RISE

RAPIDLY IMPROVING STANDARDS IN ELEMENTARY

LITERACY AND NUMERACY BOOST ASSESSMENT MIDLINE REPORT PAPUA NEW GUINEA JULY 2019

CATHERINE JOHNSTON AND KABIRA NAMIT

1 ACKNOWLEDGEMENTS

The Rapidly Increasing Standards in Elementary (RISE) project is supported through the Papua New Guinea Partnership Fund. PPF is an initiative of the Australian Government in partnership with the Government of Papua New Guinea. RISE is implemented by Save the Children in partnership with Summer Institute of Linguistics (SIL) and Callan Services for Persons with a Disabilities National Unit.

Save the Children would like to thank the elementary schools, elementary teachers, Teachers in Charge, students and parents in the Autonomous Region of Bougainville (ARoB), Eastern Highlands Province (EHP) and East Sepik Province (ESP) who took part in the RISE PNG midline study, including the enumerators, provincial education officers and project staff who administered and supported the survey. We thank the Provincial District and National Education officers for their valuable feedback and insights during data collection and baseline presentation meetings.

We also thank Save the Children United States Education Team for their technical advice in relation to administration and analysis.

The authors specifically thank the following individuals for their thorough and thoughtful technical reviews during the drafting phase:

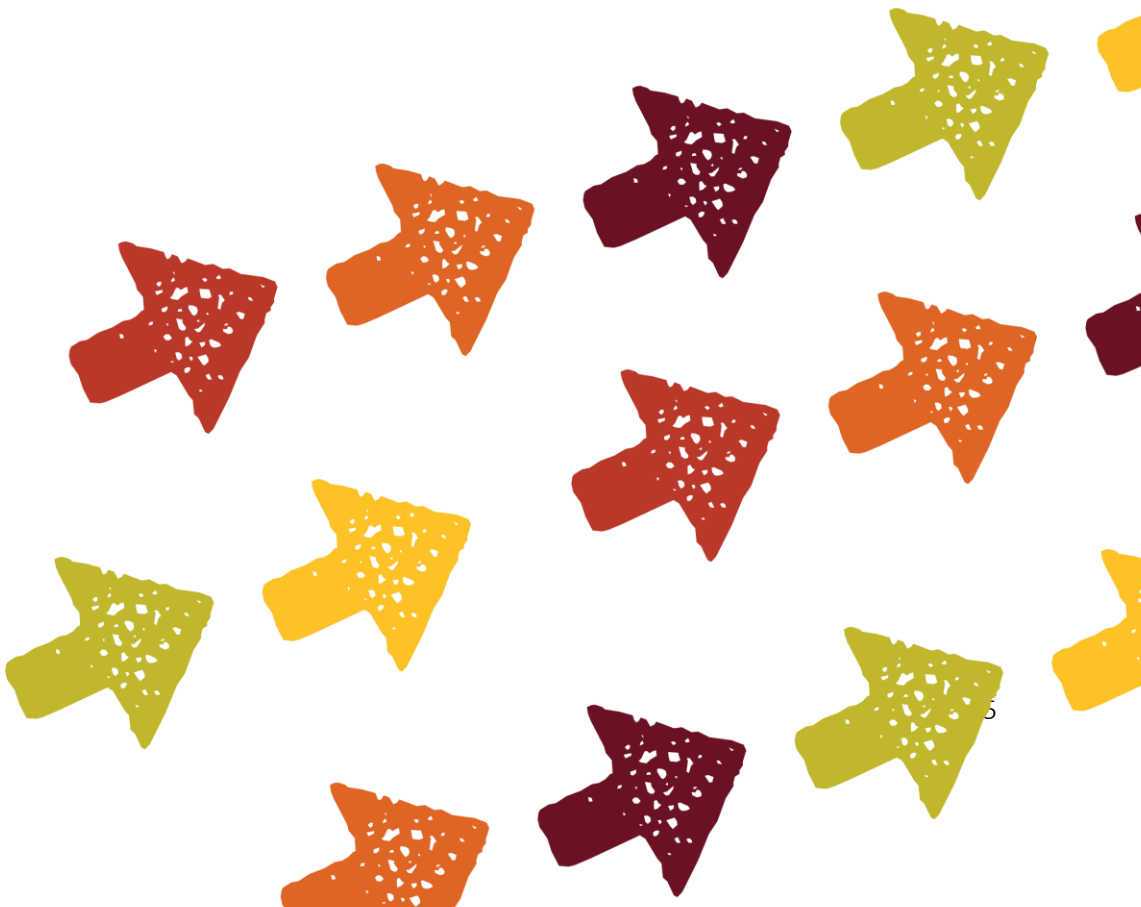
RISE PNG Project Manager, Save the Children: Joy Wong

RISE PNG Program & Development Quality/Monitoring & Evaluation, Save the Children: Richard Orengo

Principal Education Advisor, Save the Children Australia: Gary Ovington

RISE PNG provincial education officers and teacher training staff, Save the Children

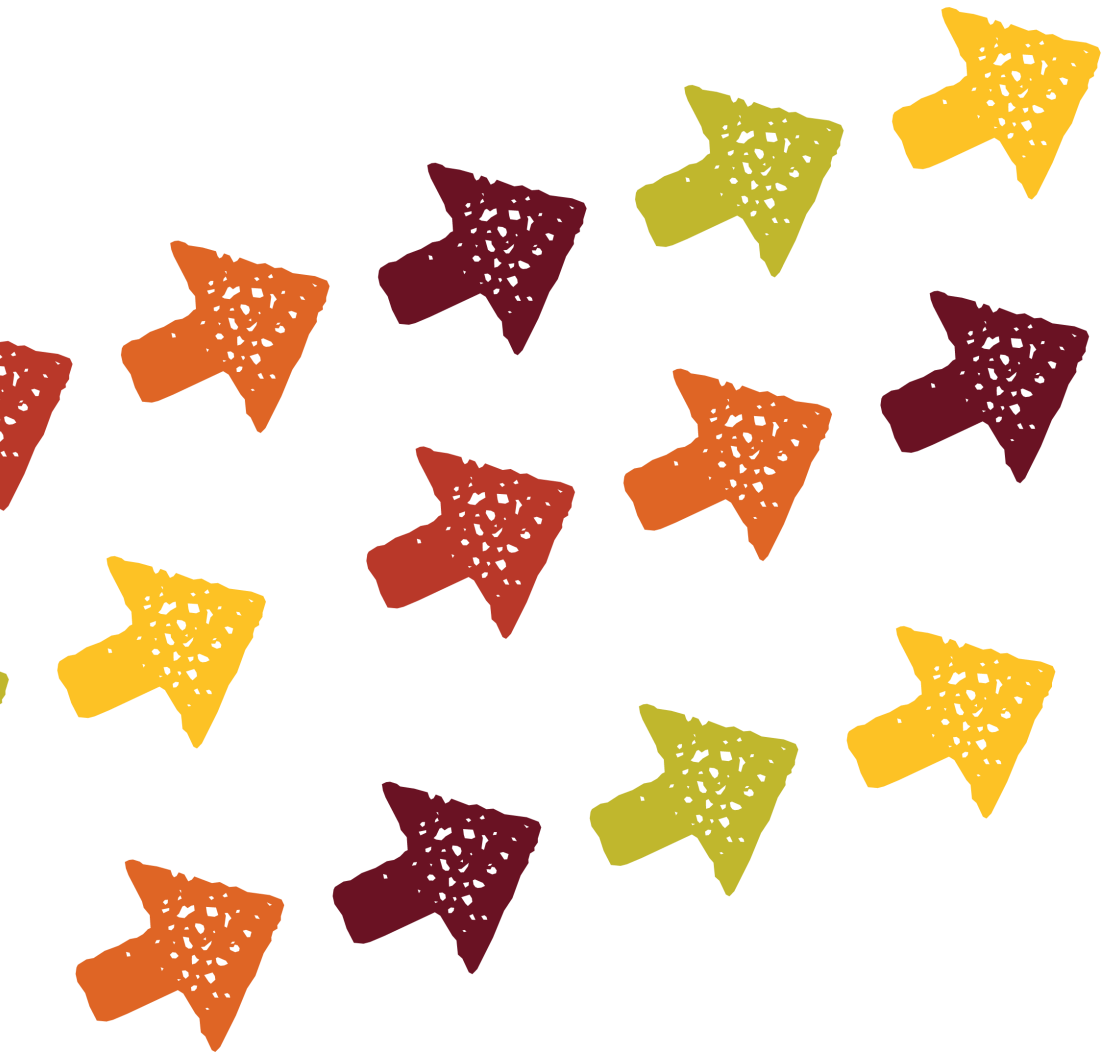
CONTENTS



ACRONYMS

ARoB	Autonomous Region of Bougainville
BOM	Board of Management
CEO	Community education officer
CLV	Community literacy volunteer
DFAT	Australian Department of Foreign Affairs and Trade
E1	Elementary One
E2	Elementary Two
ECCD	Early Childhood Care and Development
ECCE	Early Childhood Care and Education
EFL	Education for Life
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EHP	Eastern Highlands Province
EMIS	Education Monitoring Information System
EOP	End of Program Outcome
EP	Elementary Prep
ESP	East Sepik Province
GoPNG	Government of Papua New Guinea
IDELA	International Development Emergent Learning Assessment
IRR	Inter-rater reliability
KoBo	KoBo toolbox, a free online software package for surveys
LB	Literacy Boost
LLG	Local Level Government
MDE	Minimum Detectable Effect
MEP	Monitoring Evaluation Plan
NDoE	National Department of Education
NEP	National Education Plan
NB	Numeracy Boost
PILNA	Pacific Island's Literacy and Numeracy Assessment

PNG	Papua New Guinea
PPF	PNG Partnership Fund
PLC	Professional Learning Circles
RISE	Rapidly Improving Standards in Elementary
SBC	Standards Based Curriculum
SIL	Summer Institute of Linguistics
TFF	Tuition Free Fee
TIC	Teacher in Charge



TABLES AND FIGURES

GLOSSARY

Causal: expressing or indicating cause : causative; of, relating to, or constituting a cause

Cohort: a group of people with a shared characteristic; a group of people with a common statistical characteristic.

Control: a person or thing used as a standard of comparison for checking the results of a survey or experiment.

Cumulative: increasing or increased in quantity, degree, or force by successive additions.

Intervention: the project activities undertaken at the school, community, district, provincial and national level to achieve the outcome

Inter Rater Reliability: In statistics, inter-rater reliability (also called by various similar names, such as inter-rater agreement, inter-rater concordance, inter-observer reliability, and so on) is the degree of agreement among raters. It is a score of how much homogeneity or consensus exists in the ratings given by various judges.

Multivariate: Multivariate statistics is a subdivision of statistics encompassing the simultaneous observation and analysis of more than one outcome variable. The application of multivariate statistics is multivariate analysis.

Randomised Control Trial: A randomized controlled trial (or randomized control trial; RCT) is a type of scientific (often medical) experiment that aims to reduce certain sources of bias when testing the effectiveness of new treatments; this is accomplished by randomly allocating subjects to two or more groups, treating them differently, and then comparing them with respect to a measured response. One group—the experimental group—has the intervention being assessed, while the other—usually called the control group—has an alternative condition, such as a placebo or no intervention. The groups are followed under conditions of the trial design to see how effective the experimental intervention was. Treatment efficacy is assessed in comparison to the control. There may be more than one treatment group or more than one control group.

Regression: In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors', 'covariates', or 'features'). The most common form of regression analysis is linear regression, in which a researcher finds the line (or a more complex linear function) that most closely fits the data according to a specific mathematical criterion. For example, the method of ordinary least squares computes the unique line (or hyperplane) that minimizes the sum of squared distances between the true data and that line (or hyperplane). For specific mathematical reasons (see linear regression), this allows the researcher to estimate the conditional expectation (or population average value) of the dependent variable when the independent variables take on a given set of values. Less common forms of regression use slightly different procedures to estimate alternative location parameters (e.g., quantile regression or Necessary Condition Analysis) or estimate the conditional expectation across a broader collection of non-linear models (e.g., nonparametric regression).

Sample: A set of data or elements drawn from a larger population and analyzed to estimate the characteristics of that population

Statistical significance: Statistical significance is a measure of whether your research findings are meaningful. More specifically, it's whether your stat closely matches what value you would expect to find in an entire population.

2 EXECUTIVE SUMMARY

2.1 BACKGROUND

The Rapidly Improving Standards in Elementary (RISE) program aims to improve the quality of early grade education for children aged four to eight in Papua New Guinea (PNG). The program aims to reach 100,000 children in three provinces: East Sepik Province (ESP), Eastern Highlands Province (EHP) and the Autonomous Region of Bougainville (ARoB). The project is funded by the Australian Government, in partnership with the Government of PNG (GoPNG) and is running from July 2017 to April 2020. RISE PNG is administered by a consortium led by Save the Children, in partnership with Callan Services for Persons with Disabilities and Summer Institute of Linguistics (SIL) PNG.

A baseline study was conducted from October to November 2017 to assess elementary students' current literacy and numeracy competency in sampled schools in EHP, ESP and ARoB. The midline study for elementary students and elementary teachers focuses on gathering quantitative and qualitative data to measure: (1) improvements in children's literacy and numeracy after one year of RISE PNG implementation and 2) improvements in elementary teachers' competency in teaching English, Tok Pisin and mathematics.
The midline study was undertaken in 131 elementary schools in ESP, EHP and ARoB, using adapted versions of Save the Children's early grade literacy and mathematics assessments and an adapted lesson observation tool and survey instrument, which was used in the baseline study and draws upon globally recognised instruments such as Stallings and TEACH.[1]

2.2 STUDY DESIGN, METHODS AND TOOLS

The objective of the midline study was to measure the learning gains of elementary students after one year of intervention as part of a randomised control trial. More specifically, the research design purpose was threefold:

1. To assess if Literacy Boost (LB) and Numeracy Boost (NB) are viable approaches to improving elementary literacy and numeracy skills in PNG.

2. To assess the impact of the RISE PNG program on elementary students' learning outcomes.

3. To contribute to PNG and global literature on 'what works' in improving early grade literacy and numeracy outcomes.[2]

The sample is representative of elementary schools in Save the Children's three target provinces[3]. There are three groups in the midline study:

1. Students in elementary 2 (E2)
2. E2 teachers

[1] Stallings is a method that records a teacher's actions at regular intervals during the lesson. TEACH is a hybrid-lesson observation instrument designed by the World Bank that integrates high- and low-inference questions for observing teacher practice and skills in the classroom.
[2] This study will add to the rich evidence from over 25 impact studies on literacy and numeracy including Save the Children (2016) Rwanda Impact Evaluation, which showed that children in LB are 40% more likely than the control group to progress to Grade 3.
[3] See page x for details of the study's sampling method

3. Teachers in Charge (TIC)

The Intervention

During the 12-month period, intervention schools received support in a variety of ways. E2 teachers and Teachers in Charge (TICs) attended three rounds of training: literacy, numeracy and cross cutting themes with a focus on gender and inclusive education. Community literacy activities were established with children participating in reading clubs and storytelling sessions with community elders (Story Time) and parents engaging in caregiver workshops to learn strategies to support their children's development in literacy. Schools received teaching and learning resources with teaching strategies to complement the SBC curriculum and community literacy volunteers who oversaw community literacy activities received book banks and curriculum to support their work in community. Teachers were further supported through face to face monitoring and reflection meetings with other teachers.

The E2 student sample (2,035) was taken from the baseline cohort of randomly assigned intervention and control elementary schools. (pg. 25 and pg. 31) The control schools in Year 1 of implementation became intervention schools in Year 2. To avoid contamination, the midline assessment was undertaken prior to training of elementary teachers from the Year 2 intervention group.

Early school closures (by mid-November) and high rates of student absenteeism on the day of the assessment impacted the midline school, teacher and student response sample. From 197 schools assessed at baseline, 131 were reached at midline – 42 in ESP, 46 in EHP and 43 in ARoB. From a total of 2,743 students assessed at baseline,.2,035 students were assessed at midline, including 1,004 from the baseline cohort. From the 175 teachers assessed at baseline, 112 were reassessed at midline. (see detailed response sample pg. 31).

The midline assessment utilised four survey instruments:

1. A combined Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA).

2. A lesson observation instrument.

3. A teacher survey, which was adapted from the baseline instrument with additional items and focused on the LB teaching strategies covered in teacher training during term breaks in 2018.

4. A Teacher in Charge (TIC) survey instrument to help verify and confirm results from the students and teachers.

The student assessment instrument was adapted to ensure E2 standard equivalency. The lesson observation was also altered to capture teaching strategies covered in training.

During the period from October to November 2018, 2,035 eight to ten year olds from 131 elementary schools were assessed for foundational literacy (in English and Tok Pisin) and foundational mathematics in ESP, EHP and ARoB. In addition, 112 E2 classes were observed and 112 teachers and 132 TICs were interviewed.

Analysis

To test the comparability of children in intervention and control groups, from baseline to midline, the research team relied on multivariate regression and assessments of the significance of the coefficient of an intervention variable, with adjustments for the clustered sampling strategy including robust standard errors. The next section provides more details.

2.3 STUDY FINDINGS

2.3.1 Students' literacy and numeracy scores and factors that influence results

Students' literacy and numeracy results

Overall the midline assessment showed that students who received the intervention achieved higher results in literacy and numeracy than students who did not, and while the results were not statistically significant for all students combined, LB and NB are making a statistically significant difference to girls' learning, in particular in numeracy among female students in ARoB and in literacy among female students ESP. (Annex E)

Teachers skills

The midline also shows a statistically significant improvement in literacy and numeracy teaching skills, classroom management and student engagement among E2 teachers in ARoB, ESP and EHP who completed the LB and NB training. (Annex D)

There is a clear pattern across all three provinces: *RISE PNG is currently having a disproportionately positive impact on girls compared to boys.* In no province do we see the intervention havng a significant effect on boys alone. It is also evident that in no province are boys or girls performing worse in intervention schools compared to control schools. In ARoB, there are statistically significant improvements in numeracy overall and in numeracy for girls. Improvements are evident in literacy but these are not statistically significant.

In EHP, there is no statistically significant impact of Save the Children's intervention at this stage, despite students from intervention schools scoring higher in literacy and numeracy compared to control students in this province. In ESP, there are statistically significant improvements in literacy overall and statistically significant improvements in literacy for girls. This points to provincial variations in the effectiveness of the intervention on teacher quality and student learning and perhaps other variables that may be influencing results such as access to Early Childhood Care and Education (ECCE), a mother's literacy and children's reading habits at home.

When comparing provincial differences in the 2015 Pacific Islands Literacy and Numeracy Assessment (PILNA), a similar trend is observed – ESP students scored higher in numeracy (57.69%) than their peers in EHP (50.66%) and ARoB (49.75%) and also outperformed them in literacy (Educational Quality Assessment Program, 2015).

While it is evident that the RISE PNG program is having a positive effect on girls, boys are outscoring girls on many test items. Irrespective of intervention status, statistically significant differences between boys and girls can be found in relation to skip patterns, word problems, shape recognition and reading comprehension (Study Findings). The gains that girls are making, however, may begin to equalise results over time if the trend observed after a year of intervention extends to later years. Based on the Grade 5 Pacific Islands Literacy and Numeracy (PILNA) results, girls are performing higher than boys in literacy and numeracy.

There is an emerging correlation between the impact of Save the Children's teacher training and measures of teaching competency, which has been captured during the lesson observations and student literacy and numeracy results. There is strong evidence of this in ESP, where teaching competency in reading comprehension has improved considerably. (Annex D). This is probably contributing to statistically significant improvements in children's literacy scores overall and with girls in particular. At this point in the implementation cycle, the intervention appears to have shown the strongest effect in ESP, followed by ARoB and EHP.

A number of significant findings are evident across the range of literacy and numeracy test items. For literacy, the largest effect size between intervention and control students was in relation to the proportion of readers and non-readers. A higher proportion of children in intervention schools, 49.5%, could read at least five words

of the E2 English passage in 30 seconds compared to 44.7% in control schools, demonstrating an effect size of 4.75 percentage points at a 90% confidence level (see Study findings). The increased frequency and attention placed on reading activities in class and at home through reading club activities and book borrowing appears to be producing the desired effect of increasing students' reading skills.

For numeracy, the most significant gains were shown in skip patterns of five, word problems and shape identification. Children experienced greater difficulty with subtraction than addition, particularly when dealing with double-digit numbers. The word problems that children found challenging were also ones that required students to subtract items. This suggests that the strategies used by children to solve double-digit sums are not effective. During administration, it was evident that children had not grasped techniques such as grouping, which is useful when children transition from single- to double-digit sums.

Factors that influence children's literacy and numeracy results

Irrespective of whether a child received the intervention (or not) many factors influence children's learning gains in literacy and numeracy over time. The variables showing the greatest influence on elementary students' English and mathematics scores are age, gender, grade level, grade repetition, ECCE participation, mothers' literacy levels, home reading, reading books from the Bloom Reader app, socio-economic status, borrowing books from the reading club, chewing betel nut and a child's motivation to go to school/enjoyment of school. (reference to where these results are found in the document)

Children who had attended ECCE performed, on average, much higher in literacy and numeracy subtests. This is consistent with global evidence on the significant benefits ECCE has on children's readiness for learning at elementary level (ref) The socio-economic status of families, as well as a mother's literacy, are important predictors of children learning results and the effect size is considerable for children's literacy. (ref where found in the document)

The home learning environment is a powerful influence on learning outcomes for elementary children, especially the number of books that a child has exposure to at home and the frequency with which they read outside the classroom. Essentially, more books and more reading at home improves literacy and numeracy performance in the classroom (ref where found in the report)

Interestingly, the effect size and level of statistical significance is larger for home reading than for caregivers who read to their child. (ref where found in the report) A plausible explanation might be that caregivers read with their child less as the child becomes older. E2 students in PNG are typically nine to ten years of age and this might explain the tendency for children to read more on their own at home than with a parent. It is also likely that caregivers with low literacy may not feel confident reading E2 level textbooks and storybooks so tend to leave children to read on their own at this age. Associated variables such as borrowing books from the community book bank and reading books from Bloom Reader also had a positive impact on literacy

While the penetration of reading clubs and community literacy activities has only just started, and children's access to mobile technology so they can read from Bloom Reader remains limited, there is promising evidence of its contribution to children's literacy and numeracy results in the classroom. Children who borrowed books and read from bloom reader had higher scores than children who did not. The score variances were considerable with these children scoring between 7.6 and 11.25 percentage points higher than other children. (ref where found in the report), it is evident that children in PNG want reading books and have a desire to read and that children who have access to, and read books have better literacy

The midline also found that children who chewed betel nut had statistically significantly lower numeracy and literacy scores than children who did not. Children who chew betel nut scored almost 2.5 percentage points lower on overall literacy and numeracy scores and 3.5 percentage points lower on literacy alone (Annex F). This

is similar to the findings in the RISE PNG International Development Emergent Learning Assessment (IDELA) Baseline Study (2019), which assessed children's emergent literacy, numeracy, motor development and social-emotional development. The IDELA baseline study found that children who consumed betel nut scored 5.5 percentage points lower than children who did not chew betel nut. In both assessments, girls were more likely to chew than boys. The gender dimension and practice of betel nut consumption in young children warrants further investigation.

Attendance patterns are low and unchanged since the baseline study and are associated with lower literacy scores. However, absenteeism does reduce with each level – EP reported the highest absenteeism followed by E1 and then E2. There is a slightly higher absenteeism rate among boys, and male teachers are more likely to be absent than female teachers. The gender dimension to absenteeism of both teachers and students requires further investigation.

Based on the responses from students, illness was the primary reason for absence. The second reason was strongly associated with an apparent disinterest in schooling. Based on the responses made by students, parents appear ambivalent towards their children's education and readily accept requests to stay at home. Absenteeism is having a negative effect on literacy and numeracy acquisition, particularly for test items that require higher order thinking. Without focus on this issue, many children will fail to achieve expected standards in English and mathematics, irrespective of efforts to improve teaching standards and provide quality teaching and learning materials.

Approximately one quarter of children in E2 had repeated, irrespective of gender. The study found that children who repeat a grade level had lower literacy and numeracy scores. The literacy skills most affected by repetition are English and Tok Pisin word scores and comprehension. This is an important finding given the high prevalence of repetition in PNG and the belief that this is an effective strategy for children's learning. The results of this study present a strong case to revisit this assumption.

Teacher competency gains in literacy and numeracy

Overall, comparing teachers who received the LB and intervention and those who did not, there were statistically significant proficiency differences in four of the five teacher competency domains.

Large differences were evident in reading comprehension, (31.7%), classroom management (19.92%) and student engagement (17.24%) – both integrated topics in the teacher training cycles. (Annex D). The results for classroom management and student engagement were highly statistically significant $p<0.001$ and for reading comprehension, $p<0.01$. Based on the observation results however, teachers tended to favour read-a-louds rather than other strategies that encourage students to read by themselves. This means that children may not be practicing reading during class. Observations indicated less evidence of a range of teaching strategies such as group and pair work.

2.3.2 Findings relevant to the Monitoring Evaluation Plan

The findings are presented in the table below.

Table 1: Baseline and midline results for Intermediate Outcome IO1-104

End of Program Outcome (EOPO)	Intermediate Outcome (IO)	Outcome Indicator	Target	Baseline Results	Midline Results
EOPO 1:	IO1:	Percentage of teachers	400 of the 1,800 targeted teachers	No difference between	**Teaching reading comprehension:**

Improved literacy and numeracy learning outcomes for elementary students	*Elementary teachers demonstrate confidence and competence in teaching and assessing Standards Based Curriculum (SBC) English, Language and Maths*	demonstrating improved literacy and numeracy instructional practice after receiving Literacy Boost (LB) and Numeracy Boost (NB) training	(25%) demonstrate improved literacy and numeracy instructional practices after first year of training and additional improvement in subsequent years	intervention and control school teachers at baseline, measured using a modified Stallings tool	Teachers who received intervention scored 66.94% for teaching reading comprehension compared to 35.29% by control teachers (31 percentage points difference)[4] p<0.01 **Teacher capacity to assess students:** Similarly, intervention teachers who received intervention scored 25.04% for assessments as compared to 15.02% for control teachers (10 percentage points difference) P<0.05
	IO2: *Elementary schools resourced with literacy and numeracy instructional materials*	National Education Plan (NEP) indicator 20: Percentage of schools adequately resourced with relevant learning materials NEP 22: Percentage of schools at each level using relevant learning materials	90% of 650 elementary schools have access to teaching guides[5] 20% of schools involved in Education for Life (EFL) trial have used the Teacher Presenter Kit in the last month	81.71 % of schools had English Standards Based Curriculum (SBC) teaching guides 14.86% of schools had English SBC E2 readers 9.14% of schools had reading books in English 65.71 % of schools were using SBC	69.61% of intervention teachers who received intervention were up to date with syllabus compared to 63.04% of control teachers[6] 41.51 % of (T) teachers missed sections of SBC compared to 56.52% of (C) teachers EFL trial results pending
	IO3: *Parents and caregivers of early grade children demonstrate*	Percentage of caregivers and children with a reported improvement in	25% of targeted caregivers and children report an improvement in	No difference in home reading habits of children in intervention	22% of children in intervention schools engaged in reading clubs

[4] All results presented in this report, particularly where the intervention and control group are being compared, will include details of the scores achieved (usually presented as percentage figures), and an assessment of the statistical significance of the result. Statistical testing is used to assess whether there is causal relationship between an intervention and a result, that is, did a particular activity (eg training) have an impact (it could be positive or negative) on the group undertaking the activity (teachers who received the training (intervention group) compared to a group who did not (teacher who did not receive the training (control group). The significance level for this study is 5%, that is, the testing tells us that for a statistically significant result, we can be 95% confident that the result we are measuring has been achieved by the activity.

[5] This indicator will be changed to reflect the PPF recommendation to report on NEP indicator linked to use of SBC, which is more appropriate for RISE as in-service teacher training is focused on this.

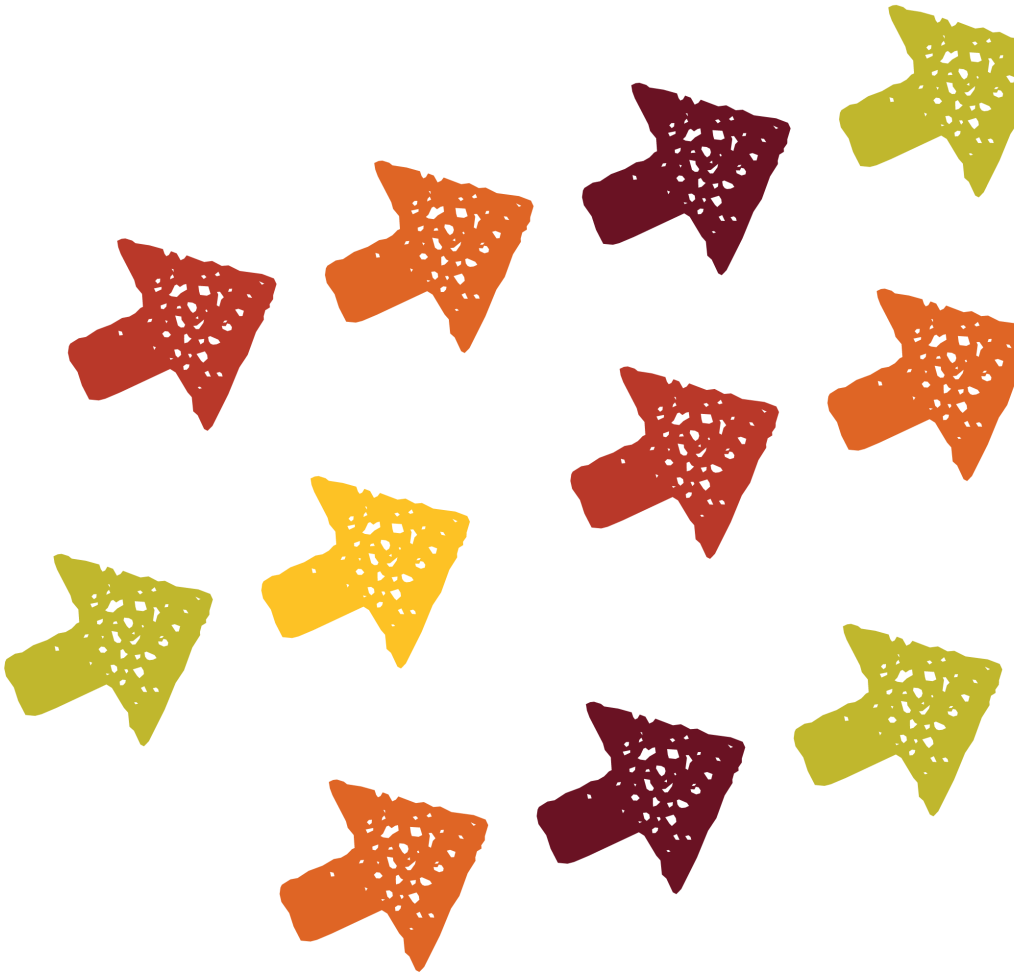
[6] This indicator is measured based on daily scripted lessons that are used and expected in elementary grades. A teacher was considered up to date if they were teaching the correct day.

	improved home reading practices and community reading culture	home reading habits and community reading culture	home reading habits	schools and control schools	compared to 12% control children 18.86% of intervention children in intervention schools borrow books compared to 11.07% control children
	I04: *Improved literacy and numeracy outcomes in elementary schools*	Percentage of girls and boys in elementary schools who demonstrate improved literacy and numeracy skills	20% of children show an improvement in literacy and numeracy skills (in comparison to control groups) for elementary students	No difference in literacy and numeracy skills of children in intervention schools and control schools at baseline	Girls in intervention schools scored 5.07% higher in numeracy than girls in control schools, and 4.53% higher in overall literacy & numeracy compared to control schools. (*p<0.01)* Boys in intervention schools show no statistically significant improvement compared to boys in control schools.

2.4 STUDY RECOMMENDATIONS

1. Share results showing the importance of home reading and book borrowing with teachers, Boards of Management (BOMs), provincial and district education officers and community members so children are encouraged to read frequently at home and in the classroom. Use every opportunity during training and coordination meetings to emphasise and prioritise reading practices among children. Organise community meetings to share the results pictorially that summarises the main findings.
2. Undertake further consultation with community and education stakeholders to explore the practice of betel nut consumption in children to identify possible explanations for the prevalence and gender dimension. Utilise this same mechanism to work with parents on the attitudinal shifts required to encourage more regular school attendance.
3. Continue to emphasise the community literacy component of the program, which is showing promising results and supporting children to read more.
4. Continue to promote Bloom Reader, given the positive results from the sample of students regularly reading from the app.
5. In the Term 3 refresher training, revisit all reading strategies and continue to practice collaborative learning strategies.
6. In the Term 3 refresher training, revisit addition and subtraction strategies – particularly for double-digit numbers – to ensure teachers have a good grasp of teaching strategies that support children's skills in these areas.

7. Undertake provincial level workshops to gain a deeper understanding of the gender and provincial dimensions arising from the midline findings.
8. Similar to the recommendation at baseline, continue to actively contribute to shared learning opportunities across the PNG Partnership Fund (PPF) program in order to stimulate joint policy dialogue with government and development partners. Teacher and student absenteeism and early school closures are seriously compromising interventions.



3 INTRODUCTION

3.1 BACKGROUND

The Government of PNG's (GoPNG) commitment to quality learning for all is illustrated by the current National Education Plan (NEP) 2015-2019. The plan acknowledges that while school enrolments have improved, children are still dropping out early, not attending regularly and not meeting the curriculum standards of their grade level.

Between 2011 and 2013, the National Department of Education (NDoE) administered the Early Grade Reading Assessment (EGRA) on Grade 3 and Grade 4 students and found children in elementary and lower primary grades had not acquired basic reading skills (Williams, 2013). In 2015, this was complemented by the Pacific Islands Literacy and Numeracy Assessment, which found that Grade 5 students in PNG were less likely to achieve the minimum standards in literacy and numeracy than their regional peers in the Pacific (Educational Quality Assessment Program, 2015). The proportion of Grade 5 students in PNG who met or exceeded the expected level was 51%, which was below the regional result of 68%.

A number of reports into the quality of education have highlighted constraints to improving learning outcomes:

- Limited English skills of elementary teachers (Drinan, 2014) coupled with poor quality initial teacher training (Manikuali, 2014).
- Weaknesses in the curriculum with limited attention to literacy and numeracy (Czuba, 2013).
- Capacity gaps in school management, in-service teacher education and system administration (Norman, 2012).

Reforms, including the development of the Standards Based Curriculum (SBC) and the focus on English language and numeracy, have attempted to address the constraints (Czuba, 2013). But, despite the curriculum reforms, utilisation of the curriculum is only just beginning and the competency of elementary teachers to teach literacy and numeracy (Manikuali, 2014) remains a challenge. As was found at baseline, elementary schools struggle to meet the needs of learners due to limited access to relevant storybooks and levelled readers, high rates of teacher absenteeism and limited monitoring (Johnston, Namit, & Reen, 2018). Although enrolment rates are high in Elementary Prep (EP), 18% of boys and 23% of girls drop out before reaching Grade 3 (NDoE PNG, 2015).

3.2 PROJECT DESCRIPTION

The RISE PNG program aims to improve the quality of early grade education for 100,000 children aged four to eight in three provinces of PNG: East Sepik Province (ESP), Eastern Highlands Province (EHP) and the Autonomous Region of Bougainville (ARoB). The project is funded by the Australian Government in partnership with the GoPNG, and will run from July 2017 to April 2020. RISE PNG is a consortium led by Save the Children, in partnership with Callan Services for Persons with Disabilities and Summer Institute of Linguistics (SIL) PNG.

RISE PNG's Literacy Boost (LB) and Numeracy Boost (NB) intervention seeks to address gaps in teacher competency and inclusive practice, supporting teachers to deliver the new SBC and build a literacy and numeracy rich environment. The teacher training intervention seeks to improve teaching and assessment skills of elementary teachers using Save the Children's internationally proven LB and NB materials (ref) adapted for the new SBC in PNG.

Each RISE PNG provincial team has six or seven teacher trainers who provide training for the Elementary 1 teachers during the 2018 and 2019 school years. Teachers are also supported through Professional Learning Circles (PLCs) where teachers meet on a monthly basis to exchange learnings and undertake planning together. The PLCs are overseen by community education officers (CEOs) and teacher trainers within the RISE PNG program.

The LB and NB intervention also includes substantial community activity with caregivers. Through community literacy action cycles, community members and caregivers are exposed to strategies that support children's learning at home and in the community. Caregivers have the opportunity to attend workshops that equip them with the skills needed to practice early grade literacy and numeracy strategies at home, particularly home reading. Communities establish reading and mathematics clubs and hold events to promote reading and the importance of education. Literacy and numeracy activities at the community level are supported by CEOs who guide volunteers to oversee activities in their communities.

RISE PNG has developed a robust research and learning agenda to ensure the effectiveness of the interventions can be measured against changes in children's learning and development outcomes and their participation rates over time. This midline report complements the initial baseline study undertaken in relation to elementary literacy and numeracy outcomes. This report presents midline information to support the following outcomes of the MEP.

Table 2: MEP End of Project Outcome and Intermediate Outcome Indicators

End of Program Outcome (EOPO)	Intermediate Outcome (IO)	Outcome Indicator	Target
EOPO 1: **Improved literacy and numeracy learning outcomes for elementary students**	IO1: *Elementary teachers demonstrate confidence and competence teaching and assessing SBC English, Language and Maths*	Percentage of teachers demonstrating improved literacy and numeracy instructional practice after receiving LB and NB training	25% of the 1,800 targeted teachers (450) demonstrate improved literacy and numeracy instructional practices after first year of their training and additional improvement in subsequent years
	IO2: *Elementary schools resourced with literacy and numeracy instructional materials*	NEP Indicator 20: Percentage of schools adequately resourced with relevant learning materials NEP Indicator 22: Percentage of schools at each level using relevant learning materials	90% of 650 elementary schools have access to teaching guides. 20% of schools involved in Education for Life (EFL) trial have used the Teacher Presenter Kit in the last month
	IO3: *Parents and caregivers of early grade children demonstrate improved home reading practices and community reading culture*	Percentage of caregivers and children with a reported improvement in home reading habits and community reading culture	25% of targeted caregivers and children report an improvement in home reading habits
	IO4: *Improved literacy and numeracy outcomes in elementary schools*	Percentage of girls and boys in elementary schools who demonstrate improved literacy and numeracy skills	20% of children show an improvement in literacy and numeracy skills (in comparison to control groups) for elementary students

4 MIDLINE METHODOLOGY

4.1 SUMMARY

The impact study for Outcome 1 of the RISE PNG project focuses on four foundational literacy and numeracy outcomes:

1. Teachers' proficiency in teaching the SBC English, Language and mathematics subjects.

2. The availability of quality teaching and learning resources for teachers and students.

3. Caregiver and parent support for elementary children's learning at home.

4. Children's results in foundational literacy and numeracy.

The objective of this research is to assess if LB and NB are viable approaches to improve literacy and numeracy skills of elementary students and teaching competencies of elementary teachers in PNG as part of a longitudinal randomised control trial. This is an impact study representative of elementary students in Save the Children's three target provinces in the RISE PNG project.

4.2 RESEARCH PURPOSE

The purpose of the study is to:

- assess if LB and NB are viable approaches to improve literacy and numeracy skills in PNG.
- assess the impact of the RISE PNG program on the learning outcomes of elementary students and the instructional practices of elementary teachers.
- understand the barriers to improving teaching practices and the learning outcomes of elementary students in PNG.
- understand the community level factors which improve learning outcomes of elementary students in PNG.

4.2.1 Midline research questions

To prepare midline data for the MEP and identify variables that are positively and negatively affecting elementary students' literacy and numeracy outcomes, the following research questions will be answered in this study:

1. *What are children's learning gains in literacy skills in English and Tok Pisin in intervention and control schools at baseline and midline?*

2. *What are children's learning gains in numeracy skills in intervention and control schools at baseline and midline?*

3. *How does gender influence the reading and numeracy skills of children in PNG from baseline to endline?*

4. *What variables in the home or community (e.g. reading books at home, parents reading to their child, attendance at reading club, betel nut chewing) are influencing children's acquisition of literacy and numeracy skills?*

5. *What variables in the school (e.g. corporal punishment, teacher absenteeism, teaching competencies in literacy and mathematics) are influencing children's acquisition of literacy and numeracy skills?*

6. *What proficiency gains in teachers' language and numeracy teaching are evident at baseline, midline and endline?*

4.3 ASSESSMENT INSTRUMENTS

Table 3 is a summary of the instruments used in the impact study.

Table 3: Instruments used in literacy and numeracy midline

Children's literacy and numeracy skills	Elementary teacher competency
Save the Children's EGRA and EGMA instrument	1. Lesson observation with low and high inference questions 2. Teacher interview 3. TIC interview

The following elementary student data was collected:

- **Student background data:** Student's sex, age, language spoken in the home, health status, socio-economic status[7], prior experience of early childhood education, and the literacy and numeracy environment at home.
- The LB and NB Assessment asks children about the **types of print** they see in their homes as well as the **reading, homework support and storytelling habits** of the people they live with.
- **Caring for siblings, chores and work** were added to the list of activities undertaken at home.

The following elementary school background data was collected:

- **Elementary school type:** service provider, education of teacher, training of teacher, number of students, student-teacher ratio.
- **Teacher background data:** sex, age, language spoken, reported or known disability, whether or not they are from the locality, whether or not they speak the same language as their students etc.
- **Qualifications:** level, subject specialisation (if any), the system they studied in, length of study, whether teaching literacy and numeracy was covered, whether they received any specific training on early years' education etc.
- **Teacher Professional Development (in-service) experiences**
- **Teacher and student attendance:** number of days they were absent in the past month and reasons for not attending.
- **Motivations for becoming a teacher**: whether or not they consider teaching is their vocation and how the role is perceived in their community.

[7] The following questions were used to assess socio-economic status: Do the students have footwear? Do they have a school bag? Does their house have a metal roof? How many siblings do they have? Do the parents have a mobile phone?

The midline study used Save the Children's Literacy and Numeracy Assessment instrument for the student learning assessment. A lesson observation tool and teacher survey were added to collect information about teaching competency. For the midline, a TIC interview was included to gather additional information about the teachers in their school, their professional development and their perspectives on the quality of teaching and learning in the elementary school. A description of each tool follows.

Literacy and Numeracy Assessment Tool

The Literacy and Numeracy Assessment Tool administered by Save the Children globally is an adaptation of the original Early Grade Reading Assessment (EGRA) and Early Grade Numeracy Assessment (EGMA) tools developed by RTI International, with modifications to how the reading comprehension subtest is undertaken. The literacy and numeracy skills assessed during the assessment are presented in Table 4 below.

Table 4: Foundational literacy and numeracy skills assessed at midline

Foundational literacy skills
- Phonological awareness – alphabet knowledge
- Vocabulary
- Fluency
- Comprehension (literal, inferential and evaluative)

Foundational numeracy skills
- Numbers and operations
- Geometry
- Functional areas, including:
  - Spatial reasoning
  - Foundational algebra

The specific components of the literacy assessment are outlined in Table 5 below.

Table 5: Literacy subtest item categories

Component	Early reading skill	Skill demonstrated by students' ability to:
1.Letter identification	Alphabet knowledge	provide the name and/or sound of letters presented in both upper case and lower case in a random order.
2. Familiar word reading	Word recognition & vocabulary	read words which are randomly ordered and drawn from a list of frequent words.
3. Oral reading fluency with comprehension	Oral reading fluency Reading comprehension	read a text with accuracy, with little effort, and at a sufficient rate and, respond correctly to different types of questions, including literal and inferential questions about the text they have read.

The specific components of the numeracy assessment are outlined in Table 6 below.

Table 6: Numeracy subtest item categories

Component	Early numeracy skill	Skill demonstrated by students' ability to:
1.Number identification	Number knowledge	identify 20 single- and double-digit numbers.
2. Number pattern identification	Spatial reasoning	identify 'skip patterns' (the correct sequence of multiples of two and five) five times.

3. Addition and Subtraction	Operations	answer 10 single- and double-digit addition questions and 10 single- and double-digit subtraction questions.
4. Word Problems	Arithmetic, foundational algebra	solve subtest of increasingly difficult four-word problems.
5. Shapes	Geometry	identify five shapes and their characteristics.

Lesson observation & teacher survey

The observation form contained questions about the composition of the classroom and characteristics of the teacher (e.g. language, education background, class size) to capture the quality of the learning environment. There were 20 questions focused on general resources, literacy and numeracy instruction, and interactions in the classroom. These were low inference questions and were mostly answered as yes or no and numerically. This was followed by a series of observations about the literacy teaching strategies used during the lesson. These questions were mostly direct observations with drop down menus for enumerators to select the corresponding answer. The questions also aligned closely with the strategies taught during teacher training. This provided a means to identify which strategies teachers were able to apply in their classroom after training. The third section of the observation included a series of questions about classroom management and child centered pedagogy and were organised as questions with responses selected using a five-point Likert scale. These were high inference questions and require the enumerator to make a judgement and score this out of five based on their observations.

Following the lesson observation, a teacher survey was undertaken. The survey covered basic information about the teacher, their education, their training in the past year, and whether they had another job in the community. The next series of questions asked for the teachers' reflections on the lesson just observed – what they thought they did well and not so well, and what strategies they find easy to apply and difficult to apply. Questions about aspects of lesson planning were asked including their application of the syllabus. Finally, a section of child centered pedagogy, classroom management and professional support were included to triangulate the observations made by enumerators. The enumerators for the teacher survey were teacher trainers within Save. Care was taken to ensure that no teacher trainer assessed his or her own trainee during the midline.

Teacher in charge (TIC) survey

The teacher in charge survey was not one of the original baseline assessment instruments. The findings at baseline suggested that additional school level information was needed to better understand the barriers and challenges faced by teachers and to triangulate responses from the teachers and students. The survey covered information about the number of students in the school, the number of teachers, the specific issues facing children's learning in the school, the challenges teachers face, the language of instruction, and the availability of SBC for all teachers, including teaching guides and readers for classrooms in elementary.

4.4 ADAPTATION OF INSTRUMENTS

The literacy and numeracy tools were adapted slightly at midline. This was based on recommendations, made after the baseline survey, to further simplify the language so enumerators could more easily understand the instructions.

Literacy and Numeracy Assessment Tool

The instrument needed to be versioned to account for the fact we were now assessing children one year older and children who were now using the SBC E2 syllabus. Many items, including letter and number identification and skip patterns were repeated at midline with no change. The frequently used words in English and Tok Pisin,

the English and Tok Pisin passage used for reading comprehension, and the addition and subtraction questions were adapted to align with the SBC standards for E2.

Lesson observation and teacher survey

The lesson observation instrument changed considerably from the baseline study. Basic information about the characteristics of the classroom and teaching and learning resources were retained, but with some variation in how the questions were asked in order to minimise inference error. This was aimed to improve the accuracy of these items on the assessment.

The 'time on task' section of the lesson observation assessment from the original baseline was removed as the results at baseline were difficult to analyse due to the lack of specificity of items the enumerators needed to select. The design did not enable detailed observations of teaching strategies to be captured and enumerators struggled with this component of the original assessment. In its place, questions that aligned directly with strategies covered in teacher training were included. The questions were designed as direct closed questions with a follow-up question that asked enumerators to select strategies that best matched their observation.

The third section of the observation included high inference questions and utilised a Likert scale. These questions were modified slightly from the original baseline instrument to include specific and targeted questions about classroom management and child centered teaching practices.

TIC survey

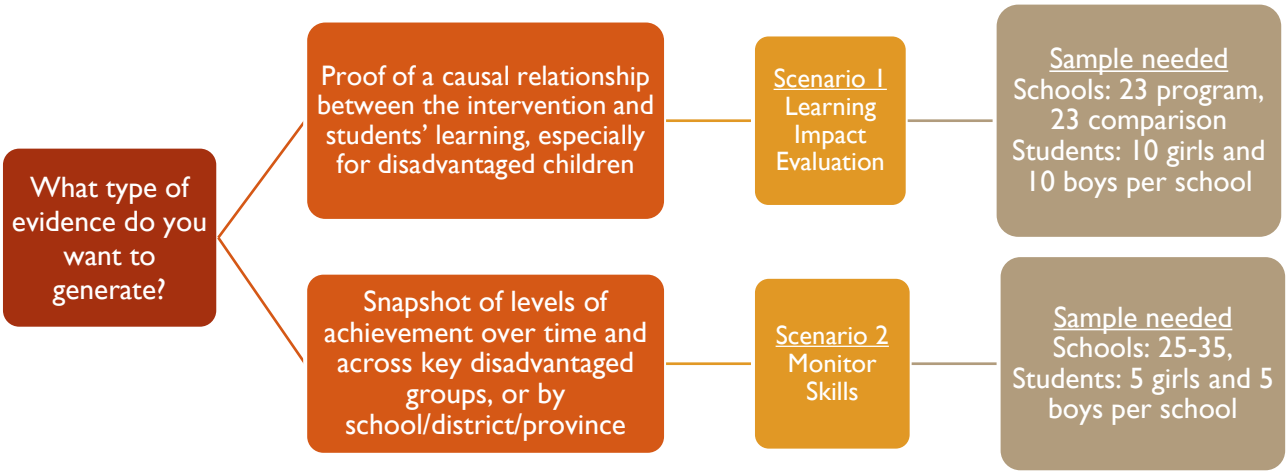
A number of questions from the original teacher survey were moved to the TIC survey as it was noted at baseline that classroom teachers did not know the full details of school enrolment and how the school was engaged in RISE PNG activities. The TIC was also likely to know more about the activities at the community level, about which the research team wanted to receive information.

4.5 SAMPLING

Our research aimed to measure the impact of LB and NB training independently in all three provinces. According to studies that Save the Children has conducted in other countries, the minimum number of schools required to assess if LB and NB training have had an impact is 46. The original sampling model utilised by RISE PNG used the recommended 23 schools for intervention and 23 schools for control as the base model in defining the initial sample. This is described in Figure 1 below.

Figure 1 shows the recommended sample size based on Save the Children's learning assessment guidance at learningassessment@savechildren.org.

Figure 1: Recommended sample size based on the type of evidence needed



To account for high attrition levels and challenges with absenteeism in PNG and to ensure we could capture the MDE (Minimum Detectable Effect) of the LB and NB teacher training, the research team planned to collect data from at least 72 schools per province (with up to 20 student assessments per school) (Table 9). The increase in sample number was due to concerns about loss to follow up based on the experience of previous EGRA[8] and SMS Story assessments[9] in PNG. Loss to follow up is refers to the students who were part of the intervention but have become lost (in this case unreachable due to absenteeism) at the point of follow up, which was at midline. It can create a bias which PNG RISE sought to avoid.

Schools: 216 (72 schools per province, 108 intervention and 108 control)

Teachers: 216 (1 teacher per school)

Students: 4,320 (2,160 boys and 2,160 girls/10 boys and 10 girls per school)

Table 7: Intervention and control schools and students by province

Schools/E1 Teachers	ESP	EHP	ARoB	Total
Intervention	36	36	36	**108**
Control	36	36	36	**108**
Total	**72**	**72**	**72**	**216**
Students	**ESP**	**EHP**	**ARoB**	**Total**
Intervention	720	720	720	**2160**
Control	720	720	720	**2160**
Total	**1440**	**1440**	**1440**	**4320**

[8] The READ PNG project was a 4-year project, implemented by the NDoE, financed by the Global Partnership for Education and supervised by the World Bank. EGRA was undertaken as part of this project in 2015.

[9] SMS Story research project was to determine if daily mobile phone text message stories and lesson plans would improve children's reading in Papua New Guinea (PNG) elementary schools (Keleebu, Gee, Jones and Watson, 2013).

Statistical software (Stata) was used to randomly assign the 216 elementary schools to an intervention or control group. Randomisation was stratified at the LLG level (each province received an equal number of intervention intervention and control schools within the same locality) and elementary schools were equally divided between intervention and control across all three provinces (36 intervention and 36 control schools in ARoB, ESP and EHP).

The sample of 216 elementary schools, which are representative of elementary schools in EHP, ESP and ARoB, were randomly assigned to the Year 1 or Year 2 intervention group. Based on this, 4,320 elementary students from E1 classes were included in the sample at baseline. The baseline response sample included 2,743 students (738 in ARoB, 930 in EHP and 1,075 in ESP) and 175 teachers from 253 schools.

The power calculation details for the original sample at baseline were as follows:

Table 8: Power calculations for the sample

Power Calculations	
Intracluster correlation	0.05
Elementary schools	216
Children per school	20
Statistical significance	95%
Minimum detectable effect size (MDES)	0.2

RISE PNG approached the NDoE to provide a detailed update about the purpose of the impact evaluation. NDoE offered their support and provided RISE PNG with the full Education Monitoring Information System (EMIS) dataset for the three provinces used in this assessment. Using this EMIS dataset, schools were randomly selected with geographical stratification. The research team stratified the data geographically to ensure that the sample was unbiased and representative of each province. Stata randomly selected certain districts, certain LLGs within those districts and certain schools within those LLGs. In order to avoid contamination and noting the logistical difficultly of organising trainings for only a few selected schools in a locality, randomisation was carried out at the LLG level (each school within an LLG was assigned to either an intervention or a control status).

Given that our sample covered at least 20% of the schools in all three provinces, the research team is confident that it accounts for the distribution of rural and urban schools, school ownership levels and other variables likely to exist at the provincial level. As a concrete example, Catholic schools make up 29% of the schools in the three provinces and 34% of our sample. The same is true of rural and urban division.

The key assessments included in the original research design for baseline-midline and endline as summarized in Table 9, below.

Table 9: RISE PNG literacy and numeracy assessment timeline

Baseline	Twenty Elementary 1 (E1) students (10 boys and 10 girls) from 72 schools (36 intervention and 36 control) in each of the three targeted provinces[10] were tested over a period of six weeks in late October and November 2017. The baseline assessment of teachers ran concurrently with the LB and NB student assessment. Elementary 2 (E2) teachers from the same 72 schools were tested over a period of six weeks in late October and November 2017.

[10] East Sepik Province (ESP), Eastern Highlands Province (EHP) and the Autonomous Region of Bougainville (ARoB). Refer to Table 8, p 23 for details of the total cohort involved in this study

Intervention	Prior to the beginning of the 2018 school year, teachers in the intervention cohort (n=108)) were provided with the first round of training as part of the LB intervention.[11] Training continued during school breaks with supplementary learning circle activities at cluster level during term. NB and gender and disability inclusion topics were covered during Term 2 and Term 3 school breaks. The teachers then integrated LB, NB and gender and disability inclusion with the SBC curriculum. Community literacy activities were designed to promote reading at home and in community. Community literacy volunteers were trained to establish reading clubs and facilitate caregiver workshops in communities.
Midline	At the end of Term 4 (November/December 2018), the literacy and numeracy skills of the total student cohort (both intervention & control) (finishing E2 by this stage) and the teaching skills of the total teacher cohort (both intervention & control) were assessed.

The midline study sought to measure the same children again to determine whether any gains were visible in literacy and numeracy as a result of the intervention. If the students weren't present, additional students were randomly assessed (to give further analytical flexibility of performing a case – control study and a cohort study).

4.6 DATA QUALITY ASSURANCE

Testing & enumerator training:

Enumerators were selected/recruited from the provinces so that they would be able to speak the local dialects and children would be able to understand their accent. A majority of enumerators for the midline were young graduates (for example, from the University of Goroka for EHP) and had experience with digitized surveys. Prior to hiring, enumerators were asked to take a reading fluency test.

In order to ensure consistent administration of the literacy and numeracy assessment instruments, enumerators went through a five-day training including two full pilots in schools that weren't part of the baseline/midline. Only enumerators who attended each day of the training were included within the data collection teams.

The assessor training included ethics and consent training. During the pilots, assessors worked through all data collection tools, including child assent, child protection and safeguarding procedures and policies. Enumerators were provided with a summary sheet of what they had learnt during the training to support them through their assessment, in case they required a refresher.

Data analysis methods: To test the comparability of children in intervention and control groups, from baseline to midline, the research team relied on multivariate regression and assessments of the significance of the coefficient of an intervention variable, with adjustments for the clustered sampling strategy including robust standard errors.

Data handling and confidentiality: Data was collected using KoBo ToolBox software (kobotoolbox.org). KoBo accounts were password protected and managed by the principal investigator. After data collection, tablets were set up so the data that had been collected was encrypted and could not be viewed on the tablet.

Access to data in the server was only available to the principal investigator. The three monitoring and evaluation coordinators for each province were able to view the data as it came in but not edit or change any values. After all data had been uploaded to the server, the principal investigator downloaded it into a password protected computer and separated student names from the rest of the dataset.

[11] The training is aimed at improving the competence and confidence of elementary teachers in teaching literacy skills.

Internal and external validity: As an experimental study, it is designed with strong internal validity. Because intervention was randomly assigned, we are confident that our intervention and control groups are balanced, on average, in all characteristics. We examined the results of our baseline balance tests and found no statistically significant differences between the intervention and control groups.

During the first year of implementation, RISE PNG paid careful attention to: (1) ensure that intervention status was preserved in the roll-out of activities so the control group remained, as much as possible, uncontaminated by the intervention, and (2) monitor the control group for any spillover or contamination from other programmatic interventions.

Causal relationships: The design of our impact evaluation will allow us to interpret midline and endline differences which may have resulted from the program's interventions. However, the nature of the relationships included in our multivariate regression analysis, while suggestive of causal relationships, should be interpreted as strictly correlational.

4.7 LIMITATIONS

Lack of qualitative data and explicit validation process: The literacy and numeracy study design did not build in a strong qualitative component and there was no allocation of time and resources for additional data collection through focus group discussions and key informant interviews, which could have helped in triangulating the findings from the literacy and numeracy assessment. As a result, some of the findings lack in-depth analysis. There may also be inaccuracy in the home learning environment subtask items because of the reliance on children rather than caregivers to relay this information. Feedback from enumerators also suggested that children were fatigued by the time they reached the home learning environment section of the assessment.

A lack of inter-rater reliability in classroom observations: The lesson observation tool did not report an inter-rater reliability (IRR) due to time and resource constraints during the midline. Only one teacher trainer observed the classroom. Ideally, we would have had two so we could compare their findings and ensure the results were unbiased and accurate – similar to the approach carried out for the student assessments (for a random 10% of the students).

Hawthorne effect: Teachers knew when they were being observed and may have been keen to demonstrate their newly acquired skills in LB and NB when enumerators were present.

Early closure of schools: In order to allow enough time for the intervention to have an effect on children's literacy and numeracy skills, the midline was conducted in October and November of 2018, near the end of Term 4. However, many schools had already shut down by then – well before their official closing date. The early closure of schools reduces our sample size by 33% contributing to a lack of statistically significant differences between intervention and control school children for certain subtests, despite the mean values for children in intervention schools being 2–3 percentage points higher on nearly all subtests.

Contamination: Teachers may have shifted from intervention schools to control schools. Moreover, anecdotal evidence suggests that a few teachers participated in the training claiming to be from Local Level Governments (LLGs) that were part of the intervention group in order to receive the training a year in advance. While not widespread, we have anecdotal evidence to suggest that this took place in a few LLGs in ARoB.

Literacy lesson: In order to compare with the baseline, the majority of observed lessons were literacy lessons. This has reduced the data available about the effectiveness of NB training on teachers' practices in mathematics.

5 STUDY FINDINGS

5.1 RESPONSE SAMPLE

The total impact response sample includes the schools, students and teachers that were assessed at baseline and again at midline. The response at midline compared to baseline is as follows:

Table 10: Baseline and midline response sample by province

Baseline & midline schools by province				
	ESP	EHP	ARoB	**Total**
Baseline	73	59	65	**197**
Midline	42	46	43	131
Total impact sample	**42**	**46**	**43**	**131**
Baseline & midline students by province				
	ESP	EHP	ARoB	**Total**
Baseline E1 students	1075	930	738	**2743**
Midline E2 students	715	797	523	**2035**[12]
Total E2 students from baseline response sample	**335**	**328**	**341**	**1004**
Baseline & midline teachers by province				
	ESP	EHP	ARoB	**Total**
Baseline E1 teachers	68	62	45	**175**
Midline E2 teachers	33	43	36	**112**
Total impact sample	**33**	**43**	**36**	**112**
Midline TICs by province				
	ESP	EHP	ARoB	**Total**
Midline E2 TICs	42	46	44	**132**

The number of schools was lower mostly due to early closures rather than purposeful action to reduce the response sample. This had an effect on the numbers of students from baseline that were assessed again at midline. The impact sample from the original 2,743 students has reduced to 1,004 students, based on the names matched at midline against those at baseline. This reduction was due primarily to absenteeism on the day of the assessment and early school closure in some schools. Drop out and transfer were not noted as reasons for absenteeism on the day. At baseline, ARoB, ESP and EHP absenteeism rate was very high (52%, 52% and 46%), and this pattern appears to have continued during the past 12 months, reducing the number of the original sample by half.

[12] This number is made up of 1,004 students who were from the original baseline sample and 1031 students who were not assessed at baseline but in the same classroom. Results were balanced across all literacy and numeracy test items so the 2,035 response sample was used to ensure we retained a representative sample for each province.

The research team used results from 839 control students and 989 intervention students when analysing differences between intervention and control, which excludes a sample of 207 students from backup schools who were assessed but not assigned intervention and control status.

As discussed later in the report, high absenteeism is likely to be one of the primary reasons why LB and NB has not been more effective. Children are simply not coming to school enough. However, it is clear that absenteeism did not bias the results. It is merely a reflection of the situation in PNG.

Table 11: Midline student response sample by gender

Child's gender	Frequency	Percentage	Cumulative
Male	997	48.99%	48.99%
Female	1038	51.01%	100.00%
Total:	**2035**	**100%**	

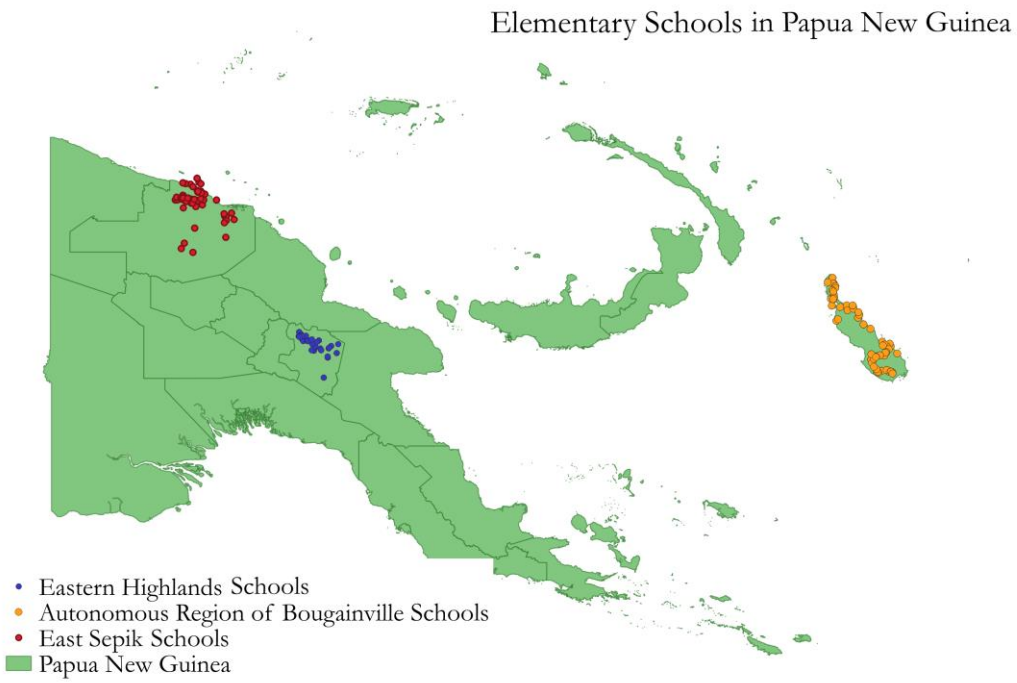
There were slightly more girls than boys assessed at midline, which mirrors the baseline numbers. As was instructed, the majority of students assessed were E2 students. An EP or E1 student was only assessed if they had been assessed at baseline and had repeated E1 so was still in E1 when enumerators visited the school for midline.

Table 12: Midline student response sample by grade level

Child's grade in school	Frequency	Percentage	Cumulative
Elementary Prep	3	0.15%	0.15%
Elementary 1	91	4.47%	4.62%
Elementary II	1,941	95.38%	100.00%
Total:	**2,035**	**100%**	

The following map highlights the school areas covered at midline. Provincial maps are available in Appendix C.

Figure 2: Map of elementary schools in Papua New Guinea in the response sample at midline



5.2 MONITORING EVALUATION PLAN BASELINE AND MIDLINE FIGURES

This report provides updated midline figures for Outcome 1 including: (1) the proportion of teachers demonstrating improved literacy and numeracy instructional practices after receiving LB and NB training, (2) the percentage of schools adequately resourced with learning materials and the percentage of schools at each level using relevant learning materials, (3) the proportion of caregivers and children with a reported improvement in home reading and community reading culture, and (4) the proportion of girls and boys in elementary schools who demonstrate improved literacy and numeracy.

Table 13: Baseline and midline figures for Outcome 1

End of Program Outcome (EOPO)	Intermediate Outcome (IO)	Outcome Indicator	Target	Baseline Results	Midline Results
EOPO 1: **Improved literacy and numeracy learning outcomes for elementary students**	IO1: *Elementary teachers demonstrate confidence and competence in teaching and assessing Standards Based Curriculum (SBC) English, Language and Maths*	Percentage of teachers demonstrating improved literacy and numeracy instructional practice after receiving Literacy Boost (LB) and Numeracy Boost (NB) training	400 of the 1,800 targeted teachers (25%) demonstrate improved literacy and numeracy instructional practices after first year of training and additional improvement in subsequent years	No difference between intervention and control school teachers at baseline, measured using a modified Stallings tool	**Teaching reading comprehension:** Teachers who received intervention scored 66.94% for teaching reading comprehension compared to 35.29% by control teachers (31 percentage points

					difference) [13] p<0.01 **Teacher capacity to assess students:** Similarly, intervention teachers who received intervention scored 25.04% for assessments as compared to 15.02% for control teachers (10 percentage points difference) P<0.05
	IO2: *Elementary schools resourced with literacy and numeracy instructional materials*	National Education Plan (NEP) indicator 20: Percentage of schools adequately resourced with relevant learning materials NEP 22: Percentage of schools at each level using relevant learning materials	90% of 650 elementary schools have access to teaching guides[14] 20% of schools involved in Education for Life (EFL) trial have used the Teacher Presenter Kit in the last month	81.71 % of schools had English Standards Based Curriculum (SBC) teaching guides 14.86% of schools had English SBC E2 readers 9.14% of schools had reading books in English 65.71 % of schools were using SBC	69.61% of intervention teachers who received intervention were up to date with syllabus compared to 63.04% of control teachers[15] 41.51 % of (T) teachers missed sections of SBC compared to 56.52% of (C) teachers EFL trial results pending
	IO3: *Parents and caregivers of early grade children demonstrate improved home reading practices and community reading culture*	Percentage of caregivers and children with a reported improvement in home reading habits and community reading culture	25% of targeted caregivers and children report an improvement in home reading habits	No difference in home reading habits of children in intervention schools and control schools	22% of children in intervention schools engaged in reading clubs compared to 12% control children 18.86% of intervention children in intervention schools borrow books compared to

[13] All results presented in this report, particularly where the intervention and control group are being compared, will include details of the scores achieved (usually presented as percentage figures), and an assessment of the statistical significance of the result. Statistical testing is used to assess whether there is causal relationship between an intervention and a result, that is, did a particular activity (e.g. training) have an impact (it could be positive or negative) on the group undertaking the activity (teachers who received the training (intervention group) compared to a group who did not (teacher who did not receive the training (control group). The significance level for this study is 5%, that is, the testing tells us that for a statistically significant result, we can be 95% confident that the result we are measuring has been achieved by the activity.

[14] This indicator will be changed to reflect the PPF recommendation to report on NEP indicator linked to use of SBC, which is more appropriate for RISE as in-service teacher training is focused on this.

[15] This indicator is measured based on daily scripted lessons that are used and expected in elementary grades. A teacher was considered up to date if they were teaching the correct day.

					11.07% control children
	I04: *Improved literacy and numeracy outcomes in elementary schools*	Percentage of girls and boys in elementary schools who demonstrate improved literacy and numeracy skills	20% of children show an improvement in literacy and numeracy skills (in comparison to control groups) for elementary students	No difference in literacy and numeracy skills of children in intervention schools and control schools at baseline	Girls in intervention schools scored 5.07% higher in numeracy than girls in control schools, and 4.53% higher in overall literacy & numeracy compared to control schools. (*p<0.01)* Boys in intervention schools show no statistically significant improvement compared to boys in control schools.

5.3 PROPORTION OF TEACHERS DEMONSTRATING IMPROVED LITERACY AND NUMERACY INSTRUCTIONAL PRACTICE AFTER RECEIVING LITERACY BOOST AND NUMERACY BOOST TRAINING

Table 14: Baseline and midline results for the percentage of teachers demonstrating improved literacy and numeracy instructional practice after receiving LB and NB training

Indicator	Baseline figure	Midline figure
Percentage of teachers demonstrating improved literacy and numeracy instructional practice after receiving LB and NB training	No difference between intervention and control school teachers at baseline, measured using a modified Stallings tool	**Teaching reading:** Intervention teachers scored 66.94% for teaching reading compared to control teachers who scored 35.29% (31 percentage points difference, statistically significant, *p<0.01*) **Assessments:** Similarly, intervention teachers scored 25.04% for assessments as compared to control teachers who scored 15.02% for this competency (10 percentage points difference, statistically significant, *p<0.05*)

5.3.1 Lesson observation results

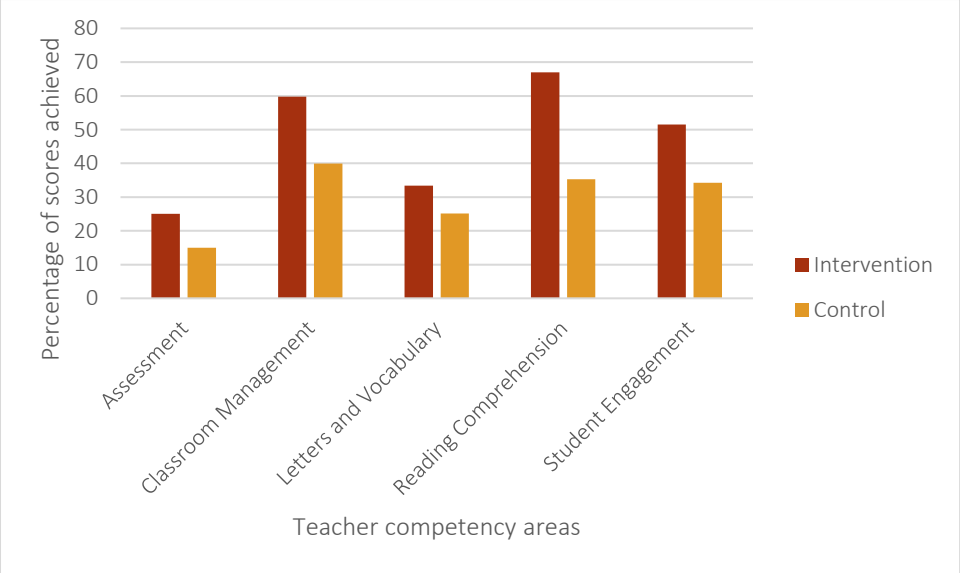
Of the teachers assessed at midline (n=112), 72(64.29%) had received training in 2018. Save the Children was the main provider, training 57% of the teachers assessed in this study, followed by the Government of PNG, which provided training to 39%. 'Other' organisations provided the remaining 4% of training to teachers. This is likely to be churches or UNICEF.

The trainings provided were LB (January 2018), NB (April 2018) and cross-cutting themes (June 2018). A total of 30 teachers (41%) attended all three, while the remaining 42 attended at least one training

At midline, using the lesson observation results, teachers in the intervention schools scored an average of 66.94% in teaching reading compared to teachers in control schools (35.29%). From a possible score of 5 for each item related to reading that was assessed, intervention teachers scored between 3 and 4. It was also found that intervention teachers were more likely to teach reading fluency and comprehension than control teachers. In intervention schools, 23.65% of observations reported a focus on these literacy areas compared to only 16.98% in control schools. This highlights the effect of efforts taken during teacher training to emphasise reading activities. The strong focus on different reading strategies and questioning shows a marked improvement.

Teachers in intervention schools scored 59.81% in classroom management and 51.48% in student engagement, which was very high compared to teachers in control schools who scored 39.89% and 34.24% respectively. These results indicate the teacher may be applying more collaborative student-centered methods of teaching literacy. As a result of introducing more interesting ways to teach reading, student engagement appears to have increased and this may be influencing classroom management. There is a strong correlation between a student's interest and motivation to learn and their behaviour – and the overall results support this. Motivation problems and behavior management problems are often intertwined; unmotivated students frequently misbehave and students who misbehave frequently do not care about learning (Curwin, 2010). Studies have shown that positive teacher–student relationships can lead to a warm classroom environment that facilitates successful adaptation in school and increases student motivation to learn (Koca, 2016).

Figure 3: Average percentage score in teacher competency domain by intervention status



5.4 National Education Plan Indicators on Teaching and Learning Resources

5.4.1 NEP Indicator 20: Percentage of schools adequately resourced with relevant learning materials

At baseline, most teachers (81.71%) had English teaching guides and slightly fewer (79.4%) had access to the English syllabus (79.4%). A lower proportion of teachers (69.7%) had the Language teaching guide and language

syllabus (70%) for Tok Pisin/Tok Ples language learning. It was noted that while basic teaching and learning resources were available, many classes are resource poor and some teachers (between 20% and 30%) did not have the required teaching resources.

Table 14: Baseline results for NEP Indicator 20

Relevant learning materials	Baseline figure
English SBC teaching guides	81.71%
English SBC E2 readers	14.86%
Reading books in English	9.14%

A much lower proportion of classes had the necessary student resources at baseline. Many classrooms were without reading books and student workbooks. This limited children's ability to complete exercise tasks and engage in classroom activities. The table below includes a list of resources available for students in typical E1 classrooms. It can be noted that there were few resources available for students in Tok Pisin and Tok Ples.

Table 15: Student learning resources recorded at baseline

Classroom environment	Baseline figure
Lots of printed materials on classroom walls	28.53%
Lots of printed materials in English	29.71%
Lots of printed materials in Tok Pisin	0.57%
Lots of printed materials in Tok Ples	5.71%
Sufficient SBC E2 readers	14.86%
Sufficient student reading books	16.57%
Sufficient storybooks in English	9.14%
Sufficient storybooks in Tok Pisin	0.57%
Storybooks in local languages	1.14%

The classroom environment was considered print poor with less than 30% of classrooms displaying print materials on walls. Most of the materials on display were in English with nominal material in Tok Pisin and Tok Ples. A very low proportion of schools displayed storybooks. Books in local languages were rarely sighted. In 40% of cases, teachers reported that they faced challenges due to the lack of teaching and learning resources available to them. At midline, observations of the classroom learning environment were not captured but are planned to be captured at endline.

The midline assessment did not collect data specifically on the availability of SBC, but instead focused on utilisation (see below). NEP Indicator 20 will be collected again at endline.

5.4.2 NEP Indicator 22: Percentage of schools at each level using relevant learning materials

The overall results for the percentage of schools at each level using relevant learning materials have improved considerably with better utilisation of the syllabus.

Table 16: Baseline and midline results for NEP Indicator 20

Use of relevant learning materials	Baseline figure	Midline figure
Up to date with the syllabus	63.04%	69.81%
Has missed sections of the syllabus	41.51%	56.52%

As noted in the baseline study, most teachers (65.7%) were following the SBC syllabus. However, 35% of children were being taught either without a syllabus at all or with one that is out of date. This is concerning, particularly if the material being used does not align to the performance standards of the SBC and does not include the content which is expected to graduate children into primary levels. In the endline study, this item will be assessed again to determine whether SBC utilisation has improved over time.

The baseline did not reveal the proportion of teachers teaching the SBC on the correct day according to the scripted lesson, but the majority of teachers (72.17%) were teaching to the correct week and grade level.

Table 17: Lesson match at baseline

Lesson match	Percentage of teachers
Week	72.17%
Term	74.78%
Grade	84.35%

At midline, as shown in Table 19, 70% of teachers in intervention schools were up to date with the syllabus based on the scripted lesson compared to 63% of teachers in control schools (a difference of nearly 7 percentage points).

Table 18: Coverage of the syllabus at midline by intervention status

Coverage of syllabus	Control	Intervention
Up to date with the syllabus	63.04%	69.81%
Behind the syllabus	36.96%	30.19%
Total:	100%	100%

Of the 30% of teachers in intervention schools and 37% of teachers in control schools who reported being behind in the syllabus, most teachers were not behind by more than two weeks. Most teachers in control schools (58.82%) were behind by less than a week. A higher proportion of teachers in intervention schools reported being behind by between one and two weeks. This suggests teachers in the control schools were less behind than teachers in intervention schools (although the sample size for this response was very low, just 17 for control group and xx for the intervention group). Absenteeism and early school closures are balanced across intervention and control so these variables cannot explain the different results here.

Table 19: Period of time the teachers reported being behind in the syllabus by intervention status

Amount of time teacher is behind in the SBC curriculum	Control	Intervention
Less than one week	58.82%	25%
One to two weeks	11.76%	31.25%
Two to three weeks	0%	18.75%
Three to four weeks	17.65%	6.25%
More than four weeks	11.76%	18.75%
Total:	100%	100%

When analysing the percentage of teachers who have missed sections of the syllabus, a higher proportion of control teachers appear to have missed sections of the syllabus. This could be why they are reporting being less behind than the intervention teachers who are less likely to miss sections of the syllabus.

Table 20: Proportion of teachers by intervention status who had missed sections of the syllabus

Teacher has missed sections of the syllabus	Control	Intervention
No	43.48%	58.49%
Yes	56.52%	41.51%
Total:	100%	100%

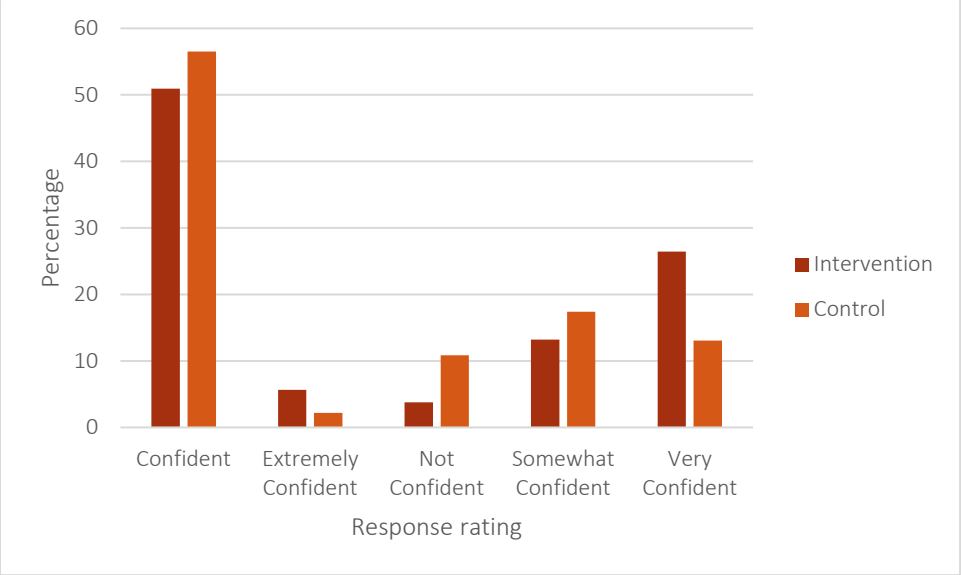
The midline also found that of the teachers missing sections of the syllabus, 42.31% from control schools and 31.82% from intervention schools reported missing at least a quarter. This is equivalent to at least one term of work. While the situation appears to be better in intervention schools, many children are not receiving the expected coverage of the syllabus and it is likely this is impacting student learning outcomes.

Table 21: Proportion of teachers by intervention status who had missed a certain amount of the syllabus

Teacher has missed this much of the syllabus	Control	Intervention
A few lessons	57.69%	68.18%
A quarter of the syllabus	23.08%	13.64%
Half of the syllabus	15.38%	9.09%
Large parts of the syllabus	3.85%	9.09%
Total:	100%	100%

At midline, teachers were asked how confident they felt teaching the SBC. While over 50% of teachers in both intervention and control schools felt confident using the teaching guides from SBC, 17% of intervention teachers and 28% of control teachers were either only somewhat confident or not confident. It is plausible that confidence and poor understanding of the SBC is related to poor SBC utilisation. Almost double the percentage of intervention teachers reported feeling very confident using the SBC teaching guides compared to control teachers.

Figure 4: A teachers' confidence using the teaching guides from SBC by intervention status



5.5 PROPORTION OF CAREGIVERS AND CHILDREN WITH A REPORTED IMPROVEMENT IN HOME READING AND COMMUNITY READING CULTURE

At baseline, just under half the children reported having storybooks and newspapers at home, while 43% reported having access to textbooks. More than three quarters of households had religious books.

At midline, as shown in Table 23, the availability of reading material at home increased for all text types. The most notable improvement was found in the availability of storybooks, with 13% more children reporting to have them at home. Promoting the value and importance of storybooks at home is a key feature of caregiver workshops and reading club activities, so this result is perhaps reflective of their impact. Given the upward trend at both baseline and midline, results indicate that access to reading material at home is increasing over time in PNG.

Table 23: Availability of reading material at home reported by children at baseline and midline

Reading material	Baseline	Midline
Textbooks	43.16%	49.95%
Religious books	75.83%	88.41%
Newspapers	47.51%	56.48%
Storybooks	48.95%	62.92%

At baseline, 72% of children claimed they had seen at least one member of the household read in the past week and 57% stated that a member of their household had read a story to them in the past week. At midline, 78% of students reported that their mother read a story to them, which shows a 21% increase in the proportion of children who have been read to at home.

Differences are evident between intervention and control responses with 76.49% of students from intervention schools reporting their mother had read to them in the past week compared to 80.65% from control schools. This infers the intervention effect is not evident for this variable. However, other changes related to the home learning environment and community reading culture are evident.

Table 24: Students report on their mother's reading practices at home by intervention status

Mother reads to the child	Control	Intervention
No	19.35%	23.51%
Yes	80.65%	76.49%
Total:	**100%**	**100%**

Table 25 shows 22% of children from intervention schools have been engaged in reading club activities outside school compared to only 12% from control schools. This is a difference of 10 percentage points. Although this is not the improvement RISE PNG was expecting, it does show improvement and hopefully a shift in community literacy actions. It also shows that reading activities are happening in some communities outside the influence of this program. This might be explained by other community-based organisations, particularly church groups, setting up community libraries in some control areas.

Table 25: Children's reported engagement in reading club activities by intervention status

Child has attended the reading club outside of school	Control	Intervention	Total
No	731	770	1,501
	87.76%	77.94%	82.43%
Yes	102	218	320
	12.24%	22.06%	17.57%
Total	833	988	1,821[16]
	100%	100%	100%

It is also evident that children have been active in other areas of community literacy with 19% of children reporting engagement in storytelling and 7% of children reporting participation at reading festivals. This indicates there is a demand for community literacy and it is something to build upon/that could be nurtured in 2019. The project faced significant challenges mobilising the community literacy component, in particular securing community buy-in so that participation rates in literacy activities were high.

The results also show that 8% more children in intervention communities are borrowing reading books from the community book bank as compared to children in the control communities. This presents a modest improvement during the first year of implementation and does present a shift in community level engagement in reading. Similar to the result for reading clubs, the presence of community book banks in control areas may be a consequence of the activities other community-based organisations or church groups.

Table 22: Reponses from children about their book borrowing activities by intervention status

Child has borrowed reading books from the community book bank	Control	Intervention	Total
No	739	800	1539
	88.93%	81.14%	84.70%

[16] A smaller response sample was included due to some enumerators going to backup schools without permission. These results were excluded from the analysis.

Yes	92	186	278
	11.07%	18.86%	15.30%
Total	831	986	1817[17]
	100%	100%	100%

5.6 PROPORTION OF GIRLS AND BOYS IN ELEMENTARY SCHOOLS WHO DEMONSTRATE IMPROVED LITERACY AND NUMERACY

The following table presents the literacy test results for E2 students from control and intervention groups. The results show that elementary students from the intervention group performed 1% to 3.5% better on all subtest items compared to students from the control group. The largest differences were in letter identification scores in English and reading comprehension scores in Tok Pisin. The improvements in word knowledge in both English and Tok Pisin were also noteworthy.

Table 27: Average literacy test scores for children by intervention status

Literacy skill test items	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
Letter recognition	839	23.7	91.1	989	24.5	94.1	3
English words	839	7.2	36.2	989	7.8	39.0	2.5
Tok Pisin words	839	6.6	33.1	989	7.1	35.6	2.5
English reading comprehension	274	2.2	37.5	360	2.3	38.5	1
Tok Pisin comprehension	196	4.5	74.8	247	4.7	78.3	3.4

In terms of measures of readers and non-readers, the improvements are more statistically significant than in other areas. The proportion of children in the intervention group who could read at least five words of the English passage in 30 seconds was 49.5% compared to 44.7% in the control group. This shows a variation of almost 5 percentage points between intervention and control. The increased frequency and attention placed on reading activities in the classroom and at home appears to be producing the desired result to improve reading skills.

Table 28: Children's ability to read at least five words of an English passage in 30 seconds by intervention status

Child read at least five words of the English reading passage	Control	Intervention	Total
No	339	367	706
	55.3%	50.48%	52.69%
Yes	274	360	634

[17] A number of children did not respond to this item.

	44.7%	49.52%	47.31%
Total	613	727	1,340[18]
	100%	100%	100%

There are notable improvements in the proportion of readers of Tok Pisin, with students from the intervention group scoring, on average, 1.9 percentage points higher than students from the control group. The difference is not as marked as the results in English, which may be explained by the low prevalence of reading materials in Tok Pisin and the likely low frequency of reading activities in Tok Pisin, at home and in classrooms. But, with positive results shown in word identification and comprehension, progress is being made in children's Tok Pisin literacy ability.

Table 29: Children's ability to read at least five words of a Tok Pisin passage in 30 seconds by intervention status

Child read at least five words of the Tok Pisin reading passage	Control	Intervention	Total
No	416	481	897
	67.97%	66.07%	66.94%
Yes	196	247	443
	32.03%	33.93%	33.06%
Total	612	728	1,340
	100%	100%	100%

5.7 RESPONSES TO THE RESEARCH QUESTIONS

To identify variables that positively and negatively correlate with children's acquisition of foundational literacy and numeracy skills, this study posed a number of research questions.

The answers will help us consider how various components of LB and NB, including teacher training and professional development, the supply of teaching and learning resources, and community activities, are contributing to changes in children's literacy and numeracy skills and teachers' competency in English and Mathematics instruction. They have been answered based on analysis of baseline and midline data and the variation between the control and intervention response samples.

The endline study will further analyse the questions to determine whether Save the Children's LB and NB have produced a measurable impact on elementary students' literacy and numeracy skills and elementary teachers' skills in teaching English and Mathematics.

5.7.1 What are the children's learning gains in literacy skills in English and Tok Pisin in intervention and control schools at baseline and midline?

[18] This response is lower than the impact sample because the 'no response' students were eliminated.

Letter identification

Table 30 highlights the difference in scores between intervention and control students. The results show that a higher proportion of students from intervention schools were able to identify one more letter than students from control schools and, on average, students from intervention schools were able to answer 94% of the letters correctly compared to 91% in control schools. This effect is statistically significant ($p<0.05$) demonstrating that the intervention has had an overall positive effect.

Table 30: Average letter identification scores by intervention status

Literacy skill test item	Control (n=839)			Intervention (n=989)			Percentage point difference	Regression Analysis
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)		
Letter recognition	839	23.7	91.1	989	24.5	94.1	3	Significant difference in letter recognition in the intervention vs control population (*p=<0.05)*

Table 31 highlights the variation in results from baseline and midline, highlighting the differences between E1 and E2 students. The age variable is likely to be the main variable influencing results in ESP, EHP and ARoB. The largest gains were evident in ARoB, followed by EHP and then ESP. The substantial gain (32 percentage points) for letter recognition in ARoB is noteworthy.

Table 31: Baseline and midline scores for letter recognition irrespective of intervention status

	Baseline results			Midline results		
	Students who knew all English letters	Students who could not identify one letter	Average number letters identified correctly (total = 26)	Students who knew all English letters	Students who could not identify one letter	Average number of letters identified correctly (total = 26)
ARoB	22.36%	2.98%	20.85	57.93%	0.00%	24.47
ESP	30.51%	7.26%	20.36	41.26%	0.70%	23.91
EHP	18.92%	3.87%	20.92	33.12%	0.00%	23.98
Total	24.39%	4.96%	20.68	42.36%	0.25%	24.08

Word knowledge

Table 32 shows the difference in word knowledge scores from E2 students in intervention and control schools. The results show the students from intervention schools were able to identify an average 8 out of 20 words in English compared to an average 7 out of 20 by the students from control schools. In percentage terms, the student average from the intervention schools was 39% of the words correctly identified compared to 36% in control schools, representing a 2.5 percentage point difference.

The variation between intervention and control for word knowledge in Tok Pisin is quite similar to English. Students from intervention schools were able to identify 7 out of a possible 22 words in Tok Pisin - an average of 33.1%, compared to slightly less than 7 words by students in control schools – an average of 35.6%. This

represents a 2.5 percentage point difference and suggests that the results for word scores are higher for intervention schools than control schools for both English and Tok Pisin.

Table 32: Average word knowledge in English and Tok Pisin by intervention status

Literacy skill test items	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
English words	839	7.249106	36.24553	989	7.799798	38.99899	2.5
Tok Pisin words	839	6.623361	33.11681	989	7.114257	35.57128	2.5

Table 33 highlights the variation in results from baseline and midline and shows the difference in results from E1 to E2 for students who were assessed. The results at baseline are comparable with the results at midline in so far as ESP has performed best in word identification, with the highest percentage of children who could identify all the English words and the highest average number of English and Tok Pisin words identified. EHP is once again showing the lowest gains between baseline and midline. Overall, EHP achieved the lowest scores in word identification at midline, which is similar to its results at baseline.

Table 33: Baseline and midline scores for English word identification

	Baseline results						Midline results					
	Children who could identify all English words	Children who could not identify one word	Average number of words identified (total = 22)	Children who could identify all Tok Pisin words	Children who could not identify one word	Average number of words identified (total = 20)	Children who could identify all English words	Children who could not identify one word	Average number of words identified at midline (total = 22)	Children who could identify all Tok Pisin words	Children who could not identify one Tok Pisin word	Average number of Tok Pisin words identified (total = 20)
ARoB	1.90%	44.44%	4.47	4.34%	57.05%	3.63	2.49%	13.00%	6.85	7.84%	34.99%	6.99
ESP	4.00%	42.14%	6.48	4.74%	49.67%	4.85	5.59%	10.21%	9.71	7.27%	21.68%	8.22
EHP	0.65%	54.84%	3.07	1.18%	72.37%	1.59	1.63%	14.43%	6.65	3.89%	26.47%	6.14
Total	2.30%	47.07%	4.78	3.43%	59.35%	3.42	3.24%	12.58%	7.78	6.09%	26.98%	7.09

Beginner readers and non-readers

As discussed earlier, the results in this subtest revealed that the E2 students from intervention schools scored 5 percentage points higher that E2 students from control schools. The results indicate that the extra emphasis placed on reading at home and in the classroom is having an effect on reading habit and reading fluency. The intervention group would have been exposed to a variety of community and classroom literacy activities and their reading scores appear to show they have responded well.

Table 34: Reading ability in English by intervention status

Child read at least five words of the English reading passage	Control	Intervention	Total
No	339	367	706
	55.3%	50.48%	52.69%
Yes	274	360	634
	44.7%	49.52%	47.31%

Total	613	727	1,340
	100%	100%	100%

The findings revealed an improvement in readers of Tok Pisin, with students from intervention groups scoring, on average, 1.9 percentage points higher than students from control groups. The difference is not as much as the results in English, which is explained by the low prevalence of reading materials in Tok Pisin and the likely low frequency of reading activities in Tok Pisin at home and in classrooms. However, progress is being made in children's grasp of Tok Pisin literacy, with positive results shown in word identification and comprehension.

Table 35: Reading ability in Tok Pisin by intervention status

Child read at least five words of the Tok Pisin reading passage	Control	Intervention	Total
No	416	481	897
	67.97%	66.07%	66.94%
Yes	196	247	443
	32.03%	33.93%	33.06%
Total	612	728	1,340[19]
	100%	100%	100%

Improvement is evident in the three provinces although the highest scores in reading ability and the lowest numbers of non-readers both come from ESP. In ARoB, non-readers of English and Tok Pisin have reduced in percentage from the baseline score. This is also evident in EHP, where the reduction in non-readers in English has been noteworthy – from 81.72% of students at baseline to 41.31% of students at midline.

Table 36: Reading ability scores at midline

	Baseline results				Midline results			
	Students who could read at least five English words per 30 seconds	Non-readers in English	Students who could read at least five Tok Pisin words per 30 seconds	Non-readers in Tok Pisin	Students who could read at least five English words per 30 seconds	Non-readers in English	Students who could read at least five Tok Pisin words per 30 seconds	Non-readers in Tok Pisin
ARoB	31.03%	68.97%	15.18%	84.82%	43.15%	56.85%	35.79%	64.21%
ESP	33.86%	66.14%	15.16%	84.84%	63.40%	36.60%	46.11%	53.89%
EHP	18.28%	81.72%	3.98%	96.02%	41.31%	41.31%	25.82%	74.18%
Total	27.82%	72.18%	11.37%	88.63%	49.67%	50.33%	35.64%	64.36%

Comprehension

The differences in reading comprehension scores between intervention and control status are shown in Table 37. There are improvements, albeit slight, in the results of intervention students – particularly in Tok Pisin reading comprehension, where students scored on average 3.4 percentage points higher than their peers in

[19] There were non-responses for this test item hence a lower number of students responded to this test item.

control schools. One percentage point difference can be seen in English comprehension scores of intervention and control students.

Table 37: Average reading comprehension scores in English and Tok Pisin by intervention status

Literacy skill test items	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
English reading comprehension	274	2.248175	37.46959	360	2.311111	38.51852	1
Tok Pisin comprehension	196	4.489796	74.82993	247	4.696356	78.2726	3.4

Provincial results for reading comprehension in English and Tok Pisin are shown in Table 38. ESP has performed best in English reading comprehension, with almost 5% of students answering 100% of the questions correctly.

ESP also performed best in Tok Pisin. Pleasingly, 21.26% of students in ESP were able to answer 100% of the Tok Pisin reading comprehension questions correctly, compared with 13% of students in ARoB and only 9.4% of students in EHP.

Table 38: Provincial results for English and Tok Pisin reading comprehension

	Baseline results				Midline results			
	English passage reading comprehension all correct	Average score (total = 8)	Tok Pisin passage reading comprehension all correct	Average score (total = 9)	English passage reading comprehension all correct	Average score (total = 8)	Tok Pisin passage reading comprehension all correct	Average score (total = 8)
ARoB	1.49%	1.27	1.08%	0.70	1.91%	0.98	13.19%	1.62
ESP	5.95%	1.82	3.63%	0.93	4.90%	1.61	21.26%	2.19
EHP	1.51%	0.87	0.11%	0.19	0.75%	0.95	9.41%	1.36
Total	3.24%	1.35	1.75%	0.62	2.51%	1.19	14.55%	1.72

It is important to note that the reading comprehension task at baseline was designed for E1 students and reflects SBC E1 standards. The reading comprehension task at midline was designed for E2 students and reflects SBC E2 standards. Children performed better in the Tok Pisin comprehension task than the English one.

5.7.2 What are the children's learning gains in numeracy skills in intervention and control schools at baseline and midline?

Number identification

E2 students in both intervention and control performed well in the number identification test, as they did when they were E1 students and measured at baseline. At least 90% of children were able to identify each of the numbers, with single-digit numbers being identified most frequently. On average, students from intervention schools scored 2% higher than students from control schools. Students from control schools experienced slightly more difficulty with the numbers 39, 69 and 88 but also found 52 a difficult number to identify.

Table 39: Average student results for number identification by intervention status

Numeracy skill test item	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	

Number identification	839	9.3	92.7	989	9.9	94.8	2

The differences between baseline and midline results are presented in Table 40. It is important to note, due to different curriculum standards, the numbers used for E1 students at baseline were different from the numbers used for E2 students at midline.

Results for each province have improved since baseline, with the highest increases seen in EHP (43 percentage points) and ARoB (30 percentage points). The highest scores were recorded in ARoB, which is consistent with baseline results.

Table 40: Provincial results for number identification at baseline and midline

	Baseline results			Midline results		
	Students who could identify all numbers	Students who could not identify any numbers	Average numbers identified correctly (total 12)	Students who could identify all numbers	Students who could not identify any numbers	Average numbers identified correctly (total = 10)
ARoB	54.34%	1.49%	10.55	84.13%	0.38%	9.43
ESP	57.30%	2.33%	10.32	81.96%	0.42%	9.43
EHP	34.73%	1.51%	9.83	77.42%	0.25%	9.33
Total	48.85%	1.82%	10.22	80.74%	0.34%	9.39

Skip pattern by two

E2 students in both intervention and control scored high in this subtest. Similar to the results for number identification, there was a 2-percentage point difference between intervention and control students.

At baseline, there was an indication that children may have memorised the sequence from repetition tasks and that understanding may not be as strong as indicated by the results. However, due to the high scores in number identification and other subtests in E2, students are likely grasping the concept of skip patterns by E2. The moderate improvement in intervention students indicates that E2 students in intervention schools may have been exposed to new teaching strategies to help them grasp the concept of skip patterns.

Table 41: Average student results for skip patterns of two by intervention status

Variable	N	Mean	N	Mean	% gain
	Control		Intervention		
Total Score	839	8.9	989	9.2	
Percentage Score	839	88.9	989	92.0	+3

There is a significant variation in results across the provinces. ARoB is the lowest performer with 64.63% of students able to count by multiples of 2 compared to ESP and EHP who both scored between 82% and 84%. On average, students scored 9.4–9.5 out of a possible 10 marks whereas ARoB students only scored 8. These findings are consistent with the baseline where students from ARoB performed poorly in skip pattern questions.

Table 42: Provincial results for skip patterns of two for baseline and midline

	Baseline results	Midline results

	Students who could count by 2	Students who could not count by 2	Average correct responses (total 10)	Students who could count by 2	Students who could not count by 2	Average correct responses (total = 10)
ARoB	46.3%	13.6%	6.8	64.6%	6.5%	8.1
ESP	67.1%	7.4%	8.2	84.5%	1.0%	9.5
EHP	53.6%	7.4%	7.9	82.4%	0.9%	9.4
Total	56.9%	9.0%	7.7	78.8%	2.4%	9.2

Skip pattern by five

Overall, students performed better in this test. There was a difference in performance between intervention and control, with students from intervention schools scoring almost 4 percentage points higher than their peers in control schools. E2 students scored higher in skip patterns by two compared to skip patterns by five. E2 students were stronger in answering skip patterns by five up to number 35, after which they faced more difficulty identifying the next number in sequence.

Table 43: Average student results for skip patterns of five by intervention status

Numeracy skill test item	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
Skip patterns of five	839	8.6	86.4	989	9.0	90.0	3.5

Addition problems

E2 students scored moderately well in this subtest, with students from both control and intervention schools scoring more than half of the 10 addition problems correctly. There was a 2.5 percentage point difference between intervention and control students, with students from intervention schools more likely to answer addition problems where answers were over 100.

The majority of students answered the single-digit questions correctly but found questions with double digits and numbers above 100 the hardest to answer. All students had more trouble with double-digit addition problems that required students to add above 10 for the 10s, such as 79 + 88. Only 11% of students from control schools could answer this question, while 12% answered it in intervention. This is compared to the sum 3 + 5, which 86% of control students and 89% of intervention students could answer.

Table 23: Average student scores for addition problems of intervention status

Numeracy skill test item	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
Addition	839	5.2	52.3	989	5.9	54.9	2.5

The addition problems that students found most difficult included double-digit numbers and where more complex calculations were required. Students in the intervention schools scored 1–3 percentage points higher on these test items than students from control schools.

Table 24: Individual addition test item scores by intervention status

Addition test items	Control	Intervention
30 + 41 (=71)	38.4%	42.2%
38 + 33 (=71)	27.5%	29.8%
79 + 88 (=167)	11.1%	12.0%
89 + 51 (=140)	11.9%	12.4%

At midline, students experienced more difficulty with addition problems than at baseline when they were E1 students and solving simpler single-digit addition problems. A much smaller proportion of students were able to answer all the addition problems at midline. The average score was just over 5 out of 10, which was much lower than at baseline where averages ranged from 6–7.

At baseline, more ESP students scored 100% than students from other provinces. At midline, however, ARoB students were more likely to get full marks for addition followed by ESP and then EHP. EHP's performance in both the baseline and midline was below other provinces, although the average score was the same as other provinces. During the assessment it was observed that students were using stick lines to work out their calculations and had not grasped the concept of calculating using multiples or clusters of 10 along number lines. These concepts may not have been grasped during teacher training.

Table 25: Provincial scores in addition at baseline and midline

	Baseline results			Midline results		
	Students who could answer all addition problems	Students who could not answer any addition problems	Average correct responses (total 10)	Students who could answer all additional problems	Students who could not answer any addition problems	Average correct responses (total = 10)
ARoB	35.4%	7.2%	7.1	4.8%	5.0%	5.4
ESP	39.5%	9.8%	7.0	3.5%	5.5%	5.4
EHP	21.8%	8.8%	6.0	3.0%	6.2%	5.4
Total	32.4%	8.8%	6.7	3.6%	5.6%	5.4

Subtraction problems

Overall, students in control and intervention schools scored lower in subtraction problems than in addition problems. The difference between control and intervention is less pronounced in subtraction compared to addition, with intervention students scoring just 1 percentage point higher overall on subtraction problems. Control students scored higher in a number of subtraction problems, which is difficult to explain without additional qualitative analysis.

Table 26: Average student scores for subtraction problems by intervention status

Numeracy skill test item	Control (n=839)			Intervention (n=989)			Percentage point difference
	Sample Size	Score	Percentage (%)	Sample Size	Score	Percentage (%)	
Subtraction	839	5.0	49.5	989	5.0	50.4	1

Similar to addition problems, double-digit subtraction problems were more difficult for students than single-digit subtraction. The lowest scores were obtained from sums such as 94 minus 63 and 68 minus 53. Students from intervention schools scored 1 percentage point higher on two of the items but for the subtraction problem 94 minus 63, students from control schools scored 1.5 percentage points higher.

Table 27: Individual subtraction problems by intervention status

Subtraction test items	Control	Intervention
35 – 26 = 9	23.2%	24.1%
68 – 53 = 15	21.8%	22.1%
94 – 63 = 31	18.5%	16.7%

Table 49 shows some variation between baseline and midline. It is evident that at baseline there was significant variation in each province. There was a high number of students who scored well and a similar number who could not answer any subtraction questions. This distinct parabola shape is not a usual distribution.

At midline, the distribution shifts to the negative with a higher proportion of students in each province scoring 0% compared to students scoring 100%. The averages for baseline and midline are similar but the distribution is very different. The results suggest that the variation in student performance has reduced with the increased difficulty of the E2 curriculum subtraction questions.

Table 28: Provincial scores in subtraction at baseline and midline

	Baseline results			Midline results		
	Students who could answer all subtraction problems	Students who could not answer any subtraction problems	Average correct responses (total 10)	Students who could answer all subtraction problems	Students who could not answer any subtraction problems	Average correct responses (total = 10)
ARoB	16.9%	26.8%	4.9	3.4%	12.1%	5.3
ESP	22.0%	24.1%	4.9	4.3%	13.4%	4.9
EHP	8.5%	32.2%	3.9	4.0%	15.7%	5.0
Total	16.0%	27.6%	4.3	4.0%	14.0%	5.0

Word problems

Students in intervention schools scored, on average, 3 percentage points higher in word problems than students from the control schools. In intervention schools, 80% of children were able to answer the first question correctly compared to 78% in control schools.

Students from the intervention and control schools struggled to answer the following two questions:

- Question 2: Vare had 30 pieces of candy. Vare ate 13 pieces of candy. How many pieces of candy did Vare have left? (17)
- Question 4: Donny had 49 chickens. He sold 24 chickens at the market. How many chickens did Donny have left? (25)

These are both subtraction problems with numbers above 20, so are likely to be less familiar to the students and more difficult to visualise.

Students from both intervention and control schools scored well in these two questions:

- Question 1: Donna had 12 pieces of chalk. The teacher took away 2 pieces of chalk. How many pieces of chalk did Donna have left? (10)
- Question 3: There are 4 dogs. Each dog has 4 legs. How many legs are there in total?' (16)

These two questions included numbers that are single digit or below 20. It is likely the children were able to identify these numbers more readily that Question 2 and Question 4.

Table 29: Average word problem results by intervention status

Variable	Control		Intervention		Percentage point difference
	Sample size	Average score	Sample size	Average score	
Question 1	839	0.8	989	0.8	
Question 2	839	0.3	989	0.3	
Question 3	839	0.5	989	0.5	
Question 4	839	0.3	989	0.3	
Total Score	839	1.8	989	1.9	
Percentage Score	839	45.9	989	48.5	3.5

The provincial scores are shown in Table 51. Overall, ESP appears to have produced the best results in word problems with the highest average score and a lowest proportion of students scoring zero. In EHP and ARoB, there were still reasonably high numbers of students who were unable to answer any word problems and while the result is better than at baseline, it remains a cause for concern.

Table 30: Student scores by province at baseline and midline

	Baseline results			Midline results		
	Students who could answer all word problems	Students who could not answer any word problems	Average correct responses (total 4)	Students who could answer all word problems	Students who could not answer any word problems	Average correct responses (total = 4)
ARoB	20.9%	27.1%	1.4	17.0%	19.7%	1.9
ESP	45.8%	9.3%	2.1	16.5%	9.4%	2.1
EHP	26.3%	17.2%	1.6	8.0%	13.3%	1.8
Total	32.5%	16.8%	1.7	13.3%	13.6%	1.9

Shapes

In shape tests, there was significant variation between intervention and control when identifying triangles and describing their key features.

Students from intervention schools scored 8 percentage points higher on both test items, which is a very large difference. In intervention schools, 42% of students were able to correctly describe the features of a triangle while only 34% of students from control schools could describe the triangle shape correctly. Interestingly, a key focus of numeracy teacher training is how to recognise shapes and their features. The results suggest that the training on triangles may have been effectively transferred to students in class activities.

Table 31: Average student scores in shape identification by intervention status

Variable	Control		Intervention		Percentage point difference
	Sample size	Average score	Sample size	Average score	
Circle	839	1.0	989	1.0	
Triangle	839	0.6	989	0.7	
Definition of triangle	839	0.3	989	0.4	
Real world circle	839	0.9	989	0.9	
Real world rectangle	839	0.8	989	0.8	
Total Score	839	3.6	989	3.8	
Percentage Score	839	72.5	989	75.6	3

The shape test item did not change from baseline to midline, so improvements were expected at midline. The results in all provinces improved, with the most significant improvement recorded in ARoB. At baseline, ARoB students were the lowest scorers in shape identification and while this remained the case at midline, the gains they made were large. It is important to acknowledge, however, that with the exception of triangle identification and definition, E2 students achieved relatively similar scores irrespective of intervention status.

Table 32: Student results for shape identification by province for baseline and midline

	Baseline results			Midline results		
	Students who could identify all the shapes	Students who could not identify any shapes	Average correct responses (total 5)	Students who could identify all the shapes	Students who could not identify any shapes	Average correct responses (total = 5)
ARoB	14.9%	11.7%	2.7	24.5%	1.9%	3.4
ESP	30.3%	3.6%	3.4	38.6%	0.3%	3.9
EHP	25.6%	5.1%	3.4	34.9%	0.8%	3.8
Total	24.6%	6.3%	3.2	33.5%	0.9%	3.7

5.7.3 How does gender influence PNG children's reading and numeracy skills from baseline to midline?

At baseline, no major gender gap was observed in literacy and numeracy scores among E1 students. The exception was alphabet knowledge where girls were more likely to know all 26 letters of the alphabet (i.e. get a full score on the letter identification component of the assessment). The baseline also revealed that girls tended to perform worse on telling the time.

In the RISE PNG International Development Emergent Learning Assessment (IDELA) Baseline Study (2019), which measured children in ECCE, girls scored lower than boys on overall IDELA, an aggregate measure of emergent literacy, numeracy, motor development and social emotional development. There were no significant differences by gender for emergent numeracy and emergent literacy. Girls tended to score lower in motor development and social-emotional development.

In later years of schooling, differences begin to widen. By Grade 5, girls outperform boys in literacy. This difference can be observed in the domain as a whole, as well as in the individual strands, as shown in Table 54. The gap between girls and boys for each year level is widest in the 'writing' strand, followed by 'language features' and then 'reading'. According to the PILNA 2015, the distribution of scores is relatively symmetrical for both boys and girls, with the range of scores in the upper half approximately the same as the range of scores

in the lower half of the data. The range of scores for girls is approximately 407 to 502, and 399 to 494 for boys. The median score is 475 points for girls and 468 points for boys (Belisle, 2016).

Table 33: Distribution of literacy scores by gender, PILNA 2015

Grade 5	Gender	Descriptive stats	Domain	Strands		
			Literacy	Reading	Language features	Writing
	Females	Mean	475.2	482.3	484.9	459.7
		Standard deviation	40.5	34.1	45.5	61.2
	Males	Mean	467.3	477.4	478.4	447.9
		Standard deviation	39.9	33.8	44.8	59.8

Source: PILNA, 2015

In relation to numeracy, Table 55 shows the mean scores of the numeracy domain with the strands disaggregated by gender. In general, boys slightly outperform girls based on the average achievement in numeracy. However, the difference between girls and boys is relatively small. The average performance of boys and girls is very similar, except for one small difference in the 'number' and 'measurement and data' strands, where boys scored slightly higher than girls. It highlights that both boys and girls have a strong proficiency in the 'operations' strand. However, the standard deviation increases for both boys and girls indicating a wider range of scores.

Table 34: Distribution of numeracy scores by gender, PILNA 2015

Grade 5	Gender	Descriptive stats	Domain	Strands		
			Numeracy	Numbers	Operations	Measurement & data
	Females	Mean	497.8	505.4	494.8	502.1
		Standard deviation	34.9	46.1	35.5	37.3
	Males	Mean	498.1	506.5	494.7	504.4
		Standard deviation	34.9	46.5	35.3	37.8

Source: PILNA, 2015

The RISE PNG midline assessment presents some interesting findings and it is evident that by E2, gender differences are starting to emerge between boys and girls, particularly in relation to numeracy. Based on the regression analysis presented in Annex F and G, and irrespective of intervention, statistically significant differences are evident in a number of literacy and numeracy subtests. Boys scored 2.727 percentage points higher in overall numeracy and 4.065 percentage points higher in shape recognition compared to girls. There were also statistically significant differences in skip patterns of two and five, addition and word problems, with boys scoring between 2.5 and 3.6 percentage points higher than girls.

In Figure 5, 37.7 % of boys were able to identify all six shapes correctly compared to only 29.5% of girls. This shows that a higher proportion of boys than girls were able to identify all 6 shapes. The average score for girls and boys was 4 out of 6 shapes and slightly more girls than boys reached this average.

Figure 5: gender differences in E2 student results in shape recognition

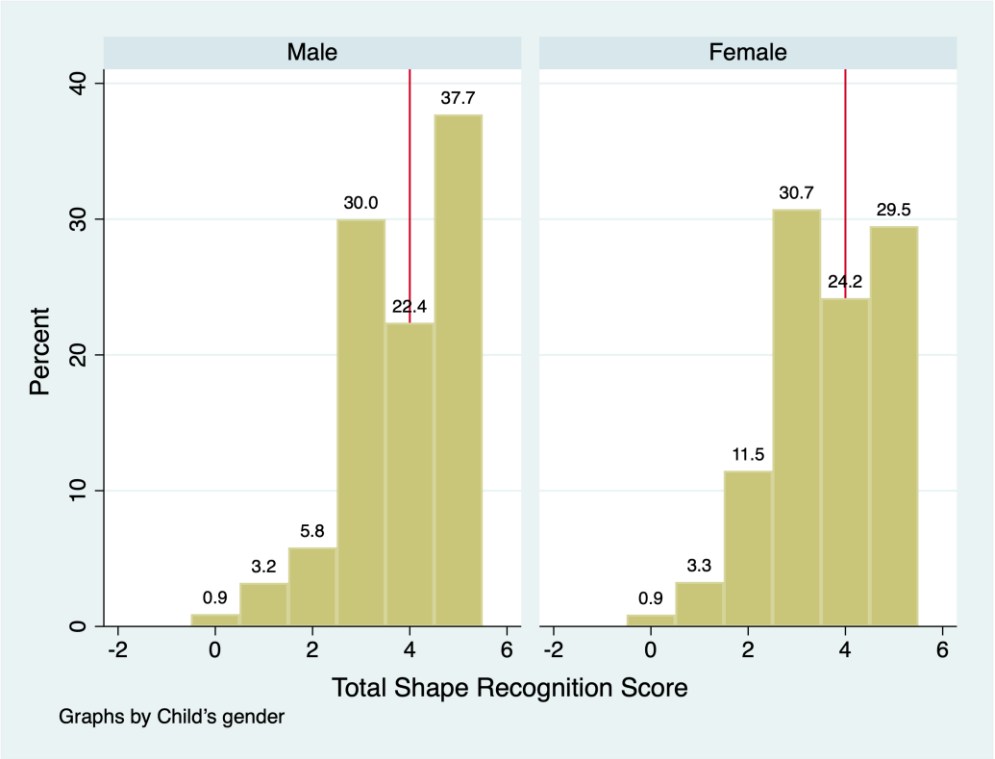
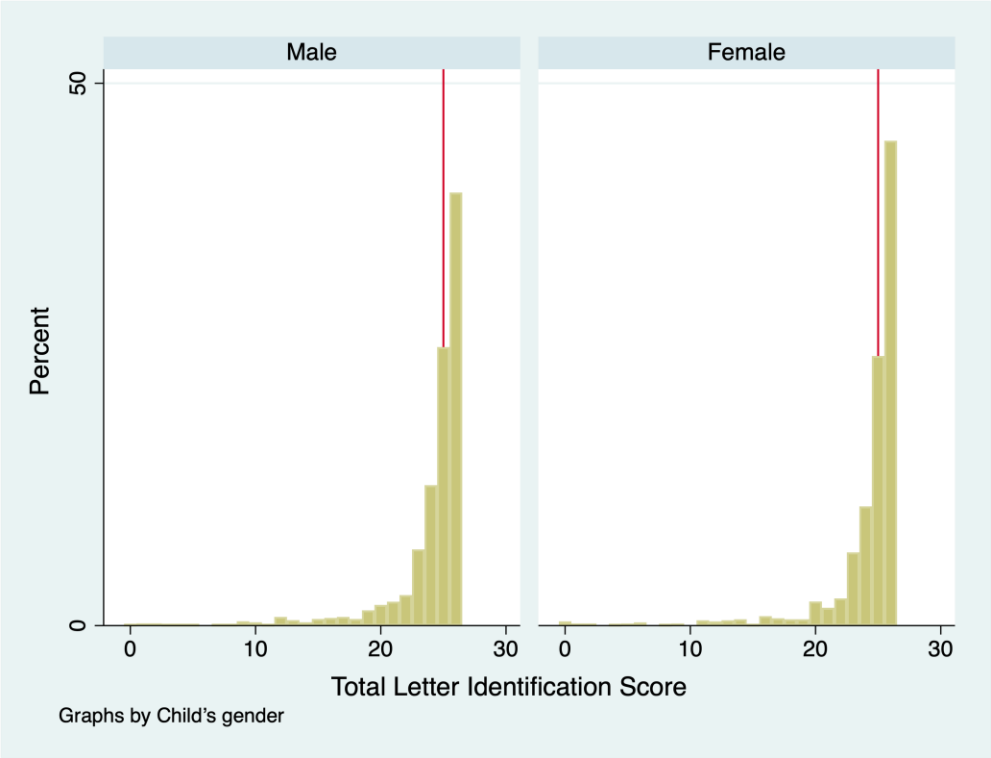


Figure 6 shows a higher proportion of boys were able to answer both two-word and four-word problems correctly. Almost 16% of girls were not able to answer any of the word problems compared to 11% of boys.

Figure 6: gender differences in E2 student results in numeracy word problems

Similar to the finding at baseline, girls are more likely than boys to know the 26 letters of the alphabet but, on average, both boys and girls tended to answer 25 of the letters correctly. Figure 7 shows the slight variation.

Figure 7: percentage of students and their total letter identification score



While it is evident that girls are not performing as well as boys in numeracy, the gains they have made as a result of the intervention are statistically significant. (Annex E). It demonstrates that the intervention is having a positive effect on girls' acquisition of foundational numeracy skills.

The gains made by boys are less significant and suggest that the intervention is yet to show an effect on boys' literacy and numeracy scores. While variation exists between control and intervention, the gains are not statistically significant for the combined sample of boys and girls. When the regression is analysed for girls only, the statistically significant gains are evident for the overall literacy and numeracy score, as well as the individual numeracy score. The analysis shows that girls in intervention schools performed 5 percentage points higher in numeracy and 4.5 percentage points higher in overall literacy and numeracy than girls in control schools.

Table 35: Impact of the intervention on girls' overall scores and numeracy

	(1)	(2)	(3)
	Literacy and Numeracy	Literacy	Numeracy
Control scores	53.7%***	40.6%***	67.2%***
Percentage point increase for intervention	4.5*	4.0	5.1**
Observations	933	933	933

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.7.4 What variables in the home or community (e.g. reading books at home, parents reading to their child, attendance at reading club, betel nut chewing) are influencing children's acquisition of literacy and numeracy skills?

Number and type of reading books at home

At baseline, more than 75% of students had access to religious books at home and 49% of students reported having storybooks at home. A slightly lower percentage (43%) of students reported having textbooks at home. Interestingly in the RISE PNG IDELA baseline study, a similar result was found, with 74% of caregivers reporting to own religious books at home and 58% reporting to own storybooks.

The IDELA baseline study also found that reading materials at home is a powerful predictor of emergent literacy scores. There was a strong association between the number of different reading materials available to a child and their total IDELA score. Based on the regression analysis, a child scored 1.4 percentage points higher for each additional type of reading material owned by the caregiver and the strongest association was in Emergent Literacy (Johnston, Namit, Seidan, & Reen, 2019).

At midline, students achieved similar scores for all text types, irrespective of intervention status. Overall, more students reported having storybooks at home than at baseline. Pleasingly, 64.5% of students now have access to storybooks for reading at home. Given the increases in both intervention and control, it is evident that access to reading materials is increasing irrespective of intervention.

Table 36: Reading books at home by type and intervention status

	Overall		Control		Intervention	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
School textbooks	1017	51.3%	437	53.8%	470	48.6%
Religious books/Bible	1799	90.9%	728	89.7%	883	91.2%
Newspapers	1150	58.0%	472	58.1%	543	56.1%
Storybooks for children	1280	64.5%	527	64.9%	614	63.4%
None of the above	57	2.9%	31	3.8%	19	2.0%

Mothers' literacy level

According to the RISE PNG IDELA baseline study, a mother's literacy level was a strong predictor of literacy and numeracy results in ECCE children (Johnston, Namit, Seidan, & Reen, 2019). A child whose mother was literate tended to score 3–4 percentage points higher in literacy and numeracy compared to a child with an illiterate mother.

During the baseline literacy and numeracy assessment (2018), more than 86% of children attested to having at least one literate adult in the household. At midline, a higher proportion (+1 percentage point) of intervention students reported that their mothers were literate compared to control students. This is not a surprising result given the intervention did not directly target mothers' literacy.

Table 37: Control and intervention students reported response to mother's literacy

	Control	Intervention	Control	Intervention	Control	Intervention
Mother can read and write	Frequency		Percentage		Cumulative	
No	237	271	31.1%	30.3%	31.1%	30.3%

Yes	526	624	68.9%	69.7%	100%	100%
Total:	**763**	**895**	**100%**	**100%**		

Based on the regression analysis presented in Appendix F and G, a mother who can read and write had a significant positive effect on children's literacy results in relation to letter identification, English word scores and English comprehension scores. A child with a literate mother scored 3.4 percentage points higher than a child with an illiterate mother. A similar variation (3.6 percentage points) was evident for the English word score with less variance (1.9 percentage points) in letter identification. However, there was a negative effect on skip patterns of five which appears a slight anomaly as the only other significant effect on numeracy was in number identification and this had a positive effect.

Home reading

In the RISE PNG IDELA study (2019), caregivers reported that 48% of mothers and 32% of fathers read to their child. Storytelling and teaching letters were more popular reading activities with over 70% of mothers and 50% of fathers reported to be engaged in these activities.

At baseline, 57% of the students stated that a member of the household had read a story to them in the past week. The baseline analysis found this result difficult to reconcile given the low reported prevalence of storybooks at home. One hypothesis to explain the result was that students may not distinguish between telling a story and reading a story. In Tok Pisin, the phrase for telling a story can be the equivalent of recounting an event or having a conversation. The language of 'stori' (sharing information) may have led to anomalous findings.

The findings vary when comparing intervention and control students' responses at midline. The table below shows more children from the control response sample reported that their mother reads to them (80.65%) compared to (76.49%) of intervention students. At baseline there was no difference between intervention and control. Overall, both response samples show a significant improvement on reading habits at home with a 20% increase in responses from intervention students and 23.5% increase in responses from control students.

Table 38: Mothers who read to their child at home as reported by response student sample

	Control	Intervention	Control	Intervention	Control	Intervention
Mother reads to the child	**Frequency**		**Percentage**		**Cumulative**	
No	101	146	19.4%	23.5%	19.4%	23.5%
Yes	421	475	80.7%	76.5%	100%	100%
Total:	**522**	**621**	**100%**	**100%**		

It does show that perhaps the caregiver workshops designed to help parents and caregivers engage in learning activities at home with their children may not have delivered the desired effect. Save the Children's community literacy volunteers (CLVs) reported that caregiver workshops were one of the most challenging of all the community literacy activities. In over 25% of reported cases, caregiver workshops had issues with administration, coordination with local leaders, or poor attendance. Attendance of female caregivers at the workshops was higher than male caregivers. The median number of women attending the caregiver workshop was seven, whereas the average number of men was only three. While 10 caregivers per session is reasonable, the original expectation was much higher.

Table 39: Caregiver workshop review provided by CLVs

CLV review of caregiver workshop	Frequency	Percentage	Cumulative
Really successful	22	20.2 %	20.2%
Successful	58	53.2%	73.4%
Some problems	28	25.7%	99.1%
Lots of problems	1	0.9%	100%
Total:	**109**	**100%**	

Interestingly, the regression analysis presented in Annex F and G shows a strong correlation between overall literacy and numeracy scores and the number of family members who read to the child at home. In all literacy subtests, there is a statistically significant positive effect on scores with the highest increases evident in Tok Pisin word and reading comprehension scores. The effect size is not large but for the literacy total score, children who are read to by family members achieve 1.26 percentage points higher than children who are not read to at home.

Similarly, the regression analysis on the effect of home reading (a child reading at home) shows a strong association with literacy and numeracy scores and interestingly in 100% of test items. The strongest associations are linked to number identification, skip patterns of five, addition and subtraction, shape recognition, English words and comprehension and Tok Pisin words. On balance, home reading has a strong effect on both literacy and numeracy scores and children can achieve 1–3.5 percentage points higher in literacy and numeracy subtests if they read at home. Interestingly, the effect size and significance rating are greater for home reading compared to caregivers reading to a child.

Table 40: Regression analysis showing the effect of home reading on E2 students' literacy and numeracy scores

	Literacy and numeracy	Literacy	Numeracy	Number identification	Skip pattern 2	Skip pattern 5	Addition	Subtraction
Student scores – percentage point difference	2.408***	2.518***	2.298***	1.542***	1.380**	2.950***	1.890***	2.479***

Word problems	Shape recognition	Letter identification	English words	English comprehension	Tok Pisin words	Tok Pisin comprehension
3.236***	2.610***	1.199***	3.273***	2.345***	3.257***	2.514**

Attendance at reading club

The findings reveal that the frequency of responses from children about their attendance at reading club in the community was higher in the intervention sample compared to the control sample. A 10-percentage point difference in the reported frequency is a positive response despite some of the difficulties experienced establishing reading clubs in the community.

Table 41: Control and intervention student responses on attendance at reading club

	Control	Intervention	Control	Intervention	Control	Intervention
Child has attended the reading club outside of school	Frequency		Percentage		Cumulative	
No	731	770	87.8%	77.9%	87.8%	77.9%

Yes	102	218	12.2%	22.1%	100%	100%
Total	**833**	**988**	**100%**	**100%**		

Based on their monitoring data, CLVs reported that most of the reading clubs were successful or very successful. Almost 17% of reading clubs faced some problems and these were most commonly associated with scheduling and location. Some reported disturbances in their community and land disputes, while others reported delays in the arrival of CLV kits, a shortage of stationery and learning materials, and challenges in understanding all CLV requirements. Many faced challenges in obtaining support from the community and parents. This issue seems to be a consistent theme emerging from community literacy activities.

Table 42: CLV data on evaluations of reading clubs

CLV review of reading club	Frequency	Percentage	Cumulative
Really successful	62	36.1%	36.1%
Successful	80	46.5%	82.6%
Some problems	29	16.9%	99.4%
Lots of problems	1	0.6%	100%
Total:	**172**	**100%**	

Despite the challenges, the reading club was the most popular community literacy activity. It was the most chosen activity by communities, followed by story time and caregiver workshops. According to feedback provided by CLVs, 92.5% of children were interested or very interested in the reading activities conducted at the community level. These results highlight children's desire to read and borrow books through avenues such as reading clubs.

Table 43: CLV feedback on elementary children's interest in reading activities in community

Elementary children's feelings about the book bank and reading activities	Frequency	Percentage	Cumulative
Very interested	102	49.5%	49.5%
Interested	89	43.2%	92.7%
Not very interested	14	6.8%	99.5%
Against the activities	1	0.5%	100%
Total:	**206**	**100%**	

Borrowing books from the community book bank

The practice of borrowing books from a community book bank or library is far higher in the intervention sample than the control sample. This suggests the injection of community literacy activities in communities in ARoB, ESP and EHP has provided access to book borrowing. Overall, almost 8% more children reported borrowing books from the intervention sample compared to the control sample. This result supports the hypothesis that children want reading books and have a desire to read.

Table 44: Control and intervention student responses on borrowing books from the community book bank

	Control	Intervention	Control	Intervention	Control	Intervention

Child has borrowed reading books from the community book bank	Frequency		Percentage		Cumulative	
No	739	800	88.9%	81.1%	88.9%	81.1%
Yes	92	186	11.1%	18.9%	100%	100%
Total:	831	986	100%	100%		

The results are significant when analysing the findings from the regression analysis. For children who had borrowed books from the community book bank, their literacy scores and overall literacy and numeracy score was substantially higher. The difference in percentage points between children who borrowed books and those that did not was as high as 12 percentage points and interestingly, the scores were highest in Tok Pisin. Children who borrowed reading books scored 11.63 percentage points higher in their Tok Pisin word score and 12.15 percentage points higher in their Tok Pisin comprehension score. In the reading clubs and the reading collection in the book banks, Tok Pisin and Tok Ples reading books are available for borrowing, but the number of Tok Pisin books available is much lower than the number in English. The results for children who borrowed books and their reading comprehension scores and word scores in English, both show a statistically significant improvement. The effect size for children who borrowed books and their English word score is 9.4 percentage points and for English comprehension, 7.6 percentage points higher than children who did not borrow books. The concept of reverse causality might apply here where children who are more likely to read and more likely to borrow books are the ones who would do better on a literacy and numeracy assessment in any case.

Table 45: Regression analysis showing significant effects on literacy scores for children who borrowed books from the community book banks

	Literacy and numeracy	Literacy	Letter identification	English words	English comprehension	Tok Pisin words	Tok Pisin comprehension
Student scores – percentage point difference	6.076***	8.697***	2.678***	9.433***	7.595**	11.63***	12.15***

Effect of the Bloom Reader app

The results show a low penetration of Bloom Reader overall. The results also show how Bloom Reader has been picked up by both intervention and control at similar levels. Being an open source app, and an initiative actively promoted throughout the RISE PNG program, it is expected that contamination will occur. Open source learning apps such as these are difficult to regulate. The proportion of parents and teachers who have access to mobile technology remains very low, so diffusion of technology is unlikely over the 12-month period.

Table 46: Response sample by intervention and control who read books from their parents' phone

	Control	Intervention	Control	Intervention	Control	Intervention
Child has read books from Bloom	Frequency		Percentage		Cumulative	

Reader on her/his parents' phone						
No	778	921	93.2%	93.4%	93.2%	93.4%
Yes	57	65	6.8%	6.6%	100.0%	100.0%
Total:	**835**	**986**	**100%**	**100%**		

A distinct problem that emerged from the monitoring data was the inaccessibility of the Bloom Reader app. Many parents did not have smartphones and/or the CLV and parents were not literate enough to use the app with an internet connection. Almost 35% of CLVs said they faced some problem when sharing the Bloom Reader app, mostly because parents did not own a smartphone. When asked what they needed the most help with, 54.37% of respondents said Bloom Reader.

Table 47: CLV data on results with sharing the Bloom Reader app with caregivers

Sharing the Bloom Reader app was:	Frequency	Percentage	Cumulative
Really successful	10	29.4%	29.4%
Successful	12	35.3%	64.7%
Some problems	11	32.4%	97.1%
Lots of problems	1	2.9%	100%
Total:	**34**	**100%**	

While only 7% of students reported using Bloom Reader, these students scored statistically higher on literacy compared to students who did not use the app. This was particularly evident in the English word score where children who had read from Bloom Reader scored, on average, 11.25% higher than children who had not used the app. There was a strong association with overall literacy with children scoring, on average, 7.76 percentage points higher in this domain than children who did not have access to the app. Access to Bloom Reader had a strong association with higher scores in Tok Pisin comprehension. It is likely that children's reading frequency has increased with a ready supply of culturally relevant local stories in languages with which they are familiar.

Table 48: Regression analysis displaying significant effect scores on literacy items for children who had read books from Bloom Reader

	Literacy and Numeracy	Literacy	English Words	English Comprehension	Tok Pisin Words	Tok Pisin Comprehension
Student scores – percentage point difference	4.759**	7.758***	11.25***	6.232*	6.640*	13.97**

Children's diet – breakfast

At baseline, children who had breakfast on the day of the assessment performed better across all literacy and numeracy skill areas. It was found that a child who had breakfast in the morning was able to identify three additional letters of the alphabet compared to a child who hadn't eaten. Their comprehension scores (in English and Tok Pisin) and numeracy scores were also higher than the children who hadn't had breakfast. Of course, this does not mean that providing breakfast is a singular answer to improving literacy and numeracy. Omitted variable bias does not allow direct attribution of higher assessment scores to eating breakfast as children who

regularly have breakfast are also more likely to come from families where there are stronger routines, greater investment in education, more literacy rich households and caregivers are able to afford a breakfast for their children.

It does however highlight the relationship between nutrition and learning and is an issue that RISE PNG is closely monitoring. At baseline, 88.54% of children had breakfast on the day of the assessment – 88.26% of boys and 88.81% of girls. There was no statistically significant difference between the two. At midline, there was a marginal difference between children who ate breakfast on the day of the assessment from the control and intervention sample.

Table 49: Children who reported they ate breakfast on the day of the assessment by control and intervention status

	Control	Intervention	Control	Intervention	Control	Intervention
Child ate breakfast that morning	Frequency		Percentage		Cumulative	
No	101	106	12.0%	10.7%	12.0%	10.7%
Yes	738	882	88.0%	89.3%	100%	100%
Total:	**839**	**988**	**100%**	**100%**		

Children's diet – betel nut consumption

At baseline, betel nut consumption was not assessed comprehensively but during the validation process, it was proposed as a possible factor to explain results. In the later RISE PNG IDELA baseline study, a question to caregivers was included in the study. A key finding from this study showed that three in four caregivers reported chewing betel nut and 22% of caregivers reported that their child studying in ECCE also chewed betel nut. Even three and four year olds were reported by their caregivers to be consuming betel nut. It was also found that girls were more likely to chew betel nut ($t = 3.0$, statistically significant at the 1% level of significance) than boys. Mothers/female carers were also more likely to chew than fathers/male carers.

In an alarming finding, there is a strong negative association between the consumption of betel nut and early learning and development scores. Controlling for all other relevant factors (age, socio-economic condition, province of residence etc.), children who consumed betel nut scored 5.5 percentage points lower than similar children who did not. The strongest associations were with motor development and emergent literacy, where the average child who chewed betel nut scored 8.7 percentage points and 5.2 percentage points lower than other children.

To explore this issue further in elementary, a question was added to the midline assessment.

Table 50: Frequency and percentage of children in intervention and control schools who chew betel nut

	Control	Intervention	Control	Intervention	Control	Intervention
Child chews betel nut	Frequency		Percentage		Cumulative	
No	469	647	56.0%	65.4%	56.%	65.4%
Yes	369	342	44.0%	34.6%	100%	100%
Total:	**838**	**989**	**100%**	**100%**		

The consumption of betel nut appears to increase as children become older. By E2, 34.6% of children in intervention schools and 44% of children in control schools chew betel nut. The reduction in consumption between intervention and control may be the result of efforts undertaken by Save the Children to raise issues concerning betel nut chewing in young children through Information, Education and Communication (IEC)

messaging and during caregiver workshops. It appears there is a 10 percent reduction in children who are chewers in intervention schools but this could be coincidental rather than by design and further study is needed to determine whether the intervention has truly had this positive effect in 12 months.

Table 51: Frequency and percentage of females and males who chew betel nut

	Females	Males	Females	Males	Females	Males
Child chews betel nut	Frequency		Percentage		Cumulative	
No	591	633	56.9%	63.6%	56.9%	63.6%
Yes	447	363	43.1%	36.5%	100%	100%
Total:	**1038**	**996**	**100%**	**100%**		

As was found in RISE PNG IDELA baseline study, girls are more likely to chew than boys with 43% of E2 girls reporting to chew compared to 36.5% of E2 boys. Validation is needed to help explain this result. Based on anecdotal sources, betel nut is harvested by children as they are often the only ones with the ability to climb the trees. This activity is often done after school. There may be a gender dimension to harvesting of betel nut which needs to be explored further through community consultation. Parents speak about giving betel nut as a treat after children help with harvesting and this may be one reason for its prevalence in some communities.

Children's socio-economic status

The finding suggests that the socio-economic status[20] of families in PNG has a significant effect on literacy scores, in particular those associated with higher levels of literacy acquisition such as comprehension tasks. The literacy skill required to solve mathematics problems also show a strong association and indicate a correlation between socio-economic status and higher order thinking. The effect size was largest for English comprehension. Students who were from a higher socio-economic status tended to score almost 3 percentage points higher in the English comprehension score than students from low socio-economic status.

Table 52: Regression analysis showing the effect of a child's socio-economic status on literacy and numeracy scores

	Literacy and numeracy	Literacy	Word problems	English words	English comprehension
Student scores – percentage point difference	1.289**	1.882**	1.676**	2.559**	2.906***

Children would have greater exposure to print, verbal and written text in households with electricity and basic communication technologies such as a television and mobile phone. It is likely they would have more media to practice their reading and develop their communication skills and this might be influencing their results in literacy, helping them achieve higher scores than their counterparts who do not have access to such facilities.

[20] Socioeconomic status was calculated by taking household assets into account. Access to electricity, a television set at home, refrigerator, car/canoe, generator and mobile phone were binary variables included in the calculation.

5.7.5 What variables in the school (e.g. corporal punishment, teacher absenteeism, teaching competencies in literacy and mathematics) are influencing children's acquisition of literacy and numeracy skills?

Children's participation in Early Childhood Care and Education

At baseline, children who had ECCE experience performed significantly better than children who did not have an ECCE experience across all three provinces. On average, a child with ECCE experience is likely to identify nearly two additional letters of the alphabet compared to a child without ECCE experience. Similar differences were seen on almost all assessment tasks.

The findings at midline are comparable to findings from baseline. When comparing the control and intervention sample, it seems that a higher percentage of children (+7%) reported attending ECCE prior to elementary in the control response sample. At baseline this was balanced so it is surprising that such a difference is found at midline. This issue needs looking into during validation and may be enumerator error.

Table 53: The number and percentage of children from control and intervention schools that report to have attended ECCE

	Control	Intervention	Control	Intervention	Control	Intervention
Child has attended ECCE	Frequency		Percentage		Cumulative	
No	437	591	52.1%	59.8%	52.09%	59.8%
Yes	402	398	47.9%	40.2%	100%	100%
Total:	**839**	**989**	**100%**	**100%**		

There was a negligible difference between boys and girls with a slightly higher percentage of boys reporting their participation in ECCE compared to girls.

Table 54: The number and percentage of female and male students that report to have attended ECCE

	Females	Males	Females	Males	Females	Males
Child has attended ECCE	Frequency		Percentage		Cumulative	
No	591	543	56.9%	54.5%	56.9%	54.5%
Yes	447	454	43.1%	45.5%	100%	100%
Total:	**1038**	**997**	**100%**	**100%**		

Based on the regression analysis presented in the tables in Annex F and G, children who attended ECCE are performing significantly higher on many of the literacy subtests, as well as a number of the numeracy subtests, than children who did not attend ECCE. The largest effects were seen in word and reading comprehension scores in English and Tok Pisin. Children who had attended ECCE scored almost 8 percentage points higher in English comprehension and almost 9 percentage points higher in Tok Pisin than children who had not attended ECCE. In all numeracy subtests – with the exception of skip patterns of two and five – there was a statistically significant effect. The result affirms the findings of the baseline assessment.

Table 55: Regression analysis showing the effect of a child's attendance at ECCE on literacy and numeracy scores

	Literacy and Numeracy	Literacy	Numeracy	Number Identification	Addition Score	Subtraction
Student scores –	4.627***	7.113***	2.141**	1.634*	3.535**	3.653**

percentage point difference						

Word Problems	Shape Recognition	Letter Identification	English Words	English Comprehension	Tok Pisin Words	Tok Pisin Comprehension
5.369***	2.284*	2.213***	7.654***	7.983***	8.729***	8.988***

These findings are consistent with the mounting evidence demonstrating the importance of early childhood development programs in shaping cognitive, social and language skills, as well as lifelong approaches to learning (Pisani, Borisova, & Dowd, 2015).

Children's early learning and development in formalised ECCE or ECCD programs has been shown to influence future academic performance. Children who begin school with weak prior knowledge and skills in relevant emergent literacy domains, most notably general verbal abilities, basic phonological awareness, familiarity with the basic purposes and mechanisms of reading, and letter knowledge, are more likely to experience difficulty with higher order thinking and reading comprehension tasks in the primary grades (Snow, Burns, & Griffin, 1998). Exposure to ECCE has also been shown to have a positive impact on other education indicators such as lower repetition and dropout rates, higher enrolment rates, better long-term cognitive development, and greater educational attainment. Reducing the number of children who enter school with inadequate early learning experiences is an important step toward preventing school dropout and later learning difficulties and enabling long-term school success (Engle et al., 2011). Improving children's readiness for school will help them take advantage of their right to education.

Children's absenteeism

The baseline assessment intensely examined student absenteeism to identify the difference between persistent absenteeism (students who are listed in the register but never attend) and transient absenteeism (students who attend school but are frequently absent). A number of techniques were used to gather data on absenteeism, as illustrated in Table 77.

Table 56: Absenteeism rate at elementary schools captured at baseline

Province	Absenteeism rate (enrolled in school – present in school)	Second measure absentee rate (enrolled in E1 class – present in E1 class)
EHP	46%	43%
ARoB	52%	29%
ESP	52%	30%

All three provinces displayed a high rate of student absenteeism (between 46% and 52%). Teachers reported lower classroom enrolment rates compared to the school enrolments reported by Teachers in Charge (TICs). This is consistent with the concept of persistent absenteeism, discussed in the following paragraph. Self-reported student absenteeism (second measure of absenteeism) was in line with the data provided by teachers.

These findings lead us to surmise that there may be two kinds of student absenteeism in PNG – persistent absenteeism and transient absenteeism. Persistent absenteeism would mean the student is technically enrolled in school but never attends. The school might be collecting Tuition Free Fee (TFF) funding for that child (and would count that child as enrolled in school) but s/he is not enrolled in any particular teacher's class. If the

teacher does not have an attendance register, it would be difficult for the teacher to cross-check attendance with enrolment figures held by the TIC.

For the purpose of this assessment, the research team has measured transient absenteeism using records from the teacher. The teacher records all students' names who are present or absent on a given day and determines attendance ratios based on the total enrolment in the classroom, irrespective of school enrolment figures. This hypothesis is supported by the fact that the total enrolment recorded by the lesson observation is only 32 students per class (whereas EMIS records suggest 39 or 40 at the E2 level for the three provinces that were studied).

At midline, the results for attendance had not improved. The rates in both response samples were higher than at baseline if we compare student reported absenteeism. There was a marginal difference between intervention and control students.

Absenteeism appears to reduce with each level. EP reported the highest absenteeism followed by E1 and then E2. For example, only 46% of girls were present at the EP level but 63% of girls were present at the E2 level. This trend may be linked to age of the children and the reasons for absenteeism described below. EP children are younger and therefore more likely than older, more independent children, to accompany their parents on trips to town, the market and relatives.

Table 57: The number and percentage of children from control and intervention schools that reported being absent in the previous week

	Control	Intervention	Control	Intervention	Control	Intervention
Child has missed a day of school in the last week	Frequency		Percentage		Cumulative	
No	443	509	52.8%	51.5%	52.8%	51.5%
Yes	396	480	47.2%	48.5%	100%	100%
Total:	**839**	**989**	**100%**	**100%**		

When comparing boys and girls, there is a slightly higher absenteeism rate among boys. Further validation is required to explore the reasons for this variance although there was no statistical difference between boys' and girls' absenteeism.

Table 58: The number and percentage of boys and girls who reported being absent in the previous week

	Female	Male	Female	Male	Female	Male
Child has missed a day of school in the last week	Frequency		Percentage		Cumulative	
No	559	524	53.9%	52.6%	53.9%	52.6%
Yes	479	473	46.2%	47.4%	100%	100%
Total:	**1038**	**997**	**100%**	**100%**		

Absenteeism was also studied at the school level based on manual student counts in 132 randomly selected schools. From these data, there is a higher absenteeism rate from students in the lower grades compared to students in E2. Girls' attendance is higher than boys, which compares with the student reported data (above).

Table 59: Manual student count of attendance in 132 randomly selected schools at midline

	Ratio	Std. Err.	[95% Conf.	Interval]

Male teachers present	0.64	0.04	0.57	0.71
Female teachers present	0.72	0.04	0.65	0.79
Overall students present	0.52	0.02	0.48	0.57
Overall students absent	**0.48**	**0.02**	**0.43**	**0.52**
Boys (EP) present	0.44	0.03	0.38	0.49
Girls (EP) present	0.47	0.03	0.41	0.53
Boys (E1) present	0.51	0.03	0.45	0.57
Girls (E1) present	0.54	0.03	0.48	0.60
Boys (E2) present	0.60	0.03	0.54	0.66
Girls (E2) present	0.63	0.03	0.58	0.69

It is also evident that male teachers are more likely to be absent than female teachers. The gender dimension to absenteeism of both teachers and students requires further investigation. The table below highlights the frequency of absenteeism and alarmingly, over 31% of children were absent for five days in the previous week. The frequency of absenteeism suggests that very few children are receiving a full complement of the curriculum or the benefit of their teachers' new skills in literacy and numeracy. The stubbornness of shifts in literacy and numeracy results, particularly when we study school factors a little later in this report, highlight this concerning matter.

Table 60: Number of days missed in the last week as reported by E2 students

No. of days missed by the child in the last week	Frequency	Percentage	Cumulative
1	299	29.8%	29.6%
2	201	19.9%	49.5%
3	125	12.4%	61.8%
4	68	6.7%	68.6%
5	318	31.5%	100%
Total:	**1011**	**100**	

Exploring the reasons provided by students for their absenteeism may help identify what factors are strongly influencing children's attendance patterns in PNG.

Illness is the most regularly reported reason for absenteeism. Even when analysing the 'other reasons', which included 327 responses, illness and injury were commonly reported. The second highest reason was teacher absence. The detailed data set on 'other reasons' is revealing on this subject. Students reported a list of factors relating to their teacher that influenced their attendance: 'Teacher died and they went to attend funeral', 'Teacher said to stay at home so child was absent for a week', 'Teacher told them to stay home whole of last week. Because he will go look for food', 'Teacher went to town to purchase presents for graduation.'

Many reasons were reported to explain teacher absenteeism. Over half were unrelated to school or teaching responsibilities and were to do with the need to find food, do other paid work, go to town for supplies etc. It is also evident that school closures and suspensions are very common. From estimates in 'other responses', 35 students reported ad hoc school closures and suspensions as reasons why they did not attend school. Other school related reasons for absence were linked to activities such as in-service teacher training, graduations, Year 8 examinations and the Asia-Pacific Economic Cooperation (APEC) meeting. Alongside these school related reasons, there were many reasons linked to parental decisions on how the child spends their time. Many

children (approximately 12 responses) reported the need to attend church activities and services or prepare for Sabbath. Less than this, reported attendance at funerals and cultural events as reasons for absenteeism. By far the most common response in 'other responses' was that 'children just did not want to go and either remarked that they were bored, lazy or didn't like school'. From estimates, this reason accounted for 60 responses. Additionally, there was an equal number of responses from children who reported that they accompanied their father or mother into town, the gardens, and to visit relatives. Dirty clothes were commonly reported as a reason for absenteeism. This highlights not only the influence of parents on school attendance but the lack of commitment by parents to send their children to school. The lack of parental commitment is strongly influencing high absenteeism which is strongly impacting elementary children's learning.

Table 61: Reasons for missing school reported by E2 students

Reason for missing school			
Observations	Frequency	Percentage of responses	Percent of cases
Child was needed to work at home	82	7.41%	8.18%
Child was needed to work (for wages)	8	0.72%	0.80%
Child was sick	**343**	**30.98%**	**34.20%**
School is too far from home	12	1.08%	1.20%
Teacher was absent	97	8.76%	9.67%
No food	74	6.68%	7.38%
Weather (rain, flood, etc.)	45	4.07%	4.49%
Death in the family	77	6.96%	7.68%
Violence/clashes	9	0.81%	0.90%
Problems with school uniform/clothing	29	2.62%	2.89%
Bullying	4	0.36%	0.40%
Other reason	**327**	**29.54%**	**32.60%**
Total	1107	100%	110.37%
Valid cases:	1003		

The regression analysis highlights the negative effect of student absenteeism on literacy and numeracy scores, particularly for test items that require higher order thinking. Overall, students who had missed at least one day of school in the previous week scored 4.5 percentage points lower in the aggregate score on literacy than students who attended five days in that same week.

The largest effects are found in English word identification (-5.6 percentage points) and Tok Pisin comprehension (-6.6 percentage points). Students with poor performance also score worse in English comprehension than other students. For numeracy, children who were absent scored 4.2 percentage points lower in addition and 6.5 percentage points lower in subtraction scores.

Without due attention on this issue, many children will continue to be below expected standards in English and Mathematics, irrespective of efforts to improve teachers and provide quality teaching and learning materials.

Table 62: Regression analysis showing the relationship between absenteeism and literacy and numeracy scores

	Literacy and Numeracy	Literacy	Numeracy	Number Identification	Skip Pattern 2	Skip Pattern 5	Addition Score	Subtraction
Student scores – percentage point	-3.455***	-4.519***	-2.391***	-0.701	-0.519	0.604	-4.269***	-6.450***

difference								

Word Problems	Shape Recognition	Letter Identification	English Words	English Comprehension	Tok Pisin Words	Tok Pisin Comprehension
-3.988**	-1.415	-2.045***	-5.601***	-4.553***	-3.804*	-6.592***

Child repeated the current grade

The baseline assessment reported a high incidence of grade repetition in elementary and analysed the effect that repetition had on literacy and numeracy scores. It found that a third of the students in E1 were repeaters. The repetition rate was highest in EHP where nearly 40% of the children reported to have repeated the current grade. There were no gender differences observed in repetition rates in the baseline assessment. ESP had the lowest repetition rate of 30.83% followed by AROB with 33.06%.

At midline, the repetition rate of E2 students is lower than the level reported by children in E1. In intervention schools, a slightly higher percentage of children reported to have repeated the grade compared to control. There was a 6–7% difference between grade levels.

Table 63: The number and percentage of children from control and intervention schools that reported having repeated their current grade

	Control	Intervention	Control	Intervention	Control	Intervention
Child repeated current grade	Frequency		Percentage		Cumulative	
No	617	714	73.54%	72.19%	73.54%	72.19%
Yes	222	275	26.46%	27.81%	100%	100%
Total:	**839**	**989**	**100%**	**100%**		

Similar to findings at baseline, there was no significant gender differences in relation to repetition. Approximately one quarter of children in E2 had repeated, irrespective of gender.

Table 64: The number and percentage of boys and girls who reported having repeated their current grade

	Female	Male	Female	Male	Female	Male
Child repeated current grade	Frequency		Percentage		Cumulative	
No	776	732	74.76%	73.42%	74.76%	73.42%
Yes	262	265	25.24%	26.58%	100%	100%
Total:	**1038**	**997**	**100%**	**100%**		

Repetition has a statistically significant negative effect on children's literacy and numeracy scores for both intervention and control students. The literacy skills most affected by repetition are English and Tok Pisin word scores and comprehension. Numeracy skills are also impacted with skip patterns and number identification showing the greatest negative effect. Students who have repeated score 6.5 percentage points lower in literacy and 2.3 percentage points lower in numeracy compared to students who have not repeated their current grade. This is an important finding given the high prevalence of repetition in PNG and the belief that this is a good strategy for children's learning. The results of this study negate this belief. It is also important to highlight the probability of reverse causality in relation to repetition. This infers that children who score poorly tend to repeat and thus repetition becomes correlated with poor scores.

Table 65: Regression analysis showing the relationship between repetition and literacy and numeracy scores

	Literacy and Numeracy	Literacy	Numeracy	Number Identification	Skip Pattern 2	Skip Pattern 5
Student scores – percentage point difference	-4.413***	-6.480***	-2.346**	-1.708*	-3.133**	-3.427*

Word Problems	Shape Recognition	Letter Identification	English Words	English Comprehension	Tok Pisin Words	Tok Pisin Comprehension
-3.486*	-2.529*	-1.517*	-8.130***	-4.571**	-8.565***	-9.620***

5.7.6 What proficiency gains in teachers' language and numeracy teaching are evident at baseline and endline?

Overall there were statistically significant proficiency gains in four of the five competency domains (reading comprehension, assessment, classroom management and student engagement) measured during lesson observation. Large gains were evident in reading comprehension where teachers who had received LB and NB training scored, on average, 31.65 percentage points higher in intervention schools in teaching reading comprehension compared to teachers from control schools. Large gains were also evident in the areas of classroom management and student engagement, both integrated topics throughout the teacher training cycles.

Table 66: Overall results in teacher competency domains by intervention status

	(1)	(2)	(3)	(4)	(5)
	Letters vocabulary	Reading comprehension	Assessment	Classroom management	Student engagement
Control scores	25.1%***	35.3%***	15.0%***	39.9%***	34.2%***
Intervention percentage point difference	8.2	31.7**	10.0*	19.9***	17.2***
Observations	99	99	99	99	99

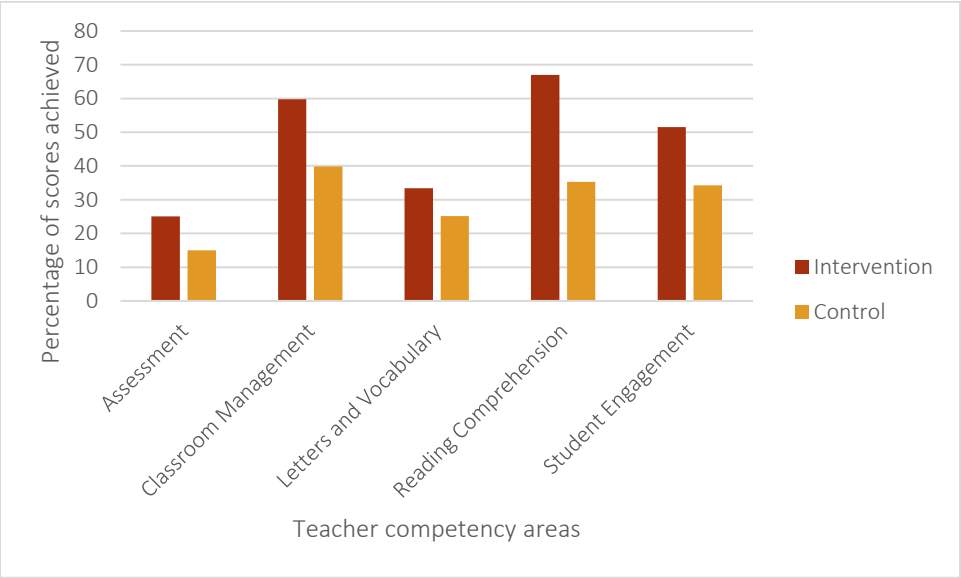
t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Overall, teachers in intervention schools scored very high (66.94%) in teaching reading comprehension compared to teachers in control schools. These results are testament to the efforts, undertaken during training, to emphasise the use of collaborative learning approaches to teach reading. The strong focus on different reading strategies and questioning shows a marked improvement as these were items assessed in this domain.

Teachers in intervention schools also scored very high in classroom management (59.81%) and student engagement (51.48%), which may have been influenced by the teachers' application of more collaborative and enquiry methods of teaching literacy. As a result of teachers adjusting their teaching methods and introducing more interesting ways to teach reading, student engagement appears to have increased. This may be influencing classroom management. There is a strong correlation between students' interest and motivation to learn and students' behaviour. Motivation problems and behavior management problems are often intertwined – unmotivated students frequently misbehave and students who misbehave frequently do not care about learning (Curwin, 2010). Studies have shown that positive teacher-student relationships can lead to a warm

classroom environment, which facilitates successful adaptation in school and thereby increases student motivation to learn (Koca, 2016).

Figure 8: Average percentage score in teacher competency domain by intervention status



Provincial variation is evident and may point to differences in the quality and effectiveness of professional development programs administered by RISE PNG provincial teacher training teams. In relation to ARoB, statistically significant improvements in classroom management and student engagement are evident. Improvements are also evident in relation to reading comprehension, but these are not significant.

Table 67: Results of teacher competency domains by intervention status in ARoB

	(1)	(2)	(3)	(4)	(5)
	Letters and vocabulary	Reading comprehension	Assessment	Classroom management	Student engagement
Control scores	34.2***	52.7*	21.2***	31.1***	39.1***
Intervention percentage point difference	7.4	31.7	9.3	28.3**	17.4*
Observations	35	12	35	35	35

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Similarly, in EHP, teachers have performed well in engaging students actively in classroom activities. This improvement is reflected in all provinces. Teachers in EHP have demonstrated their skills in teaching reading comprehension more readily than in ARoB and teachers' improvement in this area is statistically significant.

Table 68: Results of teacher competency domains by intervention status in EHP

	(1)	(2)	(3)	(4)	(5)
	Letters and vocabulary	Reading comprehension	Assessment	Classroom management	Student engagement
Control	17.4%**	36.4%**	11.5%***	52.7%***	36.8%***

Intervention percentage point difference	6.1	31.8*	5.5	8.0	14.6*
Observations	40	13	40	40	40

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Perhaps the most significant improvement has been in ESP where there have been statistically significant improvements in four competency domains including reading comprehension, assessment, classroom management and student engagement. Scores in all areas are generally higher than in other provinces. This does suggest that the training in ESP has been more effective in strengthening teachers' skills, particularly in core competency areas such as teaching reading and classroom management.

Table 69: Results of teacher competency domains by intervention status in ESP

	(1)	(2)	(3)	(4)	(5)
	Letters and vocabulary	Reading comprehension	Assessment	Classroom management	Student engagement
Control scores	22.2%*	22.1%*	10.3%	34.6%***	22.4%***
Intervention percentage point difference	17.8	26.0*	21.8*	24.3*	22.7*
Observations	24	14	24	24	24

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of lesson observation highlight the shifts in teachers' practices as a result of training and the emphasis on collaborative literacy strategies. In Table 91, the mastery of decoding is achieved through a mix of strategies that encourage a relationship between sound and sight to be established. In the control teachers, the typical strategy is sounding out letters in a word. This is the dominant strategy in control schools. In intervention schools, however, a blend of strategies is used, and phonic matching activities are evident in almost one quarter of classrooms. The expansion of this strategy will be supporting children's phonemic understanding.

Table 70: Lesson observations on decoding strategies used by intervention status

Decoding strategy	Control	Intervention
Sounding out letters of a word c-a-t = cat	76.5%	47.6%
Completing a phonics exercise matching similar sounding words together	5.9%	23.8%
Completing a written spelling exercise	17.6%	19.1%
Other	0%	9.5%
Total:	100%	100%

In Table 92, the introduction to paired reading is evident in the intervention classrooms. While not a high percentage of intervention classrooms, it is nonetheless evident and shows teachers are gradually introducing different strategies to encourage children to practice their reading fluency. Choral reading and shared reading

with the teacher remain the dominant strategies used in both intervention and control and highlights the need to keep pushing for alternative ways to teach fluency to enhance this literacy skill among students.

Table 71: Lesson observations on strategies used by teachers to encourage students to practice reading fluency by intervention status

Reading fluency teaching strategy	Control	Intervention
Students read orally with the teacher	66.7%	68.8%
Students did choral reading	33.3%	25%
Students did paired reading	0%	6.3%
Total:	100%	100%

The teachers in the intervention schools have increased their practice of questioning and report back. As shown in Table 93, 46.15% of lesson observations in intervention schools included questions and report back compared to only 8.33% of lesson observations in control schools. This is strongly promoted during LB training and is a key strategy for teaching reading with understanding, as well as developing students speaking and listening skills. The variety of strategies used, such as role play, group discussion and question and report back, will enable students to reinforce their learning.

Table 72: Lesson observations on activities used by teachers to help students listen and speak by intervention status

Activities done to help students listen and speak	Control	Intervention
Role play	33.3%	15.4%
Group discussion	25.00%	30.8%
Questions and report back	8.3%	46.2%
Other	33.3%	7.7%
Total:	100%	100%

There is a big shift in how new words are visually presented as part of a teacher's reinforcement of new learning. In control schools, 60% of classrooms visually presented new words on the blackboard or poster paper compared to 83% of classrooms in intervention schools. This is an important strategy for building vocabulary and reinforcing learning. Table 94 shows this shift in practice.

Table 73: Lesson observations on how new words were introduced to students by intervention status

A list of new words or spelling words is written on the board or poster paper	Control	Intervention
No	39.1%	17.0%
Yes	60.9%	83.0%
Total:	100%	100%

The following tables highlight some of the changes evident in intervention teachers' practice of reading stories. It presents a story of change in a teachers' ability to introduce elements of the story before reading begins so that early story meaning can take place and students can engage more meaningfully with the story once reading begins. It highlights the shift in practice and provides evidence that this shift can occur within a relatively short period of time. As shown in Table 95, in 77.27% of lesson observations in intervention schools, the teacher

introduced key words and concepts that build on the story elements before reading the story. This compared to only 29% of lesson observations in control schools.

Table 74: Lesson observations on how the teacher introduces key words and concepts before reading by intervention status

Before reading the story, the teacher introduces key words/concepts	Control	Intervention
No	70.6%	22.7%
Yes	29.4%	77.3%
Total:	**100%**	**100%**

Additionally, teachers in intervention schools are more likely to ask predictive questions before reading the story than teachers in control schools. In over 60% of classroom observations, teachers in intervention schools were observed asking predictive questions, which encourages students to think critically about the story plot and features. This compares to less than one quarter of classrooms in control schools engaging students in this way.

Table 75: Lesson observations on how the teacher asks predictive questions before reading the story by intervention status

Before reading the story, the teacher asks predictive questions	Control	Intervention
No	76.5%	36.4%
Yes	23.5%	63.6%
Total:	**100%**	**100%**

Similarly, a high proportion of intervention teachers are spending time asking questions during reading as a way of assessing students' understanding and as a strategy to engage students in the richness and complexity of the story so more meaning can be garnered. This is a familiar strategy reinforced during trainings and is fundamental to reading with comprehension. It also enables all students to follow the story more closely and check their understanding as the story is read.

Table 76: Lesson observations on how the teachers asks basic factual questions about the text during reading by intervention status

During the reading, the teacher asks basic factual questions about the text	Control	Intervention
No	70.6%	18.2%
Yes	29.4%	81.8%
Total:	**100**	

The improvements in teacher practice in reading is evident in the way teachers introduce the story and ask questions before, during and after the story. Table 98 shows the results of lesson observations of teachers asking inference questions after the story is read to encourage students to think critically about the meaning and reinforce their understanding of key message or the moral of the story. In 77% of classroom observations, intervention teachers asked inference questions about the text after reading the story compared to only 41% of classroom observations in control schools.

Table 77: Lesson observations of the teachers practice of asking inference questions after the story by intervention status

After reading the story, the teacher asks inference questions about the text	Control	Intervention
No	58.8%	22.7%
Yes	41.2%	77.3%
Total:	100%	100%

Positive discipline, while not a discrete topic of teacher training, has been integrated into the activities and strategies taught during training to reduce abuse through corporal punishment and to promote child protection. Key messages linked to holistic classroom management have been reinforced and there are positive signs that teachers in intervention schools are using positive discipline more frequently than teachers in control schools. As shown in Table 99, there was a difference of 22 percentage points between teachers in intervention schools observed using positive discipline techniques most of the time or all of the time and teachers in control schools.

Table 78: Lesson observations of teachers use of positive discipline by intervention status

Teacher frequently uses positive discipline techniques	Control	Intervention
All the time	13.0%	13.2
Most of the time	28.3%	50.94%
Sometimes	28.3%	15.1%
Rarely	8.7%	5.7%
Not observed	21.7%	15.1%
Total:	100%	100%

Teacher interaction with students also appears to have improved as a result of the intervention, with a higher proportion of teachers interacting with students in a respectful manner all the time. The frequency in which teachers interact in a respectful manner irrespective of gender, grade, age or language is significantly higher than is observed in control schools. Once again, while not a discrete topic in teacher training, the reinforcement of messages on the importance of positive teacher-student relations and the relationship to learning outcomes appears to have informed practice change.

Table 79: Lesson observations of teacher's interaction with students by intervention status

Teacher interacts with all students irrespective of gender, grade, age, language	Control	Intervention
All the time	26.1%	45.3%
Most of the time	13.0%	33.9%
Sometimes	30.4%	13.2%
Rarely	2.17%	3.8%
Not observed	28.3%	3.8%
Total:	100%	100%

6 CONCLUSION

Overall, students who were assessed at baseline achieved marked improvements in literacy and numeracy subtest scores at midline. LB and NB appear to be making a statistically significant difference to girls' learning, to students in ARoB and ESP and to the quality of literacy and numeracy teaching delivered in elementary schools in AROB, ESP and EHP.

There is a clear pattern across all three provinces: *RISE PNG is currently having a disproportionately positive impact on girls compared to boys.* In no province do we see the intervention have an effect on one gender alone nor do we see either boys or girls performing worse in intervention schools compared to control schools. In ARoB, there are statistically significant improvements in numeracy overall and statistically significant improvements in numeracy for girls. Improvements were evident in literacy, but these were not statistically significant. In EHP, there appears to be no impact of Save the Children's intervention at this stage, despite intervention students scoring higher in literacy and numeracy compared to control students. Generally, in ESP, there is statistically significant improvements in literacy overall and statistically significant improvements in literacy for girls. This points to provincial variations in the effectiveness of the intervention on teachers and students.

When analysed against improvements in teaching competency, a parallel can be drawn between the effect of Save the Children's training – and resultant measures of teaching competency – and student learning. ESP presents a strong correlation between teaching competency in reading comprehension (ESP teachers performed well in this area) and the statistically significant improvements in children's literacy scores overall and with girls in particular. It seems the intervention has shown the strongest effect in ESP at this point of the implementation cycle, followed by ARoB and EHP.

Irrespective of intervention status, there are many factors that are influencing children's learning gains in literacy and numeracy over time. The regression analysis presents a range of variables that are influencing child's performance in English and mathematics. This makes intuitive sense and there is extensive evidence to support these findings. The results also support many initial findings at baseline and appear as consistent variables for PNG.

The variables showing the greatest influence on elementary students' English and mathematics scores are age, gender, grade level, grade repetition, ECCE attendance, mothers' literacy levels, home reading, reading books from Bloom Reader, socio-economic status, borrowing books from the reading club, chewing betel nut and enjoyment of school.

Grade level has been included in the model because some of the children that were assessed at baseline were held back and still in E1 when the midline study was undertaken. Those 91 children performed poorly in the assessment and on average, children at the E2 level scored 15 percentage points higher on the assessment compared to the children in E1. Repeaters performed poorly but once we included grade into the model, most of the variation was absorbed by that variable and so only for literacy do we see a statistically significant negative effect of 3 percentage points (in addition to the grade effect) for students who have repeated.

Children who had attended ECCE performed, on average, much higher in literacy and numeracy subtests. This is consistent with global evidence of the significant benefits ECCE has on children's readiness for learning in elementary. The socio-economic status of families, as well as a mothers' literacy, are important predictors of children literacy and numeracy results and the effect size is considerable for children's literacy scores.

The home learning environment is a powerful influence on learning outcomes for elementary children. The number of books that a child has exposure to at home and the frequency in which they read outside the

classroom are especially important. Essentially, more books and more reading at home improves literacy and numeracy performance in the classroom. Both variables are highly statistically significant, as are associated variables such as borrowing books from the community book bank and reading books from Bloom Reader. While the penetration of reading clubs and community literacy activities has only started, the evidence of its contribution to children's literacy and numeracy results in the classroom is promising.

Similar to the finding in RISE PNG IDELA baseline study, chewing betel nut is highly statistically significant. Children who chew betel nut scored almost 2.5 percentage points lower on overall literacy and numeracy scores and 3.5 percentage points lower on literacy alone. The gender dimension of betel nut consumption warrants further investigation particularly given the overall results that show girls' performance is lower than boys.

Attendance patterns are as poor as they were at baseline and poor attendance revealed a negative and statistically significant impact. One theory is that parents are not pushing or encouraging children to attend school every day. When all the reasons are analysed, the majority of the 'other responses' were because the child did not want to go to school. While the illness category could be a legitimate reason for not attending, it is likely that children are expressing a reluctance and parents are too willing to let them stay at home.

The model took ARoB as the base and we find no difference between children in ARoB and children in EHP. However, children in ESP, on average, perform better than the children in ARoB by 6 percentage points (holding all the other variables in our regression like age etc. constant).

7 STUDY RECOMMENDATIONS

1. Share results showing the importance of home reading and book borrowing with teachers, Boards of Management (BOMs), provincial and district education officers and community members so children are encouraged to read frequently at home and in the classroom. Use every opportunity during training and coordination meetings to emphasise and prioritise reading practices among children. Organise community meetings to share the results pictorially that summarises the main findings.
2. Undertake further consultation with community and education stakeholders to explore the practice of betel nut consumption in children to identify possible explanations for the prevalence and gender dimension. Utilise this same mechanism to work with parents on the attitudinal shifts required to encourage more regular school attendance.
3. Continue to emphasise the community literacy component of the program, which is showing promising results and supporting children to read more.
4. Continue to promote Bloom Reader, given the positive results from the sample of students regularly reading from the app.
5. In the Term 3 refresher training, revisit all reading strategies and continue to practice collaborative learning strategies.
6. In the Term 3 refresher training, revisit addition and subtraction strategies – particularly for double-digit numbers – to ensure teachers have a good grasp of teaching strategies that support children's skills in these areas.
7. Undertake provincial level workshops to gain a deeper understanding of the gender and provincial dimensions arising from the midline findings.
8. Similar to the recommendation at baseline, continue to actively contribute to shared learning opportunities across the PNG Partnership Fund (PPF) program in order to stimulate joint policy dialogue with government and development partners. Teacher and student absenteeism and early school closures are seriously compromising interventions.

APPENDIX A: MAPS OF PNG AND TARGET PROVINCES (AROB, ESP AND EHP) SHOWING DATA COLLECTION SITES AT MIDLINE

FIGURE A1: MAP OF ELEMENTARY SCHOOLS IN AROB

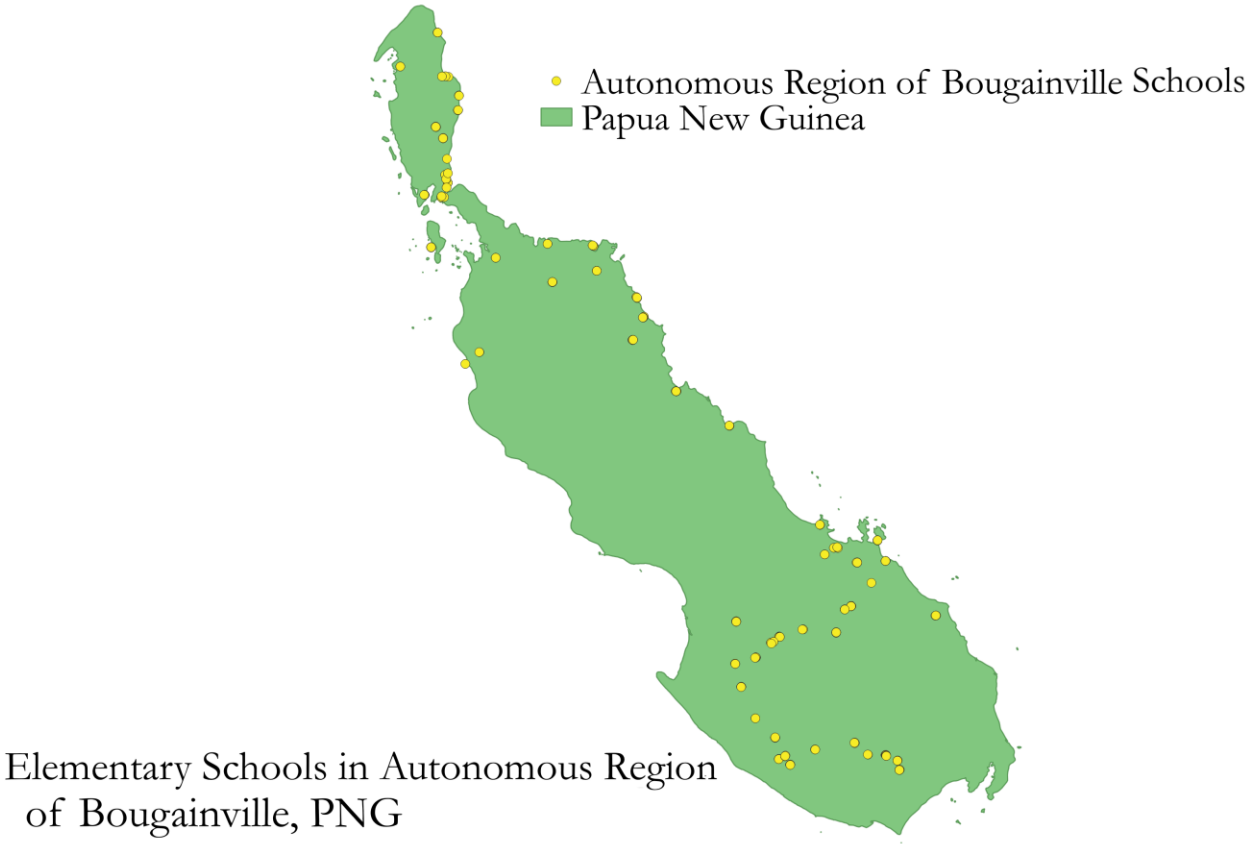


FIGURE A2: MAP OF ELEMENTARY SCHOOLS IN EHP

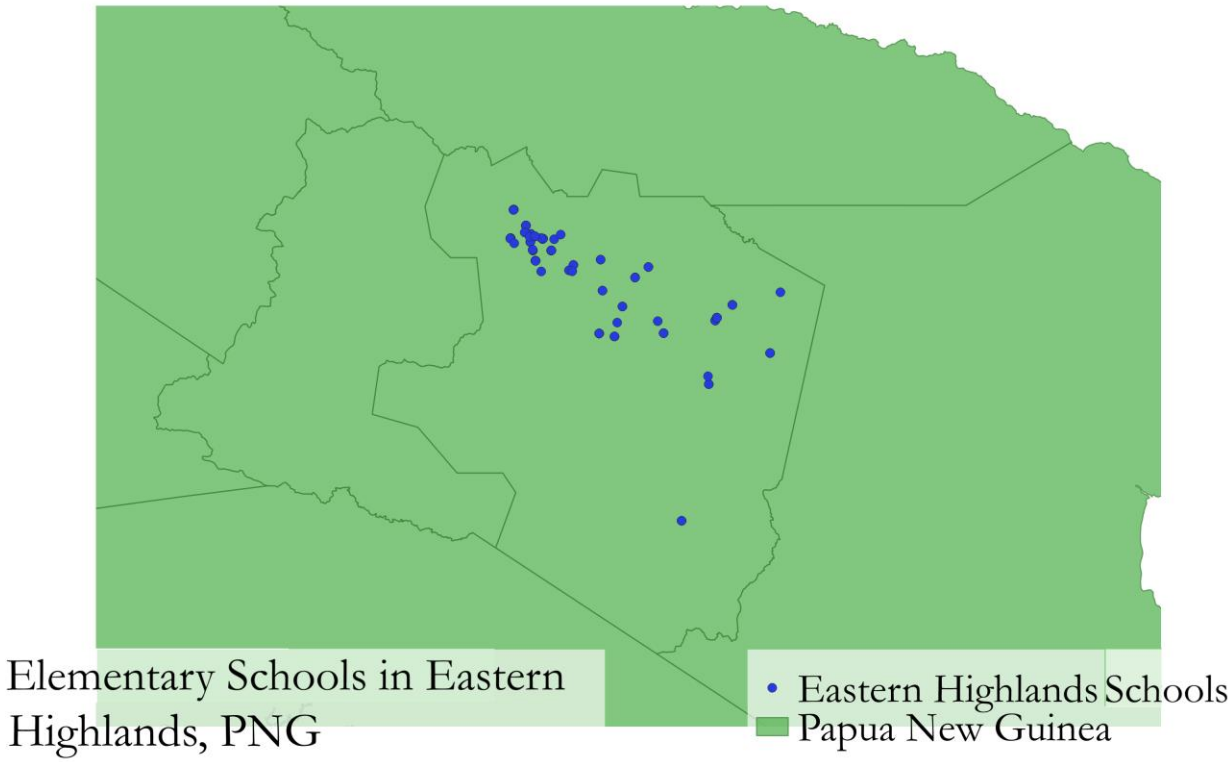
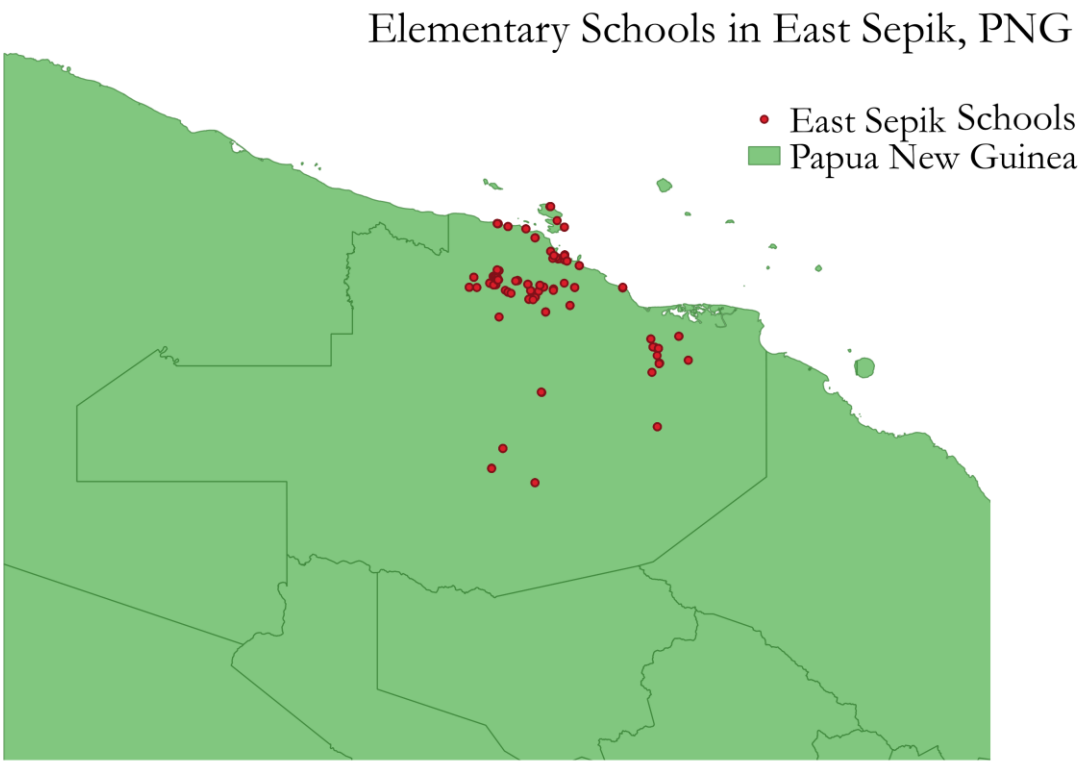


FIGURE A3: MAP OF ELEMENTARY SCHOOLS IN ESP



APPENDIX B: PRETEST AND POSTTEST DATA FROM TEACHER TRAINING

Monitoring of pretests and posttests following literacy, numeracy and cross cutting themes training showed positive results in teachers' knowledge acquisition of strategies to raise student learning outcomes in the classroom.

Literacy Boost training

LB training was conducted during the January 2018 school holidays, one to two weeks before the start of the school term. The core topics included teaching strategies and learning games that strengthen students' skills in speaking and listening, reading fluency and comprehension, and understanding different text types including fiction and non-fiction texts. The training was conducted over five days and included homework assignments and practical sessions.

Pretest and post-tests were collected for 1,781 male and female teachers. The ratio of female to male teacher trainees was 58:42, which is in line with EMIS data. The average age of the training cohort was 40.15 years and their level of education was very low. One fourth of the teachers had no high school certificate, 65% of the teachers had an elementary training certificate and only 0.34% had a bachelor's degree. On average, teachers had about 12 years of teaching experience. Female teachers tended to have more experience than males (13 years vs 11 years).

The average class size they taught in was 32, which is significantly lower than the numbers reported in EMIS (approximately 42 students). This is in line with the school data collected at baseline, which shows that EMIS numbers are likely inflated.

The Likert scores for self-reported competencies were coded from 1 (insufficient preparation) to 5 (extremely well prepared). Mean scores from the pretest suggested that most teachers felt reasonably confident and well prepared in the literacy skill areas they were being trained in for LB (responses tended to range from 3 to 4). The post-test analysis examined if there was a statistically significant shift in self-reported competencies due to the training.

The post-test analysis shows statistically significant improvement in a number of skills required for teaching literacy. The largest improvements were reported in phonemic awareness, decoding, fluency building and writing fiction and non-fiction texts. Teachers generally reported improvements by 1 Likert measure from their pretest score. Teachers responded favourably to an increased understanding of literacy and an increased ability to identify the structure and content of non-fiction texts. They also reported favourably to understanding the skills children need to decode and encode text.

Figure B1: Pretest and posttest training results on planning and teaching reading comprehension skills

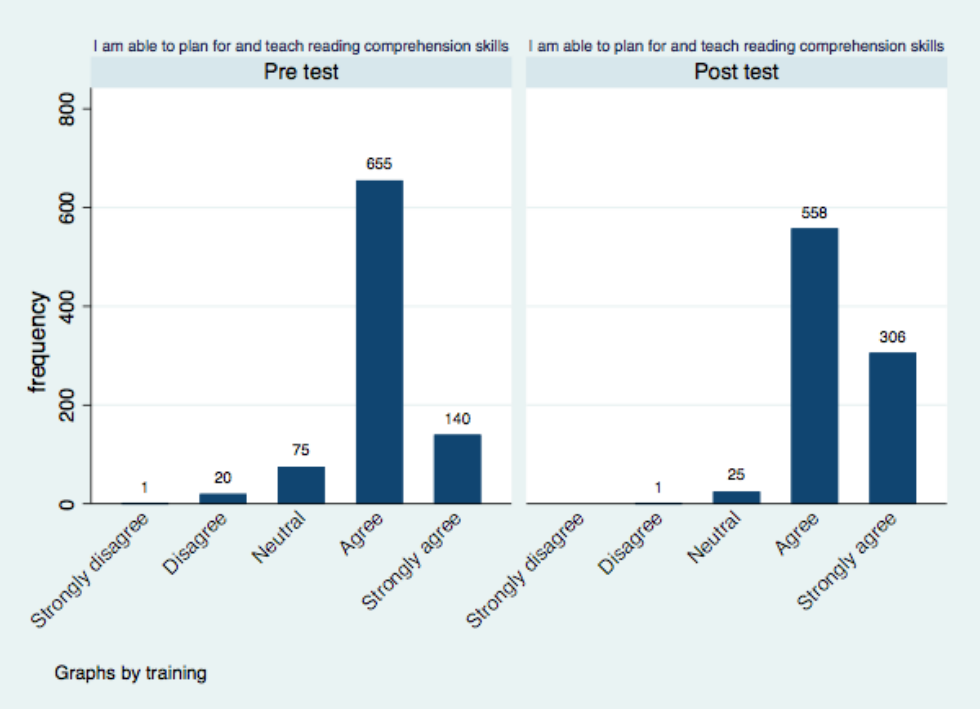


Figure B1 shows the pretest and posttest scores of teachers who received LB training. It is evident that many teachers felt more confident in their ability to plan for and teach reading comprehension skills. The proportion of responses which strongly agreed increased from 140 to 306 with only one teacher responding their lack of confidence in this area. This is explored more deeply when the results from the lesson observation are explained, particularly the observed skills of reading.

Figure B2: Pretest and Posttest training results on a teacher's ability to identify the structure and content of fiction texts

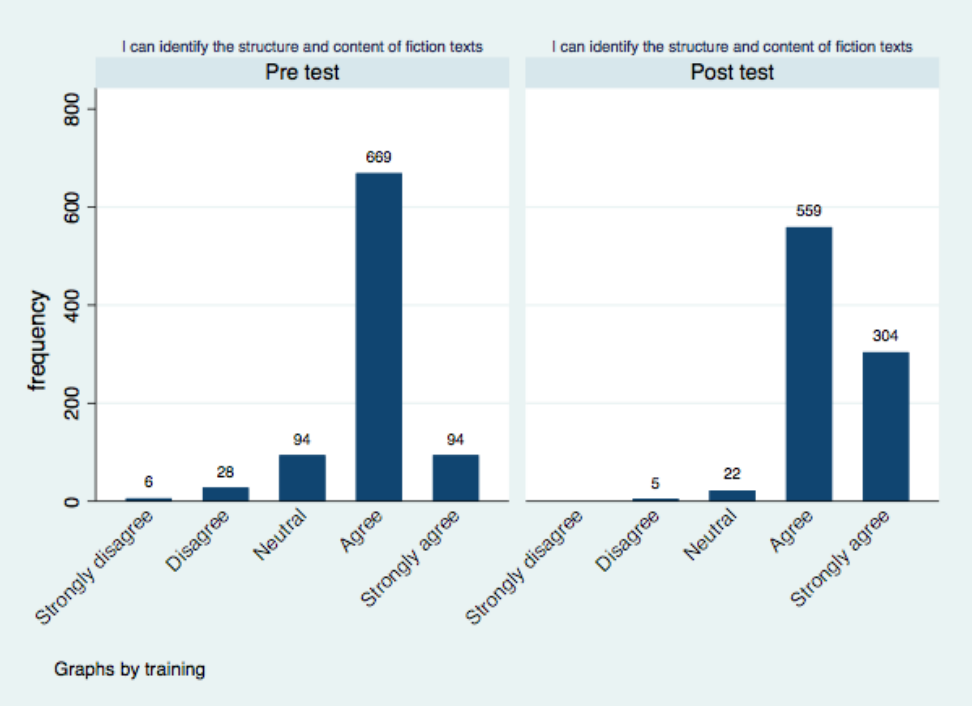


Figure B2 shows the results of teachers self-reported ability to identify the structure and content of fiction texts. This is an important skill to develop in the early grades. Similar to their confidence in teaching reading

comprehension, the proportion of responses which strongly agreed increased from 94 in the pretest to 304 in the posttest. A much smaller proportion of teachers still lack confidence in this teaching skill after training.

Figure B3: Pretest and Posttest training results on a teacher's ability to write nonfiction texts

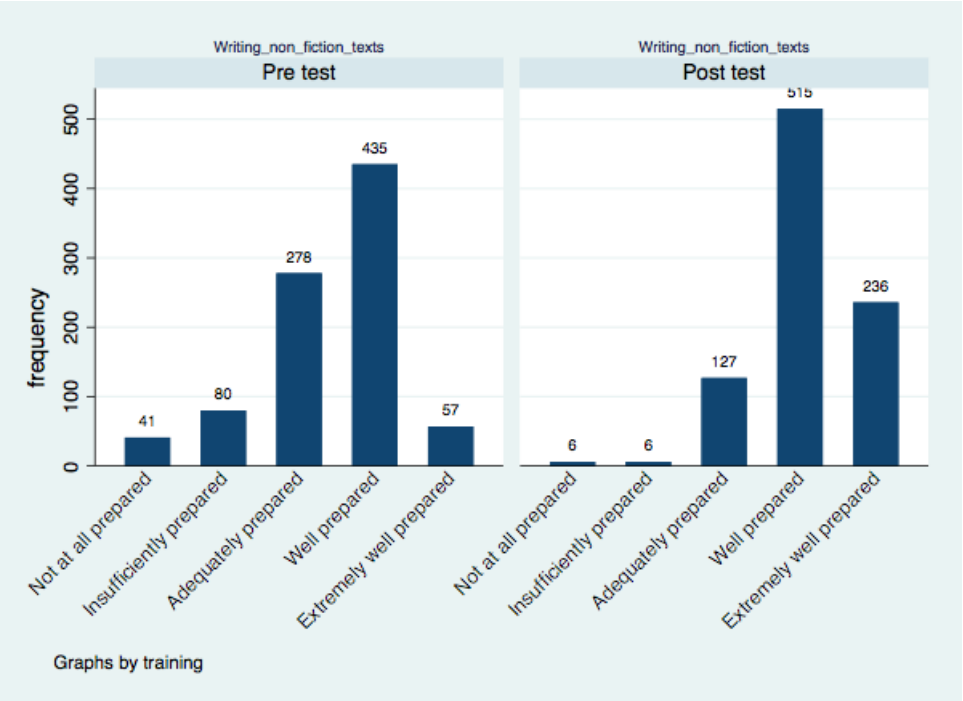


Figure B3 follows a similar trajectory where the proportion of teachers who felt extremely well prepared as a result of training increased significantly from the pretest. Similarly, there was a reduction in teachers who felt insufficiently prepared.

Numeracy Boost training

NB training was conducted in mid-2018 and consisted of training in the following strands of mathematics — numbers, operations, measurement, geometry, data and statistics.

Before receiving training, teachers were asked to fill out a pre-training survey assessing their numeracy skills and their confidence level in teaching different strands of mathematics. A total of 961 teachers were assessed before the training, of which 45.06% were male and 54.94% were female. Of the 961, 25.81% were from the ARoB, 36.52% were from EHP and 37.67% were from ESP. The highest percentage of teachers (50.16%) had a Certificate of Elementary Training, and 35.48% of teachers had completed Grade 10 as their highest education qualification. The median age of teachers was 40 years, and the median years of teaching experience was 11.

Teachers who went through the training were also asked to fill out a post-training survey to assess their confidence levels in each strand after the training. A total of 953 (male and female) teachers were assessed after the training. Eight teachers needed to leave early due to family reasons and did not submit their post-training assessment.

The pretests and posttests were analysed to see whether the training had any effect on the teachers' confidence levels in understanding and teaching the aforementioned strands of mathematics. To look at the impact training

had on teachers' attitudes, an **ordered probit regression model** was used. Teachers were asked to rate their ease with each strand on the following Likert scale.

1. *I would like more help and training to understand this strand of the mathematics curriculum.*
2. *I am able to teach some of the elements of this strand well.*
3. *I am able to teach most of the elements of this strand well.*
4. *I am able to teach this strand of the mathematics curriculum well.*

All five strands showed a statistically significant increase in teachers' confidence levels in the ordered probit regression model, with | t | values of more than 1.96. Comparing the results of both tests, we see that for the operations strand, only 23.45% of teachers surveyed before receiving training believed they could teach the strand well. After receiving training, this number increased to 53.28%. Similarly, for the data and statistics strand, only 10.30% of teachers said they could teach the strand well before the training. After the training, this number increased to 31.92%. The same trend was seen in all other strands as well. This clearly shows that the training increased teachers' confidence levels in understanding and teaching the Mathematics curriculum.

Table B1: Likert scale means for different strands of mathematics curriculum

Likert scale means before and after receiving training	Before	Before	After	After
Strand	N[21]	Mean	N	Mean
Numbers	915	2.65	946	3.00
Operations	934	2.90	946	3.30
Measurement	926	2.59	944	2.91
Geometry	927	2.39	946	2.72
Data and Statistics	922	2.41	943	2.84

The above table shows the means of the Likert scale in each parameter categorised by pre-training and post-training surveys. A statistically significant increase can be seen in the mean for each strand. For example, looking at numbers, we can see that the mean in the pre-training survey was 2.65, and increased to 3 in the post-training survey indicating an improvement in the way in which teachers were understanding how to apply numeracy strategies when teaching about numbers.

Crosscutting theme knowledge and awareness

Crosscutting theme training was conducted in June 2018 and consisted of the following topics: gender, disability inclusion, transition, my school app, formative and summative assessment methods and vernacular to English transition. The training ran for five days and included homework assignments.

In the pretest, a total of 1,180 teachers were surveyed, of which 41.19% were male and 58.81% were female. Of the 1,180 teachers, 37.37% were from ARoB, 30.93% were from EHP and 30.17% were from ESP. The highest

[21] The totals are different as some teachers did not respond to all items in the pretest and posttest.

percentage of teachers (38.73%) had completed Year 10 and did not have a High School Certificate. Only 0.76% (nine) of the teachers had a Bachelor of Education qualification. The median age of teachers was 41 years.

The two tests were combined to see whether the training had any effect on the teachers' confidence levels in using assessment tools and inclusive strategies in the classroom. To look at the impact training had on teachers' attitudes, an ordered probit regression model was used. Teachers were asked to rate questions on a Likert scale of 1 (strong disagree) to 5 (strongly agree),

All seven questions showed a statistically significant positive effect in the ordered probit regression model, with | t | values of more than 1.96. Comparing the results of both tests, we see that for the first parameter (teachers' understanding of inclusive education) only 22.37% teachers marked 'strongly agree' in the pretest, whereas in the posttest, this number increased to 45.04%. Similarly, for the second parameter, 18.73% of teachers marked 'strongly agree' in the pretest and 36.58% marked it as 'strongly agree' in the posttest. The same trend was seen in all other parameters. This shows that the training increased teachers' confidence levels in the above-mentioned areas.

Table B2: Average scores for variables assessed in pretest and posttest

	Pretest	Pretest	*Posttest*	*Posttest*
Variable	N	Mean	N	Mean
Understands what inclusive education is	1180	4.10	1028	4.42
Can identify three key barriers to participation of children with disabilities	1180	4.00	1028	4.34
Feels confident supporting girls to fully engage in school life	1180	4.31	1028	4.47
Can adjust the daily life lessons in the SBC to suit children's different abilities	1180	4.16	1028	4.34
Has five or more inclusive education strategies to apply in the classroom	1180	3.94	1028	4.28
Knows how to plan for students' transition to primary school	1180	3.99	1028	4.34
Is confident in using the assessment tools in the SBC	1180	4.11	1028	4.37

The above table shows the means of the Likert scale in each parameter categorised by pretest and posttest surveys. A statistically significant increase can be seen in the mean for each parameter. For example, looking at the fifth parameter, i.e., teacher has five or more inclusive strategies, we can see that the mean in pretest was 3.94, and increased to 4.28 in posttest.

APPENDIX C: SUMMARY HEADLINE DATA SETS

Table C1: ARoB headline numbers from school survey

Headline numbers from school survey: Autonomous Region of Bougainville		
Background	**Baseline**	**Midline**
Number of schools visited	121	44
Total number of children assessed	738	523
Total number of girls	371	295
Total number of boys	367	228
Disabilities	**Baseline**	**Midline**
Percentage of children with disabilities enrolled in school (out of total enrolled children)	3%	1.50%

Student assessment		
Letter Identification	**Baseline**	**Midline**
Percentage of children who could identify all letters (out of 26)	22.36%	57.93%
Percentage of children who could identify no letter (out of 26)	2.98%	0%
English words	**Baseline**	**Midline**
Percentage of children who could identify all English words (out of 20)	1.90%	2.49%
Percentage of children who could identify no English words (out of 20)	44.44%	13%
Tok Pisin words	**Baseline**	**Midline**
Percentage of children who could identify all Tok Pisin words (out of 20)	4.34%	7.84%
Percentage of children who could identify no Tok Pisin words (out of 20)	57.05%	34.99%
Number identification	**Baseline**	**Midline**

Percentage of children who could identify all numbers (out of 10)	54.34%	84.13%
Percentage of children who could identify no number (out of 10)	1.49%	0.38%
Skip pattern (by 2)	**Baseline**	**Midline**
Percentage of children who could answer all skip by two questions (out of 10)	46.34%	64.63%
Percentage of children who scored 0 on skip by 2 questions (out of 10)	13.55%	6.50%
Skip pattern (by 5)	**Baseline**	**Midline**
Percentage of children who could answer all skip by 5 questions (out of 10)	29.27%	55.83%
Percentage of children who scored 0 on skip by 5 questions (out of 10)	24.25%	14.34%
Addition problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all addition sums (out of 10)	35.37%	4.78%
Percentage of children who could do no addition sum (out of 10)	7.18%	4.97%
Subtraction problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all subtraction sums (out of 10)	16.94%	3.44%
Percentage of children who could do no subtraction sum (out of 10)	26.83%	12.05%
Word problems	**Baseline**	**Midline**
Percentage of children who could solve all word problems (out of 4)	20.87%	17.02%
Percentage of children who could solve no word problem (out of 4)	27.10%	19.69%
Shapes	**Baseline**	**Midline**
Percentage of children who could recognise all shapes (out of 5)	14.91%	24.47%

Percentage of children who couldn't recognise any shape (out of 5)	11.65%	1.91%

Table C2: EHP headline numbers from school survey

Headline numbers from school survey: Eastern Highlands Province		
Background	**Baseline**	**Midline**
Number of schools visited	59	42
Total number of children assessed	930	797
Total number of girls	453	383
Total number of boys	477	414
Disabilities	**Baseline**	**Midline**
Percentage of children with disabilities enrolled in school (out of total enrolled children)	1%	1.56%

Student assessment		
Letter identification	**Baseline**	**Midline**
Percentage of children who could identify all letters (out of 26)	18.92%	33.12%
Percentage of children who could identify no letter (out of 26)	3.87%	0%
English words	**Baseline**	**Midline**
Percentage of children who could identify all English words (out of 20)	0.65%	1.63%
Percentage of children who could identify no English words (out of 20)	54.84%	14.43%
Tok Pisin words	**Baseline**	**Midline**
Percentage of children who could identify all Tok Pisin words (out of 20)	1.18%	3.89%

Percentage of children who could identify no Tok Pisin words (out of 20)	72.37%	26.47%
Number identification	**Baseline**	**Midline**
Percentage of children who could identify all numbers (out of 10)	34.73%	77.42%
Percentage of children who could identify no number (out of 10)	1.51%	0.25%
Skip pattern (by 2)	**Baseline**	**Midline**
Percentage of children who could answer all skip by 2 questions (out of 10)	53.55%	82.43%
Percentage of children who scored 0 on skip by 2 questions (out of 10)	7.42%	0.88%
Skip pattern (by 5)	**Baseline**	**Midline**
Percentage of children who could answer all skip by 5 questions (out of 10)	75.05%	92.22%
Percentage of children who scored 0 on skip 5 questions (out of 10)	7.96%	0.88%
Addition problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all addition sums (out of 10)	21.83%	3.01%
Percentage of children who could do no addition sum (out of 10)	8.82%	6.15%
Subtraction problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all subtraction sums (out of 10)	8.49%	4.02%
Percentage of children who could do no subtraction sum (out of 10)	32.15%	15.68%
Word problems	**Baseline**	**Midline**
Percentage of children who could solve all word problems (out of 4)	26.34%	8.03%

Percentage of children who could solve no word problem (out of 4)	17.20%	13.30%
Shapes	**Baseline**	**Midline**
Percentage of children who could recognise all shapes (out of 5)	25.59%	34.88%
Percentage of children who couldn't recognise any shape (out of 5)	5.05%	0.75%

Table C3 ESP headline numbers from school survey

Headline numbers from school survey: East Sepik Province		
Background	**Baseline**	**Midline**
Number of schools visited	73	46
Total number of children assessed	1,075	715
Total number of girls	541	360
Total number of boys	534	355
Disabilities	**Baseline**	**Midline**
Percentage of children with disabilities enrolled in school (out of total enrolled children)	2%	3%

Student assessment		
Letter identification	**Baseline**	**Midline**
Percentage of children who could identify all letters (out of 26)	30.51%	41.26%
Percentage of children who could identify no letter (out of 26)	7.26%	0.70%
English words	**Baseline**	**Midline**
Percentage of children who could identify all English words (out of 20)	4.00%	5.59%

Percentage of children who could identify no English words (out of 20)	42.14%	10.21%
Tok Pisin words	**Baseline**	**Midline**
Percentage of children who could identify all Tok Pisin words (out of 20)	4.74%	7.27%
Percentage of children who could identify no Tok Pisin words (out of 20)	49.67%	21.68%
Number identification	**Baseline**	**Midline**
Percentage of children who could identify all numbers (out of 10)	57.30%	81.96%
Percentage of children who could identify no number (out of 10)	2.33%	0.42%
Skip pattern (by 2)	**Baseline**	**Midline**
Percentage of children who could answer all skip by 2 questions (out of 10)	67.07%	84.48%
Percentage of children who scored 0 on skip by 2 questions (out of 10)	7.35%	0.98%
Skip pattern (by 5)	**Baseline**	**Midline**
Percentage of children who could answer all skip by 5 questions (out of 10)	70.14%	91.47%
Percentage of children who scored 0 on skip 5 questions (out of 10)	9.49%	0.84%
Addition problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all addition sums (out of 10)	39.53%	3.50%
Percentage of children who could do no addition sum (out of 10)	9.77%	5.45%
Subtraction problems Baseline questions were versioned for E1 level. Midline questions were versioned for E2 level.	**Baseline**	**Midline**
Percentage of children who could do all subtraction sums (out of 10)	21.95%	4.34%

Percentage of children who could do no subtraction sum (out of 10)	24.09%	13.43%
Word problems	**Baseline**	**Midline**
Percentage of children who could solve all word problems (out of 4)	45.77%	16.50%
Percentage of children who could solve no word problem (out of 4)	9.30%	9.37%
Shapes	**Baseline**	**Midline**
Percentage of children who could recognise all shapes (out of 5)	30.33%	38.60%
Percentage of children who couldn't recognise any shape (out of 5)	3.63%	0.28%

APPENDIX D: FINAL REGRESSION ANALYSIS ON TEACHING SKILLS

Table D1: Overall regression analysis on teaching skills

Overall

	(1) Letters and Vocabulary (Percentage)	(2) Reading Comprehension (Percentage)	(3) Assessment (Percentage)	(4) Classroom Management (Percentage)	(5) Student Engagement (Percentage)
Intervention Status	8.228	31.65**	10.02*	19.92***	17.24***
	(1.41)	(3.30)	(2.54)	(4.41)	(3.92)
Constant	25.13***	35.29***	15.02***	39.89***	34.24***
	(6.14)	(4.72)	(6.08)	(12.13)	(10.27)
Observations	99	39	99	99	99

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D2: ARoB regression analysis on teaching skills

ARoB

	(1) Letters and Vocabulary (Percentage)	(2) Reading Comprehension (Percentage)	(3) Assessment (Percentage)	(4) Classroom Management (Percentage)	(5) Student Engagement (Percentage)
Intervention Status	7.425	31.69	9.269	28.30**	17.43*
	(0.79)	(1.42)	(1.16)	(3.56)	(2.08)
Constant	34.22***	52.73*	21.21***	31.11***	39.09***
	(4.94)	(2.66)	(4.50)	(5.91)	(5.82)
Observations	35	12	35	35	35

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D3: EHP regression analysis on teaching skills

EHP

	(1) Letters and Vocabulary (Percentage)	(2) Reading Comprehension (Percentage)	(3) Assessment (Percentage)	(4) Classroom Management (Percentage)	(5) Student Engagement (Percentage)
Intervention Status	6.066	31.82*	5.499	8.005	14.63*
	(0.69)	(2.36)	(1.74)	(1.33)	(2.59)
Constant	17.41**	36.36**	11.50***	52.65***	36.76***
	(2.81)	(4.11)	(5.23)	(13.29)	(8.47)
Observations	40	13	40	40	40

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D4: ESP regression analysis on teaching skills

ESP

	(1) Letters and Vocabulary (Percentage)	(2) Reading Comprehension (Percentage)	(3) Assessment (Percentage)	(4) Classroom Management (Percentage)	(5) Student Engagement (Percentage)
Intervention Status	17.82	25.97*	21.84*	24.30*	22.65*
	(1.47)	(2.45)	(2.30)	(2.60)	(2.68)
Constant	22.18*	22.08*	10.33	34.55***	22.40***
	(2.80)	(2.71)	(1.87)	(5.06)	(5.22)
Observations	24	14	24	24	24

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

APPENDIX E: FINAL OVERALL REGRESSION MODEL

Table E1: Overall regression for boys and girls

For both boys and girls: no statistically significant improvement but gains in literacy and numeracy evident overall

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	2.964	3.313	2.615
	(1.59)	(1.26)	(1.78)
Constant	55.03***	40.33***	69.73***
	(40.83)	(21.57)	(60.41)
Observations	1828	1828	1828

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E2: Overall regression for boys

For boys: no impact

	(1) Literacy Score (Percentage)	(2) Literacy and Numeracy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	2.645	1.272	-0.101
	(0.93)	(0.65)	(-0.07)
Constant	40.08***	56.32***	72.56***
	(19.80)	(40.79)	(68.71)
Observations	895	895	895

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3: Overall regression for girls

For girls: impact on overall scores and numeracy

	(1)	(2)	(3)
	Literacy and Numeracy Score (Percentage)	Literacy Score (Percentage)	Numeracy Score (Percentage)
Intervention Status	4.538*	4.009	5.067**
	(2.15)	(1.36)	(2.86)
Constant	53.86***	40.55***	67.17***
	(34.48)	(19.78)	(45.48)
Observations	933	933	933

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

APPENDIX F: OVERALL REGRESSION ANALYSIS IRRESPECTIVE OF INTERVENTION STATUS

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Age of child	0.446	0.424	0.468
	(1.22)	(0.73)	(1.81)
Female	-1.126	0.156	-2.409***
	(-1.40)	(0.14)	(-3.48)
Grade of child	14.99***	15.49***	14.50***
	(6.48)	(5.55)	(5.41)
Repeated current grade	-1.841	-3.421*	-0.261
	(-1.84)	(-2.31)	(-0.29)
Child has attended ECCE	2.644**	3.192*	2.097**
	(2.93)	(2.44)	(2.85)
Child ate breakfast that morning	2.385	3.052	1.717
	(1.47)	(1.37)	(1.30)
Socio-economic situation	1.011*	1.421*	0.601
	(2.36)	(2.34)	(1.76)
Mother can read and write	2.545**	3.401**	1.689*
	(2.91)	(2.71)	(2.01)

Home Reading Environment	2.319***	2.582***	2.056***
	(6.17)	(5.08)	(5.82)
Child chews betel nut	-2.457*	-3.405*	-1.509
	(-2.34)	(-2.42)	(-1.52)
Child likes coming to school everyday	8.304***	9.930***	6.678*
	(3.57)	(4.25)	(2.25)
Child has read books on Bloom Reader	2.505	4.939*	0.0704
	(1.70)	(2.31)	(0.06)
Child has missed a day of school in the last week	-2.337**	-2.726*	-1.948*
	(-2.68)	(-2.22)	(-2.56)
Child has borrowed reading books from the community book bank	4.834***	6.579***	3.089**
	(3.81)	(3.46)	(3.34)
Autonomous Region of Bougainville	0	0	0
	(.)	(.)	(.)
Eastern Highlands Province	0.636	-3.475	4.746**
	(0.35)	(-1.26)	(3.12)
East Sepik Province	6.697***	6.895*	6.498***

	(3.46)	(2.30)	(4.49)
Constant	2.416	-15.59	20.42**
	(0.37)	(-1.83)	(2.94)
Observations	1832	1832	1832

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

APPENDIX G: PROVINCIAL REGRESSION ANALYSIS

Autonomous Region of Bougainville

Table G1: ARoB regression for boys and girls

Overall model: Statistically significant improvements in numeracy but not in literacy

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	1.588	-2.720	5.896*
	(0.53)	(-0.68)	(2.24)
Constant	52.70***	42.60***	62.80***
	(21.49)	(13.24)	(29.44)
Observations	506	506	506

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G2: ARoB regression for boys

For boys: No improvements

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	0.590	-0.245	1.425
	(0.17)	(-0.05)	(0.42)
Constant	53.15***	39.20***	67.10***
	(18.51)	(10.25)	(24.96)
Observations	221	221	221

t statistics in parentheses* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G3: ARoB regression for girls

For girls: Substantial improvements in numeracy

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	2.370	-4.133	8.873**
	(0.67)	(-0.85)	(3.01)
Constant	52.41***	44.84***	59.97***
	(19.06)	(12.87)	(23.69)
Observations	285	285	285

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Eastern Highlands Province

Table G4: EHP regression for boys and girls

Overall model: No statistically significant improvements although literacy scores increased

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	1.578	3.235	-0.0794
	(0.61)	(0.91)	(-0.04)
Constant	55.16***	37.40***	72.93***
	(31.59)	(16.53)	(43.30)
Observations	760	760	760

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G5: EHP regression for boys

For boys: No improvements

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	0.836	3.211	-1.539
	(0.32)	(0.84)	(-0.74)
Constant	56.13***	37.55***	74.70***
	(37.70)	(17.59)	(48.55)
Observations	399	399	399

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G6: EHP regression for girls

For girls: No statistically significant improvements

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	2.368	3.256	1.479
	(0.81)	(0.88)	(0.56)
Constant	54.13***	37.24***	71.02***
	(23.86)	(14.17)	(31.17)
Observations	361	361	361

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

East Sepik Province

Table G7: ESP regression for boys and girls

Overall model: Statistically significant improvements in literacy and overall

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	8.016*	12.20*	3.834
	(2.21)	(2.32)	(1.52)
Constant	56.57***	41.42***	71.73***
	(21.26)	(10.95)	(40.55)
Observations	562	562	562

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G8 ESP regression for boys

For boys: No improvements

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	4.801	7.872	1.729
	(1.30)	(1.41)	(0.75)
Constant	58.35***	43.08***	73.61***
	(20.89)	(10.26)	(47.17)
Observations	275	275	275

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G9: ESP regression for girls

For girls: improvements in literacy

	(1) Literacy and Numeracy Score (Percentage)	(2) Literacy Score (Percentage)	(3) Numeracy Score (Percentage)
Intervention Status	11.10**	16.32**	5.867
	(2.73)	(2.82)	(1.92)
Constant	54.86***	39.81***	69.90***
	(18.31)	(9.82)	(31.46)
Observations	287	287	287

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

REFERENCES

Bakken, L., Brown, N., & Downing, B. (2017). Early Childhood Education: The Long-Term Benefits. *Journal of Research in Childhood Education*, *31*(2).

Belisle, M., Cassity, E., Kacilala, R., Seniloli, M. T., & Taoi, T. (2016). Pacific Islands Literacy and Numeracy Assessment (PILNA): Collaboration and innovation in reporting and dissemination. *Using assessment data in education policy and practice: Examples from the Asia Pacific* . Melbourne: ACER and Bangkok: UNESCO.

Chu, N. S. (2001). Effects of Betel Chewing on the Central and Autonomic Nervous Systems. *Journal of Biomedical Science*, *8*(3), 229-236. doi:10.1159/000054038.

Curwin, R. L. (2010). The Relationship Between Behavior and Motivation. In R. L. Curwin, *Meeting Students Where They Live: Motivations in urban schools.* Alexandria: Association for Supervision and Curriculum Development.

Czuba, J. E. (2013). *Report of the Task Force for the Review of Outcomes Based Education in Papua New Guinea.* Port Moresby: Government of Papua New Guinea.

Drinan, F. A. (2014). *Elementary English Teacher Development Pilot Report.* Port Moresby: Teaching Service Commission.

Educational Quality Assessment Program. (2015). *Pacific Islands Literacy and Numeracy Assessment – PNG Report.* Suva, Fiji: Pacific Community – New Zealand Foreign Affairs and Trade.

Engle, P. (2009). How to avoid the loss of potential in over 200 million young children in the developing world. *Child Health and Education*, *1*(2), 58-72.

Engle, P., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet Series, Child Development*, *378*(9799), 1339-1353.

Garg, A., Chaturvedi, P., & Gupta P. C. (2014). A review of the systemic adverse effects of areca nut or betel nut. *Indian Journal of Medical and Paediatric Oncology*, *35*(1). doi:10.4103/0971-5851.133702.Gupta P. C., and W. S. (2002). Global epidemiology of areca nut usage. *Addiction Biology*, 77–83.

Hackman, D., Betancourt, L., Gallop, R., Romer, D., Brodsky, N., Hurt, H., & and Farah, M. (2014). Mapping the Trajectory of Socio-economic Disparity in Working Memory: Parental and Neighborhood Factors. *Child Development*, *85*(4). https://doi.org/10.1111/cdev.12242.

Ho C. -G., Gee, M. -J., Tsai, C. -C., Lo, C. -I., & Hwang, M -N. (2000). Factors related to betel chewing among junior high school students in Taiwan. *Community Dentistry Oral Epidemiology*.

Johnston, C., Namit, K., & Reen, J. A. (2018). *Rapidly Improving Standards in Elementary - PNG - Baseline Study.* Port Moresby: Save the Children.

Johnston, C., Namit, K., Seidan, J., & Reen, J. (2019). *RISE PNG International Development Emergent Learning Assessment (IDELA) Report* . Port Moresby: Save the Children.

Joseph, N., Nagaraj, K., & Shashidhar Kotian, M. (2010). Arecanut and tobacco use among school children in a village in South India – A cross-sectional study. *Australasian Medical Journal*, *1*(5), 299-303 Doi 10.4066/AMJ.2010.280.

Koca, F. (2016). Motivation to Learn and Teacher–Student Relationship. *Journal of International Education and Leadership*, *6*(2).

Kaleebu, N., Gee, A., Jones, R. and Watson, A. H. A. (2013). *SMS Story Impact Assessment Report*. Papua New Guinea: VSO.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early Childhood Interventions: Proven Results, Future Promise.* St. Monica California: RAND Labor and Population. Retrieved from RAND: https://www.rand.org/pubs/research_briefs/RB9145.html

Manikuali, R. N. (2014). *Experiences of elementary pre-service teacher training.* Port Moresby: PNGEI.

NDoE PNG. (2015). *Retention EMIS.* Port Morseby: National Department of Education, Papua New Guinea.Norman, P. E., Baki, P., Nongkas, C., Jones, R., & Oliver, W. (2012). *Capacity Needs Analysis for the PNG National Education System.* Port Moresby: Divine Word University and Voluntary Services Overseas.

Pisani, L., Borisova, I., & Dowd, A. (2015). *International Development and Early Learning Assessment Technical Working Paper.* Washington DC: Save the Children.

Pratt, S. (2015). The challenge of betel nut consumption to economic development: A case of Honiara. *Asia-Pacific Journal of Rural Development*, *21*(2), 103–120. doi:10.18356/b368d584-en.

Save the Children. (2017). *International Development Emergent Learning Assessment (IDELA) Adaptation and Administration Guidance.* Washington DC: Save the Children.

Save the Children. (2017). *Windows into Early Learning and Development - Cross country International Development Emergent Learning Assessment (IDELA) findings on ECCD access, quality and equity.* London: Save the Children International.

Shah, S. M. (2002). Addicted schoolchildren: Prevalence and characteristics of areca nut chewers among primary school children in Karachi, Pakistan. *Journal of Paediatrics and Child Health*, *38*(5), 507-510. doi:10.1046/.

Snow, C., Burns, S., & Griffin, P. (1998). *Preventing Reading Difficulties in Young Children.* Washington DC: National Academy Press.Trivedy C. R., Craig, G., & Warnakilasuriya, S. (2002). The oral health consequences of chewing areca nut. *Addiction Biology*, 115-25.

UNICEF. (2013). *Papua New Guinea, Education Overiview*. Retrieved from UNICEF: https://www.unicef.org/png/activities_4369.html

UNICEF. (2016). *Module Child Functioning*. Retrieved from UNICEF: https://data.unicef.org/resources/module-child-functioningWilliams, B. (2013). *"Kirapim save bilong ol tisa"Pilot Reading Booster Program in Western Highlands Province, Papua New Guinea.* Port Moresby: Read PNG.

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & and Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, (4), 21-36.