nursery and Primary School and Awich Nursery and Primary School). The endline is the post-intervention assessment whose results are compared with the baseline (as of June/July, 2019) in order to assess the effectiveness of the EGM interventions in Palabek refugee settlement.

1.3. Purpose of the Assessment

EGMA is aimed at assessing the effectiveness of the EGM interventions in Palabek refugee settlement.

1.4. Objectives of the Assessment

The EGMA was conducted to address the following objectives:

i. To assess the learners' early grade mathematics skills in the classes of P1, P2 and P3.
ii. To ascertain the level of improvement in children's numeracy abilities as a result of the EGM intervention.

1.5. Assessment Hypotheses

The EGMA was guided by the following hypotheses:

- Ha_1: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on number identification subtask compared to those in the control group, after the EGM intervention.
- Ha_2: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on number discrimination subtask compared to those in the control group, after the EGM intervention.
- Ha_3: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on missing number subtask compared to those in the control group, after the EGM intervention.
- Ha_4: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on addition subtask compared to those in the control group, after the EGM intervention.
- Ha_5: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on subtraction subtask compared to those in the control group, after the EGM intervention.
- Ha_6: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on word problems compared to those in the control group, after the EGM intervention.

CHAPTER TWO: ASSESSMENT METHODOLOGY

2.1. Assessment Tool

The EGMA tool developed by RTI was adapted to ensure issues of validity, reliability and comparability with existing assessments. The subtask areas included in this assessment are: Number Identification, Number Discrimination, Missing Number, Addition, Subtraction and Word problems.

The tool was later translated into Acholi language to suit the demands of the thematic curriculum but also to cater for the diversities in language for the learners in the settlement. It should be noted that only the instructions were translated, and at some points some translations had to be made in Arabic to cater for the few learners who could not understand either English or Acholi.

2.2. Target Population

The target population consisted of learners in P1, P2 and P3 from 2 schools in Palabek refugee settlement, Lamwo district, as shown in Table 1.

2.3. Assessment Design

To establish a causal relationship between the intervention and changes in outcomes, the assessment adopted a Quasi Experimental Design (QED) in which measurements were done in the 2 groups (treatment and control schools) both before and after exposure to the programme. The schools were grouped as follows:

- G1: Aywee Nursery and Primary School, which was the treatment school.
- G2: Awich Nursery and Primary School, which was the control school.

2.4. Externalities and Selection Criteria for the Schools

Experimental interventions may generate spill-over effects whenever untreated learners are affected by the treatment programme. The study is cognizant of spill-over effects and recognizes the difficulty of its quantification. Attempts were made to ensure that selected learners for the group were not from the same school; and that the selected school for the control was not from the same zone. This was to minimize possibilities of spill-over effects. Therefore, the schools were selected based on the following conditions:

- Being on the double shift system
- Using temporary structures for classrooms
- Enrolled learners from various tribes hence multi-lingual teaching demands
- Being located in different zones to minimize issues of spill-over effects
- Recruited both national and refugee teachers
- Enrolled both host and refugee children with the majority being refugees

2.5. Sample Size and Number of Learners Assessed

The following records as realized from the head teachers of the respective schools in the month of May/June, 2019, were used in determining the desired sample size at baseline level.

Table 1: Population per school and class

School	Type of Assessment	P1	P2	P3	Total
Aywee Nursery and Primary School	EGMA	81	99	65	245
Awich Nursery and Primary School	EGRA+ EGMA	245	192	307	744
Total					**989**

The lowest population of 65 learners for P3 class of Aywee Nursery and Primary School was used as a benchmark in determining the minimum sample size for this assessment, to allow for comparisons among classes. Based on this, it was deemed necessary to consider an allowance of 30% to cater for issues of absenteeism, transfer and drop-out at both baseline and end-line phases. This ultimately reduced the desired sample size to at least 50 learners. This number was targeted in each of the study classes (P1, P2 and P3). Note that EGMA and EGRA were conducted at Awich Nursery and Primary School as a control school, but with a separate group of learners. This therefore necessitated dividing the learners into 2 mutually exclusive groups.

At baseline, 327 learners were assessed in all the 3 classes. At endline, all these learners were targeted, however due to some unavoidable circumstances like transfer cases and absenteeism, some learners were not found during the data collection dates. At the end of the endline data collection exercise, 315 (96.3%) learners were assessed. The characteristics of these learners are explored in the **methodology section 3.1.** The number of learners assessed at baseline and endline per group (school, sex) and class were as below.

Table 2: Number of Learners Assessed Per Group and Class

		Baseline						Endline		
		Sex		Nationality				Sex		
School	Class	Male	Female	UG	SS	Total		Male	Female	Total
Aywee Nursery and Primary School	P1	26	33	0	59	**59**		25	32	**57**
	P2	34	25	0	59	**59**		34	24	**58**
	P3	38	19	0	57	**57**		38	19	**57**
	Total	**98**	**77**	**0**	**175**	**175**		**97**	**75**	**172**
Awich Nursery and Primary School	P1	19	35	0	54	**54**		16	34	**50**
	P2	26	24	0	50	**50**		23	24	**47**
	P3	30	18	0	48	**48**		28	18	**46**
	Total	**75**	**77**	**0**	**152**	**152**		**67**	**76**	**143**

2.6. Selection of Learners

The sampling of the learners was done by LGIHE officials at baseline level. This was done through simple random sampling having obtained a list of learners for each class, from the Head teachers. Note that sampling of learners was done per target class. At endline, only learners assessed at baseline were assessed.

2.7. Training of Assessors

The same teachers who participated in the baseline assessment were invited and retrained to carry on the endline assessment. The refresher training of assessors was conducted on the 27th/10/2019 at Aywee Nursery and Primary School. This training comprised of 21 participants from Aywee Nursery and Primary School (9 teachers) and Awich Nursery and Primary School for EGMA (12 teachers). Three teachers from the treatment schools did not manage to attend the training as well as the assessment due to some unavoidable circumstances. The training was facilitated by 3 experienced facilitators from LGIHE who included: The Senior Monitoring and Evaluation Officer - for the aspects of assessment methodology; and 2 Education Officers - for the aspects of numeracy tasks included in the tool.

The training was intended to make assessors become familiar with the administration of the tool and with the specific implementation and coding practices associated with it. The training covered the following aspects:

- Self-introductions, expectations and setting ground rules.
- Assessment protocol – about ECW project, the assessment purpose, assessment methodology; and data quality control issues.
- Key points to consider when conducting the assessment – seeking consent; ensuring confidentiality, safety and security, and psychosocial wellbeing of others.
- EGMA tool – question by question explanations.
- Mock assessment for the EGMA tool – practice among assessors.

An interactive strategy was utilized during the trainings whereby participants were taken through the assessment process and tools. The facilitators emphasised how questions in each subtask were to be asked and answered. In this process, assessors practiced reading the instructions and answering the corresponding questions. Any incorrect response was addressed by the facilitators.

Having gone through all the instructions and questions in the respective tools, participants were divided into 3 groups comprising of 7 teachers and 1 LGIHE staff to practice administering the assessment. The participants in a rotational way chose 2 teachers to act as Assessors while 1 teacher acted as a learner. While others were observing and taking note of the proceedings, the 3 participants simulated the assessment process. At the end of each subtask, participants had an opportunity to discuss and correct areas where there were errors.

2.8. The Fieldwork Process

The endline assessment was conducted from 4th to 8th November, 2019, in the 2 target schools. The assessment was based on a pen/pencil and paper approach. As of the baseline which was conducted from the 2nd to 5th July, 2019, teachers were tasked to assess their learners in collaboration with an external member. This meant that each learner was assessed by a pair of Assessors, comprising of a teacher from the respective school and an AVSI/LGIHE staff. This was considered in order to ensure that:

- They become aware of how individualized assessments are conducted
- They become aware of the individual learner challenges hence able to help the child after the assessment and at an appropriate time/level.

- The child feels comfortable being assessed in the presence of their teacher.
- Good quality data is collected as one helps to assist or check on the other.

After the data collection, the test papers were taken to LGIHE offices in Kampala, for data entry. The entry was done by LGIHE experienced data entrants, using a pre-designed system – Epidata (Version 3.1) that controls for data input errors and safeguards data integrity, and also eases the process of export to analytical packages.

2.9. Data Analysis

Before any data analysis was performed, rigorous cleaning of the data was performed in order to identify erroneous records. Any missing bio-data e.g. sex, nursery attendance, repetition, class, etc. was corrected as per the baseline characteristics.

The cleaning and analysis were done using STATA (Version 13.0) statistical package with the best practice of using 'do-files' which ease the process of code correction. The results were analysed in terms of proportions, percentages, means, and difference in differences estimates (DIDE) for each class and all the classes combined.

In order to investigate the casual effects of the EGM intervention on the treatment school over the control school, difference in differences (DID) analysis was conducted. The DID method removes the difference in the outcome between treatment and control groups at the baseline. It was implemented as an interaction term between the time (where 1 is assigned to the endline period and 0 is assigned to baseline period) and treatment (where 1 is assigned to the treatment group and 0 is assigned to control group) variables in a regression model as below:

$$Y = \beta_0 + \beta_1 * [Time] + \beta_2 * [Treatment] + \beta_3 * [Time * Treatment] + \varepsilon$$

Where; β_0 is the baseline average; β_1 is the time trend in control group; β_2 is the difference between two groups (treatment vs control) at baseline and β_3 is the difference in changes over time.

And with covariates as in the regression model below:

$$Y = \beta_0 + \beta_1 * [Time] + \beta_2 * [Treatment] + \beta_3 * [Time * Treatment] + \beta_4 * [Covariates] + \varepsilon$$

Furthermore, differential impact of the EGM intervention by sex of the learners is examined. This is done by estimating the impact of treatment on the outcome variables at endline by estimating the equation below:

$$Y = \beta_0 + \beta_1 * [female] + \beta_2 * [Treatment] + \beta_3 * [female * Treatment] + \varepsilon$$

2.10. Limitations to the Assessment

Application of sampling weights to reflect the probability of learners sampled when using the difference in differences estimation commands in STATA is not supported. However, the standard errors were clustered at school-class level in order to reflect the notion that, learners were chosen from their respective classes.

Secondly, much as quasi-experimental design was appropriate for this kind of pilot (due to the project duration), it has limitations when the parallel trend assumption is not fulfilled. This assumption requires that in absence of the EGM intervention, then the difference in observed outcomes between the treatment and control groups is constant over time. In other words, we need to ask the question: would the treated learners have experienced the same outcomes as the control group learners if their teachers did not receive the EGM intervention? This is definitely the hardest assumption to fulfil when there are only 2 data points. The best practice is always to acquire more data points before and after so as to have a visual inspection of this assumption. That said, care must be taken when making conclusions about the program effects as it was not possible to investigate this assumption. Nonetheless, the strongest points for the analyses performed in this assessment are:

- Robust standard errors are used to account for autocorrelation between pre-intervention and post-intervention in the same individual;
- The characteristics of respondents in treatment and control groups were examined, before and after intervention in order to investigate whether the retained learners from the 2 groups (treatment vs control) had similar characteristics;
- Furthermore, key variables are included as covariates in the estimation of the project effects.

CHAPTER THREE: PRESENTATION AND DISCUSSION OF THE FINDINGS

3.1. Baseline Characteristics of the Learners Considered for Further Analyses

In order to match the endline with the baseline findings, only learners assessed at endline were retained for the subsequent analyses. The rest of the learners were dropped. This therefore meant the 315 learners assessed at endline. Their baseline characteristics (used as covariates) are shown below.

Table 3: Baseline Characteristics of the Retained Learners

	(1)	(2)	(3)	(4)	(5)
Variable(s)	Mean Control	Mean Treated	Difference	t-value	P-value
Primary One (Control=50; Treatment=57; Total=107)					
Female	0.68	0.56	-0.12	1.260	0.21170
Nursery attendance	0.54	0.91	0.37	4.780	0.0000***
Repeated class	0.44	0.25	-0.19	2.150	0.0339**
Age	9.50	8.26	-1.24	4.970	0.0000***
Primary Two (Control=47; Treatment=58; Total=105)					
Female	0.51	0.41	-0.10	0.990	0.32660
Nursery attendance	0.57	0.72	0.15	1.610	0.11020
Repeated class	0.45	0.43	-0.02	0.160	0.87280
Age	10.81	10.16	-0.65	2.410	0.0178**
Primary Three (Control=46; Treatment=57; Total=103)					
Female	0.39	0.33	-0.06	0.600	0.54670
Nursery attendance	0.54	0.70	0.16	1.660	0.0998*
Repeated class	0.54	0.33	-0.21	2.170	0.0322**
Age	12.46	11.04	-1.42	4.250	0.0000***
All Classes (Control=143; Treatment=172; Total=315)					
Female	0.53	0.44	-0.10	1.690	0.0920*
Nursery attendance	0.55	0.78	0.23	4.400	0.0000***
Repeated class	0.48	0.34	-0.14	2.510	0.0125**
Age	10.88	9.82	-1.06	5.010	0.0000***
*** p<0.01; ** p<0.05; * p<0.1					

- Sex: Less than a half (48%) of the 315 learners assessed were female (53% in control; 44% in treatment). The proportion of female learners assessed in P1, P2 and P3 are: 62%, 46% and 36%, respectively.

- Age: The average age of the learners assessed at was 10.3 years (10.9 for control; 9.8 for treatment), ranging from 5 to 19 years. The average ages of the learners assessed in P1, P2 and P3 are: 8.8, 10.4, and 11.7 years, respectively.

- **Nursery Attendance**: The majority (68%) of the 315 learners assessed attended nursery (55% in control; 78% in treatment). The proportion of learners who attended nursery, disaggregated by class – P1, P2 and P3 are: 74%, 66% and 63%, respectively.

- **Class Repetition**: More than a third (40%) of the 297 learners assessed have ever repeated a class (48% in control; 34% in treatment). The proportion of learners who have ever repeated a class, disaggregated by class – P1, P2 and P3 are: 34%, 44% and 43%, respectively.

In order to improve on effect size estimations, baseline values of these variables are included in the final analyses as covariates. Also, the results without the covariates are shown.

3.2. Subtask 1. Number Identification

3.2.1. Commentary on Number Identification

The subtask assesses learners' ability to understand the meaning and correctly identify numbers. A learner was presented with a piece of paper having 20 numbers, in two levels, and he/she was asked to say/name each number aloud within 60 seconds. A learner was discontinued if he/she had made four consecutive errors. Below are the numbers that were presented to learners.

Level 1				
2	9	0	12	30
22	45	39	23	48
Level 2				
91	33	74	87	65
108	245	587	731	989

Extract 1: Items on Number Identification task

3.2.2. Measure of Number Identification skill

The achievement of learners in Number Identification items is presented as: (i) the percentage of learners who correctly scored at least 50% of the items; and (ii) mean number of items correctly scored by the learners.

3.2.3. Effectiveness of the EGM Intervention on Number Identification Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on number identification subtask in 60 seconds compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on number identification subtask within 60 seconds in the treatment school (77%) was higher than that of the control school (70%). The baseline and endline percentages per class and school category are as shown in the figure below.

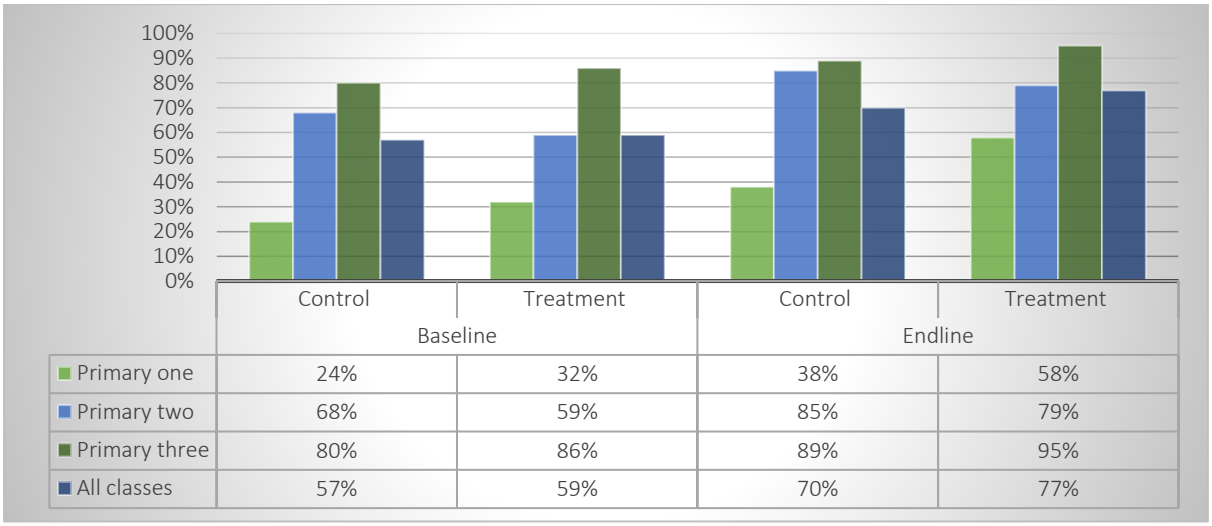


Figure 1: Number identification - percentage of learners who correctly scored at least 50% of the items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the number identification subtask is investigated using the intent-to-treat (ITT) estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (number identification) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 4: Number identification - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.12	0.04	0.00	0.05
Std. Errors	0.129	0.121	0.093	0.051
P-value	0.342	0.762	0.993	0.343
Panel 2: ITT Estimates with covariates				
DID Estimates	0.11	0.02	-0.01	0.04
Std. Errors	0.130	0.115	0.091	0.052
P-value	0.392	0.846	0.917	0.460
Observations	214	210	206	630
R-square	0.07	0.05	0.03	0.03

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items within 60 seconds (DIDE=12%). This effect is however not statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items within 60 seconds (DIDE=4%). This effect is however not statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project did not improve the number identification knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items within 60 seconds (DIDE=0%).

All classes: The results in column (4) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items within 60 seconds (DIDE=5%). This effect is however not statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the percentage of learners who correctly scored at least 50% of the test items within 60 seconds ($female * treatment\ coef = 0.08$; $P - value = 0.196$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.12	-0.06	0.05	0.02
Female	-0.08	-0.12	0.00	-0.17**
Treatment X Female	*0.12*	*-0.02*	*0.00*	*0.08*
Constant	0.44	0.91	0.89	0.79
Observations	107	105	103	315
R-squared	0.04	0.04	0.01	0.03
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

(ii) The mean number of items correctly scored by the learners within 60 seconds

At endline, the mean number of items correctly scored within 60 seconds by the treatment school learners (mean=13.9 items) was higher than that of the control school (mean=12.4 items). The baseline and endline means per class and school category are as shown in the figure below.

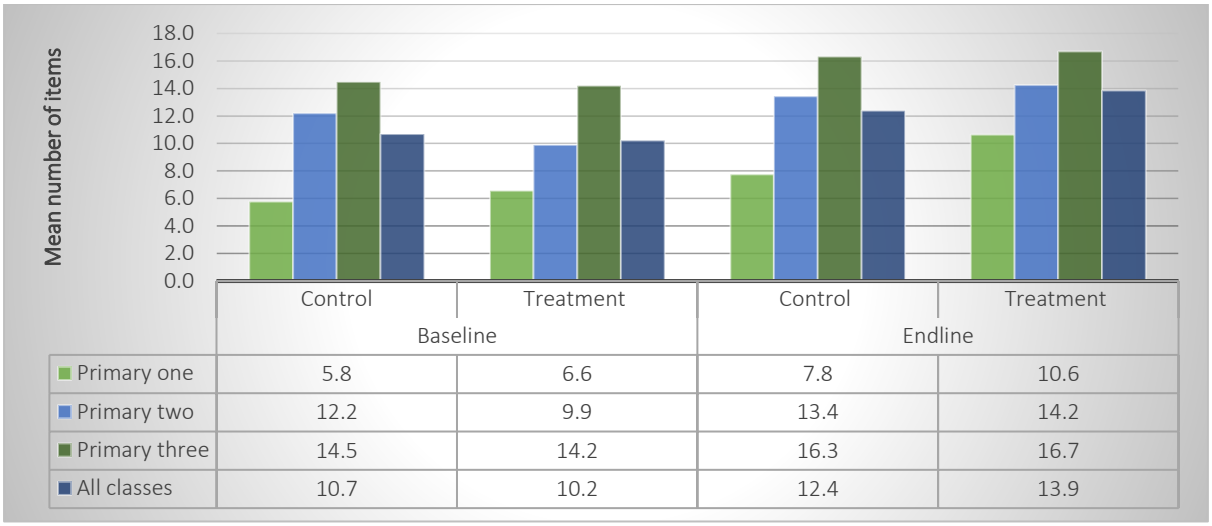


Figure 2: Number identification - mean number of items correctly scored by the learners within 60 seconds

The impact of the project on the mean number of items correctly scored by the learners on the number identification subtask is as shown below:

Table 5: Number identification - mean number of items correctly scored by the learners within 60 seconds

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	2.11	3.11	0.64	1.95
Std. Errors	1.526	1.616	1.410	0.559
P-value	0.169	0.056*	0.649	0.017**
Panel 2: ITT Estimates with covariates				
DID Estimates	2.00	2.88	0.51	1.70
Std. Errors	1.566	1.564	1.393	0.536
P-value	0.202	0.067*	0.714	0.021**
Observations	214	210	206	630
R-square	0.10	0.08	0.05	0.05

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners within 60 seconds (DIDE=2.1 items). This effect is however not statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the mean number

of items correctly scored by the P2 learners within 60 seconds (DIDE=3.1 items). This effect is statistically significant at 10% level.

Primary Three: The results in column (3) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners within 60 seconds (DIDE=0.6 items). This effect is however not statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the number identification knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners within 60 seconds (DIDE=2.0 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the mean number of items correctly scored by the learners within 60 seconds ($female * treatment\ coef = 0.72$; $P - value = 0.504$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	1.33	0.71	1.18	0.90
Female	-1.46	-2.15	0.01	-2.86***
Treatment X Female	2.47	-0.26	-2.46	0.72
Constant	8.75	14.52	16.32	13.90
Observations	107	105	103	315
R-squared	0.08	0.05	0.04	0.06
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

3.3. Subtask 2. Number Discrimination

3.3.1. Commentary on Number Discrimination

This subtask assesses learners' understanding of numerical magnitudes by comparing given pairs. This is an untimed subtask which assesses the learners' ability to compare magnitudes of numbers, with a stop rule after four consecutive errors. A learner was given a piece of paper having 10 pairs of numbers, one of which was greater than the other, and asked to tell the bigger number aloud. These items were categorized into 2 levels as shown below:

Level 1		Level 2	
7	5	94	78
11	24	146	153
39	23	287	534
58	49	623	632
65	67	867	965

Extract 2: Items on Number Discrimination task

3.3.2. Measure of Number Discrimination skill

Learners' achievement in Number Discrimination items is presented as: (i) the percentage of learners who correctly scored at least 50% of the items; and (ii) mean number of items correctly scored by the learners.

3.3.3. Effectiveness of the EGM Intervention on Number Discrimination Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on number discrimination subtask compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on number discrimination subtask in the treatment school (95%) was higher than that of the control school (84%). The baseline and endline percentages per class and school category are as shown in the figure below.

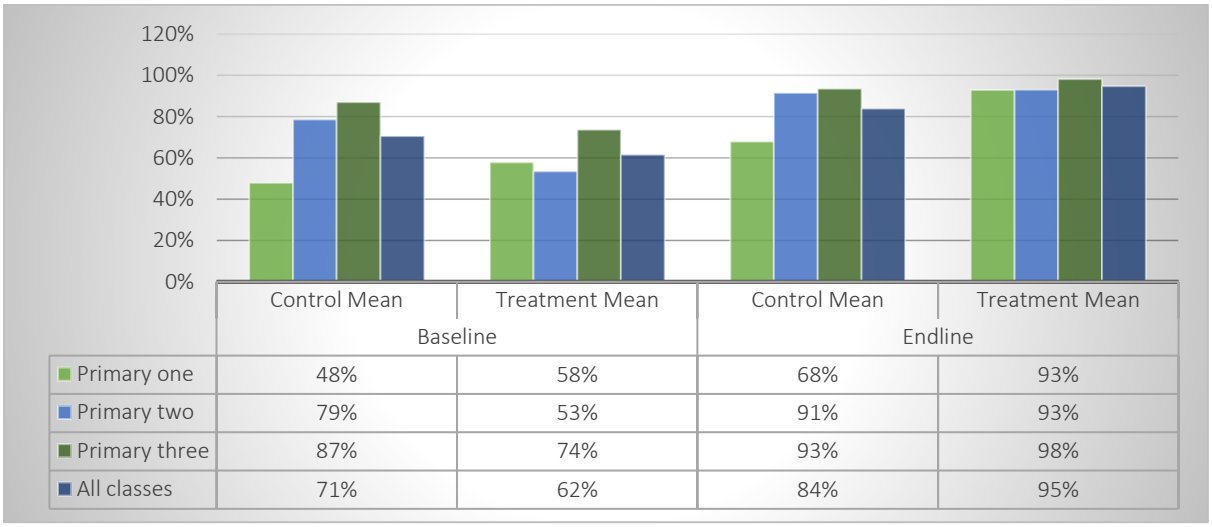


Figure 3: Number discrimination – percentage of learners who correctly scored at least 50% of the test items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the number discrimination subtask is investigated using the intent-to-treat estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (number discrimination) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 6: Number discrimination - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.15	0.27	0.18	0.20
Std. Errors	0.123	0.104	0.087	0.053
P-value	0.220	0.010**	0.040**	0.014**
Panel 2: ITT Estimates with covariates				
DID Estimates	0.16	0.26	0.17	0.19
Std. Errors	0.125	0.105	0.085	0.054
P-value	0.203	0.015**	0.045**	0.016**
Observations	214	210	206	630
R-square	0.13	0.16	0.09	0.10

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items (DIDE=15%). This effect is however not statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items (DIDE=27%). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items (DIDE=18%). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=20%). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show differential impact of the EGM intervention on the sex of the learners (in favour of males), in terms of the percentage of learners who correctly scored at least 50% of the items on number discrimination subtask ($female * treatment\ coef = -0.09$; $P - value = 0.007$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.40**	0.07	0.08	0.15
Female	0.17	0.09	0.11*	0.06***
Treatment X Female	*-0.23*	*-0.11*	*-0.08*	*-0.09****
Constant	0.56	0.87	0.89	0.81
Observations	107	105	103	315
R-squared	0.12	0.01	0.05	0.04
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

(ii) The mean number of items correctly scored by the learners

At endline, the mean number of items correctly scored by the treatment school learners in the number discrimination subtask (mean=7.6 items) was higher than that of the control school (mean=6.9 items). The baseline and endline means per class and school category are as shown in the figure below.

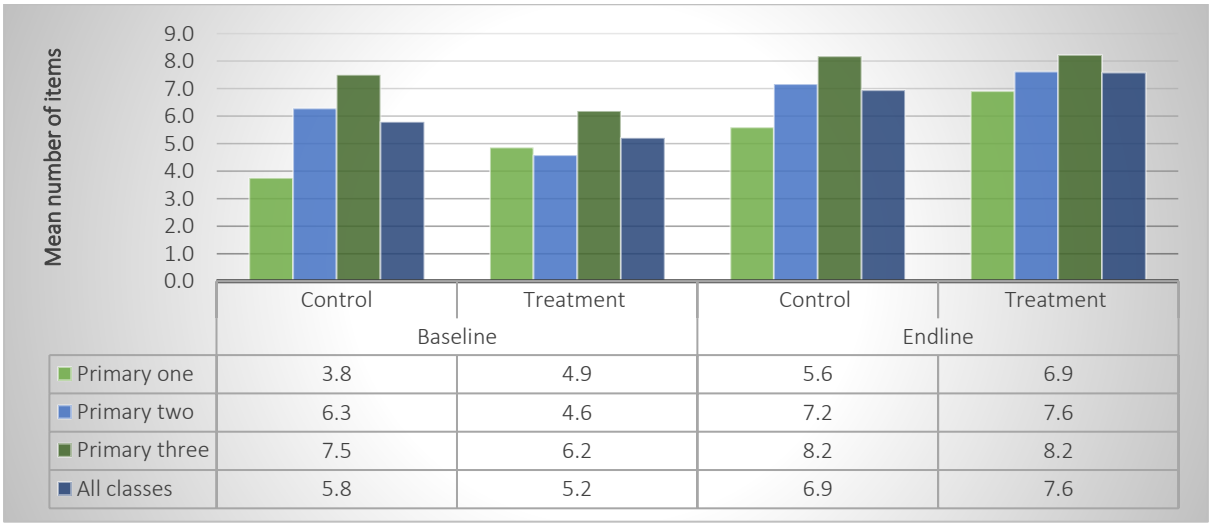


Figure 4: Number discrimination – mean number of items correctly scored by the learners

The impact of the project on the mean number of items correctly scored by the learners on the number discrimination subtask is as shown below:

Table 7: Number discrimination - mean number of items correctly scored by the learners

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.21	2.14	1.36	1.22
Std. Errors	0.688	0.693	0.631	0.442
P-value	0.757	0.002***	0.032**	0.039**
Panel 2: ITT Estimates with covariates				
DID Estimates	0.31	2.03	1.31	1.18
Std. Errors	0.691	0.688	0.624	0.428
P-value	0.651	0.004***	0.038**	0.040**
Observations	214	210	206	630
R-square	0.18	0.19	0.13	0.12

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners (DIDE=0.2 items). This effect is however not statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P2 learners (DIDE=2.1 items). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners (DIDE=1.4 items). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the number discrimination knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.2 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the mean number of items correctly scored by the learners ($female * treatment\ coef = -0.08$; $P - value = 0.267$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	1.80***	0.45	0.27	0.67
Female	0.88	-0.18	0.54	-0.13**
Treatment X Female	*-0.68*	*-0.03*	*-0.56*	*-0.08*
Constant	5.00	7.26	7.96	7.01
Observations	107	105	103	315
R-squared	0.12	0.02	0.01	0.03
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

3.4. Subtask 3. Missing Number

3.4.1. Commentary on Missing Number

The subtask assesses learners' knowledge of identifying a missing number of a sequence and their ability to count in ones, twos, fives, tens and hundreds, both in ascending and descending order. The subtask is untimed with a stop rule after four consecutive errors. A learner was presented with a sheet of paper having 10 items, categorized into two levels, where he/she was asked to tell the missing number in each item aloud. The items on this subtask are shown below:

Level 1			
5	6	7	|____|
14	15	|____|	17
20	|____|	40	50
|____|	300	400	500
2	4	6	|____|
Level 2			
348	349	|____|	351
28	|____|	24	22
30	35	|____|	45
550	540	530	|____|
3	8	|____|	18

Extract 3: Items on Missing Number task

3.4.2. Measure of Missing Number identification skill

Learners' achievement in identifying the missing number of a sequence is presented as: (i) the percentage of learners who correctly scored at least 50% of the items; and (ii) mean number of items correctly scored by the learners.

3.4.3. Effectiveness of the EGM Intervention on Missing Number Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on missing number subtask compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on missing number subtask in the treatment school (24%) was higher than that of the control school (13%). The baseline and endline percentages per class and school category are as shown in the figure below.

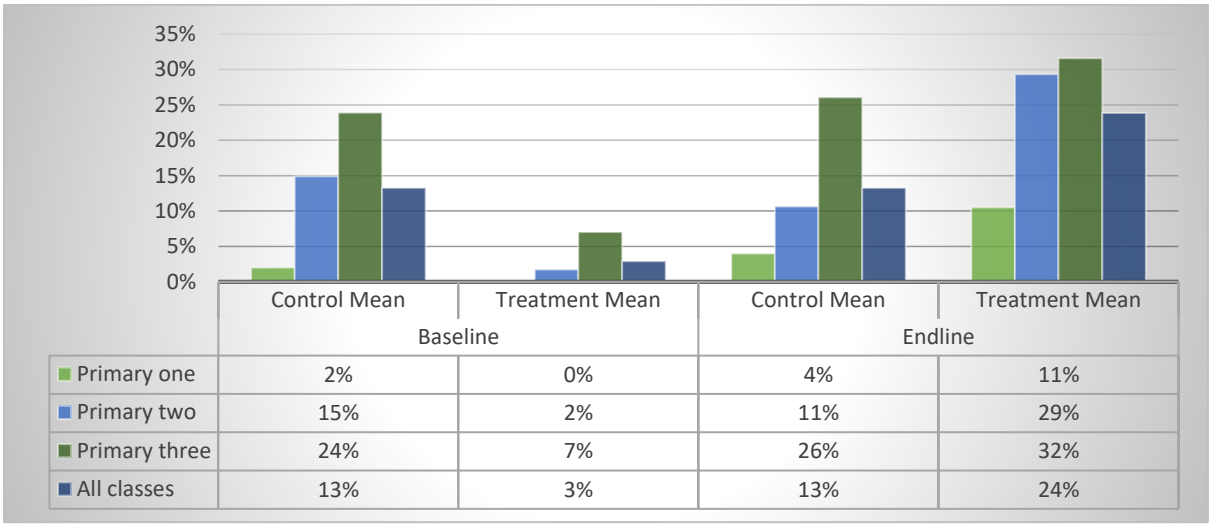


Figure 5: Missing number – percentage of learners who correctly scored at least 50% of the test items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the missing number subtask is investigated using the intent-to-treat estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (missing number) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 8: Missing number - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.09	0.32	0.22	0.21
Std. Errors	0.054	0.094	0.116	0.051
P-value	0.113	0.001***	0.054*	0.009***
Panel 2: ITT Estimates with covariates				
DID Estimates	0.09	0.31	0.22	0.20
Std. Errors	0.054	0.090	0.114	0.049
P-value	0.088*	0.001***	0.060*	0.009***
Observations	214	210	206	630
R-square	0.04	0.09	0.05	0.05

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates". *** p<0.01, ** p<0.05, * p<0.1				

Primary One: The results in column (1) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items (DIDE=9%). This effect is however not statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items (DIDE=32%). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items (DIDE=22%). This effect is statistically significant at 10% level.

All classes: The results in column (4) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=21%). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the percentage of learners who correctly scored at least 50% of the test items ($female * treatment\ coef = 0.04$; $P - value = 0.580$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.04	0.19	0.02	0.08
Female	0.06	-0.05	-0.15	-0.09
Treatment X Female	0.06	-0.03	0.08	0.04
Constant	0.00	0.13	0.32	0.18
Observations	107	105	103	315
R-squared	0.05	0.06	0.02	0.03

Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1

(ii) The mean number of items correctly scored by the learners

At endline, the mean number of items correctly scored by the treatment school learners in the number discrimination subtask (mean=3.6 items) was higher than that of the control school (mean=2.5 items). The baseline and endline means per class and school category are as shown in the figure below.

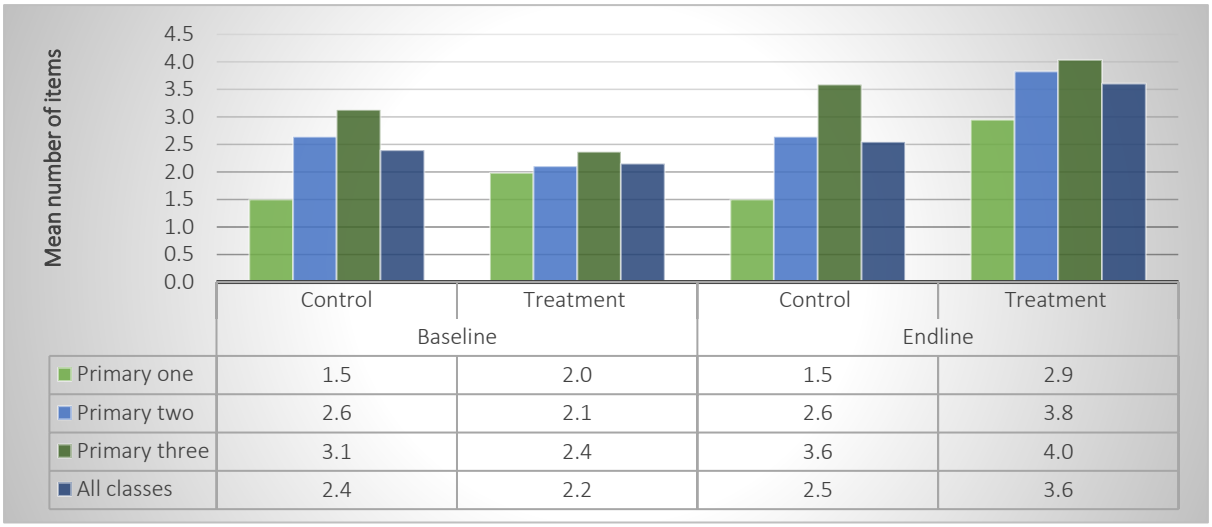


Figure 6: Missing number – mean number of items correctly scored by the learners

The impact of the project on the mean number of items correctly scored by the learners on the missing number subtask is as shown below:

Table 9: Missing number - mean number of items correctly scored by the learners

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.97	1.72	1.21	1.31
Std. Errors	0.346	0.555	0.555	0.256
P-value	0.006***	0.002***	0.030**	0.004***
Panel 2: ITT Estimates with covariates				
DID Estimates	1.01	1.65	1.16	1.28
Std. Errors	0.358	0.525	0.554	0.248
P-value	0.005***	0.002***	0.038**	0.004***
Observations	214	210	206	630
R-square	0.19	0.11	0.10	0.09
Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended				

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates". *** p<0.01, ** p<0.05, * p<0.1				

Primary One: The results in column (1) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners (DIDE=1.0 items). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P2 learners (DIDE=1.7 items). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners (DIDE=1.2 items). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the missing number knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.3 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show differential impact of the EGM intervention on the sex of the learners (in favour of females), in terms of the mean number of items correctly scored by the learners ($female * treatment\ coef = 0.63$; $P - value = 0.045$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.91***	0.88	0.48	0.69
Female	-0.46	-0.88	-0.51	-1.00***
Treatment X Female	*0.86**	*0.53*	*-0.18*	*0.63***
Constant	1.81	3.09	3.79	3.07
Observations	107	105	103	315
R-squared	0.29	0.11	0.03	0.10
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

3.5. Subtask 4. Addition

3.5.1. Commentary on Addition

This subtask assesses learners' ability to perform simple/basic and complex addition operations, ranging from one-digit numbers through two-digit numbers. This subtask was untimed with a stop rule after four consecutive errors. A learner was presented with a piece of paper having 10 items, categorized into two levels, where he/she was asked to tell the answer on each item, after carrying out some computations. The items on this subtask are shown below:

Level 1	Level 2
1 + 3 = |____|	13+6 = |____|
3 + 2 = |____|	18+7 = |____|
6 + 2 = |____|	12 + 4 = |____|
4 + 5 = |____|	22+37 = |____|
3 + 3 = |____|	38 +26= |____|

Extract 4: Items on Addition task

3.5.2. Measure of skill of Addition

Learners' achievement on Addition items is presented as: (i) the percentage of learners who correctly scored at least 50% of the items; and (ii) mean number of items correctly scored by the learners.

3.5.3. Effectiveness of the EGM Intervention on Addition Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on addition subtask compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on addition subtask in the treatment school (91%) was higher than that of the control school (70%). The baseline and endline percentages per class and school category are as shown in the figure below.

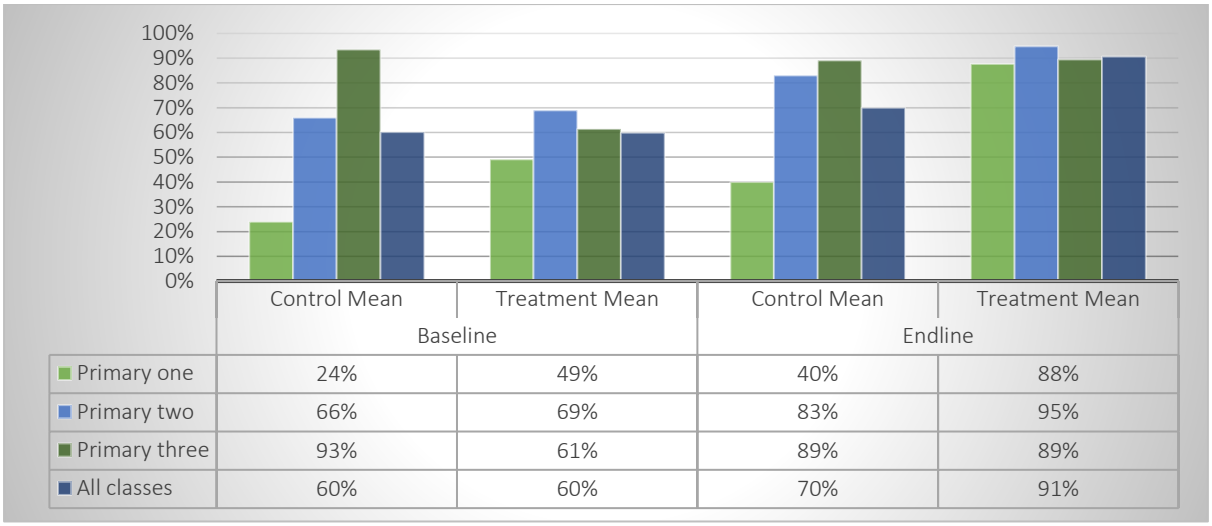


Figure 7: Addition – percentage of learners who correctly scored at least 50% of the test items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the addition subtask is investigated using the intent-to-treat estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (addition) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 10: Addition - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.23	0.09	0.32	0.21
Std. Errors	0.122	0.112	0.097	0.071
P-value	0.066*	0.431	0.001***	0.031**
Panel 2: ITT Estimates with covariates				
DID Estimates	0.25	0.06	0.32	0.21
Std. Errors	0.123	0.108	0.096	0.072
P-value	0.044**	0.571	0.001***	0.033**
Observations	214	210	206	630
R-square	0.22	0.08	0.12	0.08

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the addition knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items (DIDE=23%). This effect is statistically significant at 10% level.

Primary Two: The results in column (2) show that, the project improved the addition knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items (DIDE=9%). This effect is however not statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the addition knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items (DIDE=32%). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the addition knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=21%). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the percentage of learners who correctly scored at least 50% of the items on addition subtask ($female * treatment\ coef = 0.17; P - value = 0.314$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.30**	0.07	0.05	0.12
Female	-0.33**	-0.08	0.18**	-0.17
Treatment X Female	*0.25*	*0.10*	*-0.10*	*0.17*
Constant	0.63	0.87	0.82	0.79
Observations	107	105	103	315
R-squared	0.30	0.04	0.04	0.09
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

(ii) The mean number of items correctly scored by the learners

At endline, the mean number of items correctly scored by the treatment school learners in the addition subtask (mean=6.8 items) was higher than that of the control school (mean=5.5 items). The baseline and endline means per class and school category are as shown in the figure below.

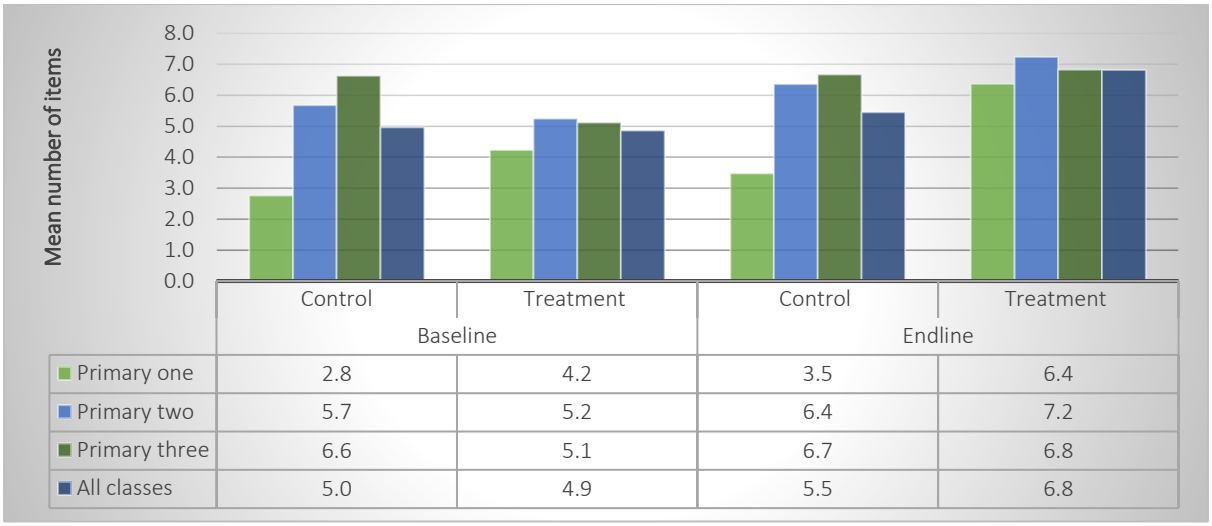


Figure 8: Addition – mean number of items correctly scored by the learners

The impact of the project on the mean number of items correctly scored by the learners on the addition subtask is as shown below:

Table 11: Addition - mean number of items correctly scored by the learners

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	1.42	1.32	1.66	1.46
Std. Errors	0.631	0.611	0.609	0.225
P-value	0.026**	0.032**	0.007***	0.001***
Panel 2: ITT Estimates with covariates				
DID Estimates	1.55	1.20	1.64	1.46
Std. Errors	0.641	0.595	0.593	0.204
P-value	0.017**	0.046**	0.006***	0.001***
Observations	214	210	206	630
R-square	0.26	0.12	0.10	0.10

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the addition knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners within 60 seconds (DIDE=1.4 items). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the addition knowledge in the treatment than control school, in terms of the mean number of items correctly scored

by the P2 learners within 60 seconds (DIDE=1.3 items). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the addition knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners (DIDE=1.7 items). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the addition knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.5 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the mean number of items correctly scored by the learners ($female * treatment\ coef = 1.08; P - value = 0.097$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	2.40***	0.58	0.05	0.80
Female	-0.76	-0.57	0.44	-0.94*
Treatment X Female	0.71	0.58	0.37	1.08*
Constant	4.00	6.65	6.50	5.96
Observations	107	105	103	315
R-squared	0.34	0.05	0.02	0.10
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

3.6. Subtask 5. Subtraction

3.6.1. Commentary on Subtraction

This subtask assesses learners' ability to perform simple/basic and complex subtraction operations, ranging from one-digit numbers through two-digit numbers. This subtask was untimed with a stop rule after four consecutive mistakes. A learner was presented with a piece of paper having 10 items, categorized into two levels, where he/she was asked to tell the answer on each item, after carrying out some computations. The items on this subtask are shown below:

Level 1	Level 2
4 - 1 = |_____|	19 - 6 = |_____|
5 - 2 = |_____|	25 - 7 = |_____|
8 - 2 = |_____|	26 - 14 = |_____|

9 - 5 = |_____|	59 - 37 = |_____|
6 - 3 = |_____|	64 - 26 = |_____|

Extract 5: Items on Subtraction task

3.6.2. Measure of the skill of Subtraction

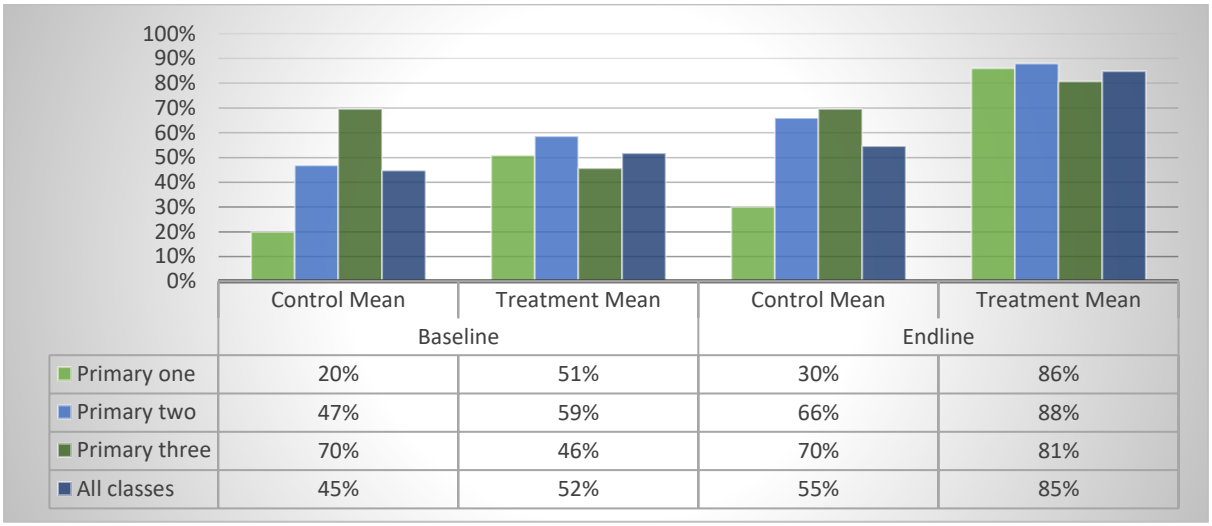
Learners' achievement on subtraction items is presented as: (i) the percentage of learners who correctly scored at least 50% of the items; and (ii) mean number of items correctly scored by the learners.

3.6.3. Effectiveness of the EGM Intervention on Subtraction Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on subtraction subtask compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on subtraction subtask in the treatment school (85%) was higher than that of the control school (55%). The baseline and endline percentages per class and school category are as shown in the figure below.



	Baseline		Endline	
	Control Mean	Treatment Mean	Control Mean	Treatment Mean
Primary one	20%	51%	30%	86%
Primary two	47%	59%	66%	88%
Primary three	70%	46%	70%	81%
All classes	45%	52%	55%	85%

Figure 9: Subtraction – percentage of learners who correctly scored at least 50% of the test items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the subtraction subtask is investigated using the intent-to-treat estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (subtraction) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 12: Subtraction - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.25	0.10	0.35	0.23
Std. Errors	0.119	0.128	0.129	0.051
P-value	0.036**	0.428	0.007***	0.006***
Panel 2: ITT Estimates with covariates				
DID Estimates	0.29	0.09	0.36	0.24
Std. Errors	0.119	0.126	0.129	0.055
P-value	0.016**	0.482	0.007***	0.008***
Observations	214	210	206	630
R-square	0.26	0.10	0.08	0.12

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items (DIDE=25%). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items (DIDE=10%). This effect is however not statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items (DIDE=35%). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=23%). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show differential impact of the EGM intervention on the sex of the learners (in favour of females), in terms of the percentage of learners who correctly scored at least 50%

of the items on subtraction subtask ($female * treatment\ coef = 0.24;\ P - value = 0.028$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.18***	0.11	0.11	0.40**
Female	-0.18	0.04	-0.16	-0.20*
Treatment X Female	0.24	0.01	0.22	0.24**
Constant	0.64	0.68	0.74	0.44
Observations	107	105	103	315
R-squared	0.13	0.02	0.09	0.34
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

(ii) The mean number of items correctly scored by the learners

At endline, the mean number of items correctly scored by the treatment school learners in the subtraction subtask (mean=6.2 items) was higher than that of the control school (mean=4.0 items). The baseline and endline means per class and school category are as shown in the figure below.

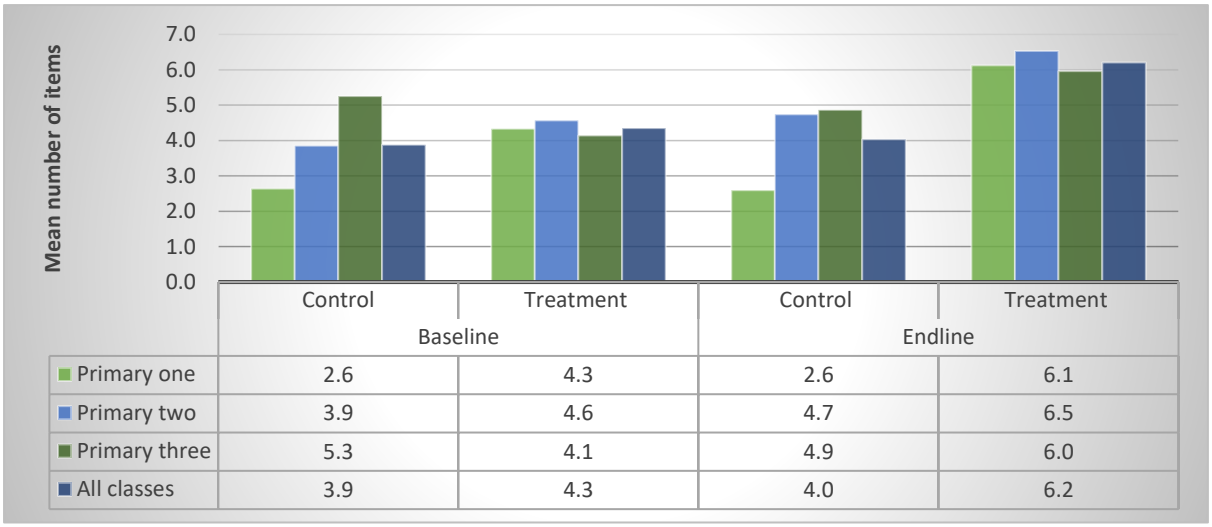


Figure 10: Subtraction – mean number of items correctly scored by the learners

The impact of the project on the mean number of items correctly scored by the learners on the subtraction subtask is as shown below:

Table 13: Subtraction - mean number of items correctly scored by the learners

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	1.83	1.07	2.22	1.71
Std. Errors	0.615	0.715	0.765	0.341
P-value	0.003***	0.135	0.004***	0.004***
Panel 2: ITT Estimates with covariates				
DID Estimates	1.97	0.94	2.25	1.71
Std. Errors	0.605	0.690	0.761	0.372
P-value	0.001***	0.173	0.003***	0.006***
Observations	214	210	206	630
R-square	0.30	0.14	0.06	0.12

Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates".
*** p<0.01, ** p<0.05, * p<0.1

Primary One: The results in column (1) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners (DIDE=1.8 items). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P2 learners (DIDE=1.1 items). This effect is however not statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners (DIDE=2.2 items). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the subtraction knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.7 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show differential impact of the EGM intervention on the sex of the learners (in favour of females), in terms of the mean number of items correctly scored by the learners on subtraction subtask ($female * treatment\ coef = 1.77$; $P - value = 0.023$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	2.68***	0.60	0.96	1.26*
Female	-1.42*	-1.86**	0.03	-1.51**
Treatment X Female	*1.21*	*2.44***	*0.42*	*1.77***
Constant	3.56	5.70	4.86	4.84
Observations	107	105	103	315
R-squared	0.48	0.20	0.05	0.21

Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1

3.7. Subtask 6. Word Problems

3.7.1. Commentary on Word Problems

This subtask assesses learners' ability to interpret and solve real world problems by applying the necessary arithmetic skills. This subtask was untimed with a stop rule after four consecutive errors. A learner was presented with a sheet of paper having 8 items, categorized into two levels, and he/she was asked to tell the answer after the assessor had read the item. The items on this subtask are given below:

Level 1

1. Okot has one orange. Ocan has three oranges. How many oranges do they have altogether?
2. Mum has three eggs. Dad has two eggs. How many eggs do they have altogether?
3. There are four tomatoes in the basket. Otim took away one tomato. How many tomatoes are left in the basket?
4. Apiyo has five mangoes. She gave away two mangoes to her friend. How many mangoes did she remain with?

Level 2

5. Two children are sitting under a mango tree. Three more children join them. How many children are sitting under the tree altogether?
6. There are six children playing football. Two are boys and the rest are girls. How many girls are playing football?
7. There are two players on Okello's team and seven players on Apiyo's team. How many players must join Okello's team so that it has the same number of players as Apiyo's team?
8. Five children joined the class. Now there are 12 children in the class. How many children were in the class to begin with?

Extract 6: Items on Word Problems task

3.7.2. Measure of Word Problems Solving Skill

Learners' achievement in solving word problems is presented as: (i) the percentage of learners who correctly scored at least 50% of the items and (ii) mean number of items correctly scored by the learners.

3.7.3. Effectiveness of the EGM Intervention on Word Problem Knowledge

Study Hypothesis: We anticipated that more learners in the treatment group will correctly score at least 50% of the items on word problem subtask compared to those in the control group, after the EGM intervention.

(i) The percentage of learners who correctly scored at least 50% of the items

At endline, the percentage of learners who correctly scored at least 50% of the items on word problem subtask in the treatment school (95%) was higher than that of the control school (69%). The baseline and endline percentages per class and school category are as shown in the figure below.

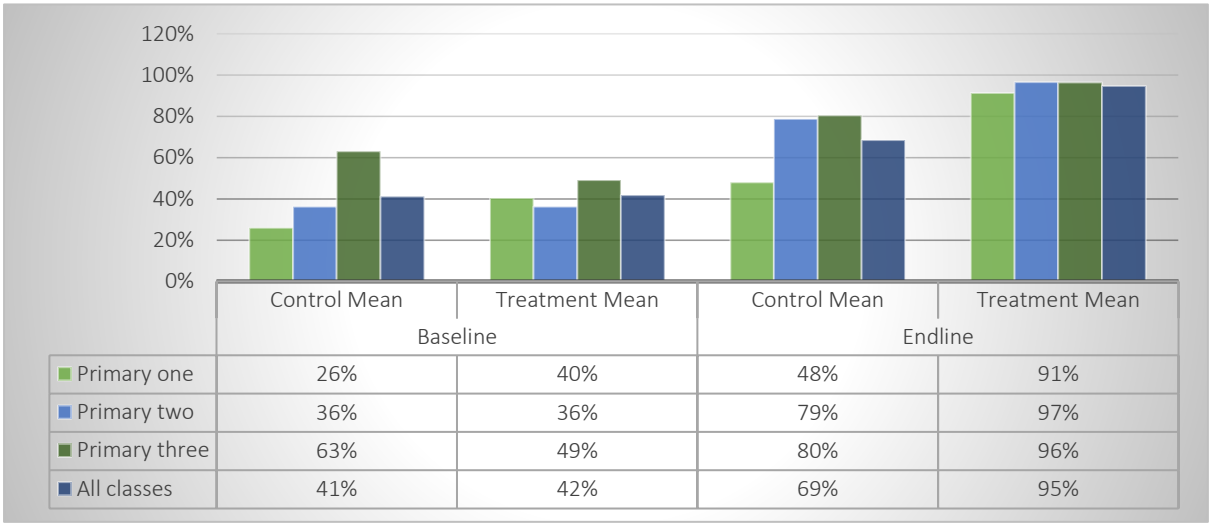


Figure 11: Word problem – percentage of learners who correctly scored at least 50% of the test items

The impact of the project on the percentage of learners who correctly scored at least 50% of the test items on the word problem subtask is investigated using the intent-to-treat estimates. The ITT estimates are calculated through a difference in differences approach of the dependent variable (word problem) on: period (baseline vs endline) and a vector of control variables (age, ever repeated a class, nursery attendance and sex).

Table 14: Word problem - proportion of learners who correctly scored at least 50% of the test items

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	0.29	0.18	0.30	0.26
Std. Errors	0.121	0.115	0.117	0.077
P-value	0.018**	0.124	0.011**	0.021**
Panel 2: ITT Estimates with covariates				
DID Estimates	0.31	0.16	0.30	0.25
Std. Errors	0.119	0.113	0.117	0.078
P-value	0.011**	0.169	0.010***	0.023**
Observations	214	210	206	630

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
R-square	0.24	0.31	0.17	0.22
Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates". *** p<0.01, ** p<0.05, * p<0.1				

Primary One: The results in column (1) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the percentage of P1 learners who correctly scored at least 50% of the test items (DIDE=29%). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the percentage of P2 learners who correctly scored at least 50% of the test items (DIDE=18%). This effect is however not statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the percentage of P3 learners who correctly scored at least 50% of the test items (DIDE=30%). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the percentage of learners who correctly scored at least 50% of the test items (DIDE=26%). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the mean number of items correctly scored by the learners ($female * treatment\ coef = 0.17; P - value = 0.096$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	0.30**	0.07	0.19**	0.17**
Female	-0.21	-0.16	0.05	-0.17*
Treatment X Female	0.20	0.22*	-0.07	0.17*
Constant	0.63	0.87	0.79	0.78
Observations	107	105	103	315
R-squared	0.25	0.11	0.07	0.14

Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1

(ii) The mean number of items correctly scored by the learners

At endline, the mean number of items correctly scored by the treatment school learners in the word problem subtask (mean=6.1 items) was higher than that of the control school (mean=4.1 items). The baseline and endline means per class and school category are as shown in the figure below.

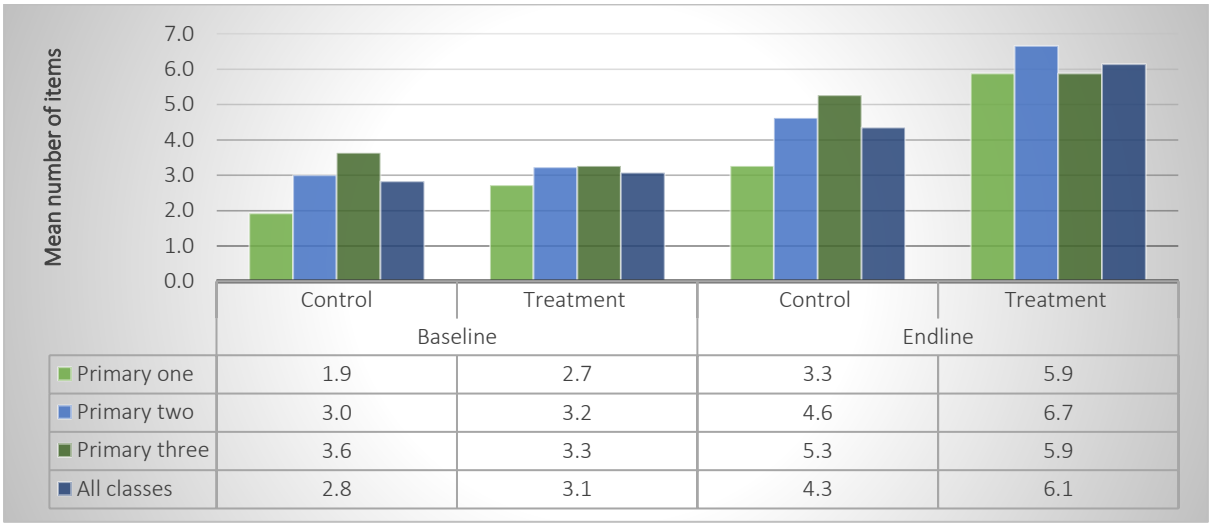


Figure 12: Word problem – mean number of items correctly scored by the learners

The impact of the project on the mean number of items correctly scored by the learners on the word problem subtask is as shown below:

Table 15: Word problem - mean number of items correctly scored by the learners

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
Panel 1: ITT Estimates without covariates				
DID Estimates	1.82	1.81	0.98	1.55
Std. Errors	0.516	0.454	0.443	0.232
P-value	0.001***	0.000***	0.028**	0.001***
Panel 2: ITT Estimates with covariates				
DID Estimates	1.95	1.74	0.99	1.55
Std. Errors	0.503	0.446	0.438	0.222
P-value	0.000***	0.000***	0.026**	0.001***
Observations	214	210	206	630
R-square	0.40	0.47	0.35	0.37
Notes: Columns (1) to (4) report the intent-to-treat estimates of the project impact. Standard errors are robust, and clustered at school-class level for estimates in Column (4). The ITT is based on the baseline and endline samples for only learners who were assessed at these 2 points. The baseline covariates in Panel 2 include: age of the learner, sex of the learner, whether learner ever attended				

	(1)	(2)	(3)	(4)
	Primary one	Primary two	Primary three	All classes
nursery and whether learner ever repeated any class. The R-squared estimates are for "ITT without covariates". *** p<0.01, ** p<0.05, * p<0.1				

Primary One: The results in column (1) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P1 learners (DIDE=1.8 items). This effect is statistically significant at 5% level.

Primary Two: The results in column (2) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P2 learners (DIDE=1.8 items). This effect is statistically significant at 5% level.

Primary Three: The results in column (3) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the P3 learners (DIDE=1.0 items). This effect is statistically significant at 5% level.

All classes: The results in column (4) show that, the project improved the word problem knowledge in the treatment than control school, in terms of the mean number of items correctly scored by the learners (DIDE=1.6 items). This effect is statistically significant at 5% level.

⇨ Impact of EGM Intervention by sex

Overall, the results show no differential impact of the EGM intervention on the sex of the learners, in terms of the mean number of items correctly scored by the learners ($female * treatment\ coef = 0.73$; $P - value = 0.084$), at 5% level of significance.

	(1)	(2)	(3)	(4)
	Primary 1	Primary 2	Primary 3	All classes
Treatment	2.02***	1.68***	0.77	1.39**
Female	-1.00	-0.75	0.21	-0.89**
Treatment X Female	*0.85*	*0.70*	*-0.42*	*0.73**
Constant	3.94	5.00	5.18	4.82
Observations	107	105	103	315
R-squared	0.33	0.29	0.03	0.20
Notes: The estimates are based on endline samples of learners who were only assessed at both baseline and endline phases, without covariates. The standard errors are robust, and clustered at school-class level for estimates in Column (4). *** p<0.01, ** p<0.05, * p<0.1				

3.8. Findings from the debrief with the teachers

In order to understand in-depths the changes in the treatment school as a result of the EGM intervention, teachers of Aywee Nursery and Primary School were engaged into a debriefing session. Their views and perceptions in regards to EGM intervention are as below:

According to most of the teachers in the treatment school, what they have learnt has facilitated them to effectively teach mathematics in their school. They now make it easy for the learners to understand the mathematical concepts being taught. For instance, one of the teachers said: "The Concrete Practice and Abstract approach has made teaching of graph work very easy. Learners now find it easy to understand this concept" (Teacher, treatment school). During the course of the intervention, the LGIHE team was able to notice most teachers successfully delivering lesson plans that they had created and this led to a positive response from the children in terms of their attitude towards mathematics and their understanding mathematical concepts.

Furthermore, the EGM intervention has enabled most teachers to control the class, make all learners engaged in the lesson and also carryout practical activities. For some of them, it was not easy to control and manage the pupils as well as involving them in practical activities before receiving the intervention. This was also observed during the course of the intervention where the LGIHE team observed that most teachers delivered their lessons in line with the CPA methodology/Singapore Math. Teachers delivered their lessons catering for all stages of lesson development that is to say: engagement, concrete, pictorial, and abstract, and assessment.

In addition to this, teachers have acknowledged the fact that they now use little energy in teaching as one of them stated: "Formally, I used a lot of energy to teacher but after receiving the EGM trainings, I now use little energy to teach and the learners understand the concepts" (Teacher, treatment school). This has been made possible through the knowledge and skills acquired on lesson planning that reflects the CPA approach. During the course of the intervention, the LGIHE team observed that teachers had up-to-date lesson plans that were matching with the respective schemes, and with relevant learning materials. However, some teachers had not commented on the self-evaluation part of the lesson plan as well as not stated the ways forward for the various lessons.

Related to this, teachers in the treatment school have appreciated the fact that, they learnt how to make teaching and learning materials, assess learners, handle learners with individual differences, and also dialogue with learners. This has enabled them to employ child centred teaching and learning approaches. Moreover, the skills gained have also enabled most teachers to teach mathematics from Abstract to Concrete. This has made learning mathematics very interesting to some learners. One of the teachers said: "The approach that we use nowadays has positively changed the attitude of learners towards mathematics. The rate of absenteeism has also reduced". This latter perception is supported by the fact that during the period of data collection, only 3(2%) learners sampled in the treatment school were absent. Related to the aspect of use of learning materials, during the course of the intervention, the LGIHE team observed that most teachers had learning materials - both concrete and pictorial which were

maximally used to develop Maths concepts among learners. However, some teachers still found difficulties in using enough learning materials attributing it to the large class sizes that sometimes hinders them from reaching all the learners.

In regards to the Community of Practice platform that was formed as a result of the EGM intervention, teachers in the treatment school perceived it as being useful since it has helped them to discuss issues regarding: the new teaching approach; classroom control; and engagement of learners. During the course of the intervention, the LGIHE team had a reflection with the teachers on the fundamental social nature of human learning as well as the role of team work as one of the avenues for the teachers' academic and social growth. Indeed, teachers during the debrief had acknowledged the COP meetings to have instilled teamwork and/or cooperation among teachers.

For continuity of the COP, teachers proposed to have weekly meetings where they will be meeting every Thursday however, they still felt need for additional support from LGIHE and AVSI in terms of materials. Teachers of the treatment school also envisage to take and keep minutes of the various COP meetings for any correspondence or reference.

In summary, teachers of the treatment school have observed the following most significant changes as a result of the EGM intervention:

- The trained teachers now use Singapore math approach that entails hands-on activities which make learners understand math concepts.
- Teachers now use child centred approaches coupled with interesting lessons that motivate the learners to comprehend the concepts.
- Teachers' confidence on EGM has improved as they view themselves with a potential of helping their learners to improve on the numeracy skills.
- Teachers stated to have improved on their scheming skills whereby they now take into consideration the various public holidays when scheming.
- The EGM intervention has instilled in the teachers the practice of always being prepared for teaching, so preparation for teaching is part of their culture.
- The EGM intervention has promoted teamwork (sharing of ideas) among teachers especially as a result of the COP.
- There is also close relationship between pupils and teachers as a result of the EGM intervention.
- The EGM intervention has made the teachers to understand their areas of weakness for improvement. This they say, was acquired during the classroom observation sessions.

CHAPTER FOUR: CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions

The results realized through difference in differences analyses reveal improvement in the treatment school compared to the control over the period of 4 months. A debrief with teachers reveals these improvements to be attributed to the EGM intervention that was implemented in the treatment school immediately after the baseline assessment. To the teachers, the CPA approach and all the aspects covered in the follow-up sessions such as classroom management, material creation, lesson planning, role of teamwork and assessment for learning have made teaching and learning of mathematics very interesting thus learners are actively engaged in the learning process.

The results therefore support the stated hypotheses with statistically significant improvements (at 5% level) being observed in the subtask areas of: number discrimination, missing numbers, addition, subtraction and word problems. Smaller project effects were realized in the subtask areas of 'number identification' in terms of the percentage of learners who scored at least 50% of the test items. However, a statistically significant improvement was realised in the mean number of items correctly scored by the learners.

Furthermore, variations in the project effects were realized across the 3 target classes in terms of percentage of learners who scored at least 50% of the test items. This could be attributed to the differences in teachers' attitude and pedagogical skills in EGM as observed during the course of the intervention.

4.2. Recommendations

Based on the findings of this assessment, the following recommendations are made:

☑ Increase on the length of the intervention

For teachers to fully comprehend and effectively implement the knowledge and skills acquired from the EGM interventions, they need more time and practice. This would allow the project implementors to have series of follow-up sessions and it would also help teachers in fostering the learning culture initiated by the project through the creation of the community of practice cycles.

☑ Training of teachers on materials development

To maximize the impact of the Singapore Mathematics Approach, there is need for an extensive training of the teachers on materials development to enable them create authentic and appropriate teaching materials which foster teaching and learning of numeracy. In addition to this, the schools also need support in terms of text books for reference.

☑ Timing of the training

The timing for the training needs to be agreed upon by the headteacher and the teachers. This should lead to suitable days when the teachers find it easy to attend the trainings. This would also reduce on cases where teachers find it difficult to attend the trainings as evidenced during one of the follow-up sessions when the turn-up of teachers was very low.

☑ Extend the intervention to other classes

During the EGM intervention, teachers recommended a similar training for teachers of transition and upper classes. This is mainly because they commended the content and approaches that were implemented in the lower classes. To them, a training of this kind in other classes would facilitate a holistic approach to improving teaching and learning of Mathematics in their schools.

4.3. Areas for future research

Further research could be done to investigate how teachers' attitudes and pedagogical skills influence acquisition, retention and application of knowledge and skills acquired from the intervention.

Further research could also be done to investigate the extent and scope of teamwork among teachers instilled by the EGM intervention and how it influences performance in numeracy in refugee contexts.