

# **INDEPENDENT EVALUATION REPORT BROAD CLASS - LISTEN TO LEARN**



<u>Disclaimer</u>: This report has been compiled by a task force of external evaluators hired through a competitive process. The views expressed in the report are based on the information and data collected during the evaluation process and do not necessarily reflect the views of Ilm Ideas. This report is an intellectual property of Ilm Ideas and the same or any part thereof my not be used in any manner whatsoever, without express permission of Ilm Ideas in writing.

# **Table of Contents**

EXECUTIVE SUMMARY	1
1 PROJECT BACKGROUND	4
1.1 INTRODUCTION	4
1.2 PROJECT OBJECTIVES	4
1.3 OUTCOMES AND OUTPUTS	4
2 EVALUATION METHODOLOGY	6
2.1 THE EVALUATION STUDY	6
2.2 EVALUATION DESIGN	6
2.3 EVALUATION METHODOLOGY	7
3 KEY EVALUATION FINDINGS	8
3.1 THE PROJECT DESIGN	8
3.1.1 RELEVANCE – RESPONDING TO A QUALITY ISSUE	8
3.1.2 IN SYNC WITH THE CURRICULUM	9
3.1.3 TEACHING ENGLISH LANGUAGE – CHALLENGES IN THE LOCAL CONTEXT	10
3.1.4 Scope of Activities	11
3.1.5 TARGETING STRATEGY	11
3.1.6 FORMATIVE ASSESSMENTS:	11
3.2 PROJECT MANAGEMENT AND IMPLEMENTATION	11
3.2.1 PROJECT PROGRESS AND RESULTS	11
3.3 STATISTICAL ANALYSIS OF LEARNING OUTCOMES	16
3.3.1 ANALYSIS OF ENGLISH ASSESSMENT SCORES OF KG STUDENTS	17
3.3.2 COHEN'S D MEASURE	18
3.3.3 ANALYSIS OF ENGLISH ASSESSMENT SCORES OF GRADE 1 STUDENTS	20
3.3.4 COHEN'S D MEASURES	21
3.4 SUSTAINABILITY AND IMPACT	22
3.4.1 Emerging impact	22
4 VALUE FOR MONEY ANALYSIS	27
5 KEY RECOMMENDATIONS	30
ANNEX 1: IN-DEPTH INTERVIEWS	31
ANNEX 2: STANDARDIZED EFFECT SIZES	33

# **List of Tables**

TABLE 1.1: KEY PROJECT OUTCOMES AND OUTPUTS	4
TABLE 2.1: EVALUATION PROCESS	6
TABLE 2.2: GRADES ASSESSED IN TREATMENT AND CONTROL SCHOOLS	6
FIGURE 3.A: HOW OFTEN DO YOU USE THE TEACHING AIDS PROVIDED BY THE PROJECT?	10
TABLE 3.1: PROGRESS AGAINST PROJECT TARGETS	13
TABLE 3.2: GENERAL ASSESSMENT IN ENGLISH - POST-INTERVENTION AVERAGE SCORES	18
TABLE 3.3: COMPARISON OF LOCATION-WISE POST-INTERVENTION MEAN SCORES	18
TABLE 3.4: COHEN'S D ESTIMATES FOR IMPROVEMENT IN SCORES	19
TABLE 3.5: COMPETENCY-WISE POST-INTERVENTION SCORES – TREATMENT VS. CONTROL	19
TABLE 3.6: POST-INTERVENTION AVERAGE SCORES	20
TABLE 3.7: COMPARISON OF LOCATION-WISE POST-INTERVENTION AVERAGES	20
TABLE 3.8: COHEN'S D ESTIMATES FOR PROGRAMME IMPACT	21
TABLE 3.9: COMPETENCY-WISE POST-INTERVENTION SCORES - TREATMENT VS. CONTROL	21
FIGURE 3.B: ARE STUDENTS MORE REGULAR IN ATTENDING ENGLISH LANGUAGE CLASSES NOW?	24
FIGURE 3.C: TO WHAT EXTENT HAS ATTENDANCE IMPROVED IN ENGLISH LANGUAGE CLASSES?	24
FIGURE 3.D: HAS ATTENTIVENESS OF STUDENTS IMPROVED WITH THE INTRODUCATION OF RADIO- ORIENTED EXERCISES, COMPARED TO TRADITIONAL LEARNING TOOLS?	25
FIGURE 3.E: WHAT IS THE LEVEL OF MOTIVATION OF STUDENTS FOR LEARNING ENGLISH AFTER THE FIRST FIVE LESSONS?	: 25
FIGURE 3.F: HAS THE NUMBER OF STUDENTS ASKING QUESTIONS INCREASED?	26
FIGURE 3.G: TO WHAT EXTENT DO PARENTS TAKE MORE INTEREST IN MONITORING THE PROGRESS THEIR CHILDREN IN THIS PROJECT, COMPARED TO INTEREST IN PREVIOUS LEARNING METHOD	SOF S?
	20
IADLE 4.1. CUSI-EFFECTIVEINESS UP INTERVENTIUNS	23

## **List of Abbreviations**

- AEO Assistant Education Officer
- **DAI** Development Alternatives Inc.
- **DD** Difference in Differences
- EDO Executive District Officer
- ES Effect Size
- ICT Information Communication Technologies
- IDI In-Depth Interview
- IRI Interactive Radio Instruction
- LFA Log Frame Analysis
- **NEGD** Nonequivalent Group Design
- **NEP** National Education Policy
- PTA Parent-Teacher Association
- SLO Student Learning Objective
- VFM Value for Money

### **Executive Summary**

This report presents the findings of the external evaluation conducted for assessing 'Broad Class – Listen to Learn," an Interactive Radio Instruction based educational project initiated in selected public schools of District Haripur in 2013, through the support of DFID. The project was implemented by POWER99. The project aims to create a classroom environment that promotes complex play, independence, socialization and problem-solving in the target schools.

The evaluation employed a mixed methods approach, using both qualitative and quantitative research methodologies. The qualitative aspect included a series of In-depth Interviews (IDIs) and Focus Group Discussions (FGDs) with various stakeholders. As part of the quantitative aspect, a student assessment was carried out for KG and Grade 1 students.

#### Key Findings:

#### a) Project

#### Design:

Examining the project design, the findings note that it derives its relevance primarily as a strategic fit given the larger picture, as it builds on priorities articulated in the National Education Policy 2009, particularly in the use of ICT. In examining the relevance of the project design, the evaluation findings note that the project adopted an Interactive Radio Instruction (IRI) based technology as a key entry point in order to achieve its goal of increasing access to quality education and improving learning outcomes in the target area. Given the context, the use of IRI technology in order to create an enabling teaching and learning environment appears to be well placed.

#### b) Project

#### Targets:

For a limited time horizon scale-up, the project objectives appear rather ambitious and spread out. From the improvement of student learning, to improvement of teaching practices and increased participation of parents and PTAs in educational activities, the project aims covered an extensive range of issues. However, the project interventions appear limited in their scope. The component focusing on engagement with parents and PTAs was mostly limited to the conducting of orientation sessions to introduce the project. The evaluation team did not find any evidence of institutional capacity building interventions carried out for the PTAs.

The findings note that overall, the project made good progress towards achieving its quantitative targets.

#### c) Project Results:

Addressing the complex issue of quality when the teachers are not adequately trained in teaching in English, the project focused on a rigorous teacher-training program. During the trainings, the teachers were introduced both to the IRI methodology as well as the subject

content. In general, the trainings were well received. Most of the teachers interviewed termed the training course as highly effective, saying that it helped them overcome some of their challenges, particularly in correct pronunciation of words in the English language. Commenting on the back-to-work application and efficacy of the training, 92% of the teachers interviewed said that they were using the techniques as well as the teaching aids acquired during the training regularly as part of their classroom practice.

The students' assessments indicated that for KG, the overall learning outcomes of students in treatment schools were marginally better than control schools. Across the eight competencies tested, children in treatment schools performed better in 4. For Grade 1, the overall learning outcomes of students in treatment schools were comparatively better than control. Across the 10 competencies tested, children in treatment schools performed better in 8.

#### d) Monitoring Processes:

Teacher training has been one of the core project components. However, in the absence of any institutional level linkages particularly with any teacher training institute, the trainings have largely remained individual focused. Interviews with the teachers note that there have been gains at the individual level in terms of knowledge/skill level. However, teachers' subsequent ability to implement what they have learned will rest on receiving institutional support, mentoring and monitoring. The findings note that the institutionalization of the project effort is not evident yet in the absence of any follow up mechanism and lack of institutional level coordination/ linkages.

#### e) Sustainability and Impact:

In assessing the project sustainability, it is noted that the project presently is heavily focused on technology rather than the educational outcomes. As noted earlier, the government department is generally supportive of the project interventions. There is an expectation that the implementing partner will continue broadcasting the program as its own contribution.

A major indicator in assessing the project impact is the willingness and ownership of the government institutions in taking for the project interventions. To this end, the findings note that while the project has been successful in developing inroads with the education department, particularly the Executive District Officer (EDO) office, the engagement has not resulted in the government taking ownership of the project by committing funds or financial support for the continuation or scaling up of the project.

The other dimension in assessing the impact of project interventions can be the measurement of actual change in knowledge, attitude and practice particularly as result of the capacity building component. While the full impact of change in practice, classroom practice and teaching methodology in this case can only be ascertained when assessed over a period of time, the findings note that in general the teachers reported to have adopted a more interactive and child-focused methodology in the classroom. The teachers also reported to adopt lesson planning more effectively as the programs required them to prepare lessons in advance. As seen in the assessment results the students in the project schools have largely performed better than the control schools.

# **1** Project Background

### **1.1 Introduction**

"Broad Class – Listen to Learn" is an Interactive Radio Instruction (IRI) based educational project initiated in selected public schools of District Haripur in 2013, through the support of DFID. It is being implemented by POWER99. The project has been piloted in selected schools of Islamabad Capital Territory and the interventions in Haripur are part of the project scale up.

The project was implemented in 198 classrooms across 59 public schools of Haripur. The project aims to create a classroom environment that promotes complex play, independence, socialization and problem-solving in the target schools.

### 1.2 Project Objectives

Key project objectives are:

- Promote child-centered practices at primary level in target area
- Improve quality of teachers through several capacity building measures
- Reduce dropouts and increase retention rate
- Increase participation of parents/members of School Councils in school activities

### 1.3 Outcomes and Outputs

Table 1.1 presents the key project outcomes and outputs:

Outcome	Outputs
Outcome 1:	Output 1.1:
Improved learning outcomes for children	Teaching and learning practice improved in schools
in urban and rural areas	Output 1.2:
	School selected from both rural and urban areas for the project
Outcome 2:	Output 2.1:
Improved attendance	Enrolment rate increased
	Output 2.2:
	Dropout level decreased in the schools
Outcome 3:	Output 3.1:
Improved teaching practices in the target schools	Teaching practices improved with provision of technical support during the lesson

#### TABLE 1.1: KEY PROJECT OUTCOMES AND OUTPUTS

	Output 3.2:
	Teachers shared their feedback, concerns and suggestions in follow-up training
Outcome 4:	Output 4.1:
Members of PTC, parents/communities are aware of program and taking part in	Increased capacity of communities to understand and advocate for their right to education
school activities leading to increased	Output 4.2:
parent and community substaction	Members of PTAs mobilized on their roles and responsibilities

### 2 Evaluation Methodology

### 2.1 The Evaluation Study

The present study is commissioned by Ilm IDEAS to carry out a detailed assessment of its project, "Broad Class – Listen to Learn" on the key evaluation parameters of Relevance, Effectiveness, Sustainability and Impact. It covers both a detailed commentary on the efficacy of the project design, as well as an assessment of project implementation and assessment processes. Furthermore, it is designed to assess the efficacy of the approaches developed and/or implemented by the project and to determine their viability for further scaling-up and long-term sustainability.

### 2.2 Evaluation Design

The evaluation employed a mixed mods approach and used both qualitative and quantitative research methodologies. The qualitative aspect included a series of In-depth Interviews (IDIs) and FGDs with various stakeholders including the implementing agency, govt. representatives from the education department – Executive District Officers (EDOs) and Assistant Education Officers (AEO)s, head teachers and teachers of the targeted schools, parents and school council members.

Stakeholders	Project staff	EDO/AEOs	Head Teachers	Teachers	Parents
No. of Interviews	3	2	18	18	
No. of FGDs				3	

As part of the quantitative aspect, a student assessment was carried out for KG and Grade 1 students. A sample with 90% Confidence Interval and 5% Margin of Error was drawn for the student assessment. The following table presents the breakdown of the sample:

#### TABLE 2.2: GRADES ASSESSED IN TREATMENT AND CONTROL SCHOOLS

	<b>Treatment Schools</b>		Control Schoo	
	KG	Grade 1	KG	Grade 1
Sample size	100	269	30	39

Oral assessments of the students from KG and Grade I were carried out. The assessment tools for the students were developed by the Assessment Expert from the evaluation team and were based on the

tools employed by the project for carrying out its baseline assessments. Considering the young age cohort of the children, however, only oral assessments were carried out.

Reliability of test was verified through employing Cronbach's Alpha, which is a measure of internal consistency and shows how closely related items in a set are as group. The Cronbach's Alpha coefficient for 46 items is 0.872. It shows the high reliability of the test. It is also indicates that items of test have consistency and inter reliability.

### 2.3 Evaluation Methodology

The methodology adopted for the evaluation followed a phased plan beginning with a preparatory stage, which included debriefing meetings with the DAI team, and a detailed review of the project document, followed by the development of an evaluation framework for the assessment. The second phase included the development of evaluation tools based on the framework and finalization of the field visit plan. The research and assessment teams were given a detailed orientation and a 2-day training before the roll-out of the field visit plan. The data collection exercise was also a phased activity, whereby the qualitative data collection was carried out first, followed by the students' assessment. Following the completion of the field-based data collection activity, the data compilation and analysis process was initiated. For the qualitative data, the analysis was carried out in multiple workshops with the field team. The analysis followed a sequential process consisting of raw data (statements made by the participants), descriptive data (summary statements of the respondents' comments), and interpretation (building on the summary statements and presenting the meaning of the data).

For the quantitative assessments the Mann-Whitney Test, a non-parametric test, was used to test for the similarity of distribution of scores of the program and comparison groups. This was because the distribution of scores follows a non-normal distribution and parametric tests such as the T- and F-test could not be applied. Cohen's D measure, based on mean and standard deviations of the samples and sub-samples, was used to calculate the cost-effectiveness of the program.

This report presents the findings of the evaluation based on the analysis carried out by the research team.

# **3** Key Evaluation Findings

This section presents the findings of the evaluation exercise. The discussion is captured along the key evaluation parameters including:

- A commentary on project design examining the validity of the design, relevance of the project and its targeting strategy
- An assessment of project progress against its targets including a comparative analysis of baseline v/s end-line findings
- An assessment of project management, key processes and project monitoring
- An examination of elements in sustainability including scalability and replicability
- The emerging impact of the project interventions

Evaluation findings against each of the above areas are presented as subsections in this chapter.

### 3.1 The Project Design

Examining the project design, the findings note that it derives its relevance primarily as a strategic fit given the larger picture, as it builds on priorities articulated in the National Education Policy 2009 particularly in its use of ICT. Specifically, it contributes to the following policy recommendations:

- ICTs shall be utilized creatively to assist teachers and students with a wide range of abilities and from varied socio-economic backgrounds.
- ICTs shall be used to strengthen the quality of teaching and educational management.

### 3.1.1 Relevance – Responding to a Quality Issue

The project rationale was built around the premise that in the given context where class rooms are dull, teaching methods are lecture based and the learning takes place primarily through rote memorization, the use of innovative technology can have a transformative effect on student learning as well as classroom environment and teaching methodology.

In examining the relevance of the project design, the evaluation findings note that the project adopted an Interactive Radio Instruction (IRI) based technology as a key entry point in order to achieve its goal of *increasing access to quality education and improving learning outcomes* in the target area. IRI technology is generally well known for its relevance in school systems where skilled teachers are in short supply and lack knowledge of instructional practices. Use of IRI can thus improve student learning while helping teachers gain skills, confidence and hands-on experience in specific pedagogies. Given the context, the use of IRI technology in order to create an enabling teaching and learning environment appears to be well placed.

The content was adapted from an earlier education project implemented through USAID's Education Sector Reform Assistance Program in 2002. In interviews with the evaluation team, the project staff shared that while they carried out some improvisations in the content, it was largely reproduced as it was developed in 2002.

While the project was designed with a view to addressing some of the key issues in terms of quality education in the local context, primarily focusing on the teaching learning methods in early grades, the project development was not preceded by an on-ground needs assessment and prioritization by the target group (public sector schools in this case). Some of the interventions, therefore, could have been developed differently, had the project carried out a detailed assessment mapping the challenges and opportunities in the operational context. For example, the project progress reports report multi grade classrooms as a key challenge to effective delivery of radio lessons in some cases. One of the reports cited 'over-crowded class rooms with *katchi* students combined with KG, Grade 1 and Grade 2 students' as a key cause of 'disturbance' during the lessons. A situational analysis preceding the implementation, therefore, could have identified these challenges and helped the project develop operational strategies to address them effectively. In interviews with evaluation team, references were made by the project implementing staff to some lessons gleaned from the pilot phase as basis for altering some of the project activities.

"Big class room size presented a challenge in carrying out interactive activities." Female Teacher – Dervish

#### 3.1.2 In Sync with the Curriculum

The overall design of the project, with its focus on teaching English language through IRI, resonates with the NEP 2009 recommendations as well as the national curriculum 2006, which emphasised the teaching of English from early grades<sup>1</sup>. According to the project staff, both the content and activities of the radio program are based on the national curriculum. The program content is thus designed to fit in with the regular school year and the prescribed syllabus for the selected grades, so that it does not become an add-on.

The interviews with the teachers, however, are revealing, as they indicate an inherent issue in the sphere of teacher education and professional development in the local context. The teachers and school administrators interviewed believed that the program content, though useful, was not based on the curriculum. This revealed that they were unfamiliar with the curriculum and had limited training regarding subject knowledge and the skills required to transfer that knowledge to students. There is therefore a fundamental flaw in the present teacher education and professional development programs, which appear to operate in isolation based on the traditional transmission paradigm, and are disconnected from the larger picture especially the curriculum and its objectives. The teachers and school heads in this case appeared to have insufficient knowledge of the curriculum and, were unable to

<sup>&</sup>lt;sup>1</sup> English language will be taught as a subject from grade I and use of English as the medium of instruction for Science and Mathematics from class IV onwards) – Education Policy 2009

relate the program content to it once the textbook was changed<sup>2</sup>. This observation is highly relevant in the current context where textbooks become very significant owing to the poor quality of teaching, and teachers are generally found to be highly reliant on textbooks.

"The radio class, even though (it) helped in improving the English language of the children... didn't follow the English syllabus in our books." Head Teacher – Sikander Pur

### 3.1.3 Teaching English Language – Challenges in the local context

Being taught academic content through a second language represents a multiple burden for the learner: understanding the abstract concept, understanding the high-level vocabulary, and understanding the language in which it is explained. In the local context this becomes even more complex as the teachers are not adequately trained in teaching English. Most teachers at primary level have a Primary Teacher Certificate (PTC). The certificate clearly states that the candidate has not been trained to teach English. Limited endowments in teaching English language were also evident in the case of the project. Almost all of the teachers interviewed said that they were initially not confident about their skills for teaching the English language. The project tried to address this issue through a rigorous teaching training program. The multiple phase training began with an orientation training and was followed up mid-course with a refresher course. Monitoring visits by the project team conducted during the course of the school year were also used as an opportunity to provide mentoring support to the teachers.

The teachers were introduced both to the IRI methodology as well as the subject content in the trainings. Most of the teachers interviewed termed the training course as highly effective, saying that it helped the overcome some of their challenges particularly in correct pronunciation of words in English. Commenting on the back-to-work application and efficacy of the training, 92% of the teachers interviewed said that they were using the techniques and the teaching aids acquired during the training regularly as part of their classroom practice.



#### FIGURE 3.A: HOW OFTEN DO YOU USE THE TEACHING AIDS PROVIDED BY THE PROJECT?

<sup>2</sup> The Project used books developed by the 'Book Group' in Karachi. These are generally considered quality books and according to the staff of the Group, as per the national curriculum, they have not been approved by any official authority.

#### 3.1.4 Scope of Activities

For a limited time horizon scale-up, the project activities appear rather ambitious and spread out. From the improvement of student learning, to improvement of teaching practices and increased participation of parents and PTAs in educational activities, the project aims covered an extensive range of issues. However, the project interventions appear limited in their scope. The component focusing on community/parents/PTA participation, for example, appears with very few focused interventions for community mobilization and institutional capacity building of the PTAs. The engagement with parents and PTAs was mostly limited to conducting orientation sessions to introduce the project.

#### 3.1.5 Targeting Strategy

In terms of a targeting strategy, the choice of district or the selection of schools does not appear to be based on pre-defined criteria. The project team explained that the schools were selected by the office of the EDO. However, there appeared to be no criteria in place except for inclusion of both urban and rural schools. The targeting strategy, however, is more specific in terms of its focus on gender. Out of the total coverage, almost 60% of the schools selected were girls' schools.

#### **3.1.6 Formative Assessments:**

Effective IRI also requires a commitment to advanced research and formative evaluation to ensure that programs are engaging and that they lead to enhanced learning<sup>3</sup>. For assessments, the project appears to have mainly relied on a baseline assessment of students' learning, followed by an end of project end-line assessment. For comparison, control schools have also been included in the sample for the two assessments. However, a clear rationale for the selection of control schools in the sample has not been articulated and as such the basis for comparison remains unclear for the baseline.

### 3.2 Project Management and Implementation

#### 3.2.1 **Project Progress and Results**

This section presents the project progress against its targets. Wherever possible, a comparative analysis with the baseline/research study has been carried to track change over the project duration.

In assessing the progress towards this output, the evaluation team used a triangulation technique and compared the project achievements presented in the LogFrame Analysis (LFA) and project documents with the data collected through field research and interviews with the project teams.

<sup>&</sup>lt;sup>3</sup> Gaible, Edmond and Mary Burns. 2005. Using Technology to Train Teachers: Appropriate Uses of ICT for Teacher Professional Development in Developing Countries. Washington, DC: infoDev / World Bank. Available at: http://www.infodev.org/en/Publication.13.html

The findings note that overall, the project made good progress towards achieving its quantitative targets. However, it is noted that the indicators developed for tracking the outcomes/outputs do not match in most cases. For example, for outcome 1: *improved learning outcomes for children in urban and rural areas*, the indicator developed by the project captures percentage of boys and girls participating in the program, whereas it should have captured the increase in learning outcomes measured through student assessment. Similarly, the target set against this outcome indicates the number of student beneficiaries targeted for outreach, which matches neither with the indicator nor the outcome it is expected to achieve.

Table 3.1 below presents the progress against the project targets.

Results Hierarchy	Objectively Verifiable Indicators	Targets	Progress	Comments
Outcome 1: Improved learning outcomes for children in urban and rural areas	% of boys and girls participating in the program	6, 000	Achieved - 8,732 (Reported)	OVIs and Targets do not match the outcome
Output 1.1: Teaching & Learning practice improved in schools	# of IRI broad class lesson on aired	Total 150 lessons 75 Lessons for KG- Grade I students 75 Lessons for	Achieved	OVI does not track the output. Output statement represents an outcome rather than an output
Output 1.2: School selected from both	# of schools with improved quality of education as a result of innovative	Grade II students 59	Achieved	OVI statement needs revision. Need to define 'improved quality of education'
the project Outcome 2:	solutions % of children increased in		Qualitative findings	OVI does not support the
Improved attendance Output 2.1:	<ul><li># of children enrolled in</li><li>class KG, Grado 1 &amp; Grado</li></ul>	20%	Support this Qualitative findings	Outcome. Increase in children indicates enrollment Difficult to attribute given the
Enrolment rate increased	2		Reported by project Approximately 5% increase in enrollment	

### TABLE 3.1: PROGRESS AGAINST PROJECT TARGETS

Output 2.2: Dropout level decreased in the schools	% of drop out decrease in the selected schools		No evidence	Difficult to an the short pro	ttribute gi oject duratio	ven n
Outcome 3: Improved teaching practices in the target schools	<pre># of participating schools that have conducted follow up teachers training</pre>	59	Achieved	OVI does outcome	not captur	re the
Output 3.1: Teaching practices	# of visits conducted for technical facilitation of	2 visits per school in a month	Achieved - 247 visits	-		
improved with provision of technical support during the lesson	teachers during the lesson					
Output 3.2:	# of male teachers	198	Achieved	Output st	tatement	Needs
Teachers shared their feedback, concerns and suggestions in follow-up training	participated # of female teachers participated			revision		
Outcome 4:	# of social mobilization and	20	Reported 59	Outcome	statement	needs
MembersofPTC,parents/communitiesareawareofprogramatkingpartinschoolactivitiesleadingactivitiesparentandincreasedparentandcommunitysatisfaction	advocacy sessions conducted		Partial achievement as per the qualitative findings	revision. N	Multiple ed together	expected

Output 4.1:	20%	Unclear target
Increased capacity of	a. # of men who	Reported: 448 men Indicator not capturing the
communities to understand and advocate for their right to education	participated in the events b. # of women who	participatedtarget1,158womenparticipated
	events	Not achieved
Output 4.2: Members of PTCs mobilized on their roles and responsibilities	<ul> <li>a. # of men PTCs 20% members mobilized</li> <li>b. # of women PTCs members mobilized</li> </ul>	Unclear target Not achieved Reported: 16 male PTC member participated 76 female PTC members participated

Overall, the logframe for the project is inadequately framed and there appears to be a disconnect between the project objectives and the interventions adopted to meet the objectives. For example, while one of the project objectives aims at "increasing participation of parents in schools through planned motivational programs" the only intervention strategy to achieve this appears to be general orientation sessions for parents to inform them about the project interventions. Similarly, the project sets out an ambitious objective to 'improve public sector service delivery by strengthening institutional capacities of education department' but none of the project intervention strategies seem to address this goal except the teacher-training component.

While the project results include improving attendance and enrolment with quantitative OVIs measuring percentage increase in each, evidence on attendance, enrolment and retention is anecdotal and is not documented in any of the progress reports whereas it has been put down as one of the project objectives. The evaluation team, however, believes that drop out/retention is a time-intensive conclusion and needs a longer period of time to prove. The underlying assumption is that if quality improves, then it will have an impact on the two indicators. However, for a limited time horizon scale-up, these should ideally not be set as project objectives or indicators.

### 3.3 Statistical Analysis of Learning Outcomes

This section presents an in-depth statistical analysis of learning outcomes of students at two class levels for General Assessment in English. Students in KG and Grade 1 of the treatment and control group were tested for General Assessment in English in the public schools of tehsil of Haripur, in the province of Khyber Pakhtunkwa. The tool for the assessment of students' learning outcomes of KG and Grade 1 were developed in the light of "Key Learning Areas and SLOs of National Curriculum of 2006". The Broad Class lessons are aimed at enhancing the listening and speaking skills of the students by giving them the basic concepts of English. The assessment was conducted through an oral test of the students to assess the improvement of 2 groups of students' learning outcomes in English due to and without intervention. The tool for grade 1 consisted of ten questions and each question contained a number of items. Reliability of test was judged by calculating the value of Cronbach's Alpha measure on 46 items which in this case was 0.872<sup>4</sup>. The scores used in the analysis are in percentages. The learning assessments are based on the data set collected by the DAI evaluation team from the treatment and control group after the intervention.

The statistical results are sensitive to the sampling strategy used to pick schools and students for treatment and control experiment. The last 10-year practice followed in the conduct of evaluations is to use randomized evaluation techniques based on random selection of treatment and control objects to assess the results. Selection of schools for treatment and control population did not theoretically and strictly follow the randomized sampling techniques, and many subjective and quantitative criteria were used to pick the schools for treatment. Moreover there was no assignment of schools in the control group. Thus a Nonequivalent Group Design (NEGD) was followed to pick schools. This sampling strategy impacts the internal validity of the program estimates. At the evaluation stage, a representative sample carrying a 5% margin of error with 90% confidence interval from the treatment population was picked to conduct tests on General Assessment of English of KG and Grade 1 students. The same test was also administered to select public schools not subject to treatment.

In the next section, the results of KG class for the treatment and control group are analysed by gender,

location and competency. The Mann-Whitney Test, a non-parametric test, is used to test for the similarity of distribution of scores of the program and comparison groups, as the distribution of scores follows a non-normal distribution and parametric tests such as T- and F-test cannot be applied. Since pre-intervention scores were not collected, a Difference-in-Differences (DD) estimate and ANCOVA model cannot be run to estimate the impact of the program in enhancing the listening and speaking English of the students. Cohen's D measure based on mean and standard deviations of the samples and sub-samples gives standardized effect size, and is used to calculate the cost effectiveness of the program.

#### 3.3.1 Analysis of English Assessment Scores of KG students

Table 3.2 gives the mean and standard deviations of the entire sample, and by gender, for program and comparison groups. Mann-Whitney tests the null hypothesis that the distribution of scores is identical for the two groups. If the p-value is less than 0.05, the null hypothesis is rejected. For the entire sample and male subsample, the p-value is less than 0.05, therefore the null hypothesis is rejected and the distribution of scores of both groups in each sample is not identical. In case of the female sub-sample, p-value > 0.05, the null hypothesis is not rejected and the distribution of scores of the program and comparison groups is identical.

	N	Average	Std-Deviations	Mann- Whitney
				Test*
Treatment	100	0.78	0.20	0.004
Control	30	0.67	0.16	
Male:				
Treatment	52	0.76	0.16	0.009
Control	16	0.61	0.22	
Female				
Treatment	48	0.81	0.16	0.068
Control	14	0.74	0.14	

#### TABLE 3.2: GENERAL ASSESSMENT IN ENGLISH - POST-INTERVENTION AVERAGE SCORES

\*p-value

<sup>4</sup> Cronbach's Alpha is a measure of internal consistency, that is, how closely related items are as a group. The estimated Cronbach's Alpha Coefficient of 0.872 indicates' high reliability of the test. It also indicates that items of test have consistency and inter-test reliability.

Table 3.3 gives the breakdown of KG post-intervention scores for the two groups by urban and rural location of students. In urban areas, the null hypothesis that the distribution of scores of students is identical is accepted, as p-value > 0.05. In rural areas, the p-value is less than 0.05, therefore the null hypothesis is rejected and the distribution of scores of the two groups is not identical. As per absolute value of means, the performance of rural students in the treatment group is marginally better than that of urban students.

	Urban		Rural				
Status	Ν	Mean	Mann- Whitney Test*	N	Mean	Mann- Whitney Test*	diff in means
Control	19	0.72	0.853	11	0.57	0.00	0.15
Treatment	30	0.75		70	0.80		-0.05

TABLE 3.3: COMPARISON OF LOCATION-WISE POST-INTERVENTION MEAN SCORES

#### 3.3.2 Cohen's D Measure

Keeping in view the limited scope of the analysis and statistical weakness in the data, Cohen's D is used to measure the impact or Effect Size (ES), in terms of gain in standard deviations in scores due to the program. Two interpretations of Cohen's D measure are as follows: a) Effect Sizes can be considered as the average percentile standing of the average treated participant, relative to the average control participant. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group. b) Effect Sizes can also be interpreted in terms of the percent of non-overlap of the treated group's scores with those of the untreated group. An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the treated group overlaps an en-overlap of 47.4 percent in the two distributions. An ES of 1.7 indicates a non-overlap of 75.4 percent in the two distributions.

Table 3.4 gives the Cohen's D for the sample and the location-wise sub-sample. For the entire sample, the improvement of scores is 0.691 standard deviations. It is fairly large and indicates a non-overlap of 43 percent of distribution of treated scores with comparison group. The small estimate for the urban sample indicates that 93 percent of both distributions overlap each other. A very large value of 1.521 for females indicates a 70 percent non-overlap of the distributions of the two groups.

	Ν	Value
Entire Sample	130	0.691
Urban	49	0.140
Rural	81	1.521

#### TABLE 3.4: COHEN'S D ESTIMATES FOR IMPROVEMENT IN SCORES

Table 3.5 compares the treatment and control group mean scores in 8 competencies tested in

KG students. As per the values of the Mann-Whitney Test, the distribution of scores is <u>identical</u> for treatment and control groups for five out of eight competencies, as p-value > 0.05 and the null hypothesis is not rejected. In other words, treatment did not change the position of the distribution of scores. The relevant competencies are Counting; Missing Letter; Sound of Alphabets, Difference between Capital and Small letters and identification of pictures.

COMPETENCY	Treatment N=100	Control N=30	Mann- Whitney test*	DIFF IN MEANS
	Mean	Mean		
Greetings	0.81	0.61	0.003	0.20
Counting	0.94	0.95	0.836	-0.01
Missing letter	0.84	0.73	0.091	0.11
Sound of alphabets	0.60	0.51	0.236	0.09
Draw structure	0.92	0.57	0.000	0.35
Difference b/w capital and small letter	0.92	0.93	0.965	-0.01
Identification of colors	0.52	0.32	0.007	0.20
Identification of pictures	0.72	0.71	0.602	0.01

TABLE 3.5: COMPETENCY-WISE POST-INTERVENTION SCORES – TREATMENT VS. CONTROL

\*p-value

#### 3.3.3 Analysis of English Assessment Scores of Grade 1 students

The p-value < 0.05 of Mann-Whitney test in Table 3.6 indicates that the null hypothesis of identical distribution of treatment and control groups for entire and gender-wise sub-sample of Grade 1 is rejected. In other words, the distributions are not identical. Similarly in Table 3.7, the distributions of scores of treatment and comparison for urban and rural sub-sample are not identical.

	N	Average	Std-Deviations	Mann- Whitney Test*
Treatment	269	0.80	0.14	0.00
Control	39	0.55	0.20	
Male:				
Treatment	106	0.75	0.16	0.00
Control	13	0.44	0.19	
Female				
Treatment	163	0.83	0.12	0.00
Control	26	0.60	0.18	

#### TABLE 3.6: POST-INTERVENTION AVERAGE SCORES

\*p-value

#### TABLE 3.7: COMPARISON OF LOCATION-WISE POST-INTERVENTION AVERAGES

	Urban			Rural			
Status	N	Mean	Mann- Whitney Test*	Ν	Mean	Mann- Whitney Test*	diff in mean s
Control	26	0.54	0.00	13	0.57	0.00	-0.03
Treatment	129	0.79		140	0.81		-0.02

\*p-value

#### 3.3.4 Cohen's D Measures

Using the above mean, standard deviations and unequal sample sizes, the impact of the program is calculated from the entire sample as well as for the urban and rural sub-sample. The results are given in Table 3.8. The impact of the program measured in standardized Effect Size is fairly large for the entire sample as well location-wise sub-samples. In case of the entire sample, nearly 75 percent of the distribution of two groups, i.e. program and comparison, do not overlap. In the case of females, the non-overlap is as high as 95 percent. In the case of males, the non-overlap of distributions is 45 percent.

#### TABLE 3.8: COHEN'S D ESTIMATES FOR PROGRAMME IMPACT

	Ν	Value
Entire Sample	308	1.681
Urban	155	1.151
Rural	153	1.746

Competency-wise mean post-intervention scores are compared for treatment and control groups in Table 3.9. Out of the 10 competencies, the distribution of scores of only two competencies are identical, while the distribution of scores of treatment and control groups of the remaining eight competencies are not identical as p-value < 0.05.

COMPETENCY	Treatment N=269	Control N=39	Mann- Whitney Test*	DIFF IN MEANS
	Mean	Mean		
Tell the Missing Alphabets	0.84	0.78	0.322	0.06
Greetings (Healthy Habits)	0.81	0.43	0.00	0.38
Match the Alphabets (Capital Letter with Small Letter)	0.97	0.88	0.00	0.09
Match the words with numbers	0.94	0.91	0.776	0.03
Identifying Colours	0.66	0.47	0.003	0.19
Draw a circle in the air	0.83	0.18	0.00	0.65
Identification of the things in home and in class room	0.80	0.60	0.00	0.20
Sound of the alphabets	0.58	0.26	0.00	0.32
Performing body actions	0.86	0.56	0.00	0.30
Reading Words	0.66	0.44	0.00	0.22

TABLE 3.9: COMPETENCY-WISE POST-INTERVENTION SCORES - TREATMENT VS. CONTROL

\*P-value

### 3.4 Sustainability and Impact

In assessing the project sustainability, it is noted that at present the project is heavily focused on technology rather than the educational outcomes. As noted earlier, while the government department is generally supportive of the project interventions, there is an expectation that the

implementing partner will continue broadcasting the program as its own contribution.

Teacher training has been one of the core project components. However, in the absence of any institution-level linkages particularly with any teacher-training institute, the trainings have largely remained individual-focused. Interviews with the teachers note that there have been gains at the individual level in terms of knowledge/skill level. However, teachers' ability to subsequently implement what they have learned will rest on receiving institutional support, mentoring and monitoring. The findings, however, note that the institutionalization of the project effort is not evident yet, due to lack of institution-level coordination/ linkages.

#### 3.4.1 Emerging impact

As the evaluation exercise follows the immediate project closure, it is difficult to fully assess the impact of various interventions. An effort by the evaluation team has, however, been made to identify the emerging trends possibly connected to impact. The discussion on impact is captured on two dimensions. The first is the immediate impact, in the shape of concrete steps which will follow the intervention. A major indicator in assessing the project impact is the willingness and ownership of the government institutions in taking for the project interventions. To this end, the findings note that while the project has been successful in developing inroads with the education department, particularly the EDO office, the engagement has not resulted in the government taking ownership of the project by committing funds or financial support for the continuation or scale-up of the project.

The other dimension of impact can be the measurement of actual change in knowledge, attitude and practice, particularly as result of the capacity building component. While the full impact of change in practice, classroom practice and teaching methodology in this case can only be ascertained when assessed over a period of time, the findings note that in general the teachers reported having adopted a more interactive and child- focused methodology in the classroom. The teachers also reported that they had adopted more effective lesson planning, as the programs required them to prepare lessons in advance. As seen in the assessment results, the students in the project schools have largely performed better than the control schools.

While interviews with the teachers and school administration as well as the project staff indicated that enrollment in the target schools had increased as a result of the project interventions, it remains to be seen if there is an actual increase in enrolment or parent were merely transferring their children enrolled in private schools to government schools. Evidence from the field appears to support this notion. Many of the parents participating in the FGDs shared that they had transferred their children to the target schools because the government schools had become 'English medium' schools.

> "Due to this project, the enrolment rate in this school has increased. Even the students of private schools are shifting to our school. Last year we enrolled 80 students." Head Teacher

The survey with teachers conducted as part of the evaluation reveals that in general attendance and interest of students in the class improved as a result of the project interventions. The results are presented in Figures 3.b-3.g.

#### FIGURE 3.B: ARE STUDENTS MORE REGULAR IN ATTENDING ENGLISH LANGUAGE CLASSES NOW?



NOT MUCH IMPROVEMENT SEEN 76-100% 51-75% 26-50% 1-25% 0% 5% 10% 15% 20% 25% 30% Series1

FIGURE 3.C: TO WHAT EXTENT HAS ATTENDANCE IMPROVED IN CLASSES?



FIGURE 3.D: HAS ATTENTIVENESS OF STUDENTS IMPROVED WITH THE INTRODUCTION OF RADIO-ORIENTED EXERCISES, COMPARED TO TRADITIONAL LEARNING TOOLS?

FIGURE 3.E: WHAT IS THE LEVEL OF MOTIVATION OF STUDENTS FOR LEARNING ENGLISH AFTER THE FIRST FIVE LESSONS?





FIGURE 3.F: HAS THE NUMBER OF STUDENTS ASKING QUESTIONS INCREASED?

**FIGURE 3.G: TO WHAT EXTENT DO PARENTS TAKE MORE INTEREST IN MONITORING THE PROGRESS OF THEIR CHILDREN IN THIS PROJECT, COMPARED TO INTEREST IN PREVIOUS LEARNING METHODS?** 



### **4** Value for Money Analysis

This section presents the Value for Money - VFM analysis. For the VFM analysis, the DFID approach comprising the 3 Es of Economy, Efficiency and Effectiveness has been adopted. The Independent Commission for Aid Impact adds a 4<sup>th</sup> dimension, namely Equity. However due to data and time constraints the dimension of equity will not be incorporated in the VFM analysis.<sup>5</sup>.

#### **Techniques and Methodology of VFM**

The evaluation literature considers the following 6 techniques categorized into 3 groups that are related to estimating the "Economy" component of VFM. These are: Cost Effectiveness Analysis; Cost Utility Analysis (Group 1); Cost benefit Analysis; Social Return on Investment (Group 2); and Rank Correlation of cost vs impact and Basic Efficiency Resource Analysis (Group 3).

The literature suggests that the 2 techniques grouped as 1 above, be adopted for evaluating programs that 'aim to reach the same goal in non-monetary terms'. In case of Broad Class the primary goal is to improve the English Language learning outcomes of KG, Grade 1 and Grade 2. A deeper insight indicates that the intervention in improving English Languages skills is different with respect to grade-wise learning outcomes. Thus the technique of cost-effectiveness for a single intervention across the three grades may not be strictly but is approximately comparable. Cost Effectiveness Analysis will compare the improvement in English Language learning outcomes measured as 0.1 gains in Std Deviations (SD) of KG and Grade 1 scores, per unit of costs measured in GBP<sup>6</sup>. For the convenience of policy makers we also measure the gains in SD of test scores per 100 GBP invested.

#### **Cost Effectiveness Analysis**

Table 4.1 summarizes the impact of interventions in terms of gains in SD of test scores. The cost calculations in the first panel are based on the impact of the intervention on the full sample of treatment students. Panel 2 and 3 in the table divide the treatment group into urban and rural samples. The standardized Cohen's D Effect Sizes directly give the gains of intervention in terms of SD for KG and Grade 1 students in Column 2. In Panel 1, the gain in scores for a treatment student is 1.681 SD in Grade 1, which is more than twice the gain of 0.691 SD estimated for KG students. For each of the two interventions we also calculate the amount of GBP needed to increase the test scores by 0.1 SD. For a KG student, GBP 1.67 is needed to increase the child's test score by 0.1 SD, and for Grade 1 the corresponding amount is GBP 0.69<sup>7</sup>. From an investment angle, to improve the quality of education, an investment of GBP 100 will yield an improvement of 5.99 SD in English Language learning skills of KG students as compared to the corresponding improvement of 14.57 SD for Grade 1 students<sup>8</sup>.

<sup>&</sup>lt;sup>5</sup> It is to be noted that we are not preparing a business case for each of these interventions. Business case involves components such as strategic analysis, Appraisal (VfM), commercial and financial viability and management case including risk assessment. VfM is one component of the business case.

<sup>&</sup>lt;sup>6</sup> Data on post-intervation learning assessments of Grade 2 was not collected for treatment and control groups, however Grade 2 beneficiaries are included in cost per student calculations

<sup>&</sup>lt;sup>7</sup> Gain of 0.2 SD moves the child from 50<sup>th</sup> to 58<sup>th</sup> percentile. As an observation children move between 0.5-0.9 SD in a year at a school. A table giving the impact on percentile improvement for selected SD gains is given in annex.

<sup>&</sup>lt;sup>8</sup> Even 1 SD per \$100 is good value for money.

From a cost effectiveness or an economy perspective it is useful to assess whether the intervention is more effective in urban or rural areas. Although the weakness in budget structure prevents us from disaggregating the budget by location to capture any economies of location, the differences in gains in SD across locations yield different cost estimates location-wise. The findings in Panel 2 of Table 4.1 indicate that in urban areas, the SD gains through intervention at KG level are quite low. The gains in SD for Grade 1 general assessment of English after the intervention are medium in conventional terms.

The third panel of Table 4.1 gives the cost calculations for the rural sample. At both levels, the gains in SD of scores are very large, suggesting that the intervention is effective and represents value for money from an economy standpoint.

In conclusion, the results for the single intervention aimed at two different class levels suggest that applying it to improve the English language skills of Grade 1 students is more cost-effective and gives higher value for money than for KG students. A caveat is in order here: in both cases, it is necessary to state that end-line assessments tend to overstate the gains in SD in the absence of pre-test data. Moreover, static one-time gains as estimated from the one-year data also tend to overstate the gains if compared to SD gains over a period of time. Across locations, there is value for money in investing in rural and urban areas at the Grade 1 level and only in rural areas at the KG level. The costs per 0.1 SD gain across the three interventions are fairly similar although it is lowest for Grade 1 in urban areas.

#### TABLE 4.1: COST-EFFECTIVENESS OF INTERVENTIONS

Data: End-Point DAI Evaluation of Broad Class						
<b>.</b>	· ·	<b>a</b> .	<u> </u>	<b>.</b> .		
Intervention	Average lest	Cost per	Cost per	Cost per	Additional	
	score gain in	student PKR	student	student in	SD per	
	SD		GBP	GBP per 0.1	GBP 100	
				SD gain		
KG	0.691	1950.08	11.54	1.67	5.99	
Grade 1	1.681	1950.08	11.54	0.69	14.57	
		Urban				
		Sample				
KG	0.14	1950.08	11.54	8.24	1.21	
Grade 1	0 506	1950.08	11 54	2.28	4 39	
	0.500	1990.00	11.5 1	2.20	1.55	
		Dunal				
		Kural				
		Sample				
KG	1.521	1950.08	11.54	0.76	13.18	
Grade 1	1.746	1950.08	11.54	0.66	15.13	
Total Budget Year 1 +						
Year 2= 15,000,000						
Total number of beneficiaries (KG + Grade 1 + Grade 2)=7692						
Cost per student PKRs=	=					
15,000,000/7692=1950.08						
Exchange rate: 1 GBP=PKRs 169/-						

With gains greater than 0.2 SD, as a rule of thumb the program is considered to be "effective"

Effect Size: S-M (Small-Modest): 0.2; S M (Modest): 0.5; L (Large): 0.8.

# 5 Key Recommendations

This section presents a set of recommendations based on the evaluation findings.

- Engagement with various stakeholders, particularly government agencies, will remain critical for the long-term sustainability and success of the project. It is therefore important to reach out to relevant stakeholders and to develop institutional linkages with government agencies. In particular, linkages with teacher training institutions would be useful in strengthening the capacity building strand of the program.
- 2. The 'Interactive Radio Instruction (IRI)' technology used by the project seems to have paid dividends. It has possibilities for scale-up for other subjects and levels of schools. Key deficiencies exist in availability of mathematics and science teachers, especially in female schools. IRI can be used to cover this deficiency, at least to an extent.
- 4. It is worthwhile for the project to explore potential partnerships for replication and scale-up, and increasing the project outreach. The project can consider engaging and leveraging support of institutions FM stations in district and Radio Pakistan for broadcasting the programs.
- 5. As the project focuses on capacity building of teachers as its key intervention and entry point, it is worthwhile to carry out a comprehensive competency mapping of teachers to be used as a baseline. Towards the end of a project an end-line can also be conducted and used for comparison with the baseline in order to assess the efficacy of the capacity building interventions.
- 6. The findings indicate an inherent issue within the teachers' education and professional development in the local context, whereby teachers are mostly not oriented towards larger curriculum objectives but trained in limited skills in the traditional transmission paradigm. While the project focus is limited and not aimed at addressing this critical gap, it does, however, provide an opportunity to address the issue in the trainings. The trainings can be used as an excellent in-road to introduce the curriculum objectives, albeit only those relevant to English Language courses, and make the connection of the larger objectives with the textbooks and course content.

# Annex 1: In-depth Interviews

	DAI-Broad Class						
	IDIs with Head teachers and teachers						
S.N							
0.	School name/Community	Head teacher	Teachers	EDO			
1	Chak Mohri	Ajmal Khan	Qazi Shafiq ur Rehman				
2	Chechian	Fakar Zaman	Wajid				
3	GPS # 4 Haripur Sheranwala Gate (Urban)	Muhammad Tariq	Sheikh Naveed Farooq				
4	GPS Jatipind	Naveed Akhtar	Fida Mahmood				
	GEMPS Sector 3 KTS, Kalabat						
5	(Urban)	Saif ur Rehman	Majid Hussain				
6	GPS KTS, Kalabat, Township	Tariq Mahmood	Matloob Ahmad				
7	GGEMPS # 1 (Urban School)	Tanvir Kausar	Nabeela Noreen				
8	GGEMPS # 3 (Urban School)	Nahida Jabeen	Kausar Parveen				
9	GGPS # 2 S.Saleh (Urban School)	Shehnaz Begam	Farhat shehzad				
10	GGPS Derwesh (Urban School)	Rashida Yasmin	Bibi Gull				
11	GGPS Kahal Bala (Rural School)	Yasmin Naqvi	Rehana				
12	GGPS Khoi Nara (Rural School)	Tanvir Akhtar	Nadia Noreen				
13	GGPS KTS # 1 Sec 4 (Urban School)	Akhirat Khursheed	Nazia Yaqoob				
14	GGPS KTS (Urban School)	Nusrat Bibi	Farukh Taj				
15	GGPS Meelam (Rural School)	Shahida Habib	Rubina Shaheen				
16	GGPS Qazian (Rural School)	Sajida Bibi	Mrs. Khalid				
17	GGPS Sikander Pur (Rural School)	Gulnaz Bibi	Sadia Nasreen				
18	GGPS Talokar (Rural School)	Tazeem Begam	Rubina Shaheen				
19	Muslim Town, G.T Road, Haripur			M. Saeed ur Rehman			

# **Annex 2: Standardized Effect Sizes**

An effect size of	Is considered	It means that
0.2	Small-Modest	The average member of the
		Intervention group had a better
		outcome than 58 percent of the
		members of the control group.
0.5	Modest-Large	The average member of the
		Intervention group had a better
		outcome than 69 percent of the
		members of the control group.
0.8	Large	The average member of the
		intervention group had a better
		outcome than 79 percent of the
		members of the control group

Source: *Planning Sample Size for Randomized Evaluations,* Abdul Lateef Jameel Poverty Action Lab