



RIGHT TO PLAY
PROTECT. EDUCATE. EMPOWER.

Enhancing Quality and Inclusive Education

Baseline Assessment: Palestine

January 2021

Prepared by: Jarret Guajardo, InformEd International

With deep gratitude to InformEd Country Manager Mohammad Shuaibi; Right to Play Country Office staff Nahed Natsheh, Kholoud Salameh, Shatha Allan, Ayman Ibrahim, Jackeline Jazrawi, Moath Mereb, and Haya Shahwan; and Right To Play HQ MEL Eric Opoku

With special thank you to the team of Ministry of Education and Higher Education Supervisors:

Manal Kharraz, Wafa'a Zubadi, Hatem Ekhdash, Amalat Ismail, Elham Al Tawil, Ruba Dawood, Salam Qawwas, Saba'a Helal, Mazen Amer, Fadwa Oudah, Reem Awwad, Yaser Ghanem, Suhair Abdullah, Riyadh Saabneh, Sama Lahloh, Mustafa Darghma, Yousef Al Teety, Hanady Abu Shkheidem, Rawda Dweik, Sawsan Qabajah, Ziyad Al Hasanat, Samah Amr, Dawlat Oweidat, Hiba Abu Radwan, Mu'taz Asfour, and Maree Al Sous

And to team of enumerators:

Khaled Hamamdeh, Montaser Khedeer, Samer Barham, Sarah Warasnah, Najwa Saadah, Najah Samamra, Saja Dawood, Salsabil Homoudah, Leena Mohammed, Maysa'a Barakat, Fardous Abu Khamis, and Nermin Al-ali

Table of Contents

List of Acronyms.....	4
Executive Summary.....	5
1. Project Background.....	9
1.1 Project Description.....	9
1.2 Context	9
2. Baseline Purpose/Objectives and Questions.....	10
2.1 Baseline Purpose & Objectives.....	10
2.2 Baseline Evaluation Questions	10
3. Methodology.....	12
3.1 Study Design.....	12
3.2 Limitations	13
3.3 Analysis.....	14
4. Baseline Evaluation Findings.....	15
4.1 Programming Area Sample Characteristics	15
4.2 EQIE Indicator Baseline Values.....	18
4.3 How similar are intervention and comparison schools?.....	19
4.4 What can the baseline assessment tell us about students' reading skills?	21
4.5 What can the baseline assessment tell us about students' life skills?	28
4.6 What supplemental reading activities are available to children (in homes and communities)?	34
4.7 What teaching knowledge, attitudes, & practices regarding inclusive PBL/literacy instruction are common among teachers at baseline?	38
4.8 What is the current state of schools?	45
4.9 What is the capacity of community members in creation and maintenance of safe learning environments?.....	50
4.10 What is the current state of community members and leaders in terms of knowledge regarding ways to address barriers to education faced by marginalized children and children with disabilities?.....	52
4.11. What might explain the higher learning outcomes of girls versus boys across grades?.....	53
5. Conclusions/Recommendations and Lessons Learned.....	54

5.1 Recommendations	54
5.2 Lessons Learned.....	57
Annex A. EQIE Project Indicators and Analysis Framework.....	60
Annex B. PMF	69
Annex C. Learner Assessment Details.....	81
Annex D. Inter-Rater Reliability (IRR).....	84
Annex E. Intra-class Correlation (ICC)	86
Annex F. Equity Regression Tables	88
Annex G. Detailed Lesson Observation Data	93
Annex H. Baseline Documentation	95
Annex I. Evaluation Methods and Limitations.....	97

List of Acronyms

CSO	Civil society organization
CWD	Children with disabilities
ECD	Early childhood development
EQ	Evaluation question
EQIE	Enhancing Quality and Inclusive Education
ICC	Intra-class correlation
IRR	Inter-rater reliability
MOE	Ministry of Education and Higher Education
MUW	Most-used words subtest
NS	Not statistically significant
PA	Phonemic awareness subtest
PBL	Play-based learning
PMF	Project management framework
PTA	Parent-teacher association
RCA	Reflect, connect, apply
RTP	Right to Play
RWC	Readers with comprehension
SDG	Sustainable Development Goal
SMC	School management committee
WG	Washington Group
WCPM	Words correct per minute

Executive Summary

Right To Play (RTP) is implementing the EQIE program in Palestine. The overall goal is to enhance quality and inclusive education, with the two main objectives to (1) improve children's minimum proficiency level in reading and life skills, focusing on children in primary Grades 1-4 in the West Bank, and (2) increase access and retention of out-of-school children with disabilities (CWD) in the Gaza Strip.

These pages summarize the key findings of the EQIE baseline assessment conducted in late November/early December 2020, aiming to (1) assess the program's starting point, (2) inform program design and strategy, and (3) establish baseline measurements against which improvement and impact will be measured during 2023 midline and 2024 endline.

Methodology

Target and achieved sample by type of respondent/tool

Tool	Target	Achieved	Notes
Grade 1 Learner Assessment	800	810	20 students per intervention and comparison schools. Plus 3 students from Teacher Interview for CWD.
Grade 4 Learner Assessment	300	320	15 students per intervention school. Plus 2 students from Teacher Interview for CWD.
Lesson Observation	40	39	Two per intervention school (same teachers as interviewed). One teacher was on sick leave.
Teacher Interview	40	41	Grade 1 & 4 teachers interviewed per intervention school. At one school two teachers were interviewed.
Principal Interview	40	40	One interview per each of intervention and comparison schools.
Caregiver Interview	400	403	Random subset of caregivers of assessed students at intervention schools.
Educ. Official Interview	3	3	Chosen by Right to Play for their relevance to the EQIE project.
Supervisor Interview	8	11	2 from Hebron, 3 from Nablus, 1 from Qalqilya, 1 from Tubas, 2 from Tulkarem, and 2 from Yatta.

^ASampling strategy: Purposeful sampling of CWD first, then random sampling up to target number per grade, striving for balance between boys and girls if in mixed-sex school.

The evaluation design is quasi-experimental, taking advantage of the phased approach to implementation among West Bank schools. This design seeks to optimize rigor of (A) impact evaluation and (B) reporting on project indicators. For impact evaluation, the baseline established an intervention and comparison sample of students (in Grade 1 at baseline) to be followed longitudinally for re-assessment in Grade 4 during 2023 midline. This allows for difference in difference analysis, comparing learner gains in intervention and comparison schools (controlling for exposure to EQIE programming, and other observable differences between groups). For reporting on indicators, the baseline also included a cross-sectional random sample

of Grade 4 students from intervention schools. This also serves as benchmark grade for learning outcome targeting.

Programming Area Sample Characteristics

Among the intervention sample, there were 9 boys schools, 7 girls schools, and 4 mixed-sex schools in Hebron and Nablus. The sample of students was 45% female, 13% refugee or internally displaced, and 2% identified by teacher as CWD. The Grade 1 sample was 6 years old on average, and Grade 4 sample was 9.2 years old. Almost all students had attended ECD or preschool.

EQ1: How similar are intervention and comparison schools?

Especially based on student-level variables, intervention and **comparison groups appear similar** enough for a solid quasi-experimental impact evaluation, but the midline analysis will need to control for the significantly different student background variables and some difference (though not statistically significant) in school-level variables when investigating the impact of the EQIE project.

EQ2: What can the baseline assessment tell us about students' reading skills?

There was no statistically significant difference in any skill between Grade 1 intervention (Int) and comparison (Comp) students. Overall performance on foundational reading skills is relatively low among Grade 1 students (for example, students could only identify 60% of letters) and moderately high among Grade 4 students (for example, students identified 80% of most used words and the 89% who qualified as readers read 54 words correct per minute on average). However, the 11% of Grade 4 students who qualify as nonreaders are far behind their peers and need remedial support. There is also a 'middle tier' of 34% of students who qualify as readers but who have not yet mastered comprehension.

EQ3: What can the baseline assessment tell us about students' life skills?

There was no statistically significant difference in any skill between Grade 1 intervention (Int) and comparison (Comp) students. Students scored lowest in concentration (44% average) and highest in critical thinking (77%). Among the Four C life skills, Grade 4 students scored significantly higher than Grade 1 students only in concentration and critical thinking. The two grades were statistically indistinguishable in communication and collaboration – this may indicate that students are not fulfilling their developmental potential in these two skills.

EQ4: What supplemental reading activities are available to children (in homes/communities)?

Students and caregivers in intervention areas reported high prevalence of textbooks and religious materials at home, but about half (or more) did not have child-friendly reading materials like storybooks, coloring books, and comics. Half or more also lacked the most interactive home literacy actions such as reading to student, telling story to student, or playing a literacy game with the student. Only one-fifth of students reported attending in-person extracurricular literacy activities in the past week, and only half reported attending media-based extracurricular literacy activities.

EQ5: What teaching knowledge, attitudes, & practices regarding inclusive PBL/literacy instruction are common among teachers at baseline?

On average, Grade 1 & 4 teachers in intervention schools had knowledge of 55% of the PBL elements. Only one teacher reported having been trained by RTP. The lesson observation tool, likely more valid than the teacher self-report, used a five-point scale to assess teachers on 27 items split into five categories.¹ Teachers scored highest on management of behavioral issues (3.9) and lowest on approach and ability to implement literacy-PBL. Based on lesson observation, the percent of teachers who are demonstrating improved teaching practices to enhance children's reading skills (Indicator 1100.1) is 18%.

EQ6: What is the current state of schools (for CWD and without) in terms of safety?

Most school staff report students come to them to address school safety concerns, and that the school has an active system for addressing student concerns. Half of the principals interviewed report that they have a strategy, plan, or policy to address school safety that is either almost fully or full implemented. These safety plan often have objectives for hygiene, CWD, and teacher behavior; but often lack objectives for exploitation, bullying, and violence. To better accommodate CWD, principals mentioned a resource or sensory room, assistive devices, and games and stationery. Overall, **2 of 20 intervention schools met the physical safety criteria** for Indicator 1300.1, compared to 6 of 20 comparison schools.

A minority of students report harsh/violent child-child or child-adult interactions in the last week at school. CWD are less likely to report almost all types of behaviors. Importantly, no CWD report anyone left out of an activity or game, an indication of inclusion.

EQ7: What is the capacity of community members in creation and maintenance of safe learning environments?

On average, caregivers could name 4.7 strategies (out of 9) for physical safety inside the classroom, 3.4 strategies (out of 6) for physical safety outside the classroom, and 3.5 strategies (out of 10) for emotional safety. However, 11-15% of caregivers could not name a single strategy in each of these categories. Caregivers were also given the option to specify 'other' strategies for emotional safety.² Most frequently, they mentioned the importance of play, games, sports, and art for both learning and relationship with the teacher.

EQ8: What is the state of community members and leaders in terms of knowledge regarding ways to address barriers to education faced by marginalized children and CWD?

Caregivers most frequently mentioned the barriers of inability to cover expenses such as assistive devices, lack of positive discipline, inaccessible school infrastructure for CWD, and lack of transport for CWD or those living far from schools.

EQ9: What might explain the higher learning outcomes of girls versus boys across grades?

Most teachers and principals reported that girls perform/learn better than boys. Those who said girls perform better explained that girls were more disciplined, calmer/quieter, better organized,

¹ 1 = Not at all, 2 = Very little extent, 3 = To some extent, 4 = To a good extent, 5 = To a great extent

² For physical safety inside and outside the classroom, it was also intended for caregivers to specify an 'other' option. However, due to a mistake in programming the KoBo forms, this option was unfortunately not included.

more focused, or have more attention and commitment to education. They also explained that girls generally spend more time at home and respond more to teachers and family, while boys spend more time outside the home and playing. Female teachers were also highlighted as more patient and loving. This may explain at least part of the regression finding that the higher reading and life skill scores of girls in this sample is explained entirely by the higher performance of girls schools versus either boys or mixed schools.

Recommendations for EQIE Programming in Palestine

- Give greater weight (though not exclusive weight) to teaching and learning strategies that emphasize meaning and understanding, starting from Grade 1.
- Increase teachers' skill and frequency in using a variety of comprehension questions/activities when reading to students. As a bonus, these techniques apply to students of any reading level, can enhance engagement/motivation of all students to read, and are also an opportunity to engage students who typically participate less.
- Provide (fun!) opportunities for remedial literacy activities in all grades to prevent and respond to students falling behind their peers.
- Consider the unique challenge of Arabic language for students who may struggle to distinguish between similar letters, especially children with learning disabilities.
- Help students see their friends and household members as assets to support them when they have problems, or to work toward common goals/coursework. Teachers can play a constructive role by organizing/assigning more group work.
- Share with parents, educators, and children themselves the fact that CWD matched the level of their peers in certain reading and life skills, as an example of how CWD are as capable of learning as their peers with the right support. This can be done through events such as parent-teacher 'learning together' sessions and outside school learning days.
- Increase teacher skill and confidence in interactive/ participatory techniques – especially free time, discussions, and child leadership opportunities.
- Take a dynamic and agile approach to meaningfully help schools accommodate the small number of CWD who each have individual needs. Ideally, draw on specialist support for diagnosis as well as thinking through solutions to individual needs.
- Find a way to enhance communication between teachers and CWD, and perhaps between CWD and their peers and caregivers.
- Enhance caregivers' use of play. Optimize play for boys and girls – for boys, leveraging their existing play toward games that enhance their reading and life skills and for girls, more opportunity for play outside school.
- Find ways to expand to all students, household literacy activities or when infeasible, access to extracurricular literacy activities.
- Help schools integrate gender considerations into their school safety plans, and to make teachers aware of the value of empowering and non-traditional roles for both boys and girls in classroom reading materials.
- Gather and share lessons learned from girls schools – what are they doing right to generate greater reading and life skills on average?

- Leverage EQIE as a nexus for bringing stakeholders together to brainstorm solutions to barriers and challenges. EQIE can and should then experiment with potential solutions and rigorously document the results through the midline of this impact evaluation.

1. Project Background

1.1 Project Description

Right To Play is implementing the Enhancing Quality and Inclusive Education (EQIE) program in Palestine. The overall goal of the project is to enhance quality and inclusive education as it relates to Sustainable Development Goal (SDG) 4, the underlying principles within the SDGs of “leave no one behind”; Palestine Government Education Sector Strategic Plan (ESSP) 2017-2022 (West Bank); and the priorities of civil society organization (CSO) partners (Gaza).

The main objectives of the project are to:

1. Improve children’s minimum proficiency level in reading and social-emotional life skills, focusing on children in primary Grades 1-4 in the West Bank (Nablus, South Nablus, Hebron, Yatta, Salfeet, Qalqiliyeh, Tulkarem and Tubas)
2. Increase access and retention of out-of-school children with disabilities (CWD) on the Gaza Strip (Rafah, North Gaza, Middle Gaza and Gaza City).

To achieve this, a total of 678 educational staff will be trained on play-based learning teaching practices to enhance children’s reading skills, 6,676 children will be provided with learning materials, 2,100 parents will be trained to implement supplemental reading activities for children outside school, 9 CSO partners will be trained and mentored on organizational development needs, and 500 out of school CWD will be supported to enroll into target educational institutions, temporary learning spaces or alternative education programs. See Annex H for Palestine EQIE Project Narrative.

1.2 Context

The EQIE project will be implemented in both the West Bank and Gaza. The project in the West Bank will report on output targets and immediate outcome achievements in relation to indicators 1100, 1200 and 1300. The Gaza project will report on output targets and immediate outcome achievements in relation to indicators 1300 and 1400. Project implementation in the West Bank will take place in close collaboration with Ministry of Education and Higher Education (MOEHE), while project implementation in Gaza will be undertaken in collaboration with CSO partners.

Table 1: # of target schools and children by geography

Geography	# of target schools	# of target children in school	# of target out-of-school children
-----------	---------------------	--------------------------------	------------------------------------

<i>West Bank</i>	78 (implemented in four phases: 21 schools in 2020, 21 in 2021, 22 in 2022, and 14 in 2023)	13,325 (6662 boys and 6663 girls)	0
<i>Gaza</i>	0	0	250

2. Baseline Purpose/Objectives and Questions

2.1 Baseline Purpose & Objectives

This report summarizes the key findings and results of the baseline assessment for Right To Play's EQIE program conducted in late November/early December 2020. As a baseline assessment for a quasi-experimental impact evaluation of the EQIE project; this report provides baseline values for the EQIE project indicators and informs EQIE programming by identifying strengths and weaknesses of students' reading abilities.

The baseline evaluation aims to:

1. Assess the program's starting point
2. Inform program design and strategy
3. Establish baseline measurements against which improvement and impact will be measured during midline and endline evaluations.

EQIE Project Indicators

A full description of the indicators that will be tracked as part of the EQIE project as well as the analysis framework for each indicator is included in Annex A.

2.2 Baseline Evaluation Questions

Aligned to the evaluation purpose and scope, this evaluation seeks to answer the following evaluation questions and sub-questions:

1. How similar are intervention and comparison schools?
2. What can the baseline assessment tell us about students' reading skills?
 - a. How do these reading skills differ by equity indicators³
3. What can the baseline assessment tell us about students' life skills?
 - a. How do these life skills differ by equity indicators?
4. What supplemental reading activities are available to children (in homes and communities)?
 - a. How do reported reading activities differ by caregivers and children?
 - b. Of those activities, which do children take part in? How frequently?
 - i. How does participation differ by equity indicators?

³ Equity indicators to be examined: sex, SES, ECD attendance, home literacy environment, language of instruction, ethnicity, disability, home language, child time patterns, school setting (dependent upon context)

- ii. How does participation in at-home reading activities differ by presence of reading materials in the home?
5. What teaching knowledge, attitudes, & practices regarding inclusive PBL/literacy instruction are common among teachers at baseline?
 - a. How do these knowledge, attitudes, and practices differ by teacher background (sex, training status (RTP training, No RTP training), # of teaching years, teacher qualification, and grade
 - b. How do teachers adapt teaching practices for CWD and gender-responsiveness?
 - c. To what extent does RTP training influence teacher knowledge, attitudes, and practices regarding inclusive PBL/literacy instruction?
 - d. How does teaching practice correlate with teacher well-being?
6. What is the current state of schools (considered for CWD and without) in terms of:
 - a. Physical safety?
 - b. Emotional safety/emotional well-being?
 - c. Materials?
7. What is the capacity (measured by knowledge, attitudes, and practice) of community members in creation and maintenance of safe learning environments?
 - a. Does capacity correlate with higher outcomes in terms of school physical safety and emotional safety/emotional well-being?
 - b. What barriers do community members face in creation and maintenance of safe learning environments, despite knowledge and attitudes?
8. What is the current state of community members and leaders in terms of knowledge regarding ways to address barriers to education faced by girls, boys, and CWD?
9. What might explain the higher learning outcomes of girls versus boys across grades?

The findings of these evaluation questions, complimented with discussion with Right To Play Country Office (CO) and global staff, will provide insights to the following program design questions for the EQIE project:

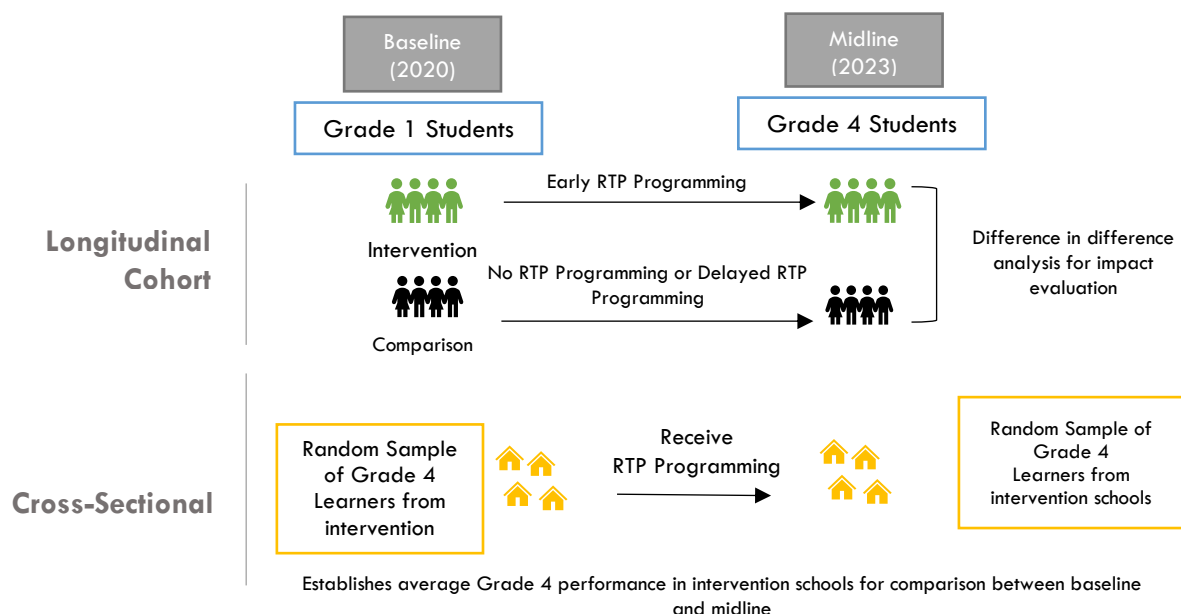
1. How should the EQIE program be tailored to meet children's current reading and life skill needs?
2. How should certain countries/context prioritize certain sections of the EQIE teacher Training curriculum to meet teachers where they are, building upon existing skills and addressing gaps?
3. How can EQIE programming support CSOs to encourage inclusive education?
4. How valid/appropriate are the implied assumptions of the logic model?
5. What specific equity indicators appear to be especially influential and ought to be paid attention to in various contexts?
6. How should the project intervention be tailored to positively influence/address the barriers that perpetuate gender inequality and social exclusion?

3. Methodology

3.1 Study Design

The evaluation employed a quasi-experimental design. This design maximizes Right To Play's ability to (1) examine the DAC criteria of effectiveness and impact at midline, while also (2) reporting on project indicators.

Figure 1: Evaluation design summary



The baseline established a sample of students to be followed longitudinally throughout the life of the program. These students were identified at baseline in Grade 1, in 20 intervention and 20 comparison schools. These same students should be followed and re-assessed in Grade 4, at the program's midline in 2023. Additionally, the project sought to establish the baseline performance of learners in Grade 4 (the age used to measure the EQIE indicator target). Given this, a cross-sectional random sample of Grade 4 learners from 20 intervention schools were assessed at baseline. At midline, another random sample of students can be assessed to compare average Grade 4 performance at baseline and midline.

In addition to the learners assessed, data were collected from other stakeholders in the education ecosystem. This includes interviews with teachers, principals, parents/caregivers, district education officials, and community leaders, as well as classroom observations.

For more information on the methodology for this evaluation, please refer to Annex I.

3.2 Limitations

Pandemic

The enumerator training and data collection period fell in November/December 2020. As such, the ongoing Coronavirus pandemic was a constant source of challenge and uncertainty, with impacts on data quality. Following government regulations, no personnel from outside MOE were allowed inside schools. **As such, the team shifted from using standard enumerators to MOE supervisors.** Some MOE struggled with touchscreen interfaces, and all supervisor training had to be conducted remotely over Zoom. In the end – with the immense effort of RTP Palestine staff, the InformEd Country Manager, and the supervisors themselves – a one-day training and half-day follow-up training were conducted to prepare the MOE to administer the Learner Assessment. While the data collection was ultimately successful, training conditions were far from ideal. Typically, training to administer a Learner Assessment of this complexity would include *at least* three days of in-person training in addition to two days of piloting/practicing at actual schools.

To mitigate risks of impact to data collection, the electronic data was closely monitored and addressed, especially during the first few days of data collection.

KoBo literacy widget

KoBo has a relatively new feature, a literacy widget, to accommodate reading skill assessments involving the timed tests of items such as words in a reading passage. However, there was one persistent bug in this feature: the data collector cannot mark an item at X seconds unless at least one word is already marked as incorrect. Annex I explains the workaround that was used. One additional challenge was that, unless data collectors marked the last item after pushing ‘stop,’ no data would be recorded. Thus, the Country Manager made videos to share with the data collectors to demonstrate the correct approach to both these issues, and in addition students were over-sampled to compensate for any data lost to these issues.

Children with disability

Disability is both relatively rare and highly individualized in terms of type and severity. Many disabilities may be ‘invisible’ if a comprehensive, technically-specialized system is not in place for testing and diagnosis. Thus, it is possible that there is some degree of mislabeling of CWD or non-CWD in the data for this report. To mitigate this challenge, information on disability is triangulated both from the school/teacher and from the students themselves.

Imperfect recall and self-reported data

Many of the tools rely on the honest and accurate memory of the respondent. Human memory is imperfect in general, and subject to self-promotion bias (both consciously and unconsciously). In at least one case – self-reported teacher practices – the data largely does not seem to be of value. Self-reported teacher PBL practices are not correlated either with classroom observation nor with student-reported teacher practices (though classroom observation and student-reported teacher practices are correlated with each other). Furthermore, there is no correlation between teacher self-reported practices and student reading and life skill outcomes. Thus, the classroom observation was used as a more valid method to calculate Indicator 1100.1.

The classroom observation, too, is also subject to potential bias, as most people will undoubtedly change their behavior when observed versus unobserved. Furthermore, a single observation captures only a limited representation of observed teacher behavior.

Quasi-experimental nature of research design

It was infeasible to randomize intervention and non-intervention schools, a second-best approach of quasi-experimental design was used for impact evaluation purposes. This means that a comparison group of schools was chosen – in this case, schools designated to receive the EQIE program during a later phase of intervention. Due to the programmatic approach, intervention and comparison schools are from different governorates. As seen in the report Section 4.3, intervention and comparison schools are similar enough for a quasi-experimental impact evaluation, especially in terms of student reading and life skills. However, there are some school-level differences, and it is possible that there are also unobserved differences between intervention and comparison schools. Controlling for observed differences will help mitigate bias in the midline impact analysis, but the quasi-experimental design cannot rule out the possibility that any midline differences in student reading and life skills is due to unobserved differences between intervention and comparison schools.

Limited sample size

Originally, a larger sample size was proposed to counter predicted attrition over the three years between baseline and midline. However, due to budget constraints the sample size was reduced to bare minimum. This leaves little room for error and increases the possibility that the sample size will be insufficient to detect impact. The risk is increased from the discovery that intra-class correlation (ICC) is closer to 0.14 rather than 0.1 initially used to calculate sample size. Possible mitigation strategies for the midline are discussed in Annexes E and I.

3.3 Analysis

Data was organized and pre-cleaned in Microsoft Excel, with further cleaning and analysis carried out using STATA. Where summary indices were created, steps were taken to ensure that the constituent pieces of each index were placed on a comparable scale before combining.

Summary statistics are used to describe the data. In some cases, the average may not summarize the data well if, for example, outcome variables are not normally distributed as is often the case for reading skills. This report highlights those cases and presents a histogram to represent the abnormal distribution more accurately.

For disaggregations, comparisons, and relationships that generalize from a sample to a larger population, statistical inference is used. Importantly, to maximize clarity this report ONLY presents disaggregations, comparisons, and relationships that were found to be statistically significant with at least 90% confidence. Exceptions include Programming Area Sample Characteristics, indicator reporting for the Project Management Framework (PMF, Annex B), and CWD – for which disaggregations are shown regardless of their statistical significance. In most cases, t-tests and regressions are used to evaluate statistical significance and the average relationship between variables. For student-level data, clustered t-tests and multilevel regression

models are used to adjust standard errors for the fact that students are grouped in schools and are likely to share some variation for being in the same learning environment.

4. Baseline Evaluation Findings

4.1 Programming Area Sample Characteristics

Programming area characteristics are summarized here to provide a description of the intervention context for the baseline assessment. These characteristics should be used to inform EQIE program design and tailor programming to the needs of the communities in which the EQIE project is working. The tables below present major characteristics for the intervention group only, since the more limited information on the comparison group is primarily for impact evaluation purposes.

School Environment

Table 2: Intervention school profile

Variable	# of schools (of 20)
Urban school	20
Boys school	9
Girls school	7
Mixed-sex school	4
Principal reports RTP has ever trained teachers	0
Principal reports other organization has trained teachers	1
Teachers facilitate reading club extracurricular activity	15
Teachers facilitate resource rooms for CWD and academically weak children	4
School has system for addressing student safety concerns	18
School has teacher code of conduct	17
School has access to trained point person for child protection referral and psychosocial support	17

Sampled Principal Profile

Table 3: Sampled principal profile at intervention schools

Variable	Average
% Female	75%
Age	46 years
Years of education	16.5 years
Years as a principal	8 years
Years as a principal at this school	4 years
Principal has received training from RTP	60%
Has considered leaving education as a career in past year	5%

Sampled Teacher Profile and Wellbeing

Table 4: Sampled teacher profile at intervention schools, by grade taught

Variable	Grade 1	Grade 4	Overall	N
% Female	81%	75%	78%	41
Age	43 years	41 years	42 years	41
Years of education	16 years	16 years	16 years	41
Permanent/civil service teacher	91%	95%	93%	41
Years as a teacher	12.6 years	13 years	13 years	41
Years as a teacher at this school	7.9 years	7.4 years	7.6 years	41
# of students in class of assessed students	28	30	29	41
# of CWD in all classes taught	1	1	1	39
# of lessons taught on average day	4.8	4.4	4.6	41
# of minutes spent teaching language/literacy on an average day	90	111	100	40
Has been trained by other organization on PBL or literacy instruction	33%	10%	22%	41
Facilitates reading club	47%	59%	53%	34
Facilitates resource rooms for CWD and academically weak children	12%	24%	18%	34
Average hours per week spent facilitating extracurricular activities for students	2.6	2.9	2.8	35
Teacher Wellbeing				
Agree or strongly agree that working as a teacher has been rewarding and worthwhile.	95%	85%	90%	41
Agree or strongly agree that there is a strong community of practice among fellow teachers.	95%	95%	95%	41
Has considered leaving education as a career in past year	19%	15%	17%	41
Has access to a trained point person for psychosocial support	50%	45%	48%	40
Has ever received training on providing psychosocial support to students	10%	25%	17%	41

Only one teacher reported having been trained by RTP, reportedly two years ago.

Sampled Student Profile

Table 5: Sampled student profile at intervention schools

Variable	Grade 1	Grade 4	Overall	N
% female	44%	46%	45%	731
Age	6 years	9.2 years	7.5 years	706
Attended ECD or preschool	98%	96%	97%	727
# of household members (including student)	6.1	6.5	6.2	724
Either refugee or internally-displaced person	12%	16%	13%	602
Teacher identified student as having a mild to moderate disability ^A	1%	3%	2%	726
Student reports at least some difficulty seeing words in books or on blackboard, even wearing glasses	1%	7%	4%	725
Student reports at least some difficulty hearing peoples' voices or music, even using a hearing aid	1%	3%	2%	725
Student reports at least some difficulty walking or climbing steps	5%	6%	5%	726
Student reports at least some difficulty remembering or concentrating on fun activities	11%	12%	11%	721
Student reports at least some difficulty feeding or dressing themselves	11%	3%	7%	726
Student reports at least some difficulty communicating with family/friends in home language	1%	1%	1%	723

^A1 student with vision impairment (does NOT wear glasses), 7 students with vision impairment (wears glasses), 1 student with physical impairment, 1 student with cognitive delay, and 1 student with speech delay

Almost no students speak a primary home language other than Arabic. Although dozens of students reported at least some difficulty on the WG Short Set questions, very few reported a lot of difficulty: 1 student for seeing words in books or on blackboard, 5 for walking or climbing steps, 6 for remembering or concentrating on fun activities, 6 for feeding or dressing themselves, and 1 for communicating with family/friends in home language.

Sampled caregiver Profile

Table 6: Sampled caregiver profile in intervention areas

Variable	Average	N
% female	48%	403
Age	38.4	403
Member of SMC	1%	403
Member of PTA	3%	403
Years of education	11.5	403

Almost no caregivers speak a primary home language other than Arabic. In response to the question: “How do you feel about your overall economic situation?”, 18% of caregivers reported ‘bad’ or ‘very bad.’ Thirty-nine percent reported ‘good’ or ‘very good.’

Sampled Supervisor Profile

Table 7: Sampled supervisor profile

Variable	Average	N
% female	64%	11
# of schools supervised	26	11
# of teachers supervised ^A	82	9
# of these teachers who have received training from RTP	6	7
Supervisor has received training from RTP	18%	11

^AThe questionnaire automatically constrained responses to maximum 99 teachers, but at least some supervisors oversee more than that

Interpretation

Intervention schools have many existing assets that EQIE programming can build upon/enhance. For example, almost all schools have extracurricular activities. Through the reportedly strong teacher communities of practice, schools that have reading clubs (15) can share valuable lessons/best practices with schools that have resource rooms for CwD and academically weak children (4), and visa-versa. Additionally, most schools have codes of conduct, systems for addressing student safety concerns, and access to trained point person for child protection referral and psychosocial support. It will be important for RTP first to listen to schools about how well these assets are working, so that EQIE can help fill any gaps.

As both principals and teachers are generally motivated and have substantial experience, it will also be important to actively involve teachers in brainstorming how EQIE programming can be adapted to their needs/implemented in their contexts as part of teacher training.

Very few caregivers are members of either SMCs or PTAs. If EQIE seeks to engage caregivers to improve the home literacy environment, SMCs and PTAs may be an insufficient channel of communication and engagement. Alternative methods may be required to reach a majority of caregivers, such as parent-teacher ‘learning together’ sessions and outside school learning days with parents, teachers, and children.

Finally, disability is a major challenge. CWD are few and far between, but have a diversity of unique needs. This means effectively addressing disability requires deep, detailed thought – there are few general needs and thus few general solutions. This report will attempt to pinpoint some of the specific needs. If RTP has not already done so, enlisting specialist support would be very helpful to ensure adequate solutions.

4.2 EQIE Indicator Baseline Values

The baseline assessment established the baseline value for each of the EQIE indicators outlined in the program. These values are presented in Annex B. These measurements provide some

insight into the level at which programming areas are starting for the program, against which to measure progress.

Importantly, however, these are mere indicators. Some are composed of a single item or construct, but many are composed of multiple items/constructs. Therefore, indicators are useful primarily as a management and reporting tool, but are limited in terms of deeper learning. More nuanced analyses of the indicators and the evaluation questions are presented in sections that follow.

4.3 How similar are intervention and comparison schools?

Out of 46 Learner Assessment variables tested for significant differences between intervention and comparison Grade 1 students, **only six variables were significantly different as shown in the table below**. The difference in sex is not surprising, given that the comparison sample of schools included 11 girls schools and only 3 boys schools (and 6 mixed schools), while the intervention sample of schools included 7 girls schools and 9 boys schools (and 4 mixed schools). Five of these variables were significantly correlated with reading or life skill outcomes, and should be controlled for during midline impact analysis.

Table 8: Statistically significant differences between Grade 1 students

Variable	Comparison	Intervention	Variable correlated with
% female	68%	44%~	Letter and phonemic awareness score*
% of students reporting at least some difficulty getting dressed	4%	11%*	Combined life skills score~
Has newspapers at home	15%	25%*	Letter score*
Has storybooks at home	41%	50%~	Combined life skills score*
# of good friends named by student	3.6	3.9~	Letter, phonemic awareness, and combined life skills score*
% of students reporting that a child in their school said something mean to another child in past week	23%	32%*	No outcomes

Significant difference or relationship with ~90% confidence, *95% confidence, **99% confidence.

Of the seven outcome variables tested for significant differences, none were significantly different between intervention and comparison Grade 1 students. These variables included % letters correct, % phonemic awareness correct, % concentration score, % communication score, % collaboration score, % critical thinking score, and whether the student knows what to do if they or a friend were being harmed.

Finally, 33 principal-reported variables were compared between intervention and comparison schools. Because this is not a case of generalizing to a larger population, no statistical inference

is used here. Instead, the difference between intervention and comparison schools appeared large enough to be considered in 12 variables. These variables are listed in the table below.

Table 9: Differences between intervention and comparison schools

Variable	Comparison	Intervention
Principal is female	90%	75%
Years as a principal	12	8
Urban school	70%	100%
Girls school	55%	35%
RTP has trained teachers	15%	0%
Another org has trained teachers on PBL or literacy instruction	45%	5%
RTP has trained principal	5%	60%
Teachers give extra activity: resource rooms for CWD and weak children in academic achievements	40%	20%
School safety plan includes: inclusion of CWD	50%	65%
School safety plan includes: bullying	5%	20%
School safety plan includes: violence	80%	45%
Teachers have access to trained psychosocial support person	85%	58%

As previously mentioned, the comparison sample includes a greater proportion of girls schools, which influences the difference in principal's sex. Six comparison schools are also rural, but all intervention schools are urban. Three comparison principals report that RTP has trained their teachers (albeit four to five years ago), and nine principals report that other organizations have trained their teachers in methodologies similar to PBL or literacy instruction. However, a majority of intervention principals report that RTP has trained them, unlike in the comparison group. A couple more comparison schools report resource rooms for CWD, although the school safety plan is slightly more likely to include CWD in intervention schools. Comparison school safety plans are less likely to include bullying but much more likely to include violence overall, and comparison teachers are more likely to have access to trained psychosocial support.

Interpretation

Especially based on student variables, intervention and comparison groups appear similar enough for a solid quasi-experimental impact evaluation. Nonetheless, the midline analysis will need to control for the significantly different student variables when investigating the impact of the EQIE project (with the exception of % of students reporting that a child in their school said something mean to another child in past week, which was uncorrelated with outcomes and may be considered an intermediate outcome itself).

Some of the principal-reported school-level variables should also be controlled for, in particular those with a strong theoretical relationship to student outcomes. This includes all variables listed above; with the exception of principal sex, years as a principal, and girls school (sex should already be controlled for at the student level).

4.4 What can the baseline tell us about students' reading skills?

This section presents figures, disaggregations, and relationships for all students in the sample, including comparison students, because no statistically significant differences were found in outcomes between intervention and comparison groups.

4.4.1 Findings Overall

The table below summarizes the average scores for the key reading skills measured, as well as the percent of students meeting the definition of 'reader' (student read at least five words correct in first 30 seconds, and therefore did not trigger the reading passage stop rule) and 'reader with comprehension' (student was a reader, and correctly answered at least 80% of reading comprehension questions). If a student did not qualify as a reader, the data collector kindly stopped the student, read the entire passage to the student, and then tested the student on comprehension questions for a measure of listening comprehension. Readers, on the other hand, were not read to at all, and therefore tested on reading comprehension.

There was no statistically significant difference in any skill between Grade 1 intervention (Int) and comparison (Comp) students. In the single reading skill tested for both grades (phonemic awareness), Grade 4 students scored significantly higher than Grade 1 students.

Table 10: Reading skill scores, by grade

Variable	Grade 1 (Comp)	Grade 1 (Int)	Grade 4 (Int)	Overall	N
% of students scoring zero on letters	7%	5%	N/A	6%	808
Average % letters correct	61%	58%	N/A	60%	808
Average % Phonemic Awareness correct	71%	71%	91%**	77%	1128
Average % MUW correct	N/A	N/A	80%	80%	318
% of students qualifying as readers ^A	N/A	N/A	89%	89%	320
Average % reading passage correct (accuracy)	N/A	N/A	92%	92%	289 readers
Average words correct per minute (fluency)	N/A	N/A	54	54 (SD =27)	289 readers
Average % reading comprehension	N/A	N/A	74%	74%	280 readers
Average % listening comprehension	N/A	N/A	41%	41%	28 nonreaders
% of students qualifying as readers with comprehension (RWC) ^B	N/A	N/A	55%	55%	322 ^C

Significant difference from Grade 1 intervention with ~90% confidence, *95% confidence, **99% confidence.

^AStudent read at least 5 words correct in first 30 seconds, and thus did not trigger the reading passage stop rule

^BStudent was a reader, and correctly answered at least 80% of reading comprehension questions

^CIncludes two students indirectly assessed through teacher interview for CWD (both were not readers with comprehension)

Interpretation

Overall performance on foundational reading skills is relatively low among Grade 1 students, especially on the letters they should know by this point in the curriculum. In contrast, overall performance is moderately high among Grade 4 students. However, the 11% of Grade 4 students who qualify as nonreaders are far behind their peers and need remedial support. The vast majority (81%) of nonreaders are boys – in other words, 17% of Grade 4 boys are nonreaders as compared to 5% of girls. Then there is a ‘middle tier’ of 34% of students – those who qualify as readers but who have not yet mastered comprehension – who need more reading practice and support on comprehension strategies. Note that the proportion of students in this middle tier of ‘readers without comprehension’ was not statistically different between boys and girls, while the likelihood of being in the top tier of readers with comprehension was much greater for girls (63%) than boys (48%). See Figures 7.2 and 7.3 below for histograms comparing the distribution of boy scores to the distribution of girl scores.

4.4.2 Skill-by-Skill Findings

The following pages ‘go beyond the average’ to take a closer look at individual reading skills and the specific items included in the subtests. As a summary statistic, the average assumes a normal (bell-shaped) distribution of scores. But in some cases, the distribution of scores is abnormal, and in these cases the average can be a poor summary of the true dynamics of student performance. The figure below illustrates an example of this for letters – although there are many students who scored around the average of 60% letters correct, there are also many students who either scored near 100% correct, or scored 0% correct (in a normal distribution, these extremes – sometimes called ‘tails’ – are assumed to taper off to a very small proportion of the population).

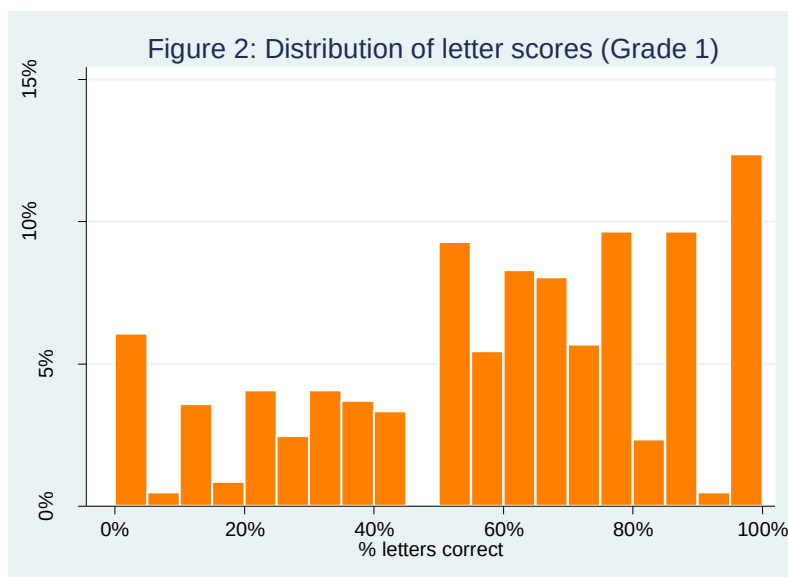
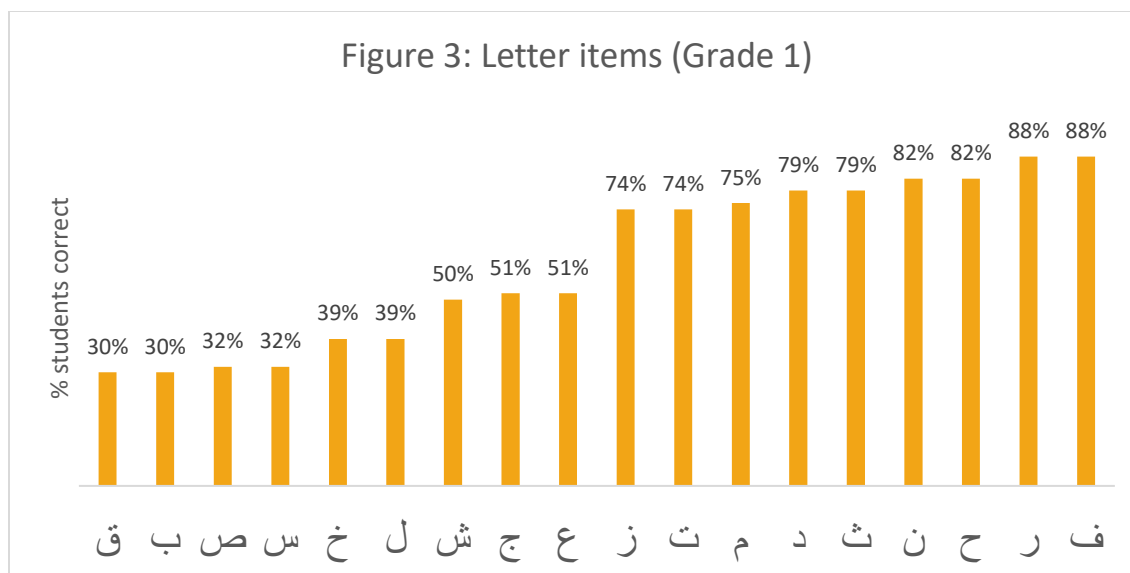
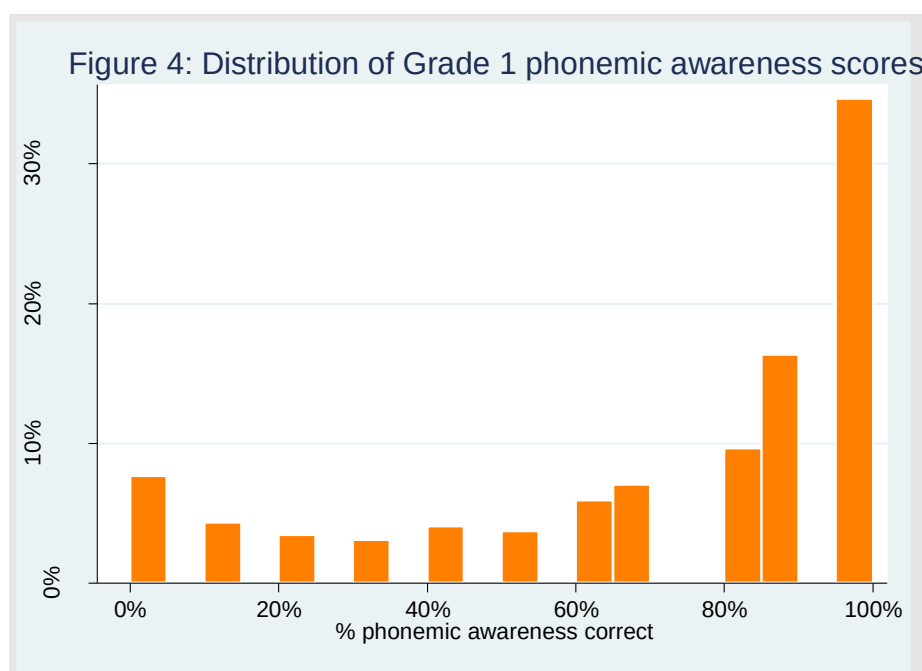


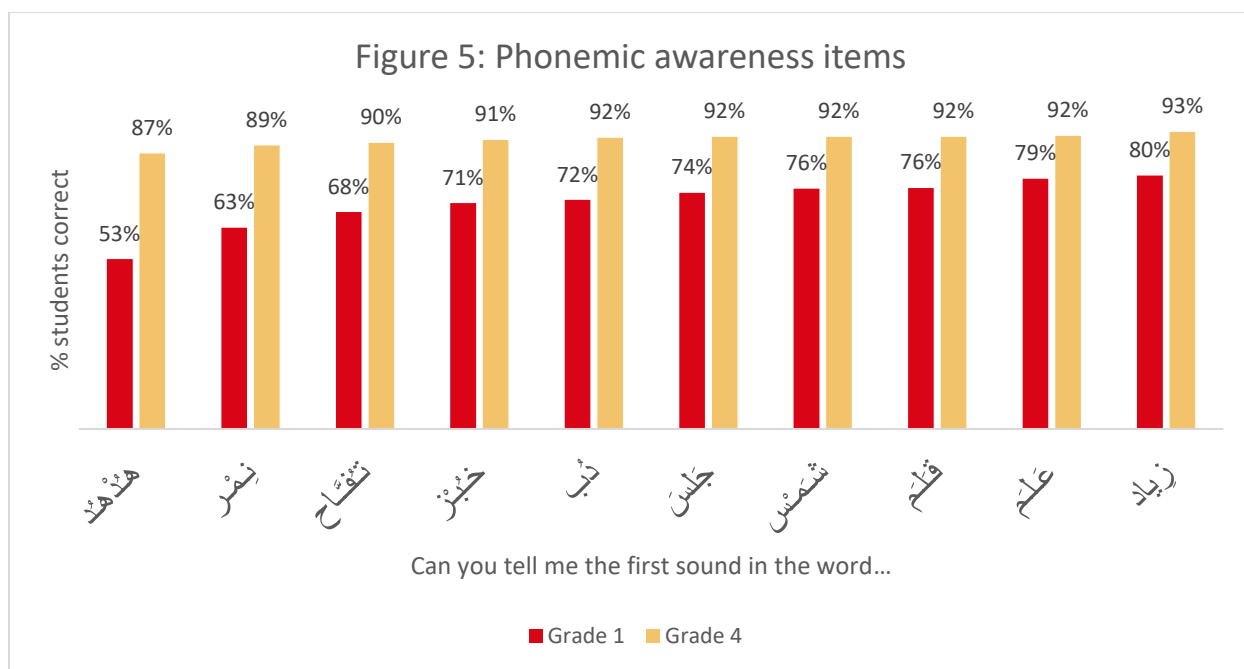
Figure 3 shows students’ average performance on specific letter items, organized from least to most often correctly recognized.



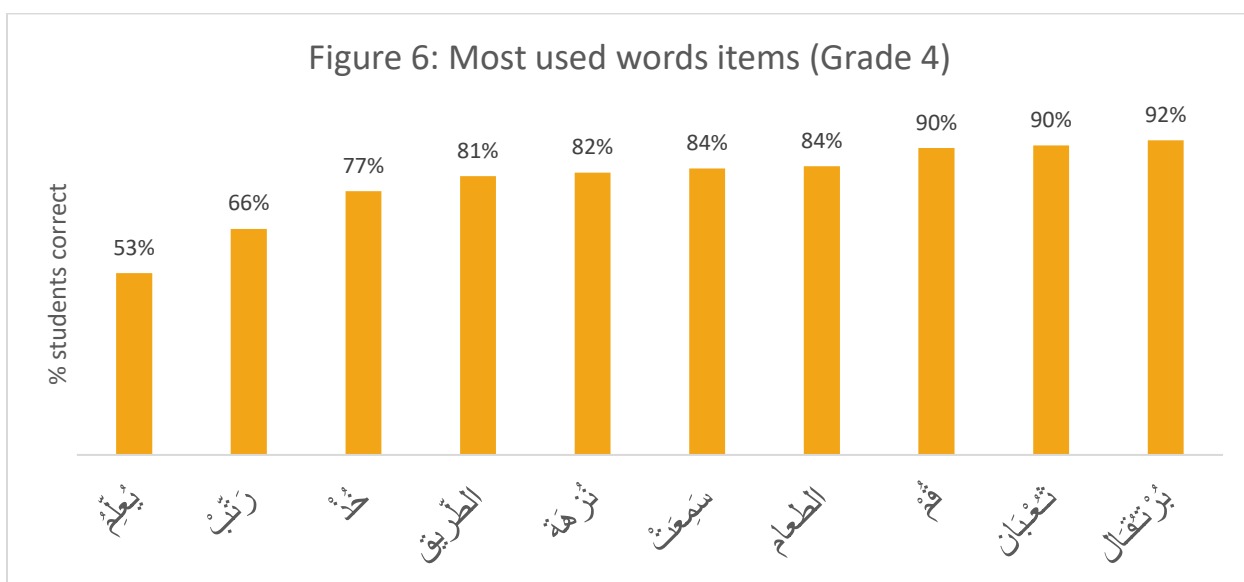
Phonemic awareness scores in Grade 1 were even more abnormally distributed than letters. As shown in the figure below, relatively few students scored near the average of 71% correct. Indeed, one-third of students scored 100% correct and 15% of students scored between 0-20% correct. For Grade 4, the average of 91% correct accurately describes how most students performed, so this report does not show the Grade 4 distribution.



The figure below shows how students performed on each words used in the phonemic awareness subtest. Student performance was more dependent on the specific word for Grade 1 than for Grade 4.

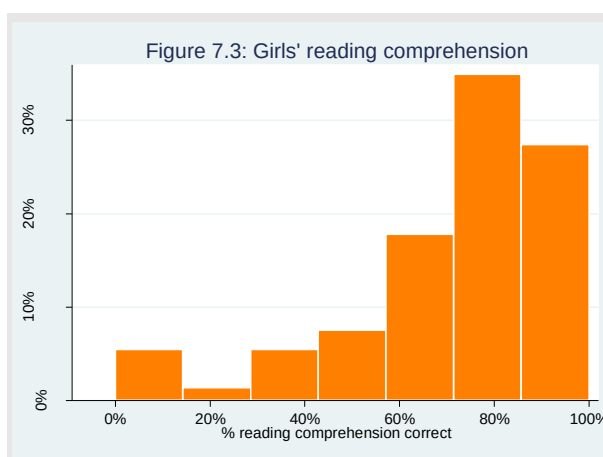
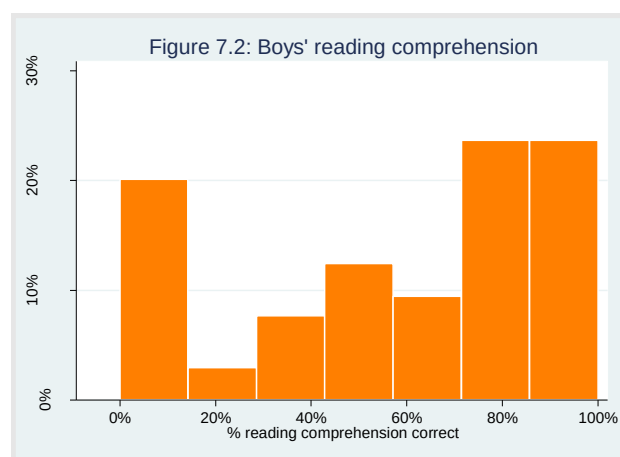
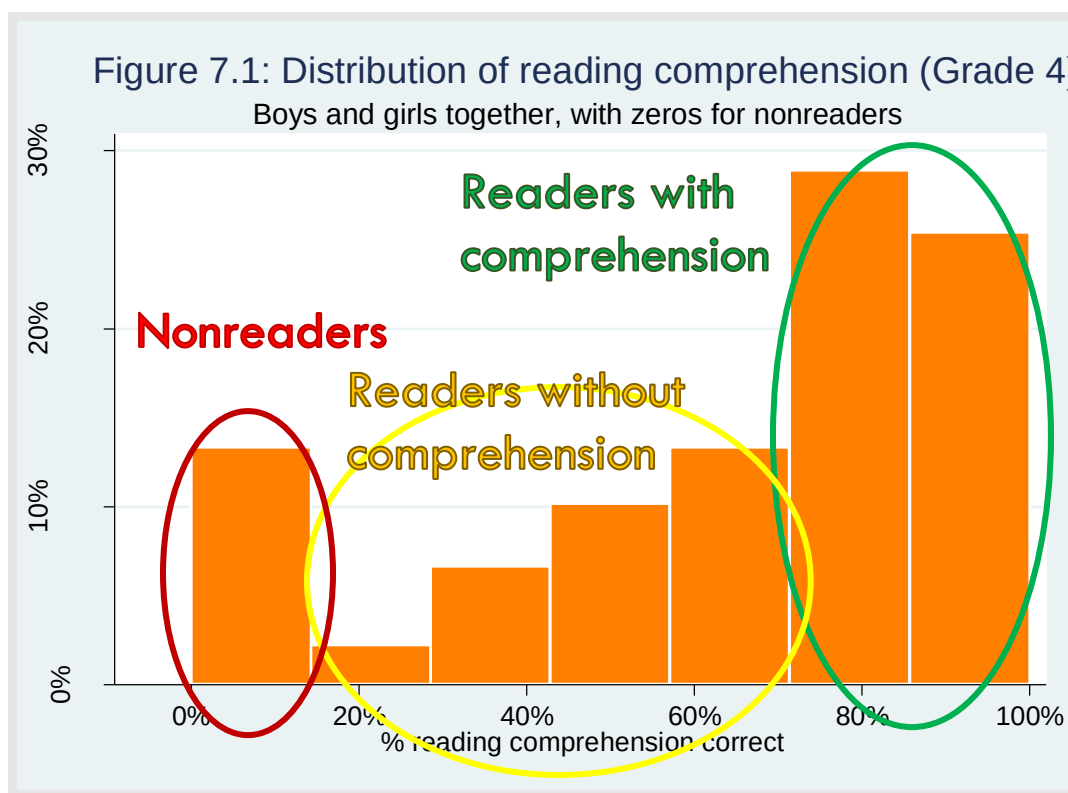


For most used words, student scores were normally distributed, thus the average describes the data well and the distribution is not shown here. The figure below shows the performance of Grade 4 students on each word in the most used words subtest. One potential pattern is that students struggled the most with words containing a shaddah.



For accuracy and fluency, the average sufficiently describes the distribution of scores. However, because reading with comprehension is the ultimate goal of reading as well as a good summary

of student's overall reading level, the distribution of reading comprehension is shown in the figures below. These figures include zero scores for students who qualified as nonreaders, and display the distribution of reading comprehension scores for all students, boys only, and girls only. Note that the category 'readers with comprehension' is the same as indicator 1000.1.



Interpretation

In many contexts around the world, reading skills are abnormally distributed – there are a sizeable proportion of students scoring very low, and a sizeable proportion scoring very high. Teachers need to be empowered to (A) accurately assess who is falling behind and (B) provide differentiated instruction – in some cases remedial instruction – to minimize these gaps before

they become larger in subsequent grades, putting students at increased risk of dropping out. Some potential strategies to address classes of diverse skills include pairing stronger with weaker students, and frequently varying instruction techniques so that students of every learning style have a chance to benefit.

There are many potential explanations for why students struggle with particular letters, but one possibility is the fact that many Arabic letters have similar shapes and are distinguished from each other only by very small differences, such as the number or position of dots. This poses a unique source of difficulty in learning to read Arabic,⁴ and might pose a particular challenge for some CWD, for example children with dyslexia.

4.4.3 Findings by Equity Indicators

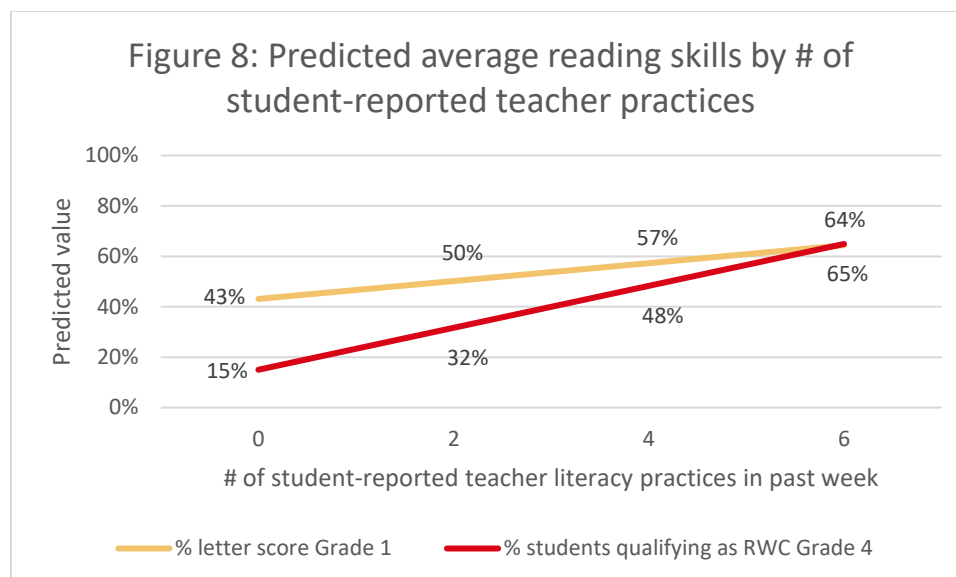
To investigate how reading skill performance varied by student background and environmental factors, a multilevel multivariate regression model was constructed (multilevel, to adjust standard errors for the grouping of students in schools). This model tests the relationship of reading skills to all student background and environmental variables simultaneously – that is, it answers the question: what is the relationship of variable X to reading skills, holding all other variables constant? Importantly, however, these are correlational relationships and should not be interpreted as a conclusive sign of causation. See Annex F for full regression table.

In general, student background variables were not significantly related to reading skills (while relationships to one or two skills were significant, there was no consistent pattern, and sometimes one relationship would be positive while the other would be negative). An important exception is CWD – this is discussed further below.

While girls did score higher on average, this is largely explained by the fact that performance in girls schools was significantly higher than either mixed or boys schools (in other words, in mixed schools girls and boys score similarly to each other). See Section 4.11 of this report for potential explanations for the greater average reading skill performance of girls versus boys.

Some environmental factors did have consistent significant relationships across reading skills. In Grade 1 and 4, the sum of student-reported teacher literacy practices in the past week was positively related to at least two reading skills. The figure below illustrates this relationship.

⁴ See, for example: <https://www.sciencedaily.com/releases/2010/08/100831102621.htm>



Additionally, the number of household members (for Grades 1 and 4) and number of environmental threats (harsh or violent child-child or child-adult interactions, for Grade 4) were negatively related to at least two reading skills. Participation in extracurricular literacy activities like reading clubs and media activities were positively correlated with reading skills in Grade 4. Interestingly, a short index of positive discipline (number of positive techniques minus number of negative techniques) was negatively correlated with two reading skills in Grade 4. The table below summarizes these relationships:

Table 7: Summary of consistent significant relationships with reading skills

Variable	Grade 1	Grade 4
Attends a girls school	1 skill (+)	6 skills (+)
# of student-reported teacher literacy practices in past week	2 skills (+)	2 skills (+)
# of student-reported household members	2 skills (-)	4 skills (-)
# of student-reported environmental threats	None	2 skills (-)
Student-reported participation in extracurricular literacy activities	1 skill (-)	2 skills (+)
Principal-reported positive discipline	None	2 skills (-)

+ positive significant relationship with reading skills, - negative significant relationship

The table below compares the average scores of CWD to their peers. Statistical inference is not used here to investigate significant differences because (A) all CWD in the target grades were sampled and (B) the number of CWD is too small to reliably use statistical inference. In a sense, all of the differences presented below could be considered significant (except where scores are practically the same, such as accuracy and reading comprehension).

Table 8: Reading skills of CWD versus peers

Variable	Non-CWD	CWD	N CWD
Grade 1			
Average % letters correct	60%	54%	9
Average % PA correct	71%	50%	9
Grade 4			
Average % PA correct	91%	80%	10
Average % MUW correct	80%	73%	10
% of students qualify as readers	89%	75%	12
Average % reading passage correct (accuracy)	92%	92%	9
Words correct per minute (fluency)	54	45	9
Average % reading comprehension correct	74%	76%	9
% of students qualify as readers with comprehension	55%	42%	12

Interpretation

With the exception of CWD (see below), there does not seem to be a strong case for programmatic focus on one or another type of student. Instead, EQIE should focus on students who are falling behind (regardless of background), as well as on ensuring existing environmental strengths are maximized and available to all students, and environmental deficits are addressed.

Interactions with people – either household members, teachers, or extracurricular activity facilitators – was much more correlated with reading skills than were number of types of reading materials in the home. This may speak to the importance of focusing on human resources to support students' learning, rather than material resources (although both are necessary and complementary). Students also need an abundance of time to practice their reading skills outside of school through extracurricular activities.

CWD indeed score lower than their peers in all reading skills, with one important exception. Those CWD who qualify as readers score equal to their peers in accuracy and comprehension. This is an important message to share – with the right supports, CWD can achieve just as much as their peers can.

4.5 What can the baseline tell us about students' life skills?

This section presents figures, disaggregations, and relationships for all students in the sample, including comparison students, because no statistically significant differences were found in outcomes between intervention and comparison groups.

4.5.1 Findings Overall

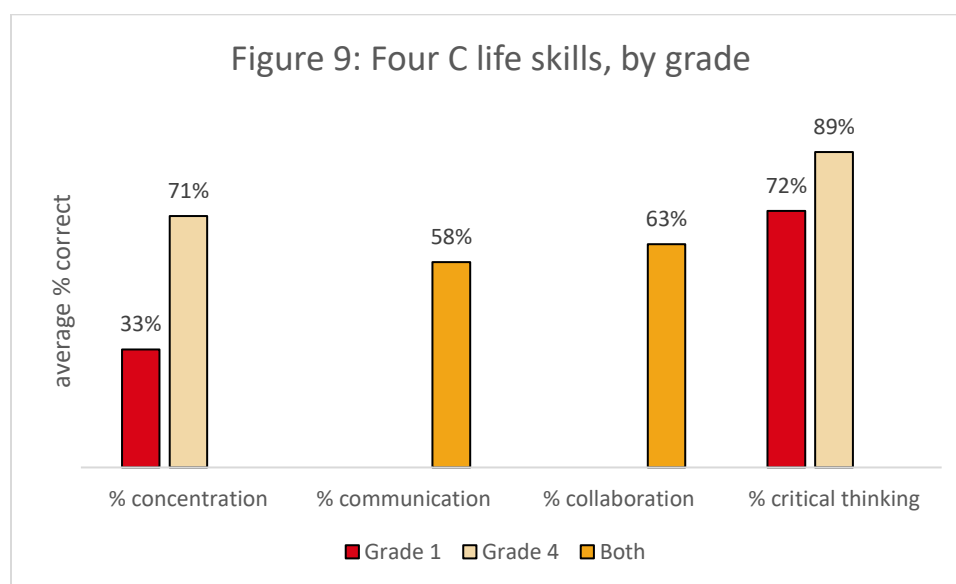
The table below presents both the core EQIE life skills for Palestine (the 'Four Cs') and secondary life skills for which some information was gathered as part of this data collection.

Table 9: Life skill scores, by grade

Variable	Grade 1 (Comp)	Grade 1 (Int)	Grade 4 (Int)	Overall	N
Core Four C Life Skills					
Average % concentration score	33%	34%	71%**	44%	1107
Average % communication score	56%	58%	61%	58%	1127
Average % collaboration score	59%	63%	66%	63%	1123
Average % critical thinking score	75%	70%	89%**	77%	1066
Average % combined Four C life skills score	55%	56%	71%**	60%	1127
Secondary Life Skills					
Child protection: Student knows what to do if they or a friend are being harmed	76%	81%	80%	79%	1095
Social network size: # of household members	6.2	6.1	6.5~	6.2	1122
Social network size: # of good friends	3.6	3.9	4.9**	4.1	1107
Social network size: # of adults at school student speaks to almost every day	1.9	2.2	3.1**	2.4	1064
Empathy: % score identifying sadness in others	49%	49%	52%	50%	1119

Significant difference from Grade 1 intervention with ~90% confidence, *95% confidence, **99% confidence.

There was no statistically significant difference in any skill between Grade 1 intervention (Int) and comparison (Comp) students. Among the Four C life skills, Grade 4 students scored significantly higher than Grade 1 students only in concentration and critical thinking. The two grades were statistically indistinguishable in communication and collaboration. The figure below presents this information graphically.

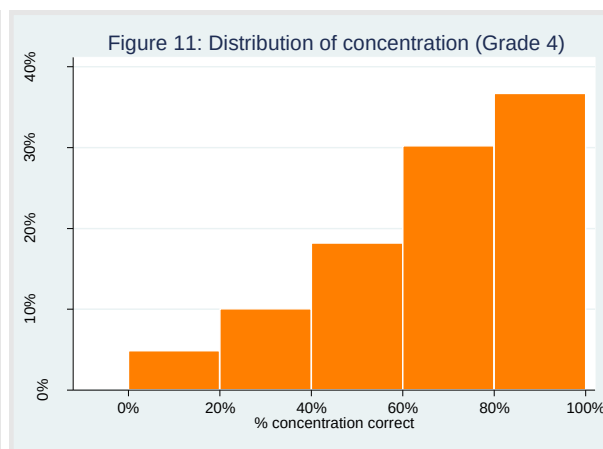
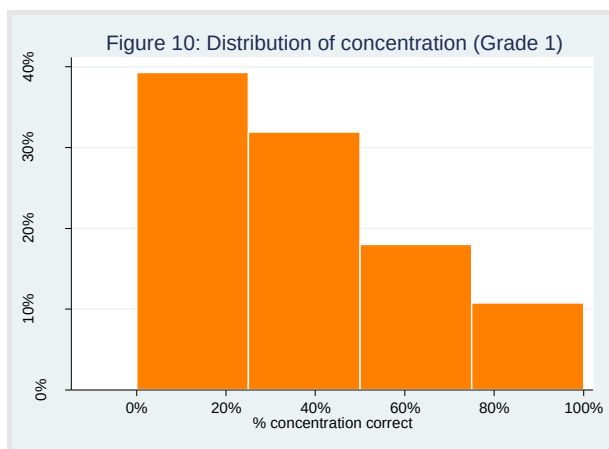


Interpretation

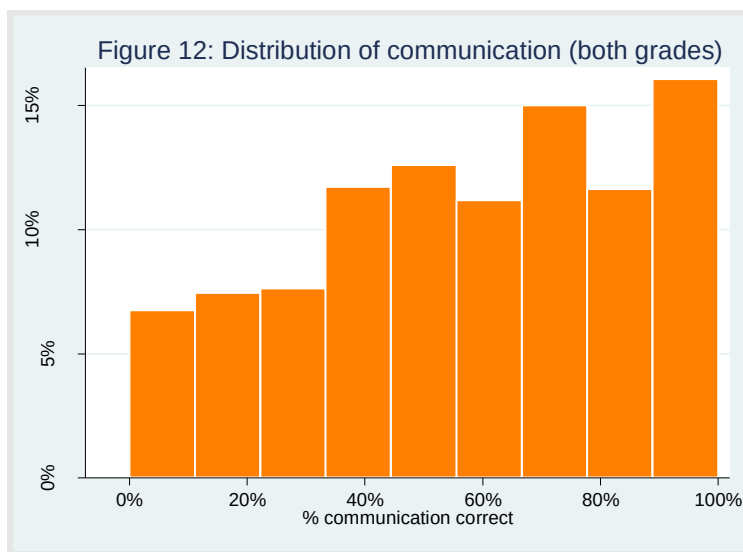
Grade 1 students struggled the most on the concentration task, even though the task was shortened (3 instead of 4 drawings) to make it easier for that grade. Thus, while there is room for improvement in all Four C skills, the need is most acute for Grade 1 concentration.

4.5.2 Skill-by-Skill Findings

The distribution of concentration score is the reverse for Grade 1 from Grade 4, as shown in the figures below. Indeed, nearly 40% of Grade 1 students could not complete even a single drawing. Interestingly, for both Grade 1 and 4 the second drawing was much easier than the third drawing.



Grade 1 and Grade 4 both had a somewhat more ‘flat’ distribution for communication, and because the scores were not significantly different by grade, this distribution is presented below for both grades combined.



In terms of items students scored the lowest on, half of Grade 1 and one-third of Grade 4 students reported not talking to anyone at home when they are sad. One-third of Grade 1 students and one-quarter of Grade 4 students do not talk to anyone at home when they have a problem with a friend. Half of Grade 1 and 42% of Grade 4 students do not talk to a friend when they are sad, and over 40% of both grades do not ask a friend for help when they have a problem with another child. When sad, Grade 4 students are more likely than Grade 1 students to talk to someone at home and much less likely to talk to an adult at school.

For collaboration and critical thinking, the distributions were skewed toward the average, and thus this report does not present their distributions. In terms of the lowest-scoring collaboration questions, half of students in both grades did not work with friends on coursework/homework in the past week, and over half of students did not work with a friend to accomplish a common goal. For the conflict questions, most students scored high. The only exception was that between 20-35% of Grade 1 students could only give one appropriate response to each hypothetical situation.

Interpretation

To boost communication and collaboration EQIE can help students see their friends and household members as assets to support them when they have problems, or to work toward common goals/coursework. Teachers can play a constructive role here, especially in terms of organizing/assigning more group work.

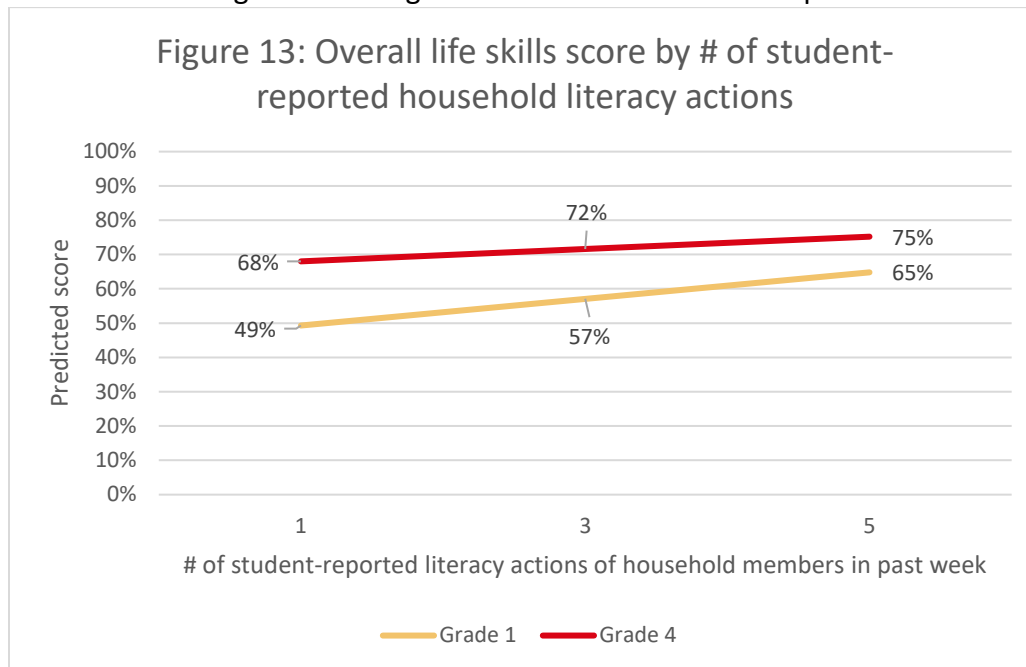
4.5.3 Findings by Equity Indicators

To investigate how life skill performance varied by student background and environmental factors, a multilevel multivariate regression model was constructed (multilevel, to adjust standard errors for the grouping of students in schools). This model tests the relationship of life skills to all student background and environmental variables simultaneously – that is, it answers the question: what is the relationship of variable (X) to life skills, holding all other variables constant? Importantly, however, these are correlational relationships and should not be interpreted as a conclusive sign of causation. See Annex F for full regression table.

Life skills were not significantly related to student background variables (while relationships to one or two skills were significant, there was no consistent pattern, and sometimes one relationship would be positive while the other would be negative). An important exception is CWD – this is discussed further below.

There was no consistent pattern between life skills and student sex – neither at an individual level, nor at the school level. However, other environmental factors did have consistent significant relationships across life skills. Interestingly, the student-reported number of

household member literacy actions in the past week⁵ was significantly correlated with many life skills, and not with reading skills. The figure below illustrates an example of this relationship.



The table below summarizes the other patterns of relationships. Student-reported teacher literacy practices in the past week were also positively correlated with life skills in both grades, as was student-reported participation in extracurricular literacy activities in Grade 1. Interestingly, the number of student-reported environmental threats (harsh or violent child-child or child-adult interactions) was positively correlated with two life skills in Grade 1.

Table 10: Summary of consistent significant relationships with life skills

Variable	Grade 1	Grade 4
# of student-reported teacher literacy practices in past week	3 skills (+)	2 skills (+)
# of student-reported literacy actions of others in household in past week	4 skills (+)	2 skills (+)
# of student-reported environmental threats	2 skills (+)	None
Student-reported participation in extracurricular literacy activities	3 skills (+)	None

+ positive significant relationship with life skills, - negative significant relationship

Count of significant relationships does not include combined life skills

⁵ In the past week: Have you seen anyone in your home reading? Has someone in your family helped you with your homework? Has a family member read to you? Has a family member told you a story? Has a family member played literacy games with you?

The table below compares the average scores of CWD to their peers. Statistical inference is not used here to investigate significant differences because (a) all CWD in the target grades were sampled and (b) the number of CWD is too small to reliably use statistical inference. In a sense, all differences presented below should be considered significant (except where scores are practically the same, such as critical thinking).

Table 11: Life skills of CWD versus their peers

Variable	Non-CWD	CWD	N CWD
Grade 1			
Average % concentration score	33%	44%	11
Average % communication score	57%	50%	12
Average % collaboration score	61%	51%	12
Average % critical thinking score	72%	76%	12
Average % overall life skill score	56%	55%	12
Grade 4			
Average % concentration score	72%	50%	11
Average % communication score	62%	50%	12
Average % collaboration score	67%	52%	11
Average % critical thinking score	89%	91%	10
Average % overall life skill score	72%	58%	12

Interpretation

With the exception of CWD (see below), there does not seem to be a strong case for focusing life skills programming on one or another type of student. Instead, EQIE should focus on students who are falling behind (regardless of background), as well as on ensuring existing environmental strengths are maximized and available to all students, and environmental deficits are addressed.

There seems to be some connection between literacy activities – at home, at school, and extracurricular – and life skills. This could signal an opportunity to improve reading and life skills simultaneously, by enhancing existing literacy-supporting interactions and especially by ensuring more students have access to them.

There are many potential explanations for the counterintuitive positive correlation between harsh/violent interactions and Grade 1 life skills (specifically, communication and collaboration). There may be reverse causation – students who are better communicators or collaborators may be more sensitive to harsh/violent interactions and more likely to report these to data collectors. It is also possible that there is an optimal, non-zero level of environmental adversity necessary for children to develop skills such as communication and collaboration. In addition, the chaotic nature of this school year (separating students into sections and minimizing interactions) may have biased student responses in complex ways.

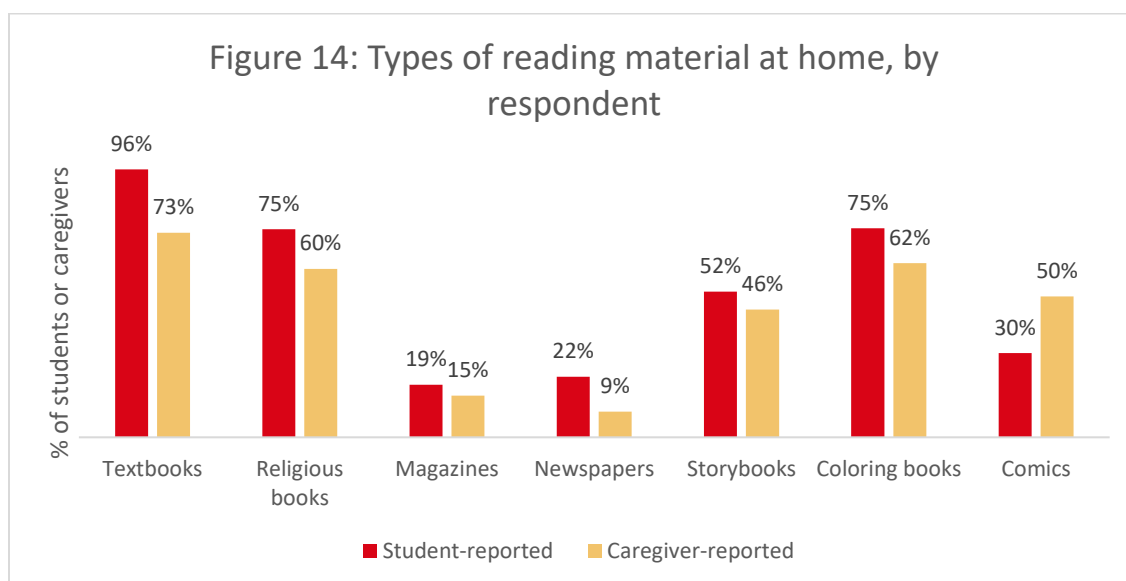
As with reading skills, it is important to acknowledge and share that while CWD lag behind their peers in many life skills, they are also capable of achieving the same level of life skills with the right supports (as in the case of critical thinking). As will be seen in the Section 4.7, CWD often struggle the most with communication (and perhaps by extension, collaboration) so it will be important to ensure all stakeholders – teachers, families, peers, etc. – are equipped for effective communication with CWD and to help support continual improvement of CWD’s communication skills. Otherwise, as seen in Table 19 above, initial gaps in lower grades can get larger over time since later knowledge and skills build upon earlier knowledge and skills.

4.6 What supplemental reading activities are available to children (in homes and communities)?

4.6.1 Findings from Learner Assessment

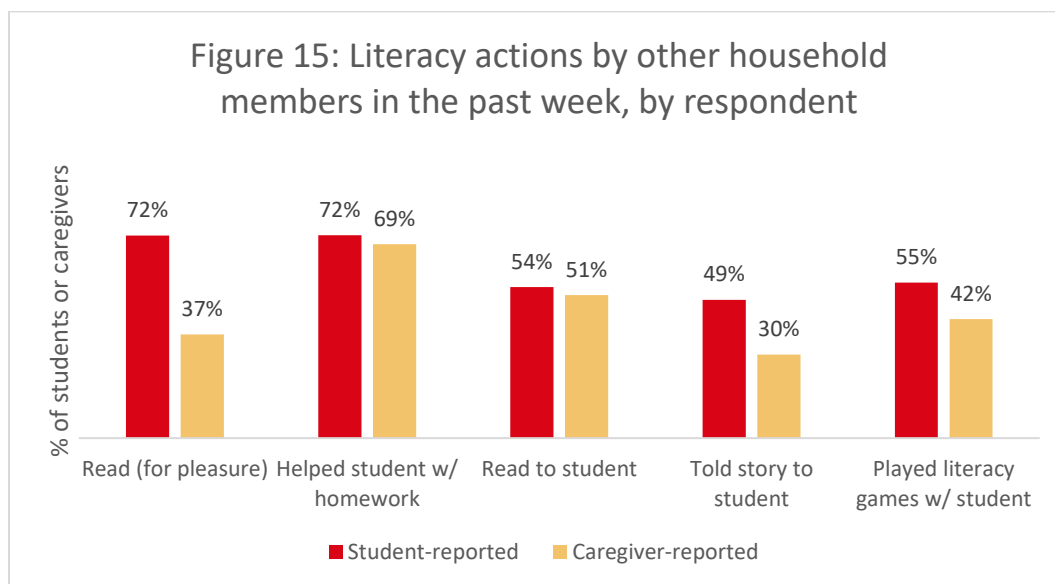
To compare between student-reported and caregiver-reported data, this section relies on data from intervention areas only. Unless specified, the figures presented are for Grades 1 and 4 combined.

First, the figure below presents the prevalence of different types of reading materials in the home as reported by students and caregivers. The pattern of reported prevalence is remarkably similar between student and caregiver respondents, with caregiver-reported prevalence somewhat lower than student-reported. The one exception is comics, which less students reported having in the home than caregivers (data collectors noted that children had difficulty distinguishing between storybooks and comics).



The figure below presents household literacy actions reported from the week prior to the data collection, as reported both by students (about someone in the household) and by caregivers (about themselves). Help with homework and reading to the student were almost equally

reported between students and caregivers, but students more frequently report seeing someone read, having a story told to them, or having someone play a literacy game with them (this may be because students were asked about *anyone* in the household, not only about their primary caregiver – indeed, the data collection team observed that children seemed to refer to their siblings more often than parents).



Students (but not caregivers) were also asked whether they had attended both in-person extracurricular literacy activities (reading camps, reading centers, reading days out, etc.) and media-based extracurricular literacy activities (radio, TV, mail, etc.). As expected during the COVID pandemic, students were more likely to report participation in media-based activities than in-person activities. Grade 4 students were more likely to participate than Grade 1 students.

Table 12: Extracurricular participation, by grade

Variable	Grade 1	Grade 4	Total	N
Student attends in-person extracurricular literacy activities	14%	28%*	20%	711
Student attended in-person extracurricular literacy activities in the past week	NS	NS	15%	710
Student attends media-based extracurricular literacy activities	46%	70%**	56%	716
Student attended media-based extracurricular literacy activities in the past week	40%	57%*	47%	715

Significant difference or relationship with ~90% confidence, *95% confidence, **99% confidence.

NS = no statistically significant difference, in which case the overall average applies to all sub-groups

4.6.2 Engagement in activities by equity indicators

To investigate how engagement in literacy-supporting actions/activities varied by student background, a multilevel multivariate regression model was constructed (multilevel, to adjust standard errors for the grouping of students in communities). This model tests the relationships with all student background variables simultaneously – that is, it answers the question: what is the relationship of variable X to engagement in literacy-supportive actions/activities, holding all other variables constant? Importantly, however, these are correlational relationships and should not be interpreted as a conclusive sign of causation.

The table below summarizes the statistically significant relationships that were most consistent across literacy-supportive actions/activities. Grade 4 students are more likely than Grade 1 peers to report access to child-friendly reading materials, household literacy actions, and extracurricular activities. Controlling for grade, younger students and those from larger households report less access, and in some cases students who are either refugees or IDPs report greater access.

Table 13: Relationships between literacy-supporting actions/activities and student background

Background Variable	Significant Correlations
Student is in Grade 4	Storybooks(+), coloring books(+), comics(+), seeing household member read(+), household member telling story to student(+), participation in media-based extracurricular activities(+)
Student age	Storybooks(-), coloring books(-), seeing household member read(-), household member telling story to student(-), participation in media-based extracurricular activities(-)
# of household members	Comics(-), seeing household member read(+), household member telling story to student(-), participation in in-person extracurricular activities(-), participation in media-based extracurricular activities(-)
Student is displaced (either refugee or IDP)	Comics(+), household member playing literacy game with student(+), participation in in-person extracurricular activities(+)

4.6.3 Engagement in activities by presence of reading materials in the home

A greater number of student-reported reading materials in the home was significantly correlated with greater likelihood of all student-reported household literacy actions and participation in all types of extracurricular activities in the past week. The one exception was a household member helping the student with homework in the past week, for which there was no significant correlation.

Interpretation

Most types of child-friendly reading materials are only available to about half of students at home. Similarly, only about half of students report being engaged in the most fun of the interactive literacy actions at home (being read to, being told a story, and playing a literacy game). Indeed, there is a strong correlation between reading materials and literacy-supportive actions/extracurricular activities.

There is scope to expand these opportunities to more students – both through making families aware of the benefits of these activities to students, as well as providing extracurricular ‘substitute’ activities for students, which will be especially important to students who come from large households where caregivers may have less time available for such activities themselves.

Importantly, these opportunities should be open to all students, even those as young as in Grade 1. As younger children benefit more from social learning,⁶ preference should be given to in-person extracurricular activities as soon as it is safe to do so.

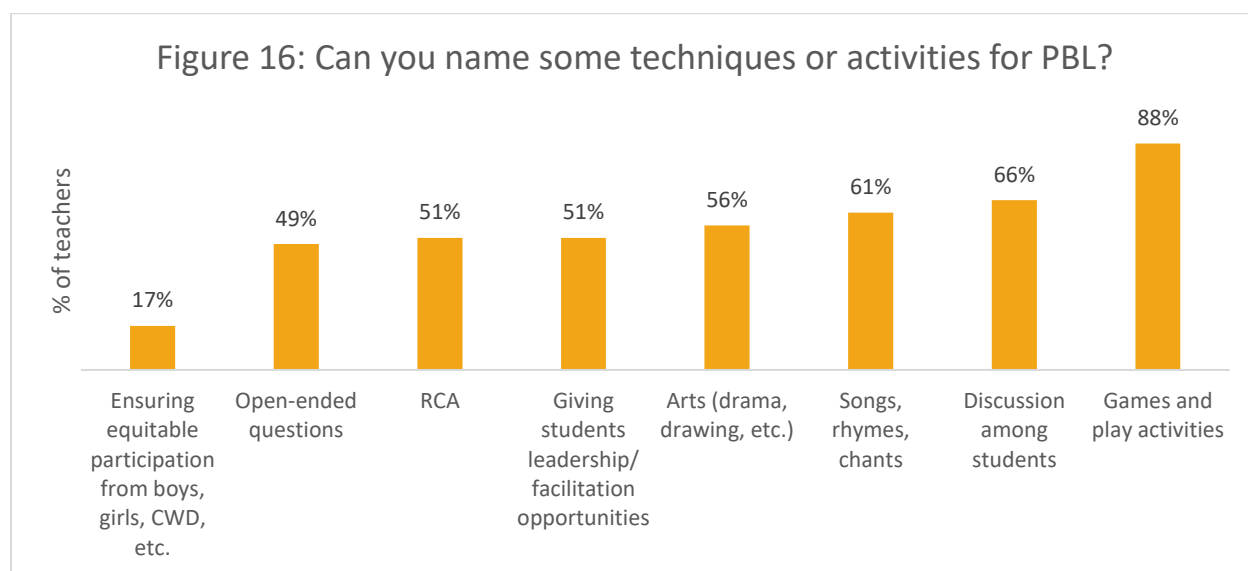
⁶ See for example: https://www.legofoundation.com/media/1063/learning-through-play_web.pdf

4.7 What teaching knowledge, attitudes, & practices regarding inclusive PBL/literacy instruction are common among teachers at baseline?

4.7.1 Overall findings: PBL Knowledge, Attitude, and Practices

See Teacher Profile and Wellbeing for background information on the 21 Grade 1 and 20 Grade 4 teachers included in this analysis (36 of whom were also observed using the Lesson Observation tool). All 41 teachers were from intervention schools.

Originally, two elements of knowledge were to be tested: knowledge of general PBL techniques and knowledge of the five reading skills. The question on knowledge of reading skills was dropped as too sensitive to ask teachers, but the PBL question was maintained: “Can you name for me some techniques or activities for play-based learning?” The figure below presents the proportion of teachers who mentioned each element of PBL. On average, teachers had knowledge of 55% of the PBL elements. There were no statistically significant differences between Grade 1 and Grade 4 teachers.



The teacher interview questions associated with attitude were asked only to teachers who had been trained by RTP. Only one teacher reported having been trained by RTP, and she saw both PBL and the five key reading skills as ‘very useful’ for children’s learning.

The teacher interview included a series of questions related to the frequency of PBL practices in the most recent full week of school. The table below presents the average frequencies for these techniques.

Table 14: In the most recent full week of school, how often^A did teacher...

Variable	Grade 1	Grade 4	Overall	N
give the students time for free play or to structure their own activities?	34%	55%~	44%	41
use subject-specific games and play activities that are linked to the lesson?	NS	NS	68%	41
give children time to ask each other questions and discuss the lesson with each other?	43%	66%*	54%	41
engage children in arts? (role play/theatre/drama, painting/drawing, etc.)	NS	NS	57%	41
engage children in singing, rhymes, or chants?	64%	44%~	54%	41
encourage the children to take leadership roles in class?	NS	NS	73%	41
use a game to teach or review letters?	72%	50%*	62%	41
use a game to teach or review sounds, rhymes, or syllables?	82%	40%**	62%	41
use a game to teach or review new words?	NS	NS	62%	41
read to students or ask them to read aloud?	NS	NS	94%	41
ask students to answer questions, summarize, or act/draw a story read in class?	NS	NS	60%	41

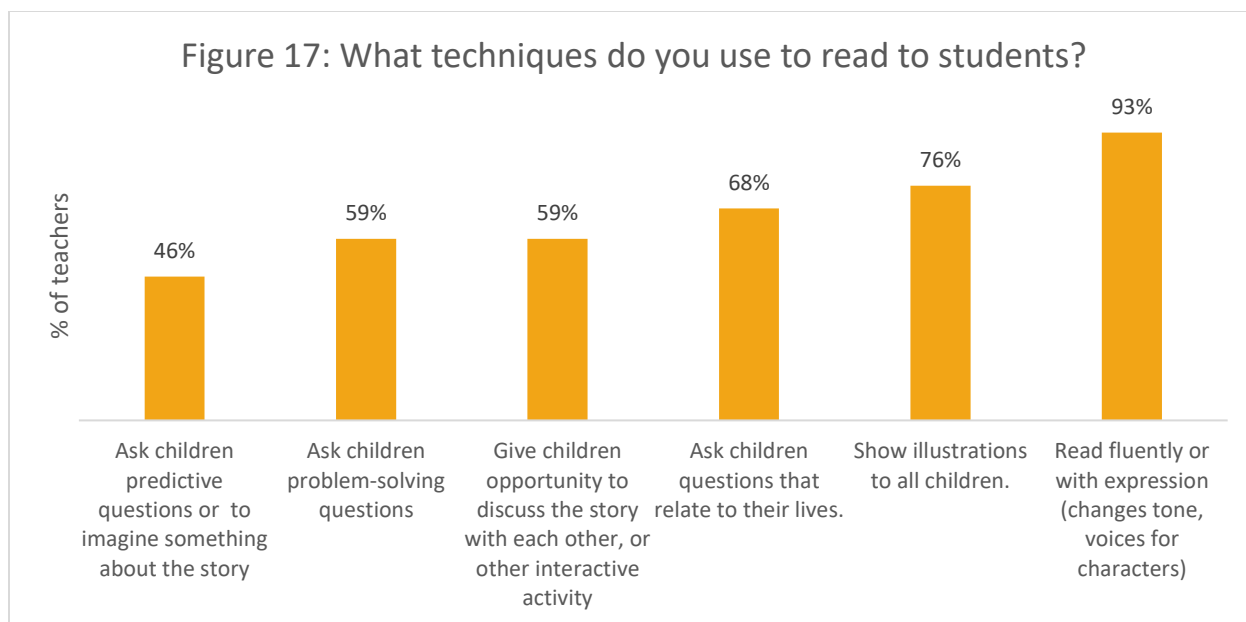
^AEvery day = 100%, Most days = 75%, Some days = 50%, One day = 25%, Not at all = 0%

Significant difference or relationship with ~90% confidence, *95% confidence, **99% confidence.

NS = no statistically significant difference, in which case the overall average applies to all sub-groups

A question about RCA was only asked of teachers who reported ever being trained by RTP. The only teacher who reported being trained by RTP also reported using the RCA technique every day of the week.

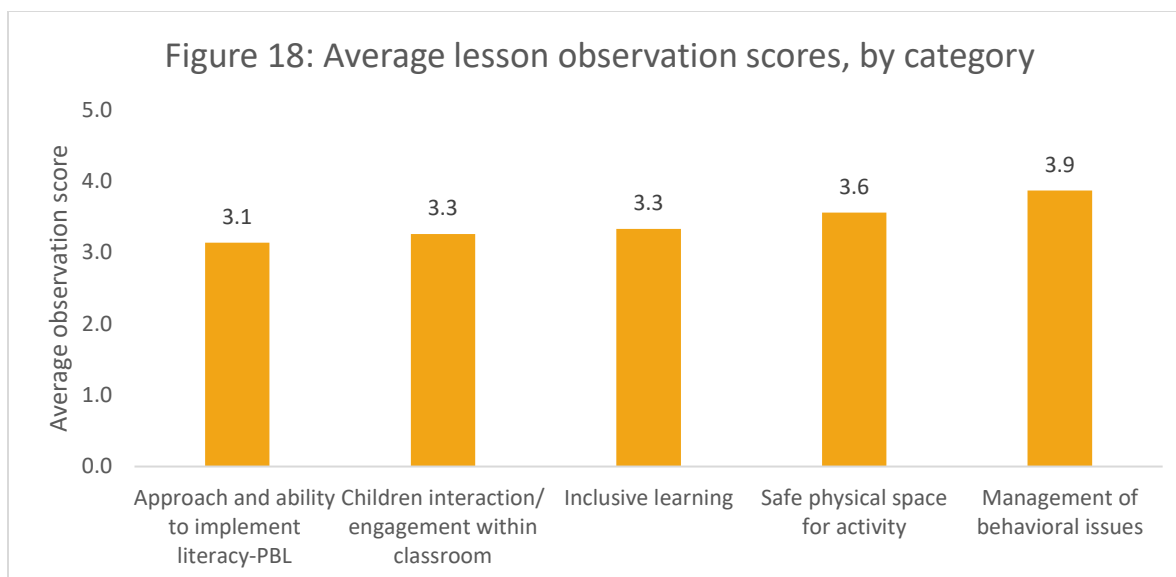
The figure below shows the proportion of teachers who reported various playful reading techniques. On average, teachers reported using 67% of these six techniques. There were no statistically significant differences between Grade 1 and Grade 4 teachers.



As noted in the Sample Teacher Profile and Wellbeing section, most teachers (all but two) also facilitate extracurricular activities, spending an average of 2.8 hours per week. Fifty-three percent of teachers facilitate reading clubs, and 18% of teachers facilitate resource rooms for CWD or academically weak students. Usually, a teacher facilitates one or the other, only four teachers facilitate both a reading club and a resource room.

In addition to the teacher interview, 39 lesson observations were conducted by MOE Supervisors (36 of these were of the same teacher from the teacher interviews). Most observations (35 of 39) were of literacy or language sessions. The observation tool used a five-point scale to assess teachers on 27 items split into five categories.⁷ The figure below shows the average score (from 1-5) in each of the five categories. Based on lesson observation, the percent of teachers who are demonstrating improved teaching practices to enhance children's reading skills (Indicator 1100.1) is 18%.

⁷ 1 = Not at all, 2 = Very little extent, 3 = To some extent, 4 = To a good extent, 5 = To a great extent



Some of the lowest-scored observation items came from the PBL section (average score in parentheses):

- Teacher uses open-ended questions to utilize RCA methodology (2.6)
- Teacher delegates roles/responsibilities to lead preparation of activity/play area, division of children groups, etc. (2.7)
- Teacher links “closing discussion” (RCA questions) with learning objective (2.7)

As well as from the inclusive learning section:

- Teacher includes CWD in all activities and discussions (1.9)
- Teacher modifies lesson and/or activity to include CWD when and/or if needed (2.2)
- Teacher encourages girls and/or boys equally to play non-traditional gender roles (2.9)

And the children interaction/engagement within classroom section:

- Children participate actively (2.2)

Very few items had a statistically significant difference between Grade 1 and Grade 4 teachers, as shown in the table below.

Table 15: Statistically significant lesson observation differences by grade

Item	Grade 1	Grade 4	N
Activity/play area is clean, and clear of equipment, obstacles, or hidden areas that may cause injuries (e.g. Desks are rearranged to ensure children safety)	3.9	3.1*	36
Teacher interjects and takes action when children display any type of bullying behavior towards their peers	4.1	3.4~	32
Children participate actively	1.9	2.6~	37

Significant difference or relationship with ~90% confidence, *95% confidence, **99% confidence.

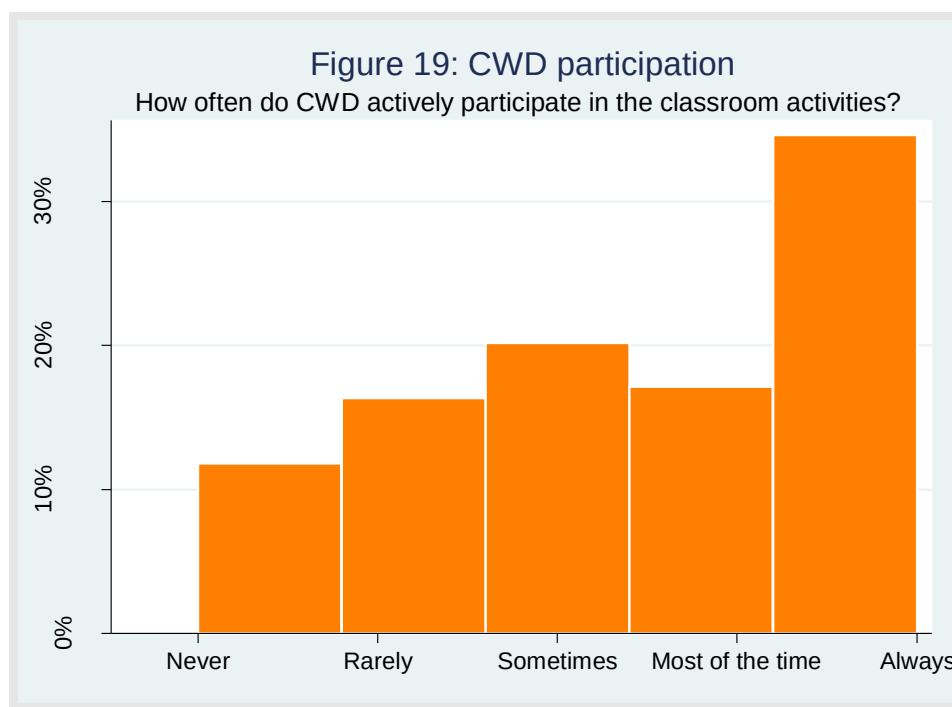
For detailed lesson observation data, see Annex G.

4.7.2 Adaptation of practice for CWD and gender-responsiveness

As previously mentioned, CWD enrolled in these schools are few and far between (at least those recognized by school staff – given the complexity of impairment/disability, there may be undiagnosed CWD who are ‘invisible’). Only 15 teachers report having CWD in their classrooms, and indeed observers only noted a grand total of 2 CWD in the lessons they observed.

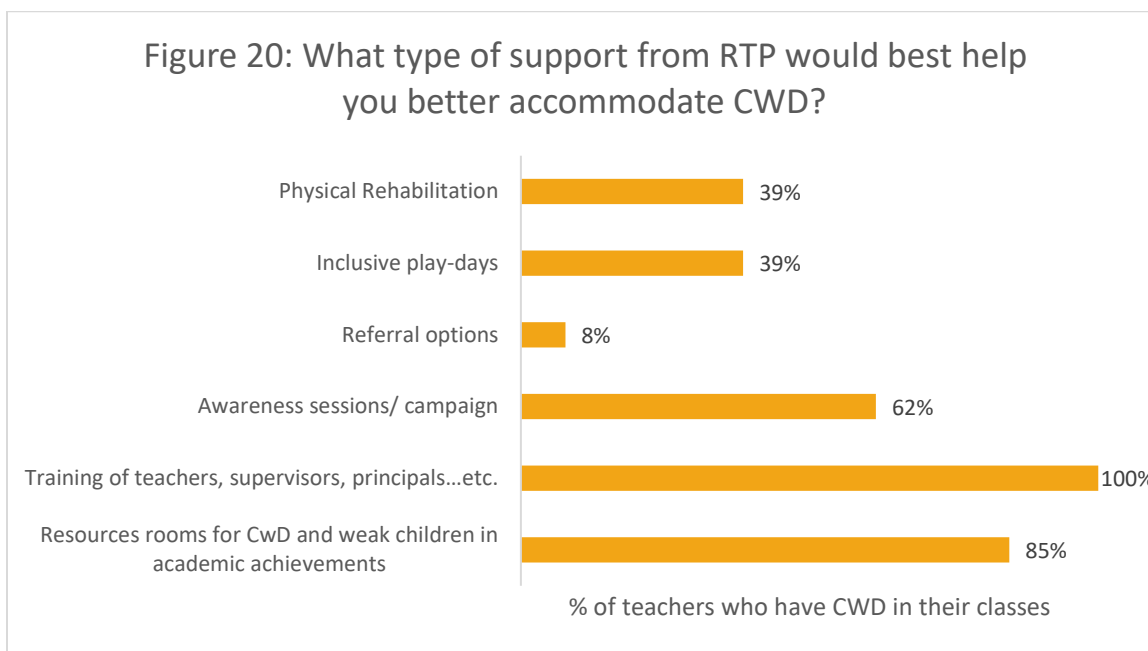
Most of the 15 teachers who report having CWD in their classrooms report being able to adapt their teaching to accommodate CWD (11 of 15). They give examples of special worksheets and toys, alternate communication strategies such as drawing and body language, communicating with parents, and integrating the CWD into activities with the other students.

The figure below shows wide variation in how often these teachers report CWD actively participate in class. When asked for an example of how CWD struggle to participate, teachers mentioned problems with communication, interaction with peers, and concentration. Of the nine teachers who have CWD and facilitate extracurricular activities, three say CWD rarely attend, and two teachers each say that CWD ‘sometimes,’ ‘most of the time,’ and ‘always’ attend.



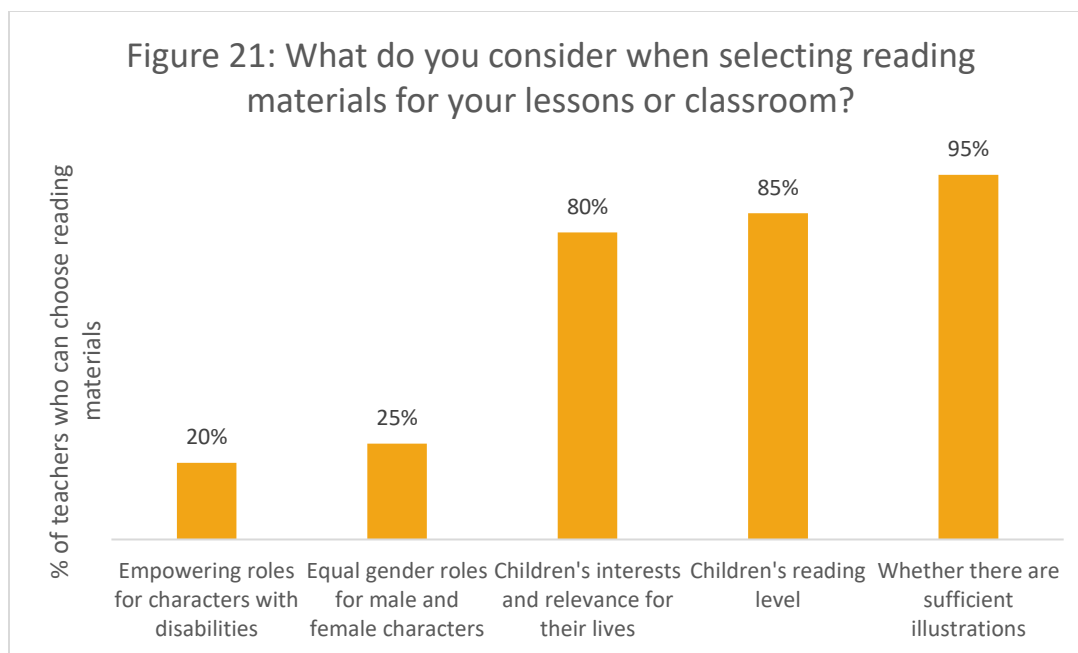
Five teachers report having no materials/equipment for CWD. Those who do have materials usually report pictures, cards, and other manipulable teaching and learning materials (TLM). In terms of the materials/equipment the teachers need, they request pictures, puzzles/shape games, play dough, cards, and word/letter games. Two teachers say they need an assistant teacher specialized in CWD, two others need a special room to teach CWD, and one teacher

needs a walking assistant. Turning from materials/equipment to other supports, the figure below presents the type of support requested of RTP to help teachers with CWD better accommodate these children in their classes.



Turning to gender-responsiveness, of the eight teachers in the four mixed-sex schools, four report that boys and girls participate equally. Three report that girls participate more than boys, and only one reports that boys participate more on average. Of the same eight teachers, five report that boys and girls perform/learn equally on average, while two report girls perform higher than boys and only one reports that boys perform higher than girls. In terms of leadership roles, two of these eight teachers report that girls take leadership roles more often, one teacher reports boys more often take leadership roles, two report equal balance naturally, and three report equal balance as ensured by the teacher. See Section 4.11 of this report for potential explanations for the disparities between girls and boys.

Relevant to both CWD and gender-responsiveness, 50% of teachers report they have a choice in selecting reading materials for their lessons/classroom. But as the figure below shows, very few take into consideration gender-responsive or CWD-empowering content.



4.7.3 Findings by teacher background and wellbeing

A teacher practice index was constructed, combining many of the self-reported Teacher Interview variables discussed above (see Annex A for details). However, none of the student reading or life skill outcomes were correlated with the self-reported teacher practices index. A separate teacher practices index was constructed using all the Lesson Observation categories except for physical safety (to stay consistent with the definition of Indicator 1100.1). This observed teacher practices index is significantly correlated with many (but not all) student outcomes, and also significantly correlated with student-reported teacher practices. Thus, the observed teacher practices index seems both more accurate and a better fit the EQIE project's theory of change, and so is the basis for calculating Indicator 1100.1.

Multivariate regression analysis was used to explore any potentially significant correlations between the observed teacher practices index and teacher background variables as well as self-reported wellbeing. Most of the background and wellbeing variables did not have any significant correlation with the observed teacher practices index, with three exceptions. Education level, number of years as a teacher, and a wellbeing index⁸ were positively correlated with the observed practices index. These are only correlations and cannot establish causality.

Interpretation

From a literacy perspective, many of the least-used playful reading techniques are related to asking a variety of comprehension questions to students or allowing students to discuss stories amongst each other. Increasing teachers' skill and frequency in using these techniques could help address the deficit in students' comprehension skills. As a bonus, these techniques apply to

⁸ The wellbeing index was constructed thusly: (rewarding/4) + (strongcop/4) + psychsupport – leavesector

students of any reading level, can enhance engagement/motivation of all students to read, and are also an opportunity to engage students who typically participate less.

From a general play-based perspective, many of both the teacher interview and lesson observation deficits revolve around the active/ interactive/ participatory level of children in the classroom – especially free time, discussions, and child leadership opportunities. These may be the greatest opportunities for the PBL approach to enhance teaching and learning.

Teachers appear to give some thought to CWD inclusion and gender-responsiveness. However, CWD inclusion can be challenging, especially given the wide variation in individual needs. For this reason, it may be important to reinforce channels for teachers to share their experience and best practices for CWD inclusion. This report lists some of their ideas, but much richer exchanges can take place directly between teachers. Regarding other RTP support, teachers themselves request training, followed by resource rooms, awareness campaigns, and materials/equipment such as pictures, puzzles/shape games, play dough, cards, and word/letter games.

With gender-responsiveness, only a minority of the intervention teachers face mixed-sex classrooms. However, all teachers can and should be more aware of the opportunity to boost all students’ sense of worth through selecting reading materials that challenge traditional gender norms and present a balance of boys, girls, and CWD in empowering roles.

4.8 What is the current state of schools?

4.8.1 Overall findings

Because much of the safety information comes from teachers and parents (as well as from principals and students), to be consistent this section presents data only from intervention schools.

Table 15: What objectives are included in school safety plan?

Objective	Principals	Teachers
Exploitation	5%	9%
Gender	10%	20%
Bullying	20%	14%
Gender-based violence	25%	36%
Other	35%	20%
Violence (general)	45%	56%
Teacher behavior	60%	56%
Inclusion of CWD	65%	43%
Hygiene	75%	81%

All 20 intervention principals report students come to school staff to address safety concerns inside this school, and 18 principals report the school has an active system for addressing student concerns. Half of the principals (10) report that they have a strategy, plan, or policy to address school safety that is either almost fully or full implemented. Nine principals report a plan that is moderately implemented, and one principal reports no plan at all. The table above presents

principal- and teacher-reported objectives that are included in the safety plan, in order of least to most mentioned objectives.

The table below shows the report of CWD and their peers about some aspects of their learning environment safety. In all cases, a minority of students report these harsh/violent child-child or child-adult interactions. However, the most prevalent behaviors are adults screaming/yelling, children teasing or saying mean things to each other, and children fighting. CWD are less likely to report almost all types of behaviors, although they report adult yelling at almost the same rate and report themselves having thrown something at another child slightly more than their non-CWD peers (though this behavior appears very rare in general). Importantly, no CWD report anyone left out of an activity or game, an indication of inclusion.

Table 16: In the last week at school...

Variable	Non-CWD	CWD ^A	N
A child teased another child to make them angry	37%	27%	698
A child left out another child from an activity or game	23%	0%	695
A child said something mean to another child?	41%	18%	692
A child did any of these things to you	22%	18%	698
Children got into a fight where someone was hit	34%	9%	698
Student (respondent) got into a fight where someone was hit	13%	9%	699
Someone threw something at a child to hurt them	23%	9%	700
Student (respondent) threw something at another child to hurt them	4%	8%	697
An adult screamed or yelled at a child in a threatening way	38%	42%	703
An adult hit or kicked a child	29%	17%	697

^A11-12 CWD in intervention schools who answered the questions.

Statistical significance not used here due to small number of CWD and the fact that all CWD were purposefully sampled from target grades.

4.8.2 Physical safety

The tables below present answers to the question “Do any of these aspects of the physical safety inside/outside the classroom environment need improvement?” for the 11 supervisors, 20 intervention principals, and 41 intervention teachers; as well as the answer to the question “What strategies can a school use to create a physically safe learning environment inside/outside the classroom?” for the 366 caregivers who answered the question. For ease of interpretation,

the tables are ordered from those aspects most identified as needing improvement on average (red) to those aspects least identified as needing improvement on average (green).

Table 17: Aspects of physical safety that need improvement inside the classroom

Safety Aspect	Supervisors	Principals	Teachers	Caregivers ^A	Average ^B
Sufficient space for games and active learning	73%	65%	48%	74%	65%
Adequate Furniture	46%	85%	53%	54%	59%
Internal design/coloring	64%	85%	28%	49%	56%
Good Lighting	46%	50%	40%	68%	51%
Painting	36%	70%	25%	36%	42%
Good Ventilation	46%	10%	38%	73%	42%
Electricity	9%	50%	20%	39%	30%
Accessible to CWD	27%	15%	33%	35%	27%
No hazards such as exposed foundation, cracked walls, exposed wires, hanging roofs, blind spots, etc.	9%	10%	10%	46%	19%
# of safety aspects identified	3.5	4.4	2.9	4.7	3.9

^AThe caregiver question was slightly different: “What strategies can a school use to create a physically safe learning environment inside the classroom?”

^BAverage with equal weights to each of the four categories of respondent

All respondents tended to identify sufficient space for games/active learning as one of the most needed intra-classroom improvements. Caregivers are much more likely to identify ventilation and hazards as strategies for intra-classroom safety, though this may be due in part to the different framing of the question. Principals are much more likely to highlight electricity, painting, internal design, and furniture as in need of improvement. Both caregivers and principals identify the most issues, while teachers identify the least issues.⁹

⁹ Note that comparisons between types of respondent are subjective judgments from the author, no statistical tests were run due to the wide variation in sample size and the fact that the sample of principals and teachers represent their entire populations, or near to their entire populations.

Table 18: Aspects of physical safety that need improvement outside the classroom

Safety Aspect	Supervisors	Principals	Teachers	Caregivers ^A	Average ^B
External area is protected from direct sunlight/rain	73%	80%	69%	72%	74%
WASH facilities / sanitation in good condition	55%	70%	41%	57%	56%
Fencing around school	46%	30%	28%	55%	40%
School location (near, safe roads, away from checkpoints...etc)	27%	40%	31%	54%	38%
External area is away from dangers such as road, deep/running water streams, exposed electric wires	36%	5%	21%	58%	30%
Accessible to CWD	18%	15%	36%	40%	27%
# of safety aspects identified	2.5	2.4	2.3	3.4	2.6

^AThe caregiver question was slightly different: "What strategies can a school use to create a physically safe learning environment outside the classroom?"

^BAverage with equal weights to each of the four categories of respondent

Caregivers also identified more strategies for exterior physical safety, in particular fencing, school location, and proximity to hazards. Principals called attention to WASH facilities as in need of improvement. However, all types of respondents most frequently highlighted protection of external area from rain/sunlight as in need of improvement.

4.8.3 Emotional safety/emotional well-being

The tables below present answers to the question "Do any of these aspects of the emotional safety of the classroom environment need improvement?" for the 11 supervisors, 20 intervention principals, and 41 intervention teachers; as well as the answer to the question "What strategies can a school use to create an emotionally safe learning environment?" for the 366 caregivers who answered the question. For ease of interpretation, the tables are ordered from those aspects most identified as needing improvement on average (red) to those aspects least identified as needing improvement on average (green).

Table 19: Aspects of emotional safety that need improvement

Safety Aspect	Supervisors	Principals	Teachers	Caregivers ^A	Average ^B
Teacher capacities overall	55%	75%	49%	70%	62%
Positive discipline	73%	45%	28%	49%	49%
Administration capacity	55%	25%	28%	47%	39%
Availability of adequate counseling service	46%	30%	31%	35%	35%
Bullying / violence	46%	20%	23%	29%	29%
Atmosphere	0%	70%	13%	31%	28%
access to qualified counselors / access to referral system	18%	30%	33%	32%	28%
Inclusion	9%	55%	18%	27%	27%
Gender	27%	15%	15%	20%	20%
Restrictions on returning to school for out-of-school children (CWD and girls specifically)	18%	25%	13%	20%	19%
# of safety aspects identified	3.2	3.5	2.3	3.2	3.0

^AThe caregiver question was slightly different: “What strategies can a school use to create an emotionally safe learning environment?”

^BAverage with equal weights to each of the four categories of respondent

All types of respondents (except supervisors) most frequently identified teacher capacity as an aspect of emotional safety in need of improvement. Principals were also likely to highlight atmosphere and inclusion, and supervisors positive discipline.

4.8.4 Materials

Principals were asked what materials schools have to accommodate CWD, and what materials they need. Almost all intervention principals report having something for CWD, although in four cases it is simply a special bathroom for CWD. Five principals mention that they have a resource room, two principals mention treatment plans, and two principals mention special equipment such as walker, wheelchair, and headphones. The most frequently mentioned materials, mentioned by six principals, were games, cards, worksheets, etc.

In terms of needs, five principals mentioned a resource or sensory room (and one of these needed a specialized teacher to oversee the room). Five principals said they need assistive devices, such as headsets, wheelchairs, and special lenses. Here again, the most frequently needed type of materials were games and stationery, mentioned by nine principals. This includes

cards, puppet theater, ‘playing hoops,’ concentration games, meaningful stories, and additional time for play.

Interpretation

Although most schools have a safety plan and CWD considerations are typically involved, objectives such as bullying, exploitation, and violence are rarely included. EQIE can help schools integrate these objectives into their existing plans. Gender is more of a priority at mixed-sex schools (50%) but could also stand to be considered in all schools regardless of sex.

CWD do not appear to experience more harsh/violent behavior than other students. In fact, they are much more likely to report no children being excluded from activities/games. Nonetheless, there is a general level of adults screaming/yelling, children teasing or saying mean things to each other, and children fighting that could be lowered.

To address this, adult stakeholders recommend improving teacher capacities and positive discipline (emotional safety). Adult stakeholders also recommend sufficient space for games and active learning for classroom safety, and a protected external area and improved WASH facilities for safety outside the classroom. Principals express a need for additional resources to better accommodate CWD, especially games, cards, stories, resource rooms, and assistive devices.

4.9 What is the capacity of community members in creation and maintenance of safe learning environments?

4.9.1 Overall Findings

Due to lack of EQIE program elements associated with this evaluation question in Palestine, the Community Leader Interview was not used. However, the Caregiver Interview gives some insights on this question for intervention areas.

When caregivers were asked about strategies to create a safe learning environment, on average caregivers could name 4.7 strategies (out of nine) for physical safety inside the classroom, 3.4 strategies (out of six) for physical safety outside the classroom, and 3.5 strategies (out of ten) for emotional safety. However, 11-15% of caregivers could not name a single strategy in each of these categories.

Caregivers were also given the option to specify ‘other’ strategies for emotional safety.¹⁰ Most frequently, they mentioned the importance of play, games, sports, and art for both learning and relationship with the teacher. A couple of caregivers mentioned the importance of love and encouragement toward students, and a couple of others mentioned the importance of considering the various levels and abilities of students.

Qualitative interviews with three MOE education officials revealed additional safety strategies, including:

Physically safe environment inside the classroom

¹⁰ For physical safety inside and outside the classroom, it was also intended for caregivers to specify an ‘other’ option. However, due to a mistake in programming the KoBo forms, this option was unfortunately not included.

- Lights, chalkboard, painted walls, and dynamic/non-traditional classroom shape
- Help from the local community
- Technology: A laptop for teachers, and a display screen
- Comfortable non-traditional seating arrangement, with appropriate distance between students
- A special place to display students' creativity and achievements
- A reading corner
- Suitable furniture, electricity, lighting, and ventilation

Physically safe environment outside the classroom

- Playgrounds and space to play
- Paintings, planting trees, and awnings
- Provide students with the ability to interact outside the classroom, such as improving their ability to communicate and think critically
- Consider the needs of each school separately
- Technical programs and projects to develop students' applied skills, without parental intervention
- A comfortable place for students to work alone, separated by a security fence, to give space for students' creativity
- Safety wall

Emotionally safe environment

- Implement the safe environment guide
- Pay attention to the psychological aspect in school, and creating a kind of family interaction
- Reduce violence
- Change teacher attitudes to be both a teacher and a guide
- Develop capabilities of teachers and principals
- Provide students with a comfortable learning space and environment
- Close contact between school leadership and the local community
- A successful parent council to enhance student participation in social events and activities
- Positive communication between teachers and the principal
- Fairness and transparency in dealing with students and the problems/issues they face

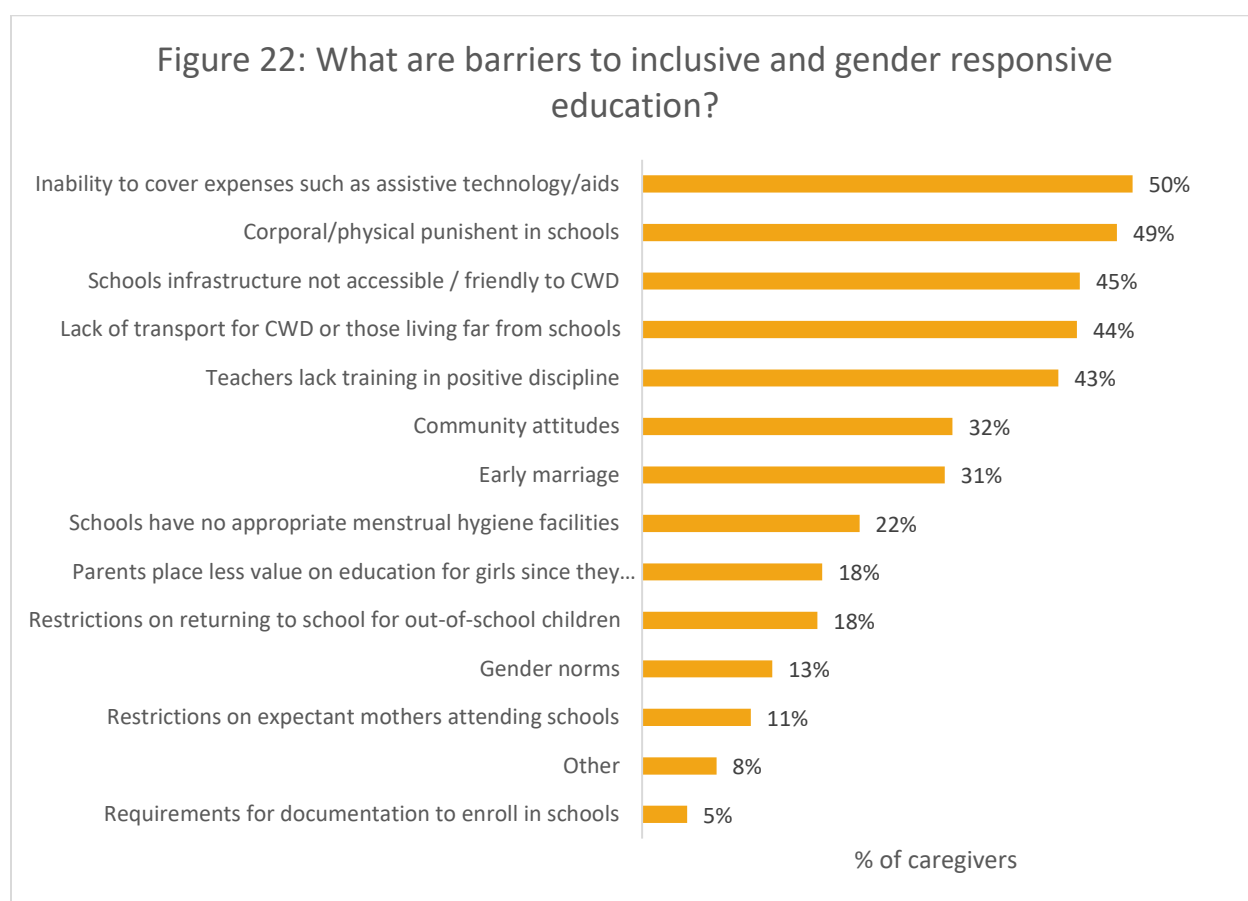
4.9.2 Correlation between capacity and the presence of safe learning environments

The number of strategies for physical safety identified by caregivers was not significantly correlated with the sum of the variables for Indicator 1300.1 (#/% of schools that are conducive to the physical well-being/safety of learners). Strangely, the number of strategies for emotional safety identified by caregivers was significantly negatively correlated with the sum of the variables for Indicator 1300.2 (#/% of schools that meet criteria for emotional safety/emotional well-being), though

this may only mean that caregivers are more aware of emotional safety strategies when the school has a greater number of emotional safety concerns.

4.10 What is the current state of community members and leaders in terms of knowledge regarding ways to address barriers to education faced by marginalized children and children with disabilities?

Due to lack of EQIE program elements associated with this evaluation question in the West Bank, the Community Leader Interview was not used. However, the Caregiver Interview and three qualitative interviews of MOE education officials give some insights on this question for intervention areas. The figure below presents the most to least mentioned barriers from the Caregiver Interview.



According to the three qualitative interviews of education officials, major barriers and challenges include large class sizes, dropout, ‘false social beliefs,’ and diversity of student skill levels. Distance learning and associated technology is also major challenge for students (especially in Grade 1 & 2), caregivers, and teachers. Ideally, distance learning would not be dependent on caregiver intervention, but in many ways it still is. To make matters worse, some caregivers

interfere in their children's learning in the wrong way, such as completing assignments for them. Teacher capacity is poor overall, and especially so in terms of working with marginalized students – in particular CWD, students from lower socioeconomic status, and students who need psychological and social support.

Education officials face challenges in terms of measuring teacher performance, especially in terms of working with marginalized students, and thus struggle to intervene to improve it. This challenge is compounded by the small number of supervisors in relation to teachers, and by the wide variation in types of challenges by regions, urban-rural, nearby, and remote schools, etc.

4.11. What might explain the higher learning outcomes of girls versus boys across grades?

The interviewed teachers at intervention schools and interviewed principals at both intervention and comparison schools were asked: "In general (in your school as a whole), do either boys or girls perform/learn better?" If the response was either boys or girls, these respondents were asked a follow-up question: "What do you think is the biggest explanation for this?"

In mixed-sex schools, most teachers (5 of 8) responded that boys and girls perform/learn equally. The two teachers who said girls perform better explained that girls were 'disciplined' and 'calm.' In contrast, only 4 of 10 principals responded that boys and girls perform equally. The four principals who said girls perform better explained that girls are better organized or have more attention and commitment to education, or that it comes from 'the nature of the family.'

Even at boys schools, 7 of 11 teachers responded that girls perform better (perhaps they were thinking beyond their school to society more generally). These teachers explained that girls study more (and are more interested in studying), are more domestic, respond more to teachers and family, and are more daring; and that boys spend more time playing and are more interested in work. Similarly, 7 of 10 principals at boys schools also said girls perform better, because girls stay at home while boys play; girls like to study; girls are more understanding, focused, and calmer; and female teachers are more patient and loving.

At girls schools, 6 of 8 teachers and 17 of 18 teachers said girls perform better, because girls are quieter, more interested in or committed to studies, pay more attention and work harder, more orderly and careful, more disciplined, etc. while boys are more active and spend a lot of time playing outside the house.

In all types of schools, some of these explanations were qualified as the nature or instinct of either sex. Many explanations were not qualified at all, so it is difficult to determine whether the respondent thought of their explanation as an immutable characteristic of either boys or girls, or whether the explanation was open to change.

Interpretation

The earlier multilevel multivariate regression analysis for Sections 4.4 and 4.5 revealed that, while in raw terms girls outscore boys in many reading skills and some life skills, these differences

disappear once controlling for other factors – namely, whether the school is girls-only, boys-only, or mixed sex. Thus, at least part of the average different performance between girls and boys may be due to the way that girls schools, boys schools, and mixed schools are structured or staffed (for one, both girls and mixed schools have all female teachers, while boys schools have a mix of male and female teachers).

The interplay between study and time spent inside/outside the home, or between study and chores and play is interesting but far from straightforward. The respondents appear to conceptualize play as something in opposition to study, and domestic time and discipline as something supportive of study. It may be important to help families and schools realize the potential benefits of both chores and play to support study and learning.

5. Conclusions/Recommendations and Lessons Learned

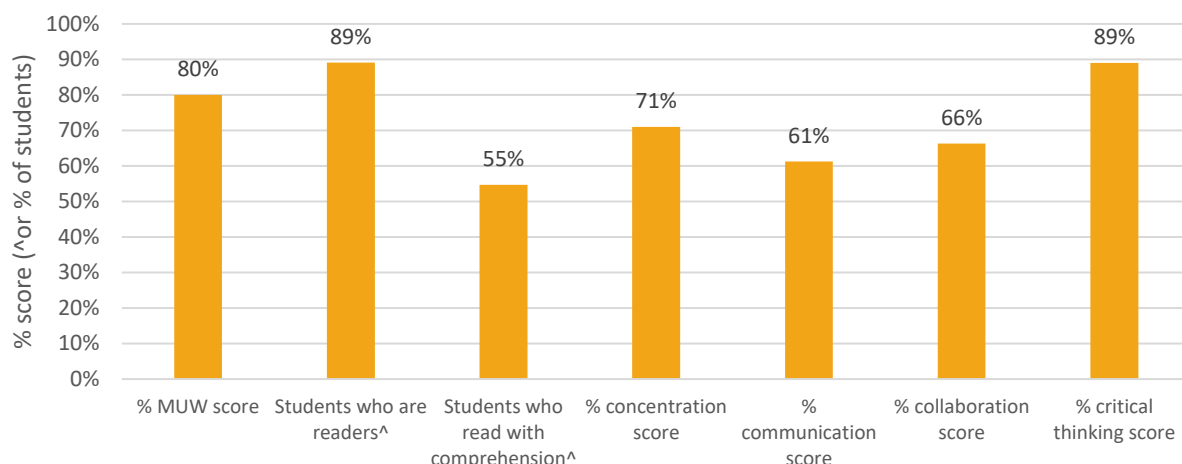
5.1 Recommendations

5.1.1 How should the EQIE program be tailored to meet children’s current reading and life skill needs?

As the table below shows, by Grade 4 most students can read but a large number struggle to read with comprehension. Thus, the EQIE program should give greater weight (though not exclusive weight) to teaching and learning strategies that emphasize meaning and understanding. This can be done for all Grades 1-4, as it is never too early to help even Grade 1 students engage in deeper thinking and imagination about stories that are read to them and the meanings of new words.

However, a minority of students (11%) cannot practically read at all by Grade 4. Therefore, (fun!) opportunities for remedial activities should be available in all grades to prevent and respond to students falling behind their peers. Potential strategies to address classes of diverse skills include pairing stronger with weaker students, and frequently varying instruction techniques so that students of every learning style have a chance to benefit. In addition, the unique challenge of the Arabic language and the small distinctions between letters should be kept in mind and addressed for students who may struggle to distinguish between similar letters, especially children with learning disabilities.

Figure 23: Select reading and life skills in Grade 4



Among the Four C life skills, by Grade 4 students are falling behind the most in communication and collaboration. EQIE can help students see their friends and household members as assets to support them when they have problems, or to work toward common goals/coursework. Teachers can play a constructive role here, especially in terms of organizing/assigning more group work.

In both reading and life skills, although CWD often scored lower than their peers, in certain skills CWD matched the level of their peers. This should be shared as an example of how CWD are capable of learning as their peers, with the right support.

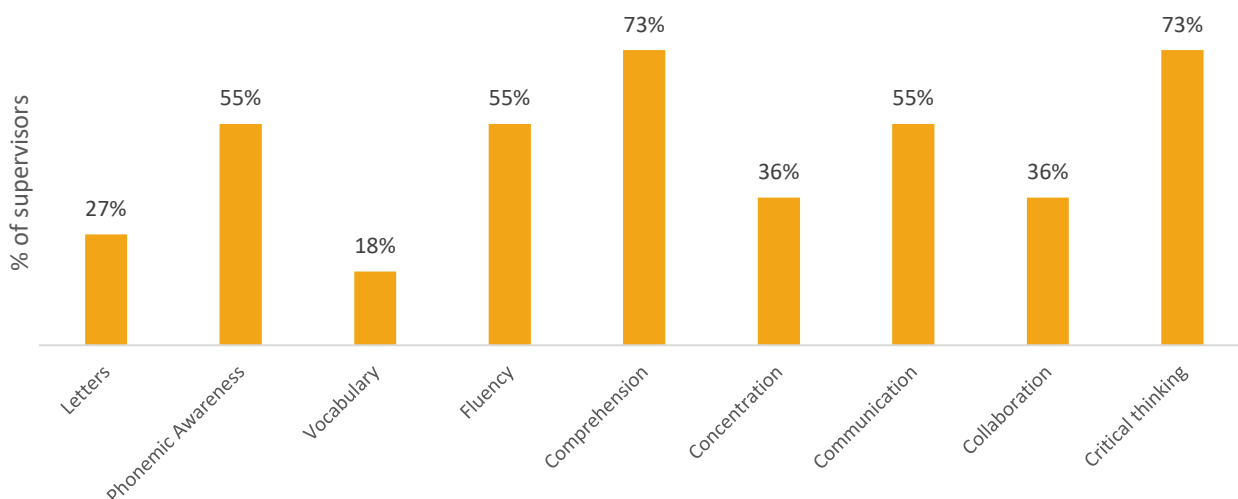
5.1.2 How should EQIE prioritize certain sections of the teacher training curriculum to meet teachers where they are, building upon existing skills and addressing gaps?

Many of the least-used playful reading techniques among teachers are related to asking a variety of comprehension questions to students or allowing students to discuss stories amongst each other. Increasing teachers' skill and frequency in using these techniques could help address the deficit in students' comprehension skills. As a bonus, these techniques apply to students of any reading level, can enhance engagement/motivation of all students to read, and are also an opportunity to engage students who typically participate less.

From a general play-based perspective, many of both the teacher interview and lesson observation deficits revolve around the active/ interactive/ participatory level of children in the classroom – especially free time, discussions, and child leadership opportunities. These may be the greatest opportunities for the PBL approach to enhance teaching and learning.

Finally, the table below presents supervisor-reported areas for teachers to improve their instruction. Supervisors tend to agree that teachers need to improve their instruction of comprehension (and secondarily, phonemic awareness and fluency). Interestingly, supervisors think teachers need to improve their instruction of critical thinking, although students performed the highest in the critical thinking portion of the Learner Assessment.

Figure 24: In what areas of instruction do most teachers need the most improvement? (according to supervisors)



5.1.3 How valid/appropriate are the implied assumptions of the logic model?

The logic model's emphasis on CWD is valid and appropriate – though the EQIE programming will have to be dynamic and agile to meaningfully help schools accommodate the small number of CWD who each have individual needs. Ideally, EQIE would draw on specialist support for diagnosis as well as thinking through solutions to individual needs. One of the more common needs is to enhance communication between teachers and CWD, and perhaps between CWD and their peers and caregivers.

The logic model's emphasis on learning through play is also valid and appropriate. Games, play time, and sufficient/safe space for play are some of the most frequently mentioned themes expressed by respondents in terms of what materials are needed for CWD and physical and emotional safety inside the classroom. There is ample opportunity to not only enhance teachers' use of play for learning, but also caregivers' use of play. Indeed, there may even be an opportunity to optimize play for boys and girls – for boys, leveraging their existing play toward games that enhance their reading and life skills and for girls, more opportunity for play outside school.

Finally, the logic model's bundling of reading skills and life skills together is valid and appropriate. There was a consistent relationship between life skills and literacy activities at home, school, and extracurricular. This signals an opportunity to improve reading and life skills simultaneously, by enhancing existing literacy-supporting interactions and especially by ensuring more students have access to them.

5.1.4 What specific equity indicators appear to be especially influential and ought to be paid attention to in various contexts?

Aside from CWD, there were no child background characteristics relevant to target certain types of students with specialized interventions. However, students with more household member literacy interactions in the past week, and those with greater participation in extracurricular literacy activities, scored higher on reading and life skills. It will be important for EQIE to explore ways of expanding these opportunities to all students.

In terms of gender-responsive programming, more can be done to help schools integrate gender considerations into their school safety plans, and to make teachers aware of the value of empowering and non-traditional roles for both boys and girls in classroom reading materials. There may also be lessons from girls schools – what are they doing right to generate greater reading and life skills on average? – that boys schools and mixed schools can learn from.

5.1.5 How should the project intervention be tailored to positively influence/address the barriers that perpetuate gender inequality and social exclusion?

Caregivers most frequently mention inability to cover expenses for assistive technologies/aids, corporal punishment, lack of school access for CWD, and lack of transport to faraway schools as barriers to inclusive and gender-responsive education. Education officials mention traditional challenges such as large class sizes, dropout, and diversity of student skill levels. But they also highlight distance learning and associated technology as a major challenge for students (especially in Grade 1 & 2), caregivers, and teachers – for whom capacity is poor overall, and especially so in terms of working with marginalized students such as CWD, students from lower socioeconomic status, and students who need psychological and social support. EQIE can serve as a nexus for bringing stakeholders together to brainstorm solutions to these challenges, and EQIE can also experiment with solutions and rigorously document the results through the midline of this impact evaluation.

5.2 Lessons Learned

This section discusses the biggest lessons learned throughout the experience of preparing for, executing, and analyzing this baseline for meaningful insights to help implement and evaluate the EQIE project.

Indicator vs. evaluation question

At times, there seemed to be a focus on indicators as the primary source of learning and data for decision-making. This resulted in multiple revisions to indicators, and more and more complex indicator definitions/means of calculation. This consumed substantial time (to revise Inception report, tools, etc. that had been drafted on an earlier version of indicators) and increases the risk that, at midline and endline, future contractors will make a mistake in attempting to replicate complex indicator definitions/calculations.

Of course, this report still presents indicators values (see Annex B). However, when using these indicators, it is important to keep their limitations in mind. Specific limitations include:

1. Many indicators are composite measures of more complex constructs (such as school safety or teacher practices). The indicator value itself communicates little – for example, a ‘low’ indicator value does not give meaningful information about what ‘part’ of the indicator is driving the low value. See the relevant section of this report for more granular information on various constructs.
2. As per reporting requirements, indicators are disaggregated many different ways. However, these disaggregations ignore statistical inference. That is, just because an indicator value differs between boys and girls, for example, does not mean that the difference is statistically significant, and thus disaggregations may be misleading in and of themselves. Where relevant (for student, parent, and teacher data) the sections of this report rigorously test which differences and relationships are statistically significant.

Quantity vs. quality (and usefulness) of data

This baseline employed seven separate tools to gather nearly one thousand variables. While the desire for comprehensive information and triangulation is understandable, there are serious trade-offs between quantity and quality (and ultimately, usefulness) of information. Resources are limited – in particular, the time and attention not only of data collectors and analysts on the data supply side, but also of program managers and other decision-makers on the data demand side.

As seen from the sheer length of this report, more could be done to slim down to the most important information. This means keeping the need-to-know and cutting the nice-to-know; optimizing data rather than maximizing data. The ‘Actionable’ principle from Innovations for Poverty Action’s Right-Fit Evidence Unit may be useful guidance for future data collections:

In theory, more information should help organizations make better-informed decisions. But in reality, the availability of more data often leads organizations to become overwhelmed by data. The actionable principle seeks to roll back this problem by calling on organizations to only collect data that they will use. Organizations should ask two questions of each and every piece of data that they want to collect: ‘Is there a specific action that we will take based on the findings?’ and ‘Do we have the resources and the commitment required to take that action?’¹¹

Quantity and complexity of baseline vs. page limits

RTP’s Baseline Report Guide emphasize both comprehensiveness and page limits. Given the amount of data discussed above, as well as the complex array of indicators and evaluation questions, meaningful comprehensive analysis required exceeding page limits. For example, the report guidelines for the Executive Summary (max four pages) state: “Should stand alone as an abbreviated version of the report. All content of the full report should be summarized.” Given the impossibility of adhering to this guidance within four pages, a nine-page Executive Summary

¹¹ <https://www.poverty-action.org/right-fit-evidence/resources/principles>

was drafted. In the future RTP should consider whether relaxing page limits to better accommodate data collections of this scope.

Complex contracting arrangements

The contract arrangement for the baseline gave the budget for InformEd staff to InformEd, and all other expenses (such as enumerators, local logistics, etc.) to RTP Country Offices. This caused confusion, resulted in mismatch between Inception Report methodology and resourcing to execute the methodology, and required substantial additional time of both InformEd and RTP staff to navigate and reconcile. In the future, if possible, RTP may consider a more simplified approach to contracting in which the budget for both contractor time and local resources reside with a single entity/contract instead of split between multiple entities.

Annex A. EQIE Project Indicators and Analysis Framework

The table below outlines how each indicator was scored in the West Bank.

Indicator #	Indicator	Description of Indicator	Data Collection Tool	Questions from Assessment Tools	Analysis Calculation
1000.1	#/% of children in target educational institutions achieving at least a minimum proficiency level in reading in grade 4	The five core reading skills are letter knowledge/phonics, phonemic awareness, vocabulary, reading fluency, and comprehension. Minimum proficiency level is defined here as the benchmark of basic knowledge in literacy assessment Task 5 (reading with comprehension). Benchmark should be contextualized in each country and project to meet MoE policy objectives.	Learner Assessment	6 comprehension questions administered to children who can read (defined as a child that can read at least 5 words correctly in 30 seconds). Note that the benchmark is 80% of comprehension questions are marked correct to be a reader with comprehension. This definition will be contextualized for each country during the tool contextualization phase (process for doing so is articulated in the Country Manager Manual).	# of grade 4 children meeting the contextualized benchmark for reading with comprehension divided by # of grade 4 children assessed
1000.2	#/% of children in target educational institutions with increased life skills	Life skills include cognitive and socio-emotional learning competencies contained in the holistic child development framework. Usually measured using validated tools such as IDELA, ISELA, and DAP. Report the aggregate profile of each competency area and individual skill value. Midline and final evaluation values will help determine whether a child has improved in a skill area or not. Using comparison group, a change in the skill value could be attributed to the project intervention.	Learner Assessment	Countries will report on competency area and skill values as follows: - Concentration (shape5 – shape8) - Communication (relate2 relate3 relate4 relate5 relate10 relate11 relate15 relate16) - Collaboration (collaboration1 – collaboration4) - Critical Thinking (conflict1 – conflict4)	Aggregated competency area and skill value, reported as an average score (disaggregated by grade)
1100.1	#/% of teachers demonstrating improved teaching practices to enhance children's reading skills	Teachers are expected to demonstrate knowledge and practice in teaching the five core skills using PBL (use of play/games & RCA). A teacher is deemed to have demonstrated improved skills in teaching reading if s/he applies literacy instruction	Classroom Observation (triangulated by Teacher Interview)	A teacher practice index was constructed, drawing from the classroom observation questions corresponding to the sections Teacher approach and ability to implement Literacy-PBL, Inclusive Learning, Behavior Management, and Children Interaction/Engagement within classroom	A threshold for the teacher practice index was set at the 75 th percentile of classroom observation scores (=0.8). To be counted in the numerator, a teacher must have a practice index score ≥ 0.8 AND must have a lesson observation in

		strategies using PBL approach relevant for the lesson observed. And also evidences these in his/her lesson plan.		Note that the self-reported teacher practice index from the teacher interview was not correlated with either the classroom observation practice index or with student outcomes. Thus, the self-reported index should be considered unreliable, and future data collections may consider cutting these items from the teacher interview. In classroom observation tool, the variable lplanpbliteracy was also used.	which the teacher's lesson plan includes literacy instruction w/ a PBL approach. The denominator is all teachers interviewed. Note: One teacher claimed to have been trained by RTP on the classroom observation form, but could not remember how long ago and the same teacher also reported not being trained by RTP during the teacher interview. Therefore, for purposes of disaggregation this teacher was considered NOT trained by RTP.
1100.2	#/% of children who report feeling that the PBL activities their teacher practiced had a positive effect on their reading and life skills (disaggregated by sex, grade and Impairment)	Children express that their teacher(s) use PBL activities in class and that the use of those activities and strategies have helped improve their reading and life skills.	Learner Assessment	Over the past week - Did your teacher read to you?, Did your teacher ask you questions about what you read?, Did you read with a classmate in school?, Did you read quietly by yourself without help?, Did you learn songs or games in class?, Did your teacher teach you the sounds of letters and words? (teachpractice1 – teachpractice6)	# of children who report teachers engaging in all PBL activities in the previous week divided by # of children assessed
1110.2	#/% of teachers who report receipt of appropriate level of support to improve their knowledge and practice of PBL approaches to enhance children's reading skills	Support and supervision mechanisms for teachers to function effectively as defined in INEE minimum standard 3 on Support and Supervision with particular emphasis on the following: • Adequate teaching and learning materials and space are available (guidance note 1). • Teachers are involved in professional development that contributes to their motivation and support (guidance notes 2-3). • A transparent, accountable supervisory mechanism provides for regular assessment, monitoring and support for teachers (guidance notes 2-3). Teachers receive coaching and	Teacher Interview	Drawn from the questions corresponding to variables rtptrain, rtpsubject, implement, and pblsupport in teacher interview. Supporting questions are rtptrainwhen, otherorgtrain, implementmonths, implementreason, rtpsharetrainin, rtpsharetrainout, support, and supporttype in teacher interview.	# of teachers trained in at least 4/5 PBL subjects, who: (a) report having at least 4/6 supports, and (b) report implementing PBL at least sometimes in their classroom; divided by all teachers interviewed

		mentoring support. • Performance appraisals for teachers are conducted, documented and discussed regularly (guidance note 4).			
1200.1	#/% of children who routinely participate in supplemental reading activities	Supplemental reading activities are reading activities outside the classroom and school hours. Happening in the home, community reading centres/clubs/reading days out. This indicator counts children who participated/attended reading club meeting/after-school reading activities in the past week. Focus on the reading club/centre/days out activities, and then ask if they have been engaged in any other activities (read-a-thons, reading to siblings etc to triangulate).	Learner Assessment	Did you attend any after-school reading activities, like reading camps, reading centers, or reading days out? Also added for COVID - Do you participate in out-of-school reading activities, like those available through radio, TV, or mail?	# of children who reported participating in supplemental reading activities in the previous week divided by # of children assessed
1200.2	#/% of children who report positive changes in their home, school, and community to support their reading	Positive changes such as parents/siblings read to child, helps child do homework, take me to community centre to read, buy my books to read, teacher use games, teacher read aloud, teacher use word cards, school has reading hour and story books for children to read, community has reading centre/club with trained facilitators and books etc that support children reading skill development. Disaggregated by setting type (home, school, and community), sex and impairment. Refer to the Learning Together Toolkit Session D for home literacy environment expectation	Learner Assessment	Home - In the past week: Have you seen anyone in your home reading? Has someone in your family helped you with your homework? Has a family member read to you? Has a family member told you a story? Has a family member played literacy games with you? School - Did your teacher read to you?, Did your teacher ask you questions about what you read?, Did you read with a classmate in school?, Did you read quietly by yourself without help? Community - Did you attend any after-school reading activities, like reading camps, reading centers, or reading days out? Also added for COVID - Do you participate in out-of-school reading activities, like those available through radio, TV, or mail?	# of children who reported engagement in reading behaviors in the previous week divided by # of children assessed; disaggregated by setting Note 1: For the community setting, this variable is the same as Indicator 1200.1. Note 2: In each of the three settings, a child is counted if they reported $\geq 50\%$ of possible behaviors. Note 3: For the overall indicator, a child is counted if they reported $\geq 50\%$ of possible behaviors in all three contexts.
1210.1	#/% of parents that report using home-based reading activities with their children	Parents refer to primary caregivers of early grade children in the project's target school community. Home based reading activities include help child do homework, read to child, tell story to	Caregiver Survey	(seenread): In the past week, have you read anything for pleasure? (helpstudy): In the past week, have you helped (childname) with homework)?	# caregivers report at least 2 activities / total # caregivers interviewed

		child, set reading corner for child, play literacy games with child, etc (refer to Learning Together Toolkit). Proportion of parents who report engaging in any of these activities with children at home in the past week are counted in this indicator.		(read2u): In the past week, have you read to (childname)? (story2u): In the past week, have you told a story to (childname)? (playgames): In the past week, have you played literacy games with (childname)?	
1210.2	# of reading initiatives provided by parents, teachers, children, and community (disaggregated by type)	Organized reading events initiated and led by parents/teacher/children/community with the aim of encouraging and enhancing the habit of reading among children in early grade (Grade 1-4). It includes reading competition, learning days, reading events, radio broadcasting, reading olympics, reading marathon, international reading days, and reading summer activities.	N/A - Measured by RTP (Project Monitoring)	N/A - Measured by RTP (Project Monitoring)	N/A - Measured by RTP (Project Monitoring)
1300.1	#/% of schools that are conducive to the physical well-being of learners	School meets INEE minimum standard 3 for facilities and services as follows: - Learning sites and structures are safe and accessible for all learners, teachers and other education personnel (see guidance notes 1-4). • Temporary and permanent learning environments are repaired, retro-fitted or replaced as needed with disaster-resilient design and construction (see guidance notes 2 and 4). • Learning spaces are marked by visible protective boundaries and clear signs. • Physical structures used for learning sites are appropriate for the situation and include adequate space for classes, administration, recreation and sanitation facilities (see guidance notes 2 and 4).	Principal Interview (triangulated by teacher interview and classroom observation)	Drawn from the questions corresponding to variables concerns, system, implement, included, and emotional/physicalin/physicalout variables in principal interview.	# of schools with (a) children coming to staff for concerns, (b) system for addressing those concerns, (c) at least moderately implemented school improvement plan, and (d) no more than 3 (out of 15) physical safety issues / total # of schools

		• Class space and seating arrangements meet agreed ratios of space per learner and teacher in order to promote participatory methodologies and learner-centred approaches (see guidance note 4). • Adequate quantities of safe water and appropriate sanitation facilities are provided for personal hygiene and protection, taking into account sex, age and people with disabilities (see guidance notes 3 and 5-6). • Skills-based health and hygiene education is promoted in the learning environment (see guidance note 6). • Schools and learning spaces are linked to child protection, health, nutrition, social and psychosocial services (see guidance note 8)			
1300.2	#/% of schools that meet criteria for emotional safety/emotional well-being	School is considered emotionally safe for children if it meets all 3 conditions 1. Education staff demonstrate participatory, inclusive, positive discipline and gender-sensitive practices in the school 2. Absence of anti-bullying behavior with clear rules against gender-based violence 3. Functional referral system with trained point person for child protection and psychosocial support Refer to INEE minimum standard 2 on protection and wellbeing Guidance notes 2-4, and 8 for details	Teacher Interview (triangulated by principal interview and classroom observation)	Drawn from the questions corresponding to variables discipline, comfortable, sexparticipate, participate, adapthow, psychtrain in teacher interview.	# of schools where both interviewed teachers (a) use only positive discipline techniques, (b) report children feeling comfortable at least most of the time, (c) report girls and boys participate equally, (d) report cwd participate at least most of the time, (e) can give an example of adaptation made for cwd, (f) have been trained on providing psychosocial support to children; divided by all teachers interviewed
1300.3	#/% of children who report feeling safe and included in the school learning environment (disaggregated M/F; impairment)	Refer to relevant sections of INEE minimum standards on Protection & Wellbeing; and Facilities & Services guidance notes 2- 4, 8 and 1-8 respectively. Children who report positively on these are counted.	Learner Assessment	In the last week, did you feel afraid in your school?; In the last week, did you feel afraid on your way to your school? Supporting questions (for interpretation): In the last week: did a child in your school	# of children reporting no to both questions / total # of children interviewed

				tease another child to make them angry?, did a child in your school leave out another child from an activity or game?, did a child in your school say something mean to another child?, did a child do any of these things to you?, did children in your school get into a fight where someone was hit?, did you get into a fight where someone was hit?, did anyone in your school throw something at a child to hurt them?, did you throw something at a child to hurt them?, did an adult in your school scream or yell at a child in a threatening way?, did an adult in your school hit or kick a child?	
1300.4	#/% of teachers with access to appropriate psychosocial support to improve their well-being	Support mechanisms for teachers function effectively as defined in INEE minimum standard 3 on Support and Supervision with specific focus on: - Appropriate, accessible and practical psychosocial support for teachers (see guidance note 6)	Teacher Interview	Teacher Interview (psychsupport): Do you have access to a trained point person for psychosocial support? Supporting questions (for interpretation): Please indicate how much you disagree or agree with each of the following statements. (rewarding) Working as a teacher has been rewarding and worthwhile. (strongcop) There is a strong community of practice amongst my fellow teachers that I can learn from, and share challenges and successes with.	# teachers that responded yes / total # of teachers interviewed
1300.5	#/% of education professionals (teachers, principals) who have considered leaving the sector over the past year	Self-reported measure from teacher interview and principal interview. Calculated as # of educators who have considered leaving in past year / # of educators interviewed	Teacher Interview Principal Interview	Teacher Interview (leavesector): In the past year, have you considered leaving teaching and education as a career? Principal Interview (leavesector): In the past year, have you considered leaving education as a career? Follow up question on both interviews is: <i>What is the primary reason for wanting to leave?</i>	# of principals and teachers that responded yes / total # of principals and teachers interviewed
1310.1	#/% school community members that exhibit knowledge in creation	School community members are SMC/PTA, School Council, School Inclusion Committee, Coaches, School	Education Official interview	Contextualized list of emotional, physical inside classroom, and physical outside strategies in questions viii, ix, & x of the	# of education officials or caregivers that are able to identify at least 3 strategies for emotional

	and maintenance of safe and inclusive learning environments	administrators/Head Teachers/Principals, Teachers, District Education Officials, and National Education Agencies/Ministries. Exhibition of knowledge in enhancing child protection, disability inclusion, promoting child rights, and dealing with corporal punishment, abuse, and/or GBV to achieve physical and emotional safety in the school. Members to consider for the calculation of this indicator are SMC/PTA/School Council Executives, Parents/Caregivers/Parent Council Executives, Grade Pre-primary - Primar 4 Teachers, and Headteachers/School Principals. Data is disaggregated by school community member type. CALCULATION: # of members in a category able to mention 2-3 strategies for creating safe learning environment divided by total # of members in that category surveyed.	Caregiver Interview	Education official Interview. Enumerators trained to map the qualitative response to these options. Within caregiver survey, question strategies will be the same.	safety, 3 strategies for physical inside safety, and 2 strategies for physical outside safety / # education officials & caregivers interviewed Note: threshold for physical outside safety set to 2 instead of 3, because only 6 possible options (instead of 9-10)
1310.2	#/% of children who know where and how to formally report a protection violation	Number of children who correctly identify where and how to formally report a protection violation (defined locally) divided by the number of learners assessed	Learner Assessment	Childprotection: If you or a friend were being harmed by someone else, what would you do? (countries are instructed to contextualize potential multiple choice responses as appropriate)	# children who accurately answer question (contextualized by country) / total # children interviewed
1400.2	# students (F/M) enrolled in target educational institutions(disaggregated by level of education and CWD)	Number of male and female students that enroll in educational institutions directly supported by the project. An institution can be counted as 'supported' when goods (books, learning materials, and infrastructure) and services (e.g. training, inspectors/supervisor visits, and training of SMC/PTA members) have been provided directly to that institution. In general, direct support to institutions are activities/interventions where the	N/A	N/A - measured and reported by RTP for Gaza programming	N/A - measured and reported by RTP for Gaza programming

		educational institution has been directly targeted. For example: A training that generally strengthens the knowledge of inspectors/supervisors in a district alone would not qualify for including all students in that given district as supported. However, if the inspectors are also provided with funds and transport to inspect particularly marginalized schools students in these specific schools can be included as 'supported'. Similarly if teachers or PTA/SMC members from a school has been supported with transport, food or accommodation to attend a training the school can be included as 'supported'.			
1410.1	% of community members and leaders that can easily identify at least 3 barriers to education faced by girls and children with disabilities (disaggregated by M/F)	School community members and leaders are SMC/PTA, School Inclusion Committee, Coaches, Junior Leaders, School administrators, CSO partners, District Education Officials, & National Education Agencies/Ministries, Political Leaders, Religious Leaders. Exhibition of knowledge in enhancing child protection, disability inclusion, promoting child rights, and dealing with corporal punishment, abuse, and/or GBV to achieve inclusive education. Community members and leaders who are able to identify any 3 of the barriers to inclusive and gender responsive education are; - Corporal/physical punishment in schools - Schools Infrastructure not accessible and friendly to persons with disability - Schools have no appropriate menstrual hygiene facilities - Teachers lack training in positive discipline	Education Official interview Caregiver Interview	Question vi. of the Education official Interview. Enumerators trained to map the qualitative response to these options. Within caregiver survey, question <i>barriers</i> will be the same. Note, however, that this indicator may not be properly measured in the West Bank, since this aspect of the EQIE project only relates to interventions in Gaza, and because of differences in MoE between West Bank and Gaza.	# of education officials or caregivers that are able to identify at least 3 barriers / # education officials & caregivers interviewed

		- Lack of means of transport for persons with disability and those living far from school - Early Marriage - Restrictions on returning to school for out-of-school children - Gender norms. - Parents place less value on education for girls since they will get married - Community attitudes - Inability to cover expenses such as assistive technology/aids - Requirements for documentation to enroll in schooling - Restrictions on expectant mothers attending school			
1410.2	#/% of children who report that they have witnessed their school taking action to make girls and CwDs feel included. (disaggregate by Sex and Marginalization type)	Actions for emotional, social and physical wellbeing of children; and gender/social inclusion taken by school.	Learner Assessment	N/A - measured and reported by RTP for Gaza programming	N/A - measured and reported by RTP for Gaza programming

Annex B. PMF

Note: For deeper analysis of disaggregations and relationships (including statistical inference), see relevant section of report.

Programme Objective	Indicators	Countries	Disaggregation	<i>n</i>	Baseline Y1-2020
PROGRAM OBJECTIVE 1000					
	1000.1: #/% of children in target educational institutions achieving at least a minimum proficiency level in reading in grade 4		Overall	322	55%
			Female	148	63%
			Male	174	48%
			Children with Disability	12	42%
			Children without Disability	310	55%
	1000.2: Oral Reading Fluency Score (WPM)	All	Overall Grade 1	N/A	N/A
			Grade 1 Female	N/A	N/A
			Grade 1 Male	N/A	N/A
			Grade 1 Children with Disability	N/A	N/A
			Grade 1 Children without Disability	N/A	N/A
			Overall Grade 4	322	47
			Grade 4 Female	148	58
			Grade 4 Male	174	38
			Grade 4 Children with Disability	12	34
			Grade 4 Children without Disability	310	48
	1000.3: #/% of children in target educational	All	Grade 1 Concentration: Overall	405	34%
			Grade 1 Female	179	35%

	institutions with increased life skills		Grade 1 Male	226	33%
			Grade 1 Children with Disability	6	47%
			Grade 1 Children without Disability	399	34%
			Grade 1 Communication: Overall	409	56%
			Grade 1 Female	180	60%
			Grade 1 Male	229	56%
			Grade 1 Children with Disability	6	52%
			Grade 1 Children without Disability	403	58%
			Grade 1 Collaboration: Overall	408	63%
			Grade 1 Female	179	66%
			Grade 1 Male	229	61%
			Grade 1 Children with Disability	6	46%
			Grade 1 Children without Disability	402	63%
			Grade 1 Critical Thinking: Overall	396	70%
			Grade 1 Female	177	73%
			Grade 1 Male	219	67%
			Grade 1 Children with Disability	6	69%
			Grade 1 Children without Disability	390	70%
			Grade 4 Concentration: Overall	308	71%
			Female	141	76%
			Male	167	67%
			Children with Disability	11	50%
			Children without Disability	297	72%

			Grade 4 Communication: Overall	317	61%
			Female	146	62%
			Male	171	61%
			Children with Disability	12	50%
			Children without Disability	305	62%
			Grade 4 Collaboration: Overall	316	66%
			Female	146	68%
			Male	170	65%
			Children with Disability	11	52%
			Children without Disability	305	67%
			Grade 4 Critical Thinking: Overall	299	89%
			Female	135	92%
			Male	164	86%
			Children with Disability	10	91%
			Children without Disability	289	89%
Outcomes & Outputs	Indicators	Countries	Disaggregation		Baseline Y1-2020
INTERMEDIATE OUTCOME 1100:					
1100: Increased integration of play-based learning by teachers to enhance children's reading skills in target educational institutions	1100.1: #/% of teachers demonstrating improved teaching practices to enhance children's reading skills	All	Overall	38	18%
			Female	29	21%
			Male	9	11%
			Teacher with RTP COTT Training Status	0	N/A
			Teacher without RTP COTT Training Status	38	18%

			Overall Grade 1 Teachers	18	28%	
			Grade 1: Female	14	29%	
			Grade 1: Male	4	25%	
			Overall Grade 4 Teachers	20	10%	
			Grade 4: Female	15	13%	
			Grade 4: Male	5	0%	
	1100.2 #/% of children that report feeling that the PBL activities their teacher practiced had a positive effect on their reading and life skills (disaggregated by sex, grade and Impairment)		All	Overall	719	35%
				Overall: Female	324	45%
				Overall: Male	395	27%
				Overall: Children with Disability	13	39%
				Overall: Children without Disability	706	35%
				Overall Grade 1	405	36%
				Grade 1: Female	179	46%
				Grade 1: Male	226	29%
				Grade 1: Children with Disability	3	33%
				Grade 1: Children without Disability	402	36%
				Overall Grade 4	314	33%
				Grade 4: Female	145	43%
				Grade 4: Male	169	25%
				Grade 4: Children with Disability	10	40%
Grade 4: Children without Disability	304	33%				
IMMEDIATE OUTCOME 1110:						
		All				

	1110:1 %/# of trained teachers scoring satisfactorily on post-COTT training tests		Sex, Grade, Years of Teaching Experience		
1110: Increased knowledge of teachers to enhance children’s reading skills through play-based learning in target educational institutions	1110.2 #/% of teachers who report receipt of appropriate level of support to improve their knowledge and practice of PBL approaches to enhance children's reading skills	All	Overall	41	2%
			Female	32	3%
			Male	9	0%
			Teacher with RTP COTT Training Status	1	100%
			Teacher without RTP COTT Training Status	N/A	N/A
			Overall Grade 1 Teachers	21	5%
			Grade 1: Female	17	6%
			Grade 1: Male	4	0%
			Overall Grade 4 Teachers	20	0%
			Grade 4: Female	15	0%
			Grade 4: Male	5	0%
INTERMEDIATE OUTCOME 1200:					
1200: Increased supplemental reading practice of children [in target locations]	1200.1 #/% of children who routinely participate in supplemental reading activities	All	Overall	710	50%
			Overall: Female	324	57%
			Overall: Male	386	45%
			Overall: Children with Disability	12	58%
			Overall: Children without Disability	698	50%
			Overall Grade 1	401	41%
			Grade 1: Female	179	49%

			Grade 1: Male	222	36%
			Grade 1: Children with Disability	3	33%
			Grade 1: Children without Disability	398	42%
			Overall Grade 4	309	62%
			Grade 4: Female	145	66%
			Grade 4: Male	164	57%
			Grade 4: Children with Disability	9	67%
			Grade 4: Children without Disability	300	61%
	1200.2: #/% of children who report positive changes in their home, school, and community to support their reading ¹²	All	Overall	714	37%
			Overall: Female	325	45%
			Overall: Male	389	30%
			Overall: Children with Disability	13	46%
			Overall: Children without Disability	701	36%
			Overall Grade 1	403	28%
			Grade 1: Female	180	37%
			Grade 1: Male	223	20%
			Grade 1: Children with Disability	3	33%
			Grade 1: Children without Disability	400	28%
			Overall Grade 4	311	48%
			Grade 4: Female	145	54%
			Grade 4: Male	166	43%

¹² Indicator 1200.2 is reporting the percentage of children reporting activities currently. This is not the baseline wording.

			Grade 4: Children with Disability	10	50%
			Grade 4: Children without Disability	301	48%
IMMEDIATE OUTCOME 1210:					
1210: Increased provision of supplemental reading activities by teachers, parents, and community	1210.1: #/% of parents that report using home-based reading activities with their children	All	Composite	403	67%
			Composite: Female	192	90%
			Composite: Male	211	45%
			Seen reading	401	37%
			Seen reading: Female	191	46%
			Seen reading: Male	210	29%
			Helped Study	403	69%
			Helped study: Female	192	95%
			Helped study: Male	211	45%
			Read to Child	402	51%
			Read to child: Female	191	73%
			Read to child: Male	211	31%
			Told story to child	402	30%
			Told story to child: Female	192	35%
			Told story to child: Male	210	24%
			Played games	403	42%
			Played games with child: Female	192	58%
			Played games with child: Male	211	28%
	1210.2: # of reading initiatives provided by parents, teachers, children, and community	All	Total		
			Reading competitions		
			Radio broadcasting		
			Reading events		

			Reading Olympics		
			Learning days		
			Learning Marathons		
			International Day		
			Summer activities (collaboration with UNRWA)		
Outcomes & Outputs		Countries	Disaggregation		Baseline Y1-2020
INTERMEDIATE OUTCOME 1300:					
1300: Improved safe and inclusive learning environments in target educational institutions.	1300.1 #/% of schools that are conducive to the physical well-being/safety of learners	All	Overall	20	10%
			Urban Schools	20	10%
			Rural Schools	0	N/A
			Mixed-Sex Schools	4	0%
			Girls Schools	7	0%
			Boys Schools	9	22%
	1300.2: #/% of schools that meet criteria for emotional safety/emotional well-being	All	Overall	20	0%
			Urban Schools	20	0%
			Rural Schools	0	0%
			Mixed-Sex Schools	4	0%
			Girls Schools	7	0%
			Boys Schools	9	0%
	1300.3 #/% of children who report feeling safe and included in school learning environment		Overall	704	56%
			Overall: Female	316	70%
			Overall: Male	388	45%
			Overall: Children with Disability	12	83%

		Overall: Children without Disability	692	56%
		Overall Grade 1	402	64%
		Grade 1: Female	178	81%
		Grade 1: Male	224	51%
		Grade 1: Children with Disability	3	100%
		Grade 1: Children without Disability	399	64%
		Overall Grade 4	302	45%
		Grade 4: Female	138	56%
		Grade 4: Male	164	36%
		Grade 4: Children with Disability	9	78%
		Grade 4: Children without Disability	293	44%
1300.4 #/% of teachers with access to appropriate psychosocial support to improve their well-being		Overall	40	48%
		Female	31	48%
		Male	9	44%
		Teacher with RTP COTT Training Status	1	100%
		Teacher without RTP COTT Training Status	39	46%
		Overall Grade 1 Teachers	20	50%
		Grade 1: Female	16	56%
		Grade 1: Male	4	25%
		Overall Grade 4 Teachers	20	45%
		Grade 4: Female	15	40%
		Grade 4: Male	5	60%

	1300.5 #/% of education professionals (teachers, principals) who have considered leaving the sector over the past year.		Overall	61	13%
			Female	47	16%
			Male	14	21%
			Overall with RTP COTT Training Status	13	8%
			Overall without RTP COTT Training Status	48	15%
			Overall Principals	20	5%
			Principals: Female	15	7%
			Principals: Male	5	0%
			Overall Teachers	41	17%
			Teachers: Female	32	13%
			Teachers: Male	9	33%
			Overall Grade 1 Teachers	21	19%
			Grade 1: Female	17	24%
			Grade 1: Male	4	0%
			Overall Grade 4 Teachers	20	15%
			Grade 4: Female	15	0%
			Grade 4: Male	5	60%
IMMEDIATE OUTCOME 1310:					
1310: Increased capacity of school community in creation and maintenance of safe and inclusive learning environments	1310.1: #/% school community members that exhibit knowledge in creation and maintenance of safe and inclusive learning environments	All	Overall	406	57%
			Female	193	56%
			Male	213	57%
			School Community Stakeholder Type 1 (SMC/PTA)	12	58%
			School Community Stakeholder Type 2 (Education officials)	3	100%

	1310.2 #/% of children who know where and how to formally report a protection violation		School Community Stakeholder Type 3 (Parent/caregivers)	391	57%
			Overall	706	81%
			Overall: Female	322	85%
			Overall: Male	384	77%
			Overall: Children with Disability	12	92%
			Overall: Children without Disability	694	80%
			Overall Grade 1	397	81%
			Grade 1: Female	177	86%
			Grade 1: Male	220	77%
			Grade 1: Children with Disability	3	100%
			Grade 1: Children without Disability	394	81%
			Overall Grade 4	309	80%
			Grade 4: Female	145	83%
			Grade 4: Male	164	77%
			Grade 4: Child with Disability	9	89%
			Grade 4: Child without Disability	300	80%
Outcomes & Outputs		Countries	Disaggregation		Baseline Y1-2020
INTERMEDIATE OUTCOME 1400:					
IMMEDIATE OUTCOME 1410:					
1410: Increased capacity of community members and leaders to	1410.1: % of community members and leaders	All	Overall	372	69%
			Female	177	68%

address barriers to access education and learning	that can easily identify at least 3 ways to address barriers to education faced by girls and children with disabilities (disaggregated by <u>M/F and School Community Stakeholder Type</u>)		Male	195	69%
			School Community Stakeholder Type 1 (SMC/PTA)	11	64%
			School Community Stakeholder Type 2 (Education officials)	3	100%
			School Community Stakeholder Type 3 (Parent/caregivers)	358	69%
	1410.2 #/% of children who report that they have witnessed their school taking action to make girls and CwDs feel included.(disaggregate by Sex and Impairment/Disability type)	All	Overall	N/A	N/A
			Overall: Female	N/A	N/A
			Overall: Male	N/A	N/A
			Overall: Children with Disability	N/A	N/A
			Overall: Children without Disability	N/A	N/A
			Overall Grade 1	N/A	N/A
			Grade 1: Female	N/A	N/A
			Grade 1: Male	N/A	N/A
			Grade 1: Children with Disability	N/A	N/A
			Grade 1: Children without Disability	N/A	N/A
			Overall Grade 4	N/A	N/A
			Grade 4: Female	N/A	N/A
			Grade 4: Male	N/A	N/A
			Grade 4: Child with Disability	N/A	N/A
			Grade 4: Child without Disability	N/A	N/A

Annex C. Learner Assessment Details

Learner Background Questions

Table C.1: Summary of Learner Assessment background items

Topic	Examples
Student background	
General	Sex, age, household size
School-related	Prior preschool attendance
Home Literacy Environment/Relationships	
Access to print	Materials present in home, types of materials
Reading activities at home	Household members whom children see read and who engage in literacy activities with children in past week
Community Literacy Environment	
Extracurricular activities	Reading clubs and out-of-school reading activities administered in response to the COVID-19 pandemic
School Experience	
Play-based learning (PBL)	Teacher practice in past week (reading, asking questions, songs, games, etc.)
Positive-learning-environment (PLE)	Child-child and adult-child interactions (teasing, fighting, yelling, etc.)

Reading Assessment

The reading assessment assessed early grade Arabic literacy proficiency through an individual oral assessment of foundational reading skills. Due to the length of the Learner Assessment, the reading assessment portion was abbreviated to include only the most important skills by grade level.

Table C.2: Summary of reading assessment subtests

Reading subtest	Grade 1	Grade 4
Letter knowledge	Name or sound of 18 of 28 Arabic letters taught by late November, according to MOEHE curriculum	N/A
Phonemic awareness	Beginning sounds of 10 common words	
Most used words	N/A	Reading of 10 common words

Reading Passage (for accuracy and fluency)	N/A	136-word grade-level story not yet covered by curriculum
Comprehension questions	N/A	6 comprehension questions about the reading passage (listening comprehension for non-readers, reading comprehension for readers)

Life Skills Assessment

The life skills assessment assessed the development of social and emotional learning (SEL) skills in children between the ages of 6 and 12 years. Students were asked questions and participated in activities that measured RTP Palestine's 'Four C' life skills of Concentration, Communication, Collaboration, and Critical Thinking using adapted items from ISELA tool. Secondly, information was gathered on students' social network size (relationship) and empathy.

Table C.3: Summary of life skills assessment subtests

Construct	Description	# of items
Concentration	Whether student is able to draw shapes of increasing difficulty with their dominant hand	4 drawings for Grade 4, 3 drawings for Grade 1
Communication	Does student talk to someone at home/a friend/an adult at school when sad? Does student ask someone for help at home/a friend/an adult at school or when student has a problem with a friend?	8 questions
Collaboration	What are some good ways to show teamwork? In the last week, has student: Played a game with friends? Worked with friends on coursework/homework? Worked with a friend to accomplish a common goal?	4 questions
Critical thinking	What to do in hypothetical situation when another child wants a toy student is playing with? What to do if another child takes the toy?	4 questions
Child protection	What would student do if they or a friend were being harmed by someone else?	1 question
Social network size (Relationship)	Number of household members, number of good friends, and number of adults at school the student talks to almost every day.	3 questions

Empathy	Can student tell when someone at home/a friend/an adult at school is sad?	3 questions
---------	---	-------------

Washington Group Short Set¹³

The Washington Group (WG) Short Set is a set of questions designed to identify (in a census or survey format) people with a disability. The questions ask whether people have difficulty performing basic universal activities (walking, seeing, hearing, cognition, self-care, and communication). Given the young ages of children in this baseline, WG Short Set was altered slightly to be more child-friendly. For example, an original Washington Group question was: Do you have difficulty seeing, even if wearing glasses? InformEd re-worded the question slightly to be: Do you have difficulty seeing words in your books or on the blackboard, even if wearing glasses?

Disability is complex and there are many levels and differences within disabilities. Disability typically requires specialists to identify and diagnose type and severity. Therefore, the WG data is considered a secondary means of identifying CWD. Primarily, the baseline relies on school identification of CWD, though even this information may be incomplete or faulty due to the aforementioned difficulty in precise diagnosis.

¹³ <http://www.washingtongroup-disability.com/washington-group-question-sets/short-set-of-disability-questions/>

Annex D. Inter-Rater Reliability (IRR)

The original plan to test inter-rater reliability (IRR) was to do so during the piloting/practicing phase of training. However, given the restricted access to schools and limitations of training (see Limitations and Lessons Learned), it was decided instead to incorporate IRR during spot checks by the Country Manager and RTP staff. As part of the spot checks, the spot checker would orchestrate 1-2 pair assessments at the school they visited. In total, 12 students (8 from grade 1 and 4 from grade 4) were assessed by two supervisors simultaneously on days 3 & 4 of the official data collection.

Long one-way ANOVA techniques were used to calculate the intra-class correlation within pairs of assessors for a measure of IRR. Using Fleiss' benchmarks for excellent ($ICC > 0.75$), good or fair ($0.75 \geq ICC > 0.4$), and poor ($0.4 \geq ICC$); all literacy outcome variables showed excellent IRR except for fluency calculated from the total duration of the story.

Table D: IRR among 8 grade 1 and 4 grade 4 students on days 3 & 4 of data collection

Construct	IRR	Notes
Letters	0.97	Grade 1 only
Phonemic awareness	0.97	
Most used words	0.81	Grade 4 only
Reader/nonreader	1.00	Grade 4 only
Reading accuracy	0.96	Grade 4 only
Reading fluency (at 30 second mark)	0.98	Grade 4 only
Reading fluency (total duration)	0.52	Grade 4 only
Reading comprehension	0.93	Grade 4 only
Communication	0.99	
Collaboration	1.00	
Concentration	0.77	
Critical thinking	1.00	

Based on these results, the following recommendations were made to the Country Manager:

1. The supervisors should focus on pressing 'finish' at the end of the story right when the child finishes (for one pair-assessed student, one supervisor pushed finish after 39 seconds and the other after 101 seconds).
2. There was some disagreement on whether some of the most used words were correct, especially رَتَبَ and ثَغْبَان. Supervisors should be reminded of the criteria for correct/incorrect. For example, every harakat must be pronounced correctly for the word to be correct. Voice recordings can also help demonstrate examples of correct/incorrect pronunciations.

3. There was some disagreement on whether some shape drawings were complete and matched the examples. Sharing with supervisors pictures of drawings that match/do not match the example can help them to visualize what 'does not match' looks like.

Given the constraints and challenges of the data collection (as documented in Limitations and Lessons Learned), a follow-up round of IRR was deemed impractical so that all stakeholders could focus on delivering the fundamentals of a successful data collection.

Annex E. Intra-class Correlation (ICC)

Long one-way ANOVA techniques were used to calculate the intra-class correlation within schools for a measure of ICC for the outcomes upon which the impact evaluation will rely. Overall, it appears that 0.14 is a good summary measure of ICC to refine sample size calculations for EQIE impact evaluation.

Table E: ICC for Learner Assessment outcome measures

Measure	ICC	Sample size (across n schools)
% letters correct	0.28	808 (40 schools)
% phonemic awareness correct	0.11	1128 (40 schools)
% most used words correct	0.08	318 (20 schools)
Reader / nonreader	0.03	320 (20 schools)
Reading accuracy	0.21	289 (20 schools)
Fluency (total duration)	0.19	289 (20 schools)
% reading comprehension correct	0.15	280 (20 schools)
Reader with comprehension	0.10	322 (20 schools)
Average ICC across reading skills	0.14	N/A
% concentration correct	0.11	1107 (40 schools)
% communication correct	0.08	1127 (40 schools)
% collaboration correct	0.09	1123 (40 schools)
% critical thinking correct	0.11	1066 (40 schools)
% combined life skills	0.14	1127 (40 schools)

This ICC is higher than the ICC assumed for original sample size calculations (0.1). If RTP wants to maintain the same probability and confidence of detecting a 0.32-0.36 effect size in its impact evaluation, this would require expanding the midline sample size. If possible, this is most efficiently done by expanding the number of schools in intervention/comparison groups – for example, 15 students from each of 25 intervention and 25 comparison schools would allow detection of a 0.32 effect size with 90% confidence and a 0.36 effect size with 95% confidence. However, there are two caveats to this approach:

1. Additional schools would have to be found that have very similar levels of programmatic exposure to existing intervention/comparison schools.
2. The additional schools would not have baseline scores and therefore no longitudinal gain scores could be calculated, thus the midline would have to rely primarily on a cross-sectional look at midline scores and only secondarily at gain scores.

Alternatively, if the original sample size is maintained (15 students from each of 20 intervention and 20 comparison schools), the minimum detectable effect sizes would increase to 0.36 (with 90% confidence) and 0.4 (with 95% confidence). RTP may decide that these are sufficiently small

minimum detectable effect sizes, accepting the risk that if the actual effect size is smaller it may fail to be documented at midline.¹⁴

¹⁴ Original sample size calculations also assumed equal variance between intervention and comparison samples. This was found to be the case for the actual baseline measures of readers with comprehension and % combined life skills.

Annex F. Equity Regression Tables

Grade 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	% letters	% PA	% Concentration	% Communication	% Collaboration	% Critical thinking	% Life Skills combined
Student is female	0.05~ (0.03)	0.00 (0.05)	-0.08* (0.04)	0.02 (0.04)	0.04 (0.03)	0.07 (0.06)	0.01 (0.02)
CWD based on teacher identification or indirect teacher interview	-0.06 (0.06)	-0.30** (0.11)	0.03 (0.17)	0.11 (0.12)	-0.01 (0.09)	-0.02 (0.13)	0.02 (0.11)
Child reports attending ECD or preschool	0.07 (0.06)	0.10 (0.08)	0.04 (0.07)	0.14* (0.07)	0.04 (0.05)	-0.02 (0.07)	0.07 (0.05)
Student's age in years	-0.03 (0.05)	0.07 (0.04)	0.04 (0.06)	-0.03 (0.05)	0.06 (0.05)	0.08 (0.05)	0.04 (0.03)
# of child-reported members of household	-0.01~ (0.01)	-0.01~ (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)
# of types of print in the home	-0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	-0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)
Sum of student-reported literacy actions of others in household	0.01 (0.01)	0.02* (0.01)	0.02~ (0.01)	0.06** (0.01)	0.03** (0.01)	0.05** (0.01)	0.04** (0.01)
Sum of student-reported teacher practices	0.04** (0.01)	0.05** (0.01)	0.00 (0.01)	0.05** (0.01)	0.06** (0.01)	0.04* (0.02)	0.04** (0.01)
Total environmental threats identified by child	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	0.02* (0.01)	0.01~ (0.01)	-0.01 (0.01)	0.00 (0.00)
Summary index of community literacy activity participation	-0.02 (0.01)	-0.03* (0.01)	0.03* (0.02)	0.03* (0.01)	0.06** (0.01)	0.03 (0.02)	0.04** (0.01)
Urban school	0.00 (0.08)	0.04 (0.07)	0.06 (0.05)	-0.04 (0.07)	0.02 (0.04)	-0.11 (0.08)	-0.01 (0.04)
Girls School	0.09 (0.06)	0.13* (0.05)	0.11~ (0.06)	0.03 (0.05)	-0.08* (0.04)	0.04 (0.06)	0.02 (0.03)
Boys School	-0.06 (0.05)	-0.02 (0.05)	-0.04 (0.05)	0.08~ (0.04)	0.02 (0.04)	0.11~ (0.06)	0.05~ (0.02)
Principal report: % positive discipline minus % negative discipline	-0.08 (0.10)	-0.05 (0.07)	0.00 (0.09)	-0.01 (0.05)	0.09~ (0.05)	-0.09 (0.12)	-0.00 (0.04)
Principal report: School has teacher code of conduct	0.11 (0.06)	0.06 (0.05)	-0.05 (0.05)	0.01 (0.04)	-0.01 (0.04)	-0.08 (0.05)	-0.03 (0.03)

Principal report: School has access to trained point person for child protection referral	-0.05 (0.07)	0.03 (0.04)	0.06 (0.06)	0.06 (0.05)	0.06 (0.05)	0.06 (0.05)	0.06** (0.02)
Constant	0.64 (0.39)	-0.14 (0.31)	-0.03 (0.39)	0.09 (0.35)	-0.35 (0.34)	0.06 (0.39)	-0.10 (0.20)
Observations	706	708	705	708	708	685	708
R-squared	0.17	0.14	0.05	0.18	0.20	0.12	0.27

Robust standard errors in parentheses, ** p<0.01, * p<0.05, ~ p<0.1

Grade 4 (Reading Skills)		(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES		% PA	% MUW	Reader	Accuracy	WCPM	% Comprehension	RWC (1000.1)
Student is female		0.09~ (0.04)	-0.01 (0.05)	0.10 (0.07)	-0.03~ (0.02)	-1.79 (2.99)	0.04 (0.03)	-0.01 (0.06)
CWD based on teacher identification or indirect teacher interview		-0.07 (0.06)	-0.01 (0.05)	0.09* (0.04)	-0.00 (0.03)	-13.91 (8.91)	0.01 (0.06)	0.01 (0.16)
Student reports attending ECD or preschool		-0.10** (0.03)	-0.06 (0.05)	0.11 (0.13)	-0.00 (0.02)	1.12 (7.32)	0.04 (0.07)	0.09 (0.13)
Student age in years		-0.00 (0.04)	-0.02 (0.02)	-0.05 (0.05)	0.03~ (0.01)	7.88 (4.58)	0.02 (0.05)	0.02 (0.07)
# of student-reported members of household		-0.02* (0.01)	-0.01~ (0.00)	-0.03* (0.01)	-0.00 (0.01)	-0.10 (0.74)	-0.01 (0.01)	-0.04** (0.01)
# of types of print in the home		0.02** (0.01)	0.01 (0.01)	0.02 (0.01)	-0.00 (0.01)	-0.10 (1.33)	0.02 (0.02)	0.03 (0.03)
Sum of student-reported literacy actions of others in household		-0.00 (0.01)	-0.02* (0.01)	-0.02 (0.01)	0.00 (0.01)	0.48 (1.50)	0.02 (0.01)	0.01 (0.02)
Sum of student-reported teacher practices		0.03 (0.02)	0.03** (0.01)	0.03 (0.02)	0.00 (0.00)	2.00 (1.20)	0.03 (0.02)	0.08** (0.03)
Total environmental threats identified by student		0.00 (0.00)	0.01 (0.00)	0.01 (0.01)	-0.00 (0.00)	-1.28* (0.47)	-0.01~ (0.01)	-0.00 (0.01)
Summary index of community literacy activity participation		0.02 (0.01)	0.02~ (0.01)	0.02 (0.02)	0.01* (0.01)	1.71 (1.41)	0.00 (0.02)	0.01 (0.03)
Girls School		0.09** (0.02)	0.16** (0.03)	0.03 (0.05)	0.09** (0.02)	21.86** (4.07)	0.11* (0.05)	0.24* (0.09)
Boys School		0.08* (0.03)	0.04 (0.04)	-0.02 (0.05)	0.00 (0.03)	2.44 (4.07)	0.12~ (0.06)	0.03 (0.10)
Principal report: % positive discipline minus % negative discipline		-0.12 (0.11)	-0.12~ (0.06)	-0.22* (0.10)	-0.00 (0.07)	-2.25 (11.17)	0.02 (0.11)	-0.21 (0.18)
Principal report: School has teacher code of conduct		-0.04 (0.04)	0.00 (0.02)	0.10 (0.06)	-0.04 (0.03)	-0.76 (4.91)	-0.03 (0.08)	-0.03 (0.11)
Principal report: School has access to trained point person for child protection referral		0.03 (0.03)	0.01 (0.02)	-0.13* (0.06)	0.01 (0.02)	7.16 (4.99)	0.14* (0.06)	0.17~ (0.09)
Constant		0.96* (0.41)	1.02** (0.22)	1.41** (0.47)	0.64** (0.15)	-40.77 (52.19)	0.12 (0.47)	-0.05 (0.60)
Observations		295	293	295	268	264	263	295
R-squared		0.16	0.18	0.12	0.15	0.20	0.15	0.15

Robust standard errors in parentheses, ** p<0.01, * p<0.05, ~ p<0.1

Grade 4 (Life Skills)	(8)	(9)	(10)	(11)	(12)
VARIABLES	% Concentration	% Communication	% Collaboration	% Critical thinking	% Life Skills (1000.3)
Student is female	0.13 (0.08)	0.05 (0.08)	-0.04 (0.05)	0.00 (0.09)	0.04 (0.06)
CWD based on teacher identification or indirect teacher interview	-0.03 (0.14)	-0.01 (0.05)	-0.15* (0.05)	0.10** (0.03)	-0.02 (0.05)
Student reports attending ECD or preschool	0.05 (0.06)	0.07 (0.07)	0.13 (0.09)	-0.06 (0.04)	0.05 (0.05)
Student age in years	0.06~ (0.03)	-0.00 (0.04)	-0.02 (0.04)	0.05~ (0.03)	0.02 (0.02)
# of student-reported members of household	-0.00 (0.01)	-0.01 (0.01)	-0.01* (0.01)	0.00 (0.01)	-0.01* (0.00)
# of types of print in the home	-0.00 (0.01)	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)
Sum of student-reported literacy actions of others in household	-0.01 (0.02)	0.05** (0.01)	0.02* (0.01)	0.01 (0.01)	0.02* (0.01)
Sum of student-reported teacher practices	0.03 (0.02)	0.05** (0.02)	0.06** (0.01)	0.02 (0.01)	0.04** (0.01)
Total environmental threats identified by student	0.01 (0.01)	0.02** (0.01)	0.00 (0.00)	-0.01 (0.00)	0.01* (0.00)
Summary index of community literacy activity participation	-0.01 (0.02)	-0.02 (0.02)	0.05** (0.01)	0.00 (0.02)	0.01 (0.01)
Girls School	-0.02 (0.05)	-0.01 (0.03)	0.13** (0.03)	-0.00 (0.07)	0.02 (0.03)
Boys School	0.05 (0.07)	0.05 (0.08)	0.03 (0.06)	-0.04 (0.07)	0.03 (0.04)
Principal report: % positive discipline minus % negative discipline	0.22** (0.06)	0.02 (0.09)	-0.13 (0.09)	-0.01 (0.10)	0.02 (0.06)
Principal report: School has teacher code of conduct	0.01 (0.03)	-0.01 (0.03)	-0.05 (0.03)	0.02 (0.05)	-0.01 (0.02)
Principal report: School has access to trained point person for child protection referral	-0.06* (0.03)	0.02 (0.03)	0.02 (0.03)	0.04 (0.03)	0.00 (0.01)
Constant	-0.20	0.19	0.44	0.36	0.21

	(0.34)	(0.35)	(0.36)	(0.28)	(0.18)
Observations	295	295	295	291	295
R-squared	0.07	0.16	0.27	0.04	0.20

Robust standard errors in parentheses, ** p<0.01, * p<0.05, ~ p<0.1

Annex G. Detailed Lesson Observation Data

Note: The N varies by item because 'N/A' was an option and often used by observers.

Table G: Detailed lesson observation data, in order of lowest to highest average for each area

Variable	Average	N
Safe physical space for activity		
Teacher provides and explains safety instructions	3.1	36
Activity/play area is accessible to children with disabilities	3.3	36
Specifying activity area clearly	3.3	36
Activity/play area is clean, and clear of equipment, obstacles, or hidden areas that may cause injuries (e.g. Desks are rearranged to ensure children safety)	3.5	36
Activity/play area is leveled, and without cracks and holes, not slippery, etc.	3.7	32
Activity/play area is protected from direct sunlight/rain	3.8	32
Activity/play area is in a safe zone away from any possible danger such as near a road, deep/running water streams, exposed electric wires, etc.	3.8	11
Approach and ability to implement literacy-PBL		
Teacher uses open-ended questions to utilize RCA methodology	2.6	39
Teacher delegates roles/responsibilities to lead preparation of activity/play area, division of children groups, etc. (before/during/after activity)	2.7	33
Teacher links "closing discussion" (RCA questions) with learning objective	2.7	38
Teacher clearly identifies session objectives	3.3	39
Teacher encourages and acknowledges children's participation even if they provide incorrect answers	3.5	38
Teacher ensures that children understand questions and instructions (reiterates questions and instructions, or provides additional explanation if and when needed, uses language that is suitable for the children's age group, etc.)	3.5	39
Teacher communicates with children using positive verbal and body language, example good eye contact, smiling, positive explanatory gestures, listening attentively, calm speaking tone, etc.)	3.6	39
Inclusive learning		
Teacher includes children with disability in all activities and discussions	1.9	11
Teacher modifies lesson / activity to include CWD when / if needed	2.2	11
Teacher encourages girls and/or boys equally to play non-traditional gender roles (ones that don't conform to social construct)	2.9	33

Teacher provides girls and boys with equal opportunities for participation in planned activity (ex. Asking questions, providing answers, self-expression, etc.)	3.1	15
When a behavioral issue occurs, Teacher responds according to the behavior of the child and his/her development needs, and not their gender	3.2	34
Teacher does not use derogatory language when speaking to / about a child	3.7	38
Management of behavioral issues		
Teacher interjects and takes action when children display any type of bullying behavior towards their peers	3.8	32
Teacher manages child's behavior in a calm, patient and effortless manner	3.8	39
Teacher does not resort to physical/corporal punishment, does not disrespect or intimidate the children	3.9	38
Children interaction/engagement within classroom		
Children participate actively	2.2	37
Children share their ideas, opinions and feelings with their peers with encouragement and support from the Teacher	3.4	39
Children ask questions freely and without any hesitation, fear or worry	3.5	39
Children seem to be comfortable and relaxed ex. Smiling, laughing, don't show any signs of worry	3.8	39

Annex H. Baseline Documentation

Tools

Learner Assessment	 Learner Assessment_Final Pa	(Arabic and English)	
Life Skills and Reading Teacher Interview for CWD	 Life Skills and Reading Teacher Int	(Arabic and English)	
Caregiver Interview	 Parent Survey_Palestine Kol	(Arabic and English)	
Teacher Interview	 Teacher Interview_Global Ko	(Arabic and English)	
Principal Interview	 Principal Survey_Palestine Kol	(Arabic and English)	
Supervisor Survey	 Supervisor Interview_Palestine	(Arabic and English)	
Classroom Observation	 PS. Classroom Observation ARABIC	 Ps. Classroom Observation ENGLIS	(English)
Education Official Interview	 District Education Official Interview_Pa	 District Education Official Interview_Pa	(English)

Supporting Documents

Palestine EQIE Narrative  !FINAL-Palestine Country Specific Nar	Inception Report  RTP EQIE Baseline_Inception F
Enumerator Manual  Enumerator Manual_Final_v3 Pal	Country Manager Manual (see Annex for Palestine Baseline ToR)  Country Manager Manual Palestine v3

Annex I. Evaluation Methods and Limitations

Right to Play report guidelines and page limits require that the evaluation methods and limitations be a maximum of 3 pages with all remaining details documented in this annex.

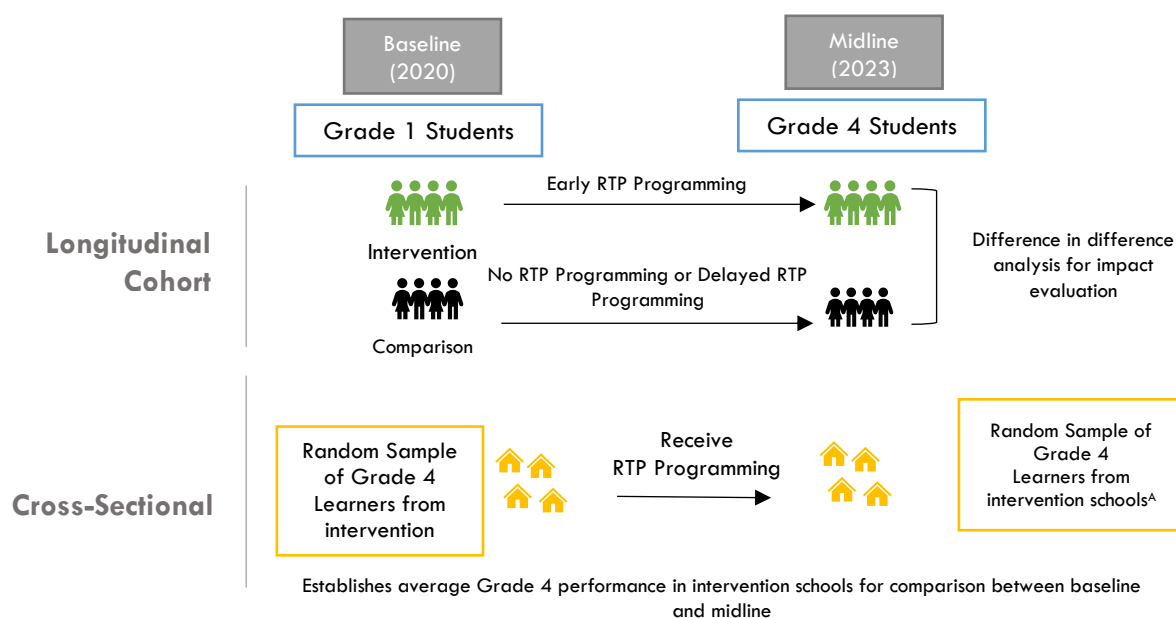
Study Design

As part of this baseline, InformEd worked closely with Right to Play to design as rigorous an overall evaluation as possible within various constraints. The resulting evaluation is quasi-experimental, taking advantage of the phased approach to implementation among West Bank schools. This design maximizes Right To Play's ability to **(A) examine the DAC criteria of effectiveness and impact at midline**, while also **(B) reporting on project indicators**.

The baseline established an intervention and comparison sample of students (in Grade 1 at baseline) to be followed longitudinally throughout the life of the program. These same students can be followed and re-assessed in Grade 4, at the program's midline in 2023. This will allow for difference in difference analysis, comparing learner gains in intervention schools to learner gains in comparison schools (controlling for exposure to EQIE programming).

Additionally, the baseline measures the performance of students in Grade 4 (the age used to measure the EQIE Program Objective indicators). A cross-sectional random sample of Grade 4 students from intervention schools were assessed. At midline, the baseline cohort of Grade 1 students will be in Grade 4, to compare performance at baseline and midline.

Figure I.1: Evaluation design summary



^AThe midline cross-sectional sample will be most effective if the assessment takes place at the same time of year as the baseline (Nov/Dec). However, the baseline-midline comparison will be made much less precise because 2020 was an exceptional school year due to COVID disruptions.

This evaluation design allows for flexibility at the program endline in 2024, depending on which evaluation questions are prioritized at that point. For example, RTP has the option of reassessing a random sample of Grade 4 students in all or a subset of intervention schools in order to add to the cross-sectional comparisons in Grade 4 student performance between baseline, midline, and endline.

Sample Size and Selection

This section provides a summary of sample size and selection, with a focus on the Learner Assessment. For full details, including rationale for sample size/selection of other tools, see InformEd's RTP EQIE Baseline Inception Report (Annex H).

Learner Assessment Sample Size

As the ultimate beneficiaries of the EQIE project, a sufficient sample size of students was needed to achieve (A) the impact evaluation objective of documenting a reasonable magnitude of project impact (if one exists) and (B) the baseline (and midline, endline) objective of providing population point estimates with reasonable accuracy.

- A. For impact evaluation, it was decided it would be worthwhile to document effect sizes as low as 0.32-0.36.¹⁵ Given informed assumptions and calculations (see Inception Report); the longitudinal sample during the 2023 midline would need to average 15 students per each of 20 intervention and 20 comparison schools. This sample allows detection of a 0.32 effect size with 80% probability and 90% confidence, or a 0.36 effect size with 95% confidence.¹⁶ This calculation assumed an intra-school correlation (ICC) of 0.1, however given actual baseline data, the ICC is closer to 0.14 (see Annex E for the actual ICC and discussion of sample size implications). With a baseline sample size of 20 students per each of 20 intervention and 20 comparison schools, this calculation also assumed 25% attrition by midline. This is likely an underestimate of actual attrition, and so RTP will need to randomly replace students at midline if less than 15 of the original 20 students can be found in 2023.
- B. For population point estimates at a given point in time, 15 students per each of 20 intervention schools provides a point estimate of +/- 7.3 percentage points with 90% confidence (or, in the case of baseline Grade 1 with 20 students per school, +/-6.9 pp with 90% confidence). In planning for the baseline, RTP found this an acceptable level of precision. However, given the actual ICC of 0.14, the precision of point estimates is now lowered to +/- 8.1 pp with 90% confidence (or, in the case of baseline Grade 1 with 20 students per school, +/-7.8 pp with 90% confidence).¹⁷

¹⁵ Given the agreed-upon sample size, if the actual program impact is of lesser magnitude than this, there is a substantial (greater than 20%) probability that such impact would not be documented with at least 90% confidence. However, the ability to document lesser impact would have required a much greater sample size.

¹⁶ The following program/commands were used in Stata to make this calculation: `clustersampsi, detectabledifference mu1(1) sd1(1) m(15) k(20) rho(0.1) beta(0.8) alpha(.1)`

¹⁷ Point estimate sample size calculations performed using [UKSamples online calculator](#).

Table I.1: Learner Assessment baseline sample size

Population	# Schools	# Students	Strategy
Grade 1 students (intervention)	20	400 (20 per school)	Purposefully sample children with disabilities first, then randomly sample up to target number, striving for balance between boys and girls if in mixed school
Grade 4 students (intervention)		300 (15 per school)	
Grade 1 students (comparison)	20	400 (20 per school)	

Learner Assessment Sample Selection

There were two levels of sample selection: the school level and individual level. At the school level, most of the 2020 implementation phase schools were selected for the intervention sample. To minimize exposure to the EQIE program over the life of the evaluation, the comparison sample contains as many schools designated for later phases as possible. Nonetheless, as discussed in the Limitations section, it will be necessary to control for program exposure (at both school and individual levels) during midline impact analysis, as well as for statistically significant differences observed between the two groups of schools (see EQ1).

Table I.2: School-level sample selection

Intervention sample	Comparison sample
20 of the 21 schools from the 2020 project implementation phase (in Hebron and Nablus governorates)	All 14 schools from 2023 phase, plus 6 schools from 2022 phase (in Qalqilya, Tubas, Tulkarem, and Yatta governorates)

As the 2020 implementation phase contained 21 schools, one school was excluded from the intervention sample because it had an existing library program (the other 20 schools did not). At the time of school sample selection, schools were not yet officially designated for the later implementation phases. Thus, RTP worked with MOEHE to identify 20 schools to be implemented in 2022 and 2023.

At the individual sample selection level, students were selected using a two-step process:

1. Given the EQIE project's focus on CWD, all CWD (if any were enrolled in the target grade) were first purposefully selected. According to RTP Palestine CO, children with glasses and hearing aids were also considered CWD.
2. Next, students were randomly selected up to the target number (20 for Grade 1, 15 for Grade 4). If there was more than one section within a grade, students were to be randomly selected from multiple sections to guard against the possibility that some schools had sorted students by performance. If the school was mixed-sex, a roughly equal number of boys and girls were to be randomly selected. However, in practice there were three mixed schools in which the number of either boys or girls far outnumbered the other sex, which was reflected in the final sample.

Sample Size and Selection for other Respondents

In addition to students, the table below summarizes the additional respondents from which complementary or triangulation data was gathered (for detailed rationale of sample size/selection, see Inception Report). The next section summarizes the various tools used to gather information from the students and other respondents.

Table I.3: Target and achieved sample by type of respondent/tool

Tool	Target	Achieved	Notes
Grade 1 Learner Assessment	800	810	Plus 3 students from Teacher Interview for CWD
Grade 4 Learner Assessment ^A	300	320	Plus 2 students from Teacher Interview for CWD
Grade 1 Lesson Observation ^A	20	19	One observation per intervention school. One teacher was on sick leave, and no alternative could be found before school closure from COVID lockdown
Grade 4 Lesson Observation ^A	20	20	One observation per intervention school.
Grade 1 Teacher Interview ^A	20	21	One interview per intervention school. However, at one school two teachers were interviewed.
Grade 4 Teacher Interview ^A	20	20	One interview per intervention school.
Principal Interview	40	40	One interview per each of intervention and comparison schools.
Grade 1 Caregiver Interview	200	206	Random subset of caregivers of assessed students.
Grade 4 Caregiver Interview	200	197	Random subset of caregivers of assessed students.
Education Official Interview	3	3	Chosen by RTP for their relevance to the EQIE project.
Supervisor Interview	8	11	2 in Hebron, 3 in Nablus, 1 in Qalqilya, 1 in Tubas, 2 in Tulkarem, and 2 in Yatta

^ARespondents from intervention schools only (see Data collection tools table for intervention/comparison split)

In general, the strategy was to target respondents as closely associated with the sample of students as possible. That means the sample of caregivers was randomly selected from among those of the assessed students, the teachers are the reading/language teachers of the assessed students, the principals are those from the schools of assessed students, and the education officials and supervisors are those who work in the governorates containing the schools of the assessed students.

In the case of teachers, 36 of the 41 interviewed teachers were also those observed (in three cases, a substitute teacher was observed). Most, but not all observed lessons were reading/language lessons (34 of 39 lessons). Because many schools had more than one section of students per grade, the match between teacher and student samples is close, but incomplete (a portion of the sampled students may be taught by a teacher other than the interviewed/observed teacher). Similarly, because not all caregivers of the assessed students were interviewed, the match between caregiver and student is close, but also incomplete. Thus, caution should be exercised when making conclusions about the relationship between teacher/caregiver variables and student variables.

Data Collection Tools

In total, seven data collection tools were drafted or adapted from authoritative sources for use in this baseline. Five tools were primarily quantitative and as such were digitized using KoBo software. The Education Official Interview was primarily qualitative and thus kept on paper, and the nonlinear nature of the Lesson Observation meant that paper was deemed to be more efficient for that tool as well. Copies of final tools are available in Annex H.

Table I.4: Data collection tools

Tool	Intervention	Comparison	Administered by	Modality
<i>Learner Assessment</i> ^A	Grade 1 & 4	Grade 1 only	MOEHE Supervisors	In-person, using KoBo
<i>Lesson Observation</i>	Grade 1 & 4	N/A		In-person, on paper
<i>Teacher Interview</i>	Grade 1 & 4	N/A	Enumerators	Phone, using KoBo
<i>Principal Interview</i>	Yes	Yes		
<i>Caregiver Interview</i>	Yes	N/A		
<i>Education Official Interview</i>	Yes		InformEd Country Manager	Phone, on paper
<i>Supervisor Interview</i>	Yes		Self-administered	KoBo

^AThe Learner Assessment had two versions. The master version was used as a one-on-one oral assessment with the vast majority of students. However, a very small minority of students had a disability so severe that the master version was not appropriate to their needs. For these students (5 total), their teacher was interviewed to gather similar information through a 'Reading and Life Skills Teacher Interview for CWD.'

Learner Assessment

The Learner Assessment assessed students' early grade reading skills and social and emotional learning skills. The tool included a student background questionnaire, a reading assessment (subtests of letter knowledge, phonemic awareness, most used words, oral reading passage, and comprehension), a socioemotional skill assessment (subtests of concentration, communication, collaboration, and critical thinking), and a child-friendly version of the Washington Group short set of disability questions. For details of the reading and life skill assessment, see Annex C. For inter-rater reliability, see Annex D.

Classroom Observation

RTP's existing classroom observation form was used, with some additional items on teacher background information and lesson plan. The observation covered five main areas: Safe physical space for activities; Teacher approach and ability to implement literacy through play-based learning (PBL); Inclusive learning; Management of behavioral issues; Child interaction/engagement with classroom.

Teacher, Principal, and Supervisor Interview

Teachers and principals were interviewed over the phone, and supervisors self-administered the survey to themselves. Data from all three types of respondents was gathered electronically using Kobo.

Teacher Interview	Principal Interview	Supervisor Interview
General/background information PBL Training and support PBL knowledge, attitude, & practices Child behavior Learning environment Disability Teacher Wellbeing	General/background information School-level information (type of school, location, etc.) PBL Training and support Discipline techniques Disability School Safety	PBL Training and support PBL Teaching methods Discipline techniques Disability

Caregiver Phone Survey

Caregivers of a subset of students were interviewed over the phone, with information recorded electronically via KoBo, on: caregiver background, household literacy activities, barriers to inclusive and gender-responsive education, and knowledge of safe and inclusive learning environments (emotionally, physically inside the classroom, and physically outside the classroom).

Education Official Interview

Three MOE-officials relevant to the EQIE project were qualitatively interviewed about education priorities and challenges, hopes for EQIE, barriers faced by CWD and other marginalized children, and strategies for creating a safe learning environment (emotionally, physically inside the classroom, and physically outside the classroom).

COVID Precautions

The Learner Assessment (and the Classroom Observation) were the only two tools to be administered in-person. As such, data collectors (MOEHE Supervisors – see Limitations section for more information) were trained remotely and took precautions to minimize the possibility of spreading any illness in schools. Supervisors wore protective gear such as masks and gloves, assessed students in a semi-private location away from other people, maintained as much distance from assessed students as reasonably possible, and sterilized laminated materials between assessments. When any Supervisors or anyone in their household had any indication of illness, that Supervisor immediately ceased participation in the data collection.

Analysis and Interpretation

Data was organized and pre-cleaned in Microsoft Excel, with further cleaning and analysis carried out using STATA. Where summary indices were created, steps were taken to ensure that the constituent pieces of each index were placed on a comparable scale before combining.

Summary statistics are used to describe the data. In some cases, the average may not summarize the data well if, for example, outcome variables are not normally distributed as is often the case for reading skills. This report highlights those cases and presents a histogram to represent the abnormal distribution more accurately.

For disaggregations, comparisons, and relationships that generalize from a sample to a larger population, statistical inference is used. Importantly, to maximize clarity this report ONLY presents disaggregations, comparisons, and relationships that were found to be statistically significant with at least 90% confidence. Exceptions include Programming Area Sample Characteristics, indicator reporting for the Project Management Framework (PMF, Annex B), and CWD – for which disaggregations are shown regardless of their statistical significance. In most cases, t-tests and regressions are used to evaluate statistical significance and the average relationship between variables. For student-level data, clustered t-tests and multilevel regression models are used to adjust standard errors for the fact that students are grouped in schools and are likely to share some variation for being in the same learning environment.

Interpretation

Throughout this report, there are sections labeled ‘Interpretation.’ These sections mark where objective presentation of results ends, and where subjective interpretation begins. The purpose of the Interpretation sections is to offer a starting point for bridging the widespread gap in development program evaluation & research: the gap between analysis and action. This gap appears when analysts are often hesitant to provide more than a detached overview of results, and practitioners are sometimes hesitant to attempt a vigorous interpretation of results.

Thus, the Interpretation sections seek to encourage enthusiastic use of results for data-driven decision-making. These sections are NOT intended as the final word or ultimate interpretation of results, but rather ONE possible interpretation of results to spark the process of data-driven programmatic action. The interpretation is informed by the author’s knowledge and experience but also subject to the author’s personal biases and incomplete knowledge of context, programming, etc.

Limitations

Pandemic

The enumerator training and data collection period fell in November/December 2020. As such, the ongoing Coronavirus/COVID pandemic was a constant source of challenge and uncertainty, with impacts on data quality.

In accordance with government regulations, no personnel from outside MOE were allowed inside schools. As such, instead of using standard enumerators, the Learner Assessment relied on MOE supervisors. Fortunately, most supervisors were familiar with early-grade reading assessments.

However, some were less familiar with touchscreen interfaces, and all supervisor training had to be conducted remotely over Zoom. In the end – with the immense effort of RTP Palestine staff, the InformEd Country Manager, and the supervisors themselves – a one-day training and half-day follow-up training were conducted that sufficiently prepared the supervisors to administer the Learner Assessment. Thus, while the data collection was ultimately successful in general, training conditions were far from ideal. Typically, training to administer a Learner Assessment of this complexity would include at least three days of in-person training in addition to two days of piloting/practicing at actual schools.

The data collection, originally planned for ten nonconsecutive days between 29 November and 20 December, had to be truncated when the West Bank began another COVID lockdown on 11 December. Again, only the quick and dedicated work of RTP Palestine staff, the InformEd Country Manager, and the supervisors themselves made it possible to finish the work before lockdown began.

To mitigate risks of impact to data collection, the electronic data was closely monitored, especially during the first few days of data collection. Any mistakes or strange patterns were immediately reported to the Country Manager to address with the individual supervisors in question (or with the team as a whole, if the issue was widespread enough).

KoBo literacy widget

KoBo has a relatively new feature, a literacy widget, to accommodate reading skill assessments involving the timed tests of items such as words in a reading passage. However, there was one persistent bug in this feature: the data collector cannot mark an item at X seconds unless at least one word is already marked as incorrect. Thus, a workaround was employed in which the name of the Country Manager (Mohammad) was placed at the beginning of the reading passage, and data collectors were trained to always mark this initial word as incorrect. One additional challenge that was not quite a bug was that, unless data collectors marked the last item after pushing ‘stop,’ no data would be recorded. Thus, the Country Manager made videos to share with the data collectors to demonstrate the correct approach to both these issues, and in addition students were over-sampled to compensate for any data lost to these issues.

Children with disability

As noted and evidenced in the body of this report, disability is both relatively rare and also highly individualized in terms of type and severity. Many disabilities may be ‘invisible’ if a comprehensive, technically-specialized system is not in place for testing and diagnosis. Thus, it is possible that there is some degree of mislabeling of CWD or non-CWD in the data for this report. To mitigate this challenge, information on disability is triangulated both from the school/teacher and from the students themselves (using the Washington Group Short Set questions).

Imperfect recall and self-reported data

Many of the tools rely on the honest memory of the respondent. Human memory is imperfect in general, and subject to self-promotion bias (both consciously and unconsciously). Thus, the data gathered should not be seen as a perfectly precise representation of reality, but either a rough representation or one in which respondents may have minimized information that reflects negatively on them and emphasized information that reflects positively on them. To minimize

this bias, triangulation was used where reasonable (either from additional respondents, or from direct observation).

In at least one case – self-reported teacher practices – the data largely does not seem to be of value. Self-reported teacher PBL practices are not correlated either with classroom observation nor with student-reported teacher practices (though classroom observation and student-reported teacher practices are correlated with each other). Furthermore, there is no correlation between teacher self-reported practices and student reading and life skill outcomes. Thus, the classroom observation was used as a more valid method to calculate Indicator 1100.1. In any case, it is important to keep in mind that this does not necessarily mean teachers were consciously giving false responses – psychology has documented many types of unconscious bias in human nature. Not only that, but teachers have been under substantial strain as the pandemic has opened and closed schools and sometimes forced an abrupt transition to remote teaching over the past year.

The classroom observation, too, is also subject to potential bias, as most people will undoubtedly change their behavior when observed versus unobserved. Furthermore, a single observation captures only a limited representation of observed teacher behavior. Nonetheless, overall the classroom observation appears to give a better idea of what teachers are capable of, and does significantly correlate with student reading and life skill outcomes.

Quasi-experimental nature of research design

As RTP found it infeasible to randomize intervention and non-intervention schools, a second-best approach of quasi-experimental design was used for impact evaluation purposes. This means that a comparison group of schools was chosen – in this case, schools designated to receive the EQIE program during a later phase of intervention. Due to the programmatic approach, intervention and comparison schools are from different governorates. As seen in the report Section 4.3, intervention and comparison schools are similar enough for a quasi-experimental impact evaluation, especially in terms of student reading and life skills. However, there are some school-level differences, and it is possible that there are also unobserved differences between intervention and comparison schools. Controlling for observed differences will help mitigate bias in the midline impact analysis, but the quasi-experimental design cannot rule out the possibility that any midline differences in student reading and life skills is due to unobserved differences between intervention and comparison schools.

Limited sample size

Originally, a larger sample size was proposed – especially to counter predicted attrition over the three years between baseline and midline. However, due to budget constraints the sample size was reduced to the bare minimum. This leaves little room for error and increases the possibility that the EQIE project will have a (small) impact, yet the sample size will be insufficient to detect this impact at midline. The risk is increased from the discovery that intra-class correlation (ICC) is actually closer to 0.14 rather than 0.1 initially used to calculate sample size. One mitigation strategy is to consider options for increasing sample size at midline. The Methodology section and Annex E discuss the sample size implications in detail. Another mitigation strategy is to maintain 90% confidence as a minimum level of statistical inference for the midline impact

analysis – if RTP is comfortable with this level of confidence, which implies a 10% likelihood of falsely concluding impact when no impact exists (this baseline report uses 90% confidence as a minimum level of statistical inference, additionally indicating when results are significant with 95% and 99% confidence).