

ENHANCING QUALITY AND INCLUSIVE EDUCATION BASELINE



RIGHT TO PLAY
PROTECT. EDUCATE. EMPOWER.

August
2021

Baseline Report Lebanon - FINAL

Acknowledgments

Prepared by: **Ashley Craft and Celia Kiene**, Cynosure Evaluation Limited

*With deep gratitude to Right to Play Global Senior MEL Manager **Eric Opoku**, Lebanon Country Office MEL Officer **Ghinwa Monzer**, and Country Office staff **Ma'amoun Salhab**, and **Milia Jebran**.*

With special thank you to the team of enumerators:

Batoul Othman, Perla Ayoub, Marwa Ibrahim, Mohammad Ghalayini, Rita Ayoub, Samar Akiki, Wajih Saadeldine, Petra Moussa, and Rita A

Table of Contents

Acknowledgments.....	2
Table of Contents.....	3
List of acronyms and abbreviations	8
1. Executive summary	10
Project background	10
Methodology.....	10
Key findings and recommendations	11
1.1 EGRA Arabic scores are relatively high at baseline, but averages mask that 10% are struggling	11
1.2 Children who do not correctly report harmful behavior also have the lowest Life Skills scores	11
1.3 Life Skills levels can benefit from specific approaches for high and low-scoring students.....	12
1.4 Availability of selected reading materials at home is significantly correlated with attendance of extracurricular literacy activities	13
1.5 Cultural attitudes towards girls may hinder RTP's implementation and further gender stereotypes	13
1.6 Parents' attitudes are key to inclusive classrooms, and an opportunity to seize for RTP	14
1.7 Teachers self-reporting usage of playful methodologies should be triangulated using classroom observations.....	14
1.8 There is evidence that around half of children with disabilities may drop out from Grade 1 to Grade 4 and that teachers may need help identifying children with disabilities in their classroom	15
2. Background and context	16
Right to Play	16
Cynosure Evaluation	16
Lebanon Context.....	16
RTP's involvement in Lebanon.....	18
EQIE program in Lebanon	19

Direct beneficiaries	19
Indirect beneficiaries	19
Main intervention pathways.....	19
Changes from original proposal.....	20
3. Approach.....	20
3.1 Evaluation purpose	20
3.2 Evaluation questions.....	20
3.3 Marginalization and equity indicators	21
3.4 Overall evaluation design.....	22
3.5 Ethical considerations	23
3.6 Training	24
3.7 Dissemination plan.....	24
3.8 Challenges and limitations.....	25
3.8.1 Implementation and data collection sites	25
3.8.2 COVID-19.....	25
3.8.3 Conflict	25
3.8.4 SurveyCTO and JavaScript EGRA tool	25
3.8.5 Selective administration of EGRA tasks	26
3.8.6 Imperfect recall and self-reported data	26
3.8.7 Reliability of child respondents	26
3.8.8 Quasi-experimental nature of research design	27
3.8.9 Limited sample size and representativeness.....	27
4. Evaluation Methodology.....	28
4.1 Tools.....	28
Learner assessment.....	28
4.1.1 Early Grade Reading Assessment (EGRA)	28
4.1.2 International Social and Emotional Learning Assessment (ISELA)	29
4.1.3 Washington Group Short Set	31
4.1.4 Teacher classroom interviews	32
4.1.5 Director interviews	32

4.1.6 Parent phone survey	32
4.1.7 Lesson observations	32
4.2 Sampling Framework	33
4.3 Representativeness of the sample	34
4.3.1 Representativeness of EGRA tasks administered with JavaScript plugin	34
4.4 Cohort tracking	34
4.5 Data quality and Security Assurance	34
5. Key characteristic subgroups and barriers at baseline	35
5.1 Respondent demographics	35
5.1.1 Students	35
5.1.2 Teachers	35
5.1.3 Directors	36
5.1.4 Parents	37
5.2 Demographics by marginalized group	37
5.2.1 Girls: their lives, social context, and demographics	37
5.2.2 Girls' daily patterns of activities	37
5.2.3 Girls' dropout rates	38
5.2.4 Children with disabilities: demographics, access, and presence in learning centers	38
5.3 Barriers to participation	40
5.3.1 Barriers affecting girls' attitude and performance in learning centers	40
5.3.2 Children with disabilities: conditions and social context affecting access and presence in centers	40
6.2.1 EGRA - What can the baseline tell us about learners' reading skills?	41
6.2.1.1 Task 1: Letter Identification	42
6.2.1.2 Task 2: Phonemic Awareness	43
6.2.1.3 Task 3: Most Used Words	45
6.2.1.4 Task 4a: Reading Passage	46
6.2.1.5 Task 4b: Reading Comprehension	49
6.2.1.6 Task 5: Listening Comprehension	51
6.2.1.7 EGRA results Summary	51

6.2.2 How do these reading skills differ by equity indicators?.....	52
6.2.2.1 Gender	52
6.2.2.2 Other equity dimensions	53
6.3 Life skills scores in intervention learning centers	54
6.3.1 What can the baseline tell us about learners’ life skills?.....	54
6.3.1.1 Critical thinking	55
6.3.1.2 Communication.....	56
6.3.1.3 Collaboration	56
6.3.1.4 Concentration	57
6.3.2 Life Skills and gender	58
6.3.3 Life Skills and poverty levels	59
6.3.4 Life Skills and disability	60
6.3.5 Life Skills and daily activity patterns.....	60
6.3.6 Reading activities with significant correlation to Life Skills scores.....	61
6.3.7 Risk variables with significant correlation to Life Skills scores	62
6.3.8 Positive answers to “Relate” variables significantly influences Life Skills scores.....	62
5.4 READING ACTIVITIES IN THE COMMUNITY: PARENT SURVEY	64
6.4.1 What supplemental reading activities are available to children (in homes and communities)?	64
6.4.2 How do reported reading activities differ by parents and children?	64
6.4.3 Of those activities, which do children take part in? How frequently?	65
6.4.4 How does participation differ by equity indicators?	66
6.4.5 How does participation in at-home reading activities differ by presence of reading materials in the home?	66
6.4.5.1 Extracurricular literary activities.....	66
6.4.5.2 EGRA scores	67
5.5 TEACHER PROFESSIONAL DEVELOPMENT.....	67
6.5.1 How does teaching practice correlate with teacher well-being?	67
6.5.2 How does teachers’ perception of their self-efficacy affect their ability to adapt to new teaching practices?	67

6.5.3 What teaching knowledge, attitudes, & practices regarding inclusive PBL/literacy instruction are common among teachers at baseline?	68
6.5.4 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.	69
6.5.5 How do teachers adapt teaching practices for CWD and gender-responsiveness?	69
2.5 To what extent does RTP training influence teacher knowledge, attitudes, and practices regarding inclusive PBL/literacy instruction?	70
6.6 BUILDING A CULTURE OF READING	70
6.6.1 Knowledge and practice of PBL and inclusive education in the learning centers (Lebanon).....	70
6.6.2 Knowledge and practice of PBL and inclusive education in the community.....	72
6.6.3 How do parents' beliefs about teaching support or conflict with RTP's approach and teacher's approach to learning?	73
6.7 SAFE AND INCLUSIVE LEARNING ENVIRONMENTS	74
6.7.1 Emotional and physical safety in the learning spaces	74
6.7.2 What is the baseline capacity (measured by knowledge, attitudes, and practice) of community members in creation and maintenance of safe learning environments?	75
6.7.3 What barriers do community members face in creation and maintenance of safe learning environments, despite knowledge and attitudes?	75
6.7.4 What is the current state of community members and leaders in terms of knowledge regarding ways to address barriers to education faced by girls and children with disabilities? ..	77
6.8 LEAVING NO ONE BEHIND.....	77
6.8.1 What are the beliefs held by parents, teachers and principals regarding gender and disability and access to education?	77
6.8.2 How much do all the actors support inclusive teaching versus differentiated teaching for CWD?.....	78
6.9 INNOVATION LABS AND ICT USAGE	79
6. KEY FINDINGS AND RECOMMENDATIONS.....	79
7.1 EGRA Arabic scores are relatively high at baseline, but averages mask that 10% are struggling.....	79
7.2 Children who do not correctly report harmful behavior also have the lowest Life Skills scores	80
7.3 Life Skills levels can benefit from specific approaches for high and low-scoring students.....	80

7.4 Availability of selected reading materials at home is significantly correlated with attendance of extracurricular literacy activities	81
7.5 Cultural attitudes towards girls may hinder RTP’s implementation and further gender stereotypes	81
7.6 Parents’ attitudes are key to inclusive classrooms, and an opportunity to seize for RTP	82
7.7 Teachers self-reporting usage of playful methodologies should be triangulated using classroom observations.....	82
7.8 There is evidence that around half of children with disabilities may drop out from Grade 1 to Grade 4 and that teachers may need help identifying children with disabilities in their classroom	83
Annex 1 – PMF Baseline indicators – Lebanon	84
Annex 2 – Terms of Reference.....	85
Annex 3 – Analysis framework.....	86
Annex 4 – ISELA competencies mapping	87
Annex 5 – Baseline survey research instruments.....	88
Annex 6 – Inception report	89

List of acronyms and abbreviations

4Cs	Concentration, Communication, Collaboration, Critical Thinking
COTT	Continuum of Teacher Training
CSO	Civil Society Organization
CWD	Children with disabilities
ECD	Early Childhood Development
EGRA	Early Grade Reading Assessment
EQIE	Enhancing Quality and Inclusive Education
FCDO	UK Foreign, Commonwealth, and Development Office
ISELA	International Social and Emotional Learning Assessment
LAL	Lebanese Alternative Learning
MDE	Minimum Detectable Effect
MLA	Monitoring Learning Achievement
MUW	EGRA Task 3: Most Used Words
OECD-DAC	Organization for Economic Cooperation and Development - Development Assistance Committee

PA	EGRA Task 2: Phonemic Awareness
PMF	Program Management Framework
PPI	Progress out of Poverty Index
PTA	Parents and Teachers Association
RTP	Right to Play
SDG	Sustainable Development Goals
SEL	Social and Emotional Learning
SES	Socio-economic status
SMC	School Management Committee
UNDP	United Nations Development Program
UNICEF	United Nations Children's Fund
UNRWA	United Nations Relief and Works Agency for Palestinian Refugees in the Near East

1. Executive summary

Project background

The overall objective of the project is to improve holistic learning outcomes of boys and girls, including children with disabilities, in primary and pre-primary institutions in the Palestinian refugee camps in Lebanon. This includes the number and percentage of children in schools achieving minimum reading proficiency levels and life skills in lower primary Grade 4 (OBJ 1000). To achieve this the project will increase teacher's knowledge and pedagogical skills that will enhance the reading skills of children through play-based learning (INT 1100; IMM 1110; OUT 1111). Teaching and learning materials will be provided to further support teachers delivery of child-centred learning (OUT 1112). The project will also increase children's participation in supplemental reading activities, using play-based learning to motivate and ignite an interest in reading, as well as ensuring that children increase the time they spend reading inside and outside school (INT 1200; IMM 1210; OUT 1213). To secure local ownership the project will also increase community and parental engagement to support home-based reading activities, and work in close collaboration with SMCs and PTAs to facilitate activities that support opportunities for children to read outside school (OUT 1211; OUT 1212).

To address the high rates of out-of-school children, the project will increase the enrolment of children with disabilities (INT 1400; IMM 1410). In close collaboration with CSO partners and supporting them in strengthening organisational development needs (OUT 1411), the project will implement initiatives to address barriers to education for children with disabilities (OUT 1413). This will include community sensitization events and training teachers on inclusive education practices. The project will also connect CSO partners with UNRWA Education Department officers, SMCs and PTAs to form School Inclusion Committees (OUT 1313), support construction and rehabilitation work (OUT 1311) and conduct sensitization sessions on child protection, child rights, corporal punishment, abuse, and/or GBV (OUT 1312). This will act as a means to increase the capacity of school communities to create and maintain safe and inclusive learning environments (INT 1300; IMM 1310). Kindergartens will also be supported to deliver quality pre-primary education and increase pre-primary enrolment of children with disabilities (INT 1400).

Methodology

The baseline evaluation of Lebanon follows the overall five-country EQIE baseline strategy. It uses a set of quantitative surveys adapted from the global tools to gather baseline data from key target groups of students, teachers, parents, principals, and classrooms which can be compared at midline and endline. Data gathered are reported against the analysis framework and the Program Management Framework (PMF), and to answer a series of evaluation questions set out in the baseline inception report. The program tracks two key sets of outcomes, performance on the Early

Grade Reading Assessment (EGRA) and on selected Socio-Emotional Learning (SEL) life skills scores. Lebanon, along with Palestine, have opted for the '4Cs' socio-emotional learning scores (See Section 4 below).

In Lebanon, the program was not able to use comparison groups at baseline due to limited comparability and difficulty obtaining permission to access UNRWA refugee camps. Therefore, the Lebanon program evaluation will track individual students from Grade 1 in target learning centers to assess changes at midline and endline. In case too many individual students are lost due to attrition, the program can switch to a cohort approach and top up numbers at midline and endline. See Section 3 for further details.

Key findings and recommendations

1.1 EGRA Arabic scores are relatively high at baseline, but averages mask that 10% are struggling

For learning outcomes, Lebanon learners have relatively high levels of Arabic and by Grade 4 some are achieving the maximum score on some EGRA tasks which will make it harder to show improvement at midline and endline (ceiling effect).

We recommend lengthening the Task 4 reading passage to 100 or more words at midline and endline so learners can show improvement beyond the ceiling of the current assessment. Analysis at midline and endline would need to draw a fair comparison accounting for any adjustments to these tasks.

This being said, the baseline survey identified that around 10% of learners are struggling with the basics of literacy, which may hold them back on more complex tasks. The country office chose not to administer more basic Task 1 and Task 2 to Grade 4 learners in the interest of saving time on the survey, but this may have hidden that some Grade 4 learners still have not mastered the fundamentals of letter names or sounds which are holding them back on the more advanced tasks. We recommend that RTP consider whether they can deliver remedial content targeted at the bottom 10% of learners in Lebanon, as this is a way of boosting non-zero EGRA scores at midline and endline.

We recommend that RTP identify and focus on underperforming students –the 10% or so who are struggling with Arabic literacy basics.

1.2 Children who do not correctly report harmful behavior also have the lowest Life Skills scores

We found that one question, where students are asked what they would do if themselves or a friend were harmed and to which they can either answer to “tell a trusted adult” (which stands for the “correct answer”), “don’t know” (which stands for the “wrong answer”) or chose not to respond, was the best marker of low life skills scores. This question was significantly positively correlated with all four life skills measures, $p < .01$, with students answering correctly scoring 12-16% higher on the life skills items than those who did not. 64% of Grade 1 and 40% of Grade 4 students did not correctly report what to do when they or a friend were harmed in learning centers, and these children are lower performers across all Life Skills scores.

It is not clear whether this is a marker of overall childhood development or there is a causal connection, but we recommend using it as a quick proxy to identify which students are the most likely to have the lowest Life Skills scores. It can be used as part of ongoing monitoring efforts and whenever some discussions are led with students to automatically flag and target the likely lowest performers.

1.3 Life Skills levels can benefit from specific approaches for high and low-scoring students

At baseline, before the intervention has started, Grade 4 Life Skills scores are already mostly above 50%. Although the PMF has set a target of an average score, this focus masks different needs of high and low scoring students.

We found several other demographic markers for lower life skills scores within grades:

- In Grade 1, boys had significantly lower communication skills (2.4 points) than girls (3.6 points), $p < .05$, and children with an identified or self-reported disability had significantly lower conflict skills (0.5 points) than those without (1.0 points), $p < .05$.
- In terms of activities, Grade 1 children who reported doing chores in the morning or afternoon had significantly lower concentration scores (1.8 to 2.1 points) than those who did not (3.0), $p < .05$, while children who reported playing with friends in the morning had significantly lower collaboration (1.3 vs. 2.0) and concentration scores, (1.8 vs. 2.9), $p < .05$.
- In Grade 4, children with disabilities had significantly lower scores versus those without on collaboration (1.4 vs. 2.4) and communication (2.9 vs. 4.4), $p < .05$.
- In Grade 4, children who reported living in a home where people used physical aggression towards each other had higher scores versus those who did not on collaboration (2.9 vs. 2.2) and communication (5.8 vs. 4.1), $p < .05$, while children who reported someone being away for six months or more had significantly lower scores on conflict compared to those who did not (0.8 vs. 1.7), $p < .01$.
- There were no reliable socio-economic scale (SES) score predictors for life skills for either grade.

Moreover, we found that life skills scores were not always normally distributed. Some distributions had high and low scoring clusters (See Section 6.3, e.g. Grade 4 critical thinking, Grade 1 collaboration, Grade 1 Concentration).

We therefore recommend that RTP consider the idea of cooperation, where peer work and “mentorship” could be organized between higher- and lower-scoring students to support development. This approach will also promote collaboration, which is one of the life skills supported by RTP in Lebanon.

1.4 Availability of selected reading materials at home is significantly correlated with attendance of extracurricular literacy activities

Although there is some discrepancy between learners and parents in the reported presence of these materials in the household, it seems that most homes do not have magazines, newspapers, storybooks, coloring books, or comics in the household.

The baseline found evidence of significant positive correlations between some home reading materials and attendance at literacy extracurricular activities. There was some evidence that students reporting storybooks, comic books, or magazines in their home were more likely to have attended literacy extracurricular activities, and that students who reported having storybooks or coloring books had higher EGRA and life skills scores than those who did not (Section 6.4.5).

We recommend that RTP explore whether there is a causal relationship between providing child-friendly reading materials in the home and increased attendance at extracurricular literacy events, increased life skills and EGRA outcomes. If shown, providing students with child-friendly reading materials may have a significant impact on their life skills and reading scores at midline and endline.

1.5 Cultural attitudes towards girls may hinder RTP’s implementation and further gender stereotypes

The baseline survey highlighted traditional attitudes among the target population, particularly along gender lines. The survey showed a clear shift in expectations around girls’ and boys’ daily patterns of activities outside of school. Whereas in Grade 1 there were no significant differences between behaviors of girls and boys, by Grade 4 girls were around one-third as likely to report they played with friends in the morning, almost twice as likely to report doing housework in the afternoon, and never reported working outside. Validation activities confirmed that due to increasing concerns about safety, parents are less likely to let girls out of the house. With good intentions, these parental attitudes may risk retrenching behavioral differences between boys and girls, which may among other things affect outside play.

In Section 6.8.1, we found that 60-70% of parents hold traditional gendered attitudes towards boys and girls in the classroom. These attitudes may present a challenge to Right to Play’s model of inclusive play, especially outside the home.

We recommend that Right to Play reassess the potential cultural barriers to implementing its model

in a relatively conservative Arabic target population, and consider the extent to which it will need to challenge and change existing norms - with the resistance that comes with it – rather than accommodate them.

1.6 Parents' attitudes are key to inclusive classrooms, and an opportunity to seize for RTP

Parents mostly display inclusive attitudes towards gender and disability, but also show agreement with traditional directive attitudes towards learning as opposed to RTP's playful constructivist approach.. We recommend involving parents as much as possible when advocating for change, as they are ultimate decision makers when it comes to children dropout and inclusive access to education.

Indeed, research shows that reading to children 1) makes them feel safer and more secure (which ties in with the previous finding on child protection), 2) improves future reading skills, given listening is a precursor for reading, developmentally, and 3) improves interaction, which, like cognitive priming techniques (asking what will happen next, etc.) are great predictors of ability scores.

We also recommend that Right to Play focus on measuring parents' and teachers' attitudes towards directive versus constructivist teaching methods (Section 7.4) and attitudes to boys' and girls' roles both inside and outside the classroom. We believe that acceptance or resistance of the Right to Play model by both parents and teachers will be an important factor determining the uptake of the model at midline and endline.

1.7 Teachers self-reporting usage of playful methodologies should be triangulated using classroom observations

Classroom observation data were not available of inclusion in the baseline report, as due to COVID and ongoing security issues in Lebanon classes were being held mostly virtually. However, InformEd's previous evaluation of the Palestine program called the validity of teachers' self-reported measures of using playful methodologies into question. Data from the baseline teacher surveys suggest that already nearly all teachers consider multiple integrated learning criteria for reading materials, display strong knowledge and adherence to play-based learning techniques, and report all safety concerns. If so, then this undermines the case that further support is needed. However, these responses may be subject to a strong positive response bias as they were aware enumerators were acting on behalf of the RTP program. In practice RTP's playful and constructivist methods may go against a hierarchical and traditional structure within the established educational system. Teachers from the teachers' survey were relatively young (average age 29 years), female (80%), and less experienced (average 8.5 years as a teacher and 4.8 years at the center).

We recommend that Right to Play conduct in-person classroom observations of all teachers included in the teachers' survey when it is feasible to do so and match the data to see if observed playful practice matches self-reported practice.

1.8 There is evidence that around half of children with disabilities may drop out from Grade 1 to Grade 4 and that teachers may need help identifying children with disabilities in their classroom

The percentage of self-reported CWD in random samples of Grade 1 and Grade 4 classrooms suggests that around half of CWD may drop out from Grade 1 to Grade 4. This should be confirmed through other methods and RTP should then consider how the program can address reasons for dropout.

The number of children who self-reported having “a lot of difficulty” or “cannot do at all” any of the six Washington Group questions is five or six times the number of learners identified by teachers. This means that teachers may need support identifying children with disabilities in their classroom, and especially recognizing ‘invisible’ disabilities such as speech or hearing impairments, which were not identified by any teacher.

Right to Play has focused on children with disabilities and whether they are disadvantaged on learning outcomes or life skills at baseline. However, the limited number of children with disabilities in the sample means that quantitative analysis may not be the best method for this. In fact, baseline survey data seemed to suggest that children with disabilities may perform better and show more resilience than their peers, at least in Grade 1.

We recommend that Right to Play explore drivers for children with disabilities dropping out of school through qualitative and participatory methods, so that RTP can understand what the program can do to address their needs.

2. Background and context

Right to Play

Right to Play is a global organization committed to improving the lives of children and youth affected by conflict, disease, and poverty. Established in 2000, Right to Play has pioneered unique play-based approach to learning and development which focuses on quality education, life skills, health, gender equality, child protection and building peaceful communities. With programming in 16 countries, Right to Play transforms the lives of more than 2.3 million children each year, both inside and outside of the classroom. In addition to work with children, Right to Play advocates with parents, local communities, and governments to advance the fundamental rights of all children.

Right to Play is implementing the Enhancing Quality and Inclusive Education (EQIE) program in Ethiopia, Lebanon, Mozambique, Palestine, and Tanzania. The EQIE program is a five-year education initiative being implemented from 2020-2024. The program aims to:

1. Improve learning outcomes of boys and girls, specifically minimum proficiency level in reading and life skills in lower primary Grade 1-4;
2. Increase retention of out-of-school girls and children with disabilities

This will be achieved by building the capacity of education staff, teachers, and parents to give children better and more opportunities to practice reading skills, both inside and outside school, using child-centered, play-based, and active learning methodologies. Right to Play will also build the capacity of Civil Society Organization (CSO) partners to identify out-of-school children and to develop follow-up and support system, together with schools and local communities, which ensure access and sustained retention in school. Through these interventions, the program will contribute to achieving the Sustainable Development Goals (SDGs), particularly SDG 4, ensuring 'inclusive and equitable quality education and promoting lifelong learning opportunities for all', including its underlying principle of 'leave no one behind'.

Cynosure Evaluation

Cynosure Evaluation is a UK-based employee-owned company specializing in monitoring and evaluation of international development programs, with an emphasis on education programs. Cynosure has conducted evaluations of education programs for the UK Foreign, Commonwealth, and Development Office (FCDO, formerly the Department for International Development) and UNICEF. The lead evaluator was involved in the evaluation of the Girls' Education Challenge Fund.

Lebanon Context

Lebanon has been plagued by regional conflict and instability through decades of civil war, the Arab-Israeli conflict and ongoing factional violence. This has resulted in an influx of refugees, initially from

the Palestinian Territories, Iraq and more recently from Syria. Consequently, Lebanon's already limited education system is under significant strain, exemplified by high numbers of out-of-school children and poor academic performance.

Data from the Monitoring Learning Achievement (MLA) survey (2016) showed that the mean score for Grade 4 students' in Arabic was 61.2%, with girls having mean score of 68% and boys mean score being 54%, demonstrates a substantial difference in performance in favour of girls. Even though UNRWA has demonstrated great improvement in Grade 4 Arabic scores since the 2013 MLA, there is still a significant portion of students, especially marginalized boys, underperforming and not achieving the required level. This suggests the need for a focused approach to improve reading skills in UNRWA elementary schools in Lebanon, with specific emphasis on closing the gap in student achievement and by supporting children, especially boys, who are underperforming to meet required levels.

Low levels of academic achievement in Lebanon have been attributed to the low qualification of teachers (nearly 55% of teachers do not hold a university degree) and the absence of a positive learning environments for students (MEHE, 2015). The qualification level of teachers, their subject knowledge and pedagogical skills are key aspects to improving proficiency levels in reading, particularly in Grades 1-4. The impact of these skills gaps has been shown in how subjects are taught. For example, 58% of Grade 2 teachers in UNRWA schools in Lebanon rarely or never used open-ended questions (3% used this consistently), 38% rarely or never probed a pupil's answer for further explanation (2% probed consistently), 24% of teachers rarely or never encouraged pupils to ask questions (4% did this consistently), and group work was not widely encouraged (69% of teachers rarely or never used group work in class) (UNRWA, 2014). This suggests the need for UNRWA teachers in Lebanon to broaden their pedagogical approach beyond the use of teacher-led rote recitation that limits the extent to which pupils can develop their oral skills and critical thinking, especially for low performing children. Inadequate learning approaches may also partly contribute to high levels of grade repetition rates in UNRWA schools, where as many as 38% of 13-year-old students are still enrolled in elementary school (UNDP, 2018). Development of reading skills is further undermined by a lack of opportunities for children to continue reading and learning with parental support outside school and at home (Right to Play Needs Assessment, 2019).

Barriers to accessing education and staying in school are also prevalent in the target locations. Although enrolment levels are high with 97.2% for Palestinian Refugees in Lebanon (PRL) and 93.7% for Palestinian Refugees from Syria (PRS) residing inside the camps, dropout rates are 0.98%, of which 0.85% were girls and 1.1% were boys (UNRWA 2018). The majority of students who dropped out expressed low school performance, bullying, early marriage, child labour and school infrastructure as the main reasons (UNRWA, 2011). When it comes to children with disabilities, according to a Protection Brief of Palestine Refugees Living in Lebanon (UNRWA, 2017), 29% of Palestinian refugee children with disabilities are not enrolled in any educational institutions.

In response to this context, UNRWA has taken important steps towards ensuring that Palestinian refugees with disabilities have equal and meaningful access to its services in line with the Disability Policy (2010) and Disability Inclusion Guidelines (2017). In Lebanon, UNRWA is the only large service provider of assistive devices for Palestinian refugees with disabilities, due to their limited access to Lebanese public services. UNRWA has endorsed the Charter on Inclusion of Persons with Disabilities in Humanitarian Action (2016), and working to ensure that good practices are documented, shared, and used for cross-learning. Despite the achievements made to date, there are still a number of challenges to further the work on disability inclusion in UNRWA schools, such as: construction work related to accessibility; awareness and training of teachers on disability inclusion and disability rights, as well as knowledge and skills on how to deal with persons with disabilities; and identification of persons with disabilities and their needs. In addition, the unprecedented financial crisis faced by UNRWA in 2018, resulted in reduction of staff and resources to further scale the implementation of disability inclusion related activities (UNRWA, 2018).

Many of the issues affecting the quality of primary education also affect pre-primary education (Kindergarten (KG)). Most KG teachers lacked pre-primary education qualifications, with many teachers having only graduated from upper elementary or middle school. Low salaries for KG teaching means it is very hard to attract suitably qualified teachers. Enrolment rates for KG children in the Palestinian refugee camps in Lebanon targeted by the proposed project are 2,165, of which 49% are girls and 51% are boys (Right to Play Needs Assessment, 2019). RTP's needs assessment only identified one child with disability among the 2,165 enrolled, demonstrating a significant gap in access and early learning opportunities for children with disabilities within the camps. Additionally, RTP found that kindergartens lacked sufficient learning aids and materials to support children's learning and development during the pre-primary years (Right to Play Needs Assessment, 2019).

There was a flare-up of hostilities between Israel and Palestine in May 2021 which affected baseline data collection in Lebanon. Because of artillery fire directed at the south of Lebanon, the RTP team was unable to travel to observe classrooms in person as anticipated and conducted virtual classroom observations during the last two weeks of June 2021.

RTP's involvement in Lebanon

Right to Play has been serving at-risk children and youth in Lebanon since 2006. RTP's work began by providing support to Palestinian refugees, and have since expanded to also support displaced Syrians, and Lebanese children and youth from vulnerable communities. The RTP Lebanon Office work with local partners to promote positive educational outcomes, psychosocial well-being, physical health, and critical life skills. Right To Play in Lebanon also build the capacity of civil society organizations and grassroots youth groups to use different forms of play to foster and sustain social change.

In 2019, Right to Play reached 25,000 children and youth in vulnerable communities across the North, South, Beqaa, Beirut, and Mount Lebanon governorates.

EQIE program in Lebanon

Direct beneficiaries

Direct beneficiaries include educational staff and learners in Grades 1-4 in 17 learning centers targeted by the program. They also include parents of the children, SMC and PTA members of the learning centers, community members and teachers, CSO partners, and out-of-school children with disabilities.

A total 176 educational staff will be trained on play-based learning teaching practices to enhance children's reading skills, 1,750 children will be provided with learning materials, 350 parents and community members will be trained to implement supplemental reading activities for children both inside and outside learning centers, 125 SMC and PTA members will be trained to support learning center development and inclusion initiatives, 8 CSO partners will be trained and mentored on organizational development needs, and 160 out-of-school children with disabilities will be supported to enroll into target educational institutions, temporary learning spaces or alternative education programs.

Indirect beneficiaries

Indirect beneficiaries include other students outside of Grade 1 and Grade 4, and students in learning centers other than those targeted by the program, and parents and caregivers who will benefit from the *Learning Together* toolkit training.

Main intervention pathways

The overall objective of the project is to improve holistic learning outcomes of boys and girls, including children with disabilities, in primary and pre-primary institutions in the Palestinian refugee camps in Lebanon. This includes the number and percentage of children in schools achieving minimum reading proficiency levels in lower primary Grade 4 (OBJ 1000). To achieve this the project will increase teacher's knowledge and pedagogical skills that will enhance the reading skills of children through play-based learning (INT 1100; IMM 1110; OUT 1111). Teaching and learning materials will be provided to further support teachers delivery of child-centred learning (OUT 1112). The project will also increase children's participation in supplemental reading activities, using play-based learning to motivate and ignite an interest in reading, as well as ensuring that children increase the time they spend reading inside and outside school (INT 1200; IMM 1210; OUT 1213). To secure local ownership the project will also increase community and parental engagement to support home-based reading activities, and work in close collaboration with SMCs and PTAs to facilitate activities that support opportunities for children to read outside school (OUT 1211; OUT 1212).

To address the high rates of out-of-school children, the project will increase the enrolment of children with disabilities (INT 1400; IMM 1410). In close collaboration with CSO partners and supporting them

in strengthening organisational development needs (OUT 1411), the project will implement initiatives to address barriers to education for children with disabilities (OUT 1413). This will include community sensitization events and training teachers on inclusive education practices. The project will also connect CSO partners with UNRWA Education Department officers, SMCs and PTAs to form School Inclusion Committees (OUT 1313), support construction and rehabilitation work (OUT 1311) and conduct sensitization sessions on child protection, child rights, corporal punishment, abuse, and/or GBV (OUT 1312). This will act as a means to increase the capacity of school communities to create and maintain safe and inclusive learning environments (INT 1300; IMM 1310). Kindergartens will also be supported to deliver quality pre-primary education and increase pre-primary enrolment of children with disabilities (INT 1400).

Changes from original proposal

The EQIE program in its original design was supposed to start in January 2020 but was postponed for several months to allow for some urgently needed hygiene activities due to the COVID pandemic. The overall goal of the project is to enhance quality and inclusive education as it relates to SDG 4, the underlying principles within the SDGs of “leave no one behind”, the UNRWA Medium Term Strategy (2016-2021), and the priorities of CSO partners.

Due to several factors, implementation inside UNRWA schools was delayed, and therefore Learning Centers inside the Palestinian refugee camps in Lebanon were considered as an alternative implementation setting for this project, keeping on the same outcomes as these are the same children that attend UNRWA schools.

Data collection for the sake of this baseline study, was also conducted through Learning Centers and not through UNRWA schools, which provided a limitation to the sample size, and modality of teaching when some classroom sessions were observed.

3. Approach

3.1 Evaluation purpose

The baseline evaluation aimed to:

- Assess the program’s starting point within each of the programming contexts
- Inform program design and strategy
- Establish baseline measurements against which improvement and impact will be measured during midline and endline evaluations.

3.2 Evaluation questions

Aligned to the evaluation purpose and scope, this evaluation seeks to answer agreed-upon evaluation questions and sub-questions. A full list of these questions can be found in Annex 6.

Furthermore, the overall evaluation purpose articulated has been considered in relation to the OECD DAC criteria¹. The DAC criteria enable the determination of merit, worth or significance of an intervention. Each criterion is a different lens or perspective through which the intervention can be viewed. Together they provide a more comprehensive picture of the intervention, the process of implementation, and the results.

This baseline equips Right to Play to design the program with relevance and coherence for the most vulnerable populations.

As described in the evaluation design and sampling framework, the evaluation questions establish the baseline values of key indicators from which effectiveness and impact of the program can be measured at baseline and endline. There is also an opportunity to examine sustainability through the longitudinal study design, increased in capacity of teachers and community leaders.

3.3 Marginalization and equity indicators

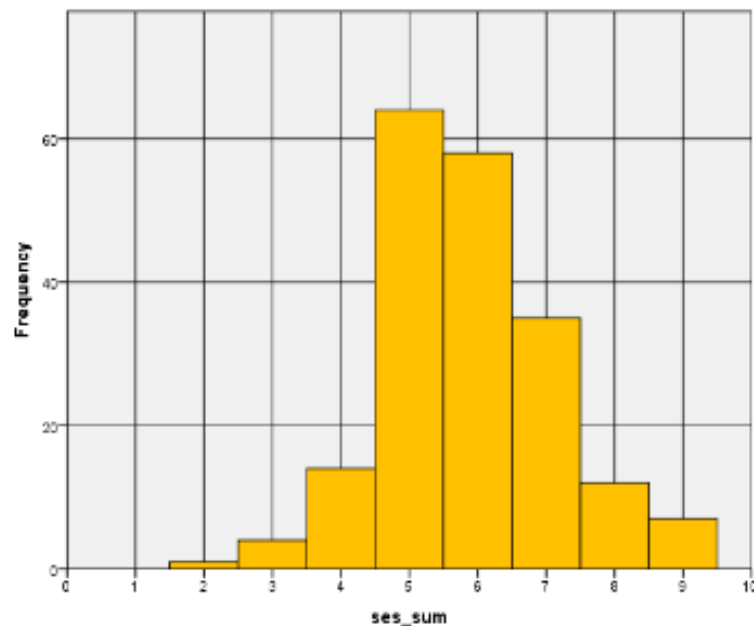
The overall EQIE program examined equity indicators of sex, SES, ECD attendance, home literacy environment, language of instruction, ethnicity, disability, home language, child time patterns, school setting, as appropriate to the contexts of the individual countries.

In Lebanon, respondent households seemed to be homogenous, so analysis by marginalization was mostly done by sex and disability status.

Almost 90% of Grade 4 respondents reported attending early childhood development classes and all but one respondent reported that Arabic was the main language used at home. Moreover, all learning centers were located in urban or semi-urban areas and all learning centers used Arabic as the language of instruction. These questions do not provide a sufficient range of cases for meaningful analysis and so were not used. Questions about language of instruction and ethnicity were not asked in Lebanon, presumably because there was thought to be very little variation amongst respondents.

¹ OECD DAC Criteria (2020): <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

Figure 1: Distribution of socio-economic status (SES) scores measuring presence of 10 household assets



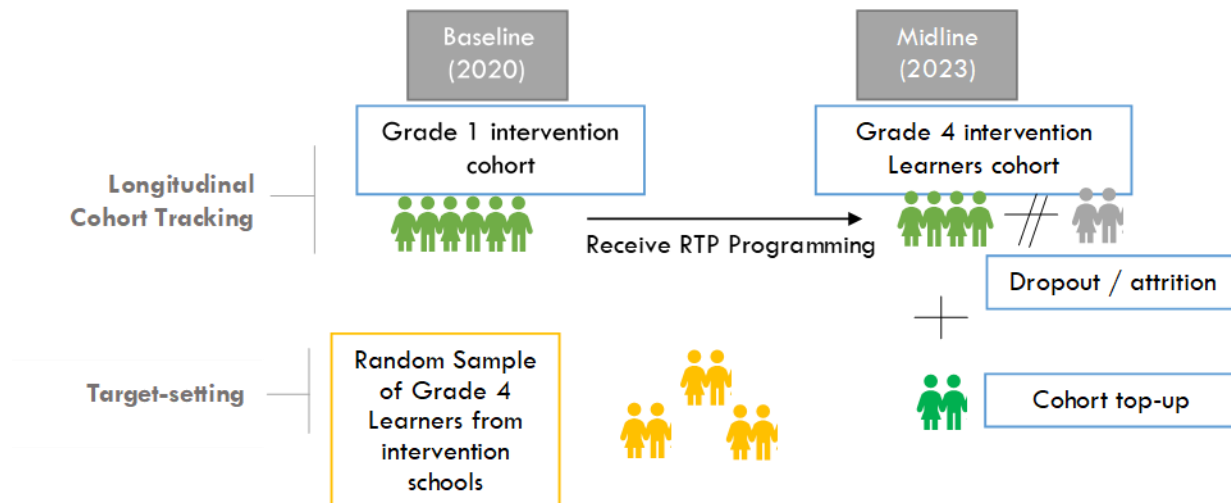
Socio-economic status (SES) scores are often used as a proxy of household poverty. However, there is no Progress out of Poverty Index (PPI) report identifying household factors most correlated with poverty in Lebanon. In fact, there was very little variation in socio-economic status (SES) scores; these were normally distributed within a narrow range, with 75% of households having an SES score of between 5 and 7, meaning that learners mentioned their households owned 5-7 of 10 possible assets. Even taking the extremes (first quartile compared to fourth quartile of SES score), we found no statistically significant differences by SES quartile for outcome or attitudinal variables of interest.

3.4 Overall evaluation design

The overall program evaluation employs a quasi-experimental design, allowing Right to Play to measure effectiveness and impact of their intervention at midline and endline across five countries. The overall program evaluation design seeks to establish a baseline sample of Grade 1 students in each country to be followed longitudinally throughout the life of the program. The program will attempt to follow and reassess the same students in Grade 4. The overall program will draw samples from both intervention and comparison schools, which will allow for a difference in difference analysis, comparing learner gains in intervention schools to learner gains in comparison schools from baseline to endline. In addition, the program seeks to establish the baseline performance of learners in Grade 4, which will be used to set the end-of-program target.

In Lebanon, the program was not able to use comparison groups at baseline due to limited comparability and difficulty obtaining permission to access UNRWA refugee camps. Therefore, the

Lebanon program evaluation will examine only change at baseline versus midline and endline in the target learning centers.



3.5 Ethical considerations

The evaluation team followed the UN Evaluation Group's (2008) 'Ethical Guidelines for Evaluation' and UK FCDO's (formerly DFID's 2011) 'Ethics Principles for Research and Evaluation' as the fundamental ethical guidelines for the evaluation. The team used the following as guidelines:

- Evaluation standards: intentionality; managing conflict of interest; appropriate interaction with participants; accuracy, transparency & reliability; critical friendship response mechanisms.
- Practical measures: informed consent; researcher screening, child and young people friendliness and awareness of protocols related to children and young people and vulnerable people; adequate reporting mechanisms; procedures for responding to participant harm; application of appropriate evaluation toolkits.
- Procedural principles: reference to existing data; linkage to program objectives; implementation of specific evaluation methodology; respect for principles of participation and for protection of rights, welfare, and confidentiality.

Right to Play staff contacted the Learning Centers and got their permissions to have access and interview their students, parents, teachers, principals and observe their classrooms.

Consent forms were sent to parents of students prior to data collection, and children who arrived to the data collection site were further asked if they want to take part of the study. Children, parents, teachers, and principals were informed that their participation in the study was voluntary and that

their refusal – if any – would not affect their relationship with Right to Play or their participation in the program.

3.6 Training

Positions for hiring enumerators and the Data Collection Coordinator were opened for 2 weeks prior selection on LinkedIn and Daleel Madani (a job-search platform widely used in Lebanon).

10 enumerators were chosen based on their CVs and previous experience in data collection. All enumerators had academic degrees in education or psychology, or previous experience working with children in schools or in community centers. For the Data Collection Coordinator position, 3 shortlisted applicants with sufficient experience in leading data collection and supervising enumerators, were interviewed, by 2 Right to Play staff members, until the selected candidate.

A four-day training was conducted by the Data Collection Coordinator for the 10 enumerators, with direct supervision from the Monitoring Evaluation and Learning Officer, on the principles of data collection and the tools to be used. Exact words and translation to be used during data collection were unified during the training. An additional training session on Child Safeguarding policies and Referrals was given by the Monitoring Evaluation and Learning Officer.

One pilot data collection day was conducted, and then evaluated and day to day follow up was ensured by the Data Collection Coordinator.

3.7 Dissemination plan

Baseline findings, lessons and recommendations are disseminated in the following ways with varied stakeholders.

- Validation workshop with project partners and participants in Lebanon via PowerPoint presentation, and gallery walk of data placemats.
- Presentation meetings, narrative report, and project implementation planning (PIP) workshop held with internal RTP project implementation team to inform delivery and adaptation for better performance.
- Final report shared with the learning centers involved in the project and with UNRWA officials that function inside Palestinian refugee camps chosen for the activities for system wide learning and action.
- Final narrative report shared with NORAD (project donor) for accountability and contribution to the wider international education development portfolio learning.

3.8 Challenges and limitations

3.8.1 Implementation and data collection sites

Due to several factors, implementation inside UNRWA schools was delayed, and therefore Learning Centers inside the Palestinian refugee camps in Lebanon were considered as an alternative implementation setting for this project, keeping on the same outcomes as these are the same children that attend UNRWA schools.

Data collection for the sake of this baseline study, was also conducted through Learning Centers and not through UNRWA schools, which provided a main limitation to the sample size, and modality of teaching when classroom sessions were observed.

3.8.2 COVID-19

The Covid pandemic did not seem to result in much actual loss of learning time but did force teachers to re-focus on re-teaching content. 9 out of 10 teachers said that the Coronavirus pandemic has resulted in 2 months or less of lost learning time due to delays. 7 out of 10 teachers reported that remote instruction during the Coronavirus pandemic has focused all or mostly about reviewing material taught before the pandemic as opposed to new material.

Most of the learning centers were still conducting online teaching, by the time data was collected. The modality of online teaching was not live interactive sessions, rather through WhatsApp group chats that limited any observation of interaction between students and the teachers. This modality was used due to the bad Internet connectivity inside Palestinian camp as explained by learning centers.

3.8.3 Conflict

Baseline fieldwork was affected by conflict between Palestine and Israel in May 2021, and enumerators were unable to visit learning centers in South area due to artillery shelling. Instead, classroom sessions in centers who were doing in-person teaching, were again shifted online and observations were conducted virtually in June 2021.

3.8.4 SurveyCTO and JavaScript EGRA tool

Right to Play used SurveyCTO to manage survey materials and incorporated an open-source JavaScript tool to facilitate collection of data for selected EGRA tasks. This was the first time the tool was used, and both enumerators and evaluators experienced problems with using and modifying the tool. For Lebanon, the tool was not suitable for use with Arabic because it assumed passages were read left-to-right. A work-around was implemented but this resulted in the first 65 cases being incorrectly recorded. The evaluators note that Right to Play had difficulties with another open-

source EGRA tool, the KoBo literacy widget, in the first round of country baselines. We recommend that Right to Play choose and develop the tool of their choice ahead of the midline with support from a programmer who can adapt the tool for full use with right-to-left languages.

3.8.5 Selective administration of EGRA tasks

Country offices made the decision to administer only certain EGRA tasks to Grade 1 and Grade 4 under the belief that some tasks, such as reading comprehension, were inappropriate for Grade 1 while other tasks, such as letter recognition, were too basic for Grade 4. This meant it was not possible to establish the ‘normal’ course of improvement from Grade 1 to Grade 4 on these tasks and set baselines which the program could use to set targets.

In future, we recommend administering all EGRA tasks to both Grade 1 and Grade 4, adjusting the difficulty of the more advanced tasks as necessary. We realize that adding tasks may lengthen the EGRA administration, particularly for Grade 4 learners, and administering more advanced tasks to younger learners may cause some discomfort. However, tasks which are too difficult can be quickly discontinued and administering the full range of tasks will allow the program to capture the full range of development in reading.

3.8.6 Imperfect recall and self-reported data

Many of the tools rely on the honest and accurate memory of the respondent. Although response periods were limited (i.e. ‘in the last week...’), self-reported data is still subject to recall bias and positive response bias, including self-promotion. The InformEd baseline for Palestine noted that in at least one case – self-reported teacher practices – the data largely did not seem to be of value, and slides from the first baseline countries suggested that perceptions from four groups of respondents on physical and environmental safety did not triangulate well.

Because classroom content was delivered virtually, mostly via WhatsApp lessons, during the time of the baseline, we were unable to observe classrooms for the Lebanon project and so were unable to triangulate the use of play-based learning practices in the classroom against teachers’ self-reported usage. We recognize the risks associated with taking teachers at their word and recommend that Right to Play prioritize this when in-person teaching resumes.

3.8.7 Reliability of child respondents

Much of the data used in analysis are self-reported by children. In Grade 1, many children are 6 years old and some as young as 5. While we recognize that children are the best sources of information for outcome variables including EGRA and life skills, Washington Group disability question and safety questions, we question their validity for other questions. In Lebanon, responses on household asset ownership (such as smartphones and televisions), children’s reading materials (and magazines and newspapers), and literary actions were significantly higher than their parents’ responses and do not triangulate. This indicates that children may not be the most reliable source of data for questions about household assets, children’s reading materials, household risk factors, or children’s after-hours

activities. Children may be particularly susceptible to positive response bias for these questions, especially when at the learning centers and being asked questions by an adult, and it is not clear that this can be mitigated.

We recommend that Right to Play determine whether parents or children more reliably report household assets and reading materials and use only a single source of information for these indicators. We further recommend that at midline Right to Play add questions about children's activities, household risk factors, and Washington Group questions to the parents' survey to compare the results and consider whether children or parents are a better source for this information.

3.8.8 Quasi-experimental nature of research design

Because it was not possible to randomly assign students to control and intervention groups, the program pursued a quasi-experimental approach. In Lebanon, due to the limited number of learning centers and differences with UNRWA schools, no comparison schools were identified. The evaluation of the Lebanon program therefore cannot make use of a difference-in-differences approach which would account for changes in policy and context during the course of the evaluation. Major changes in context and policy should be identified by the country office team and changes in EGRA and life skills scores at midline and endline should be interpreted in the context of these changes. If available, national level examinations or other year-on-year data should be examined for any broader trends.

3.8.9 Limited sample size and representativeness

The Lebanon country office chose to target 100 Learner Assessments in each of Grade 1 and Grade 4 at baseline across around nine learning centers, due to present difficulties operating in Lebanon. The target sample is insufficient to achieve the desired minimum detectable effect (MDE) of 0.25 standard deviations, does not account for clustering effect, and left no room for attrition of the cohort from Grade 1 to Grade 4.

We identified an issue around inability to show a statistically significant increase to the PMF indicator on reading with comprehension, and laid out a series of options for addressing these issues, including topping up the baseline sample, especially for the EGRA task. We recommended an absolute minimum of 200 students for a cohort-based approach (independent samples t-test) or 70 students to be followed to midline and endline (paired samples t-test). Against similar numbers at midline and endline, this will achieve a minimum detectable effect of 0.25 standard deviations with an alpha of 90% (below what is conventionally used, but outlined in the EQIE inception report for other RTP programs), beta/power of 0.8, two-tailed hypothesis. It does not account for the design effect of clustering by learning centers.

RTP responded to these concerns and elected for an individual cohort approach. The baseline field team went back in the field and gathered 55 valid EGRA Grade 1 reading exams, out of a target of 66 students. This sample does not meet the minimum recommended for EGRA and already indicates problems following up with students at midline and endline, with 16% of students uncontactable several weeks after the baseline.

Although a random sampling approach is appropriate for large population sizes, its usefulness is limited with populations of 300 or less. The data for teachers and directors presented in the report are based on samples which are not necessarily representative of the overall target populations. We recommended distributing baseline surveys to all 200 teachers and 17 learning center heads and continue to recommend doing so at midline. This will allow evaluators to draw more valid conclusions than from a random sampling of 10 teachers and 3 directors.

4. Evaluation Methodology

4.1 Tools

Learner assessment

The learner assessment contains questions on student background, household literacy activities and relationships, community literacy, and safety of the learning environment. It also contains an assessment of learning outcomes (the Early Grade Reading Assessment) and life skills (the International Social and Emotional Learning Assessment). The entire assessment took on average 40 minutes to complete.

4.1.1 Early Grade Reading Assessment (EGRA)

The Early Grade Reading Assessment (EGRA) measures students' progress towards reading. EGRA gauges early literacy skills through a 15-minute individual oral assessment composed of five tasks measuring fundamental reading skills. Given the increasing difficulty of the tasks, the Country Office decided to administer certain tasks to certain grades. The tasks are as follows.

Figure 2: Early Grade Reading Assessment tasks administered in the baseline

Task #	Task Name	Task Description	Grade 1	Grade 4
Task 1	Letter Identification	Learners give names of 50 Arabic letters	√	
Task 2	Phonemic Awareness	Learners say initial sound of 10 common words from Grade 1 curriculum	√	
Task 3	Familiar Words	Learners read 50 common words from Grade 1 or Grade 4 curriculum, with diacritical marks	√	√
Task 4a	Reading Passage	Grade 1 learners read a 31-word passage and Grade 4 learners	√	√

		read a 50-word passage without diacritical marks		
Task 4b	Reading Comprehension	Learners answer five oral questions, both factual and inferential, about the passage they just read	√	√
Task 5	Listening Comprehension	Learners answer five oral questions, both factual and inferential, about a passage the enumerator reads aloud to them		√

Given students will be starting at learning centers at the time of the baseline assessment and considering the long period of center closure from COVID-19, EGRA assessments will be equated to the grade prior to that being assessed.

RTI International, the developer of EGRA, cautions that the tool should be used as a diagnostic rather than a means of assessment, however EGRA has become commonly used as the latter. Moreover, they caution not to make comparisons across countries, as the timeline of language acquisition varies according to context and the language itself. EGRA provides insights into country-specific strengths and weaknesses of children in terms of reading skills and will be used to assess whether the program can be seen to have strengthened these skills at midline and at endline.

4.1.2 International Social and Emotional Learning Assessment (ISELA)²

The International Social and Emotional Learning Assessment (ISELA) helps program staff understand the development of Social and Emotional Learning (SEL) skills in children between the ages of 6 and 12 years. ISELA maps out skills across five different Social and Emotional Learning (SEL) competency areas put forth by the Collaborative for Academic, Social, and Emotional Learning. Annex 4 shows the full mapping of SEL skills.

Country Offices chose the ISELA indicators most adapted to their context. In Lebanon, the country office chose the '4 Cs' used in Palestine. These are:

Figure 3: ISELA sub-indicators making up the 4 Cs used in Lebanon

Measure	Variable	Question
Concentration	Shape1	Child completed 1st drawing OR still drawing after 60 seconds
	Shape2	Child completed 2nd drawing OR still drawing after 60 seconds
	Shape3	Child completed 3rd drawing OR still drawing after 60 seconds

² Save the Children, 2019. *International Social & Emotional Learning Assessment (ISELA)*.

	Shape4	Child completed 4th drawing OR still drawing after 60 seconds
Communication	Relate2 Relate3 Relate4 Relate5 Relate10 Relate11 Relate15 Relate16	Do you talk to anyone at home when you are sad? Do you talk to anyone else at home when you are sad? Do you ask anyone at home for help when you have a problem with a friend? Do you ask anyone else at home for help when you have a problem with a friend? Do you talk to any of your friends when you are sad? When you have a problem with another child, do you ask any of your friends for help? Do you talk to any adult in your school when you are sad? Do you ask any adult in your school for help when you have a problem with a friend?
Collaboration	Collaboration1 Collaboration2 Collaboration3 Collaboration4	What are some good ways to show teamwork? In the last week, have you played a game with friends? In the last week, have you worked with friends on coursework/homework? In the last week, have you worked with a friend to accomplish a common goal?
Critical thinking	 Conflict1 Conflict2 Conflict3 Conflict4	**I want you to imagine a situation where you are playing with a toy that you like. Another child wants to play with that same toy, but there is only one toy. ** What would you do in this situation? Is there anything else you would do? **Now, I want you to imagine the same situation: you are playing with a toy that you like, another child wants to play with the same toy, but there is only one toy. This time the other child takes the toy away from you and starts playing with it without asking you.** What would you do in this situation? Is there anything else you would do?

We identified an issue about how evaluators InformEd had calculated percentages for life skills indicators for other RTP country baselines. Life skills questions tend to be a series of questions which ask for more examples of a concept. Under the previous calculation, only those questions which were answered were included in the calculation and those which were discontinued were excluded. We thought this distorted the desired life skills measures.

As an example, the concentration/perseverance indicator is constructed from four *shape* questions. A child is asked to draw a shape for 60 seconds. If s/he finishes drawing the shape or is

still drawing when the time expires, s/he is given another, harder shape to draw, and so on. If s/he gives up, the exercise is discontinued. Because the ‘% correct’ indicator is divided by the number of non-missing values, a child who draws one shape and refuses to draw another will score 100%, whereas a child who draws two shapes and gives up on the third will score 67%.

Instead we have calculated and reported life skills scores as a percentage of the maximum possible points of the cluster or index. In the example above, a child would get a quarter of a point up to a maximum of one point for each shape s/he draws or is still drawing at the end of the exercise. Using this method, the child who draws one shape and refuses to draw another would score 25%, whereas the child who draws two shapes and gives up on the third would score 50%. This approach rewards children for drawing more shapes, giving more examples, etc. rather than ‘rewarding’ them for giving up and penalizing them for not getting all items attempted correct. SEL scores reported for Mozambique and Lebanon are therefore about 10 percentage points lower than those reported for the other EQIE baseline countries Ethiopia, Tanzania, and Palestine.

4.1.3 Washington Group Short Set³

The Washington Group 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). The questions are designed to be part of a larger survey.

Previous evaluators InformEd modified the Washington Group Short Set somewhat for inclusion in the learner assessment to be more child friendly. For example, an original Washington Group question was: *Do you have difficulty seeing, even if wearing glasses?* The question was slightly reworded to be: *Do you have difficulty seeing words in your books or on the blackboard, even if wearing glasses?*

We used the InformEd modified questions for continuity with other EQIE country baselines.

A comparison of the two sets of questions is shown in the table below.

Figure 4: Comparison of original and modified Washington Group questions

Questions endorsed by the Washington Group	Questions used on the survey
Do you have difficulty seeing, even if wearing glasses?	Do you have difficulty seeing words in your books or on the blackboard, even if wearing glasses?
Do you have difficulty hearing, even if using a hearing aid?	Do you have difficulty hearing peoples' voices or music, even if using a hearing aid?
Do you have difficulty walking or climbing steps?	Do you have difficulty walking or climbing steps?

³ <http://www.washingtongroup-disability.com/washington-group-question-sets/short-set-of-disability-questions/>

Do you have difficulty remembering or concentrating?	Do you have difficulty remembering or concentrating on fun activities/activities you enjoy?
Do you have difficulty (with self-care such as) washing all over or dressing?	Do you have difficulty feeding or dressing yourself?
Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?	In (insert child's home language), do you have difficulty communicating with your family or friends?

4.1.4 Teacher classroom interviews

Teachers were interviewed using Right to Play's Teacher Survey Tool.⁴ The survey includes questions on teacher background, training and support, learning environment, and current use of play-based learning techniques. The same Grade 1 and grade 4 teachers who teach the section of students for Learning Assessment were interviewed.

4.1.5 Director interviews

The Director of each learning center was interviewed using Right to Play's Principal Survey tool. The tool includes questions on principal background, play activities within the learning center, teaching methods and discipline, gender dynamics within the center and center safety. The interview collects information on pupil-teacher ratio, pupil-textbook/reading book ratio, supplemental reading initiatives in the centers, proportion of trained/qualified teachers, and proportion of teachers who have participated in RTP trainings previously. Because learning centers were grouped under directors, only three directors were interviewed, and in hindsight the supervisors in each learning centers may have been better placed to answer questions on this tool.

4.1.6 Parent phone survey

Parents of a subset of learners in Grade 1 and 4 were interviewed regarding student learning achievement and home literacy environment.

Phone interviews allowed contact with parents who were working or away from the home during the day/unable to come to the center to respond. Carrying out phone surveys also protects Right to Play staff, community members, and enumerators from in-person contact which will ensure everyone is following COVID-19 public health protocols. Phone surveys require that the centers have phone numbers of parents and are able/willing to share with the evaluation team and Right to Play. Any phone numbers used for the survey will be kept private and only used by project team members. They will be stored separate from names, to ensure privacy.

4.1.7 Lesson observations

As shown in the analysis framework in Annex 3 several indicators require a lesson observation. Lesson observations require significant training and expertise to be done well. Building upon this,

⁴ This tool will be adjusted and adapted as needed to fit the context.

each country has different requirements/restraints on who can carry out classroom observations in the country.

In the case of Lebanon, classrooms observations were to be done through learning centers and not UNRWA schools. These centers provide after-hours learning for students that attend UNRWA schools. By the time data was collected, most learning centers were still doing online learning, additionally the unrest inside Palestinian refugee camps during the last week of data collection prevented any in-person visits to the learning centers who did in-person sessions. This method prevented live interaction, which was a main limitation to filling classroom observation data. According to learning centers, the main reason for not conducting live video call sessions was the bad Internet connection services inside the camps.

The Country Office team agreed that classroom observations would be conducted when classes were again in-person, but that this would fall outside the scope of the baseline.

4.2 Sampling Framework

The following sampling framework was achieved in Lebanon.

Figure 5: Sampling Framework

Tool	Target	Achieved (Valid)	Total population	Notes
Grade 1 Learner Assessment	100	90	6,000 total students across all grades	10 students in each of 10 learning centers.
Grade 4 Learner Assessment	100	107		10 students in each of 10 learning centers.
Lesson Observation	TBD	0	17 learning centers	Postponed due to conflict
Teacher Interview	10	10	100 teachers	One interview per learning center.
Principal Interview	10	3	17 principals	Three directors of the learning centers were interviewed. Each director was responsible for about three learning centers.
Parent Interview	100	132	Est. 12,000 parents	Parents of half the children interviewed.

4.3 Representativeness of the sample

Students were randomly selected from targeted grades without quotas for girls or children with disabilities. For Lebanon, given logistical issues going into refugee camps, the target sample size was 100 Grade 1 and 100 Grade 4 Learner Assessments. This sample size was chosen to achieve a minimum detectable effect of 0.25 standard deviations between groups by following the same cohort to Grade 4 (power of 0.8, alpha of .05, one-tailed)

4.3.1 Representativeness of EGRA tasks administered with JavaScript plugin

Due to an unforeseen issue with the JavaScript plugin, the first 65 cases for EGRA did not record valid metadata for Task 1, Task 2, or Task 4a. The metadata included numbers of questions attempted, questions correct, and time remaining on task, and these were needed to calculate a valid task score. There are therefore only 63 valid cases for the EGRA Reading Task 4a used to calculate the PMF.

After discussing the options, RTP has decided to go back to schools and administer a grade-level reading Task 4 to Grade 1 students. This will enable RTP to use a cohort-based sampling approach at midline and endline, following the same Grade 1 students and re-administering the reading task appropriate to their current grade. As outlined in the methodology section above, the minimum required sample is 60 students tracked from baseline to endline. Gathering data from all 90 Grade 1 students at baseline will allow for 50% attrition of the sample, or else adjusting for design effect.

4.4 Cohort tracking

Student IDs were recorded for students in Lebanon, along with identifying information including name, village, and teacher's name. 84% of parent interviews were linked to Learner Assessments; however, it was unclear how Learner Assessments could be linked to teachers through classrooms, as multiple teachers per learning center were interviewed.

4.5 Data quality and Security Assurance

Data validation was done initially through SurveyCTO, then cleaned and reviewed using SPSS. 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 95% 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 clustered by learning center and are likely to share some variation for being in the same learning environment.

5. Key characteristic subgroups and barriers at baseline

5.1 Respondent demographics

5.1.1 Students

Figure 6: Student demographics

Variable	Learning Centres (Intervention)				Sig
	Grade 1	N	Grade 4	N	
% female	67%	90	59%	107	*
Age (years)	6.8	85	9.9	105	*
Attended ECD or preschool	-	-	87%	106	N/A
# of household members (including student)	5.7	87	5.7	106	NS
Child identified or self-identified as having a disability	22%	89	9%	106	**

N/A = Not applicable (question not asked of both groups), NS = Not significant, * $p < .05$, ** $p < .01$, *** $p < .001$

5.1.2 Teachers

Figure 7: Teacher demographics

Variable	Average value	N
% Female	80%	10
Age	29 years	10
Years of education	16 years	10
Permanent/civil service teacher	20%	10

Years as a teacher	8.5 years	10
Years as a teacher at this school	4.8 years	10
# of students in class of assessed students	22.1	10
# of CWD in all classes taught	2.8	10
# of lessons taught on average day	4.6	10
# of minutes spent teaching language/literacy on an average day	120	10
Has been trained by other organization on PBL or literacy instruction	70%	10
Facilitates reading club	0%	10
Facilitates resource rooms for CWD and academically weak children	90%	10
Average hours per week spent facilitating extracurricular activities for students	8.1	10
Agree or strongly agree that working as a teacher has been rewarding and worthwhile.	100%	10
Agree or strongly agree that there is a strong community of practice among fellow teachers.	100%	10
Has considered leaving education as a career in past year	30%	10
Has access to a trained point person for psychosocial support	100%	10
Has ever received training on providing psychosocial support to students	100%	10

5.1.3 Directors

At baseline, the three Directors of the learning centers were interviewed using the Principal survey.

Figure 7: Summary statistics of RTP Lebanon Learning Center Directors

Variable	Value
% Female	33%
Average age	44 years
Average years of education	16 years
Average years in the position	8.3 years
Years as a Director at this center	6.7 years
Principal has received training from RTP	67%
Has considered leaving education as a career in past year	33%

5.1.4 Parents

Figure 8: Sampled caregiver profile in intervention areas

Variable	Average	N
% female	73%	134
Age	37.0	120
Member of SMC	2%	128
Member of PTA	4%	128
Years of education	6.5 years	126

Three-quarters of parents could be matched to students using student IDs. Three-quarters of respondents to the parent survey were female, with an average age of 37 years. Membership in SMC and PTA was low, at 2-4%. Two-thirds of parents had completed primary school or lower.

5.2 Demographics by marginalized group

5.2.1 Girls: their lives, social context, and demographics

Girls represent 67% of Grade 1 students and 59% of Grade 4 students across learning centers sampled.*** 85% of the overall sample (n=62) attended pre-school, and 2% reported repeating any grade; these numbers are not significantly different from boys.

5.2.2 Girls' daily patterns of activities

At Grade 1, where the overall average age is 6.8 years, there are no significant differences (or borderline significant differences – see footnote below) between boys and girls across any of the SES indicators, morning, after school, or evening activities.

At Grade 4, where the overall average age is 9.9 years, some borderline significant differences between genders appear.

- In the morning, girls report being around one-third as likely[†] to play with friends (5% vs. 15% for boys), less than half as likely[†] to watch TV (11% vs. 25% for boys), and around one-fifth as likely[†] to work outside (2% vs. 9% for boys).
- After school, girls report being almost twice as likely^{*} to do housework (60% vs. 36% for boys), and no girl reported working outside^{**} (0% vs. 11% for boys).

[†] $p < .10$. In fact, most of the values of p reported as borderline significant are $p < .07$ or $p < .08$. While this is not a conventional level of reporting statistical significance, the small size of the sample and proximity to conventional levels merits consideration.

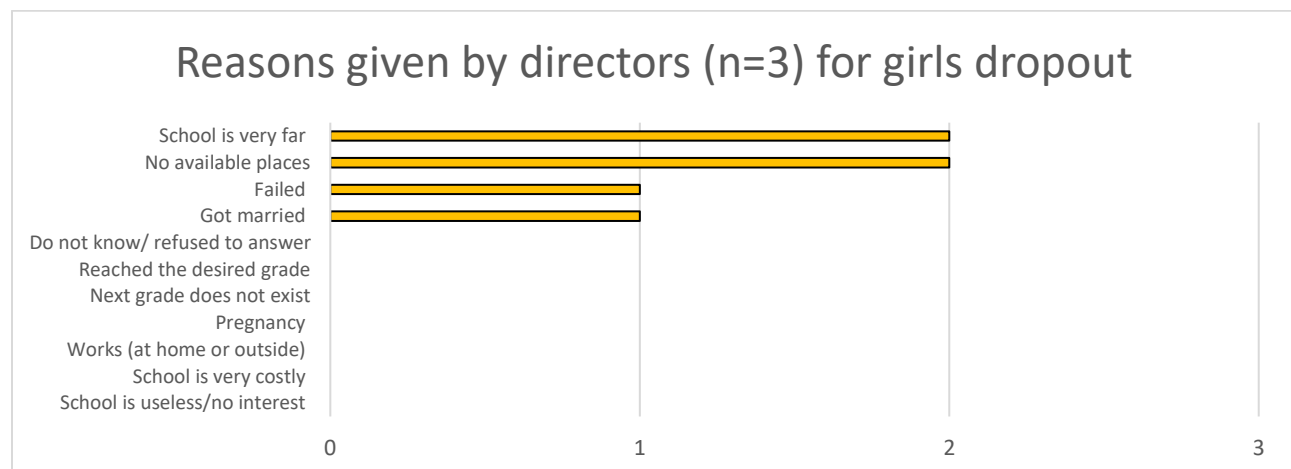
- In the evening, girls report being around one-third as likely to play with friends[†] (6% vs. 18% for boys).

This suggests that as girls grow older, their ability to play with peers and go outside are hindered by gender norms. Nearly two-thirds of girls responded that they were expected to do housework after school, which may cut into their time for homework, socializing, play, and other activities.

5.2.3 Girls' dropout rates

Two of the three Directors mentioned girls dropping out from the learning centers over the past year. On average, 17 girls dropped out per cluster of centers. Both directors mentioned distance to the center and lack of places as reasons for girls dropping out, while one mentioned marriage and one mentioned failing.

Figure 9: Reasons given by directors for girls' dropout



By comparison, two of three Directors mentioned boys dropping out from the learning centers over the past year. On average, 20 boys dropped out per cluster of centers. Both directors mentioned work at home or outside as a reason for boys dropping out, while one mentioned distance to the center, and one mentioned lack of places.

While these differences are slight and not significant in the statistical sense, they suggest that distance, lack of places, and marriage may be more of a factor for girls dropping out of the learning centers than for boys.

5.2.4 Children with disabilities: demographics, access, and presence in learning centers

Teachers identified three children as having a mild to moderate disability that should be accommodated for the learner assessment. Using the Washington Group questions, 24 learners self-identified as having 'a lot of difficulty' or 'cannot do at all' one or more of the areas. This suggests that teachers are not currently able to identify up to 90% of children with disabilities in their classrooms.

Figure 10: Disabilities identified by teacher preventing students from receiving normal learning assessments

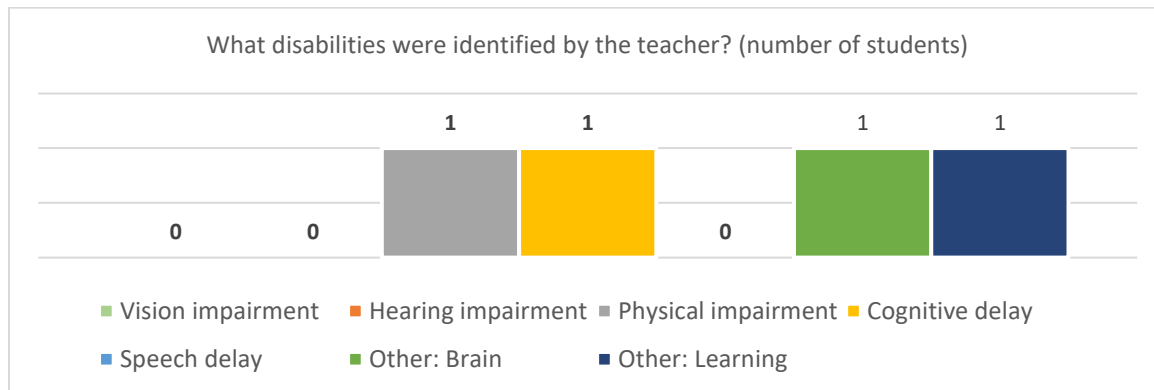
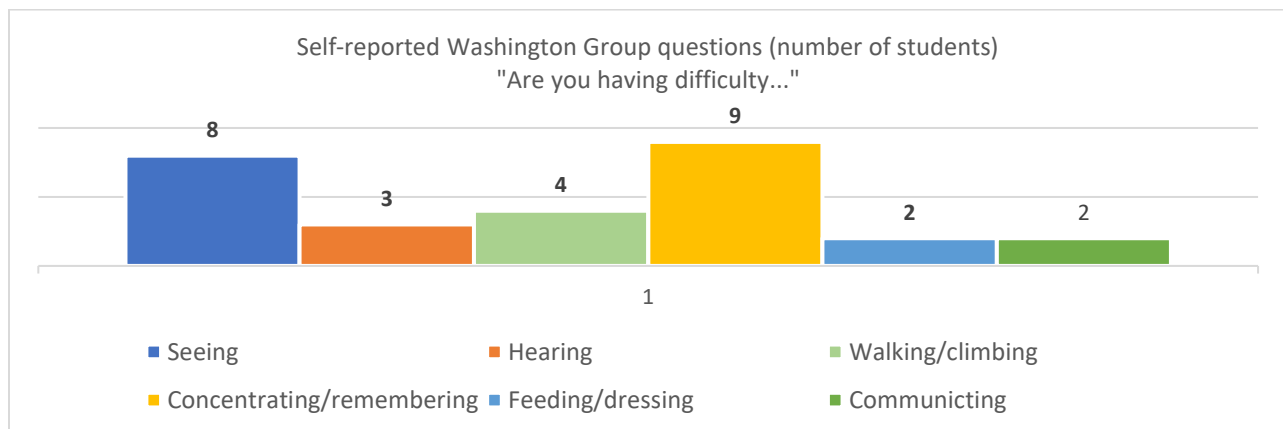


Figure 11: Disabilities self-reported by students during the learning assessment

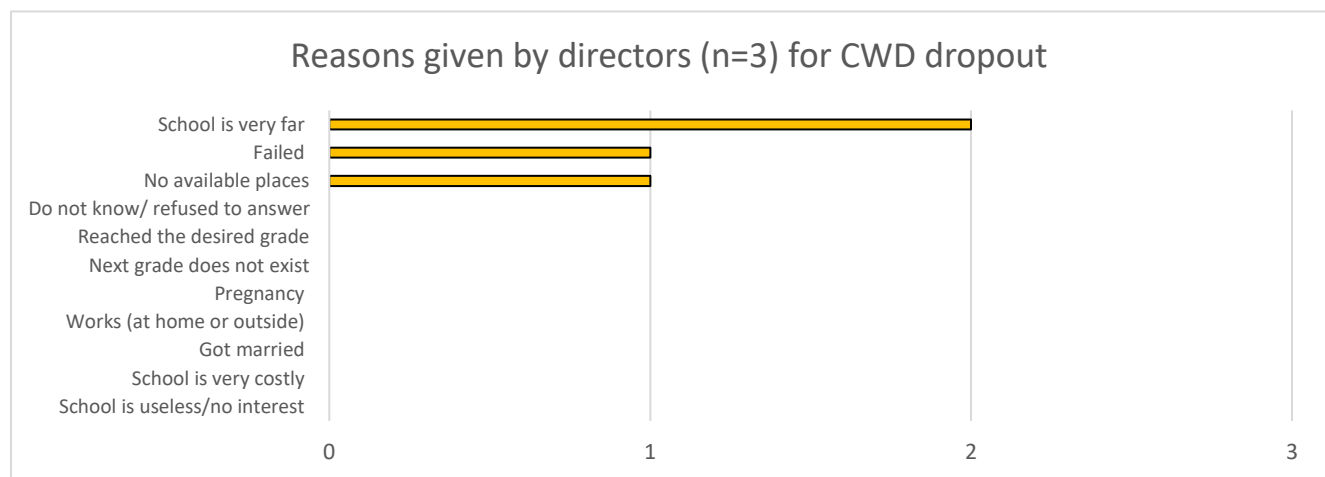


The number of learners who report having 'a lot of difficulty' or 'cannot do at all' is 5-6 times the number of learners identified by teachers. Note that numbers exceed totals because multiple disabilities could be checked.

Teachers did not successfully identify any of the children who self-identified as having a visual or hearing impairment. They were able to identify around one in four children with a mental disability.

Children with disabilities in our intervention district do not live in family and socio-economic contexts any different from the other children of the sample; they attend pre-school in similar proportions, do not repeat grades more than other children. Risk indicators do not differ for them, and their daily patterns are also similar.

Figure 12: Reasons given by directors for CWD dropout



5.3 Barriers to participation

5.3.1 Barriers affecting girls' attitude and performance in learning centers

Principals and teachers were asked to expand on barriers affecting boys' and girls' performance. Two of three directors and 30% of teachers said that girls performed better than boys, with the remainder saying they performed equally. When asked to explain barriers to equal class participation, two teachers mentioned religious obligations, another mentioned not having personal phones as a barrier. Two teachers mentioned girls were shy, two teachers mentioned boys were hindered by work obligations, one teacher mentioned girls were more daring, and one mentioned that the teacher's style affected their performance.

When it comes to leadership, 40% of teachers said that girls took leadership more often, 40% said it was equal, 10% said boys took leadership roles more often, and 10% didn't know.

5.3.2 Children with disabilities: conditions and social context affecting access and presence in centers

60% of teachers interviewed who answered the questions about children with disabilities mentioned that children with disabilities always participate actively in the classroom. Of teachers answering the question, 71% said children with disabilities always attend remedial classes, with the remainder saying they attended most of the time. 80% of teachers said they were able to adjust their approach to better accommodate children with disabilities in the classroom.

When asked to identify barriers to children with disabilities participating in the classroom, all 10 respondents referenced the actual disabilities rather than attitudes, lack of materials, or anything else. One teacher mentioned that there were no barriers because the teachers were able to provide for the disabilities.

BASELINE FINDINGS AND ANALYSIS

6.2.1 EGRA - What can the baseline tell us about learners' reading skills?

The table below summarizes the average scores for EGRA tasks administered to Grade 1 and Grade 4 learners in learning centers in Lebanon at baseline. It also presents selected indicators used in the PMF document, including the % of students qualifying as 'readers' (student read at least five words correct in first 30 seconds, and therefore did not trigger the reading passage stop rule) and 'readers with comprehension' (student correctly answered at least 80% of reading comprehension questions).

Due to access issues around UNRWA refugee camps, the baseline did not use comparison groups.

Figure 13: Early Grade Reading Assessment (EGRA) scores, baseline

	Variable	Intervention centers			
		Grade 1	N	Grade 4	N
Task 1: Letter recognition	Average # correct letters	17.7	43		
	Average % letters correct	66%	43		
Task 2: Phonemes	Average % phonemic awareness correct	78%	87		
Task 3: Most Used Words	Average # most used words correct	14.6	36	33.3	64
	Average % most used words correct	70%	37	87%	64
Task 4a: Reading	% of students qualifying as readers ⁶			92%	63
	Average % reading passage correct (accuracy)			87%	63
	Average words correct per minute (fluency)			36.5	63
Task 4b: Reading Comprehension	Average % reading comprehension			75%	91

⁶ Student read at least five words correct in the first line, and therefore did not trigger the reading passage stop rule.

	% of students qualifying as readers with comprehension ⁷			58%	91
Task 5: Listening Comprehension	Average % listening comprehension			58%	106

Learners demonstrate good levels of Arabic literacy. Grade 1s showed 66% accuracy in identifying letter names and 78% accuracy in identifying letter sounds. However, at Grade 1 speed remains a problem, with Grade 1 learners on average unable to complete two lines of letter recognition. The one ‘common’ task,⁸ Task 3 most used words, shows improvement from Grade 1 to Grade 4, with number of words doubling and accuracy increasing. In fact, Grade 4 learners exhibited a ceiling effect for Task 3 as a quarter of the sample got the maximum possible score.

By Grade 4, 92% of learners were classed as readers with average 36.5 words per minute read and 87% accuracy. However, there is a gap between fluency and comprehension. One-third of learners demonstrated reading without comprehension, failing to meet the target of 80% or more of comprehension questions.

The contextual documents and responses from teachers and directors suggest that girls outperform boys in school. In the baseline survey, girls and boys performed about equally on EGRA assessments in Grade 1 but by Grade 4 girls had higher average scores than boys in every task. However, for the baseline sample these differences were not statistically significant. This may reflect the relatively small sample size.

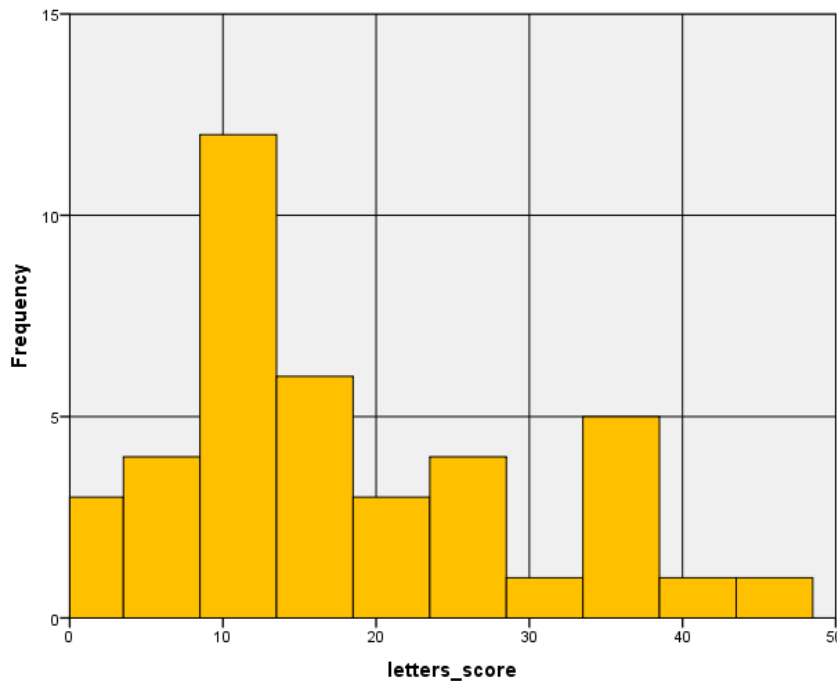
6.2.1.1 Task 1: Letter Identification

The table below shows the distribution of EGRA scores for Task 1: Letter Identification. This was a timed task (1 minute) administered with a stimulus to Grade 1 students only. Although the average is 17.7 letters, the normal curve is centered around the median value of 13 letters. Note that EGRA Task 1 had only 50 letters instead of the usual 100, and no student achieved the maximum score of 50 letters in one minute. Scores were not adjusted for time remaining.

Figure 14: EGRA Task 1 Letter Identification scores (Grade 1 only)

⁷ Student was a reader, and correctly answered at least 80% of reading comprehension questions.

⁸ It should be noted that Grade 1 and Grade 4 had different word lists.

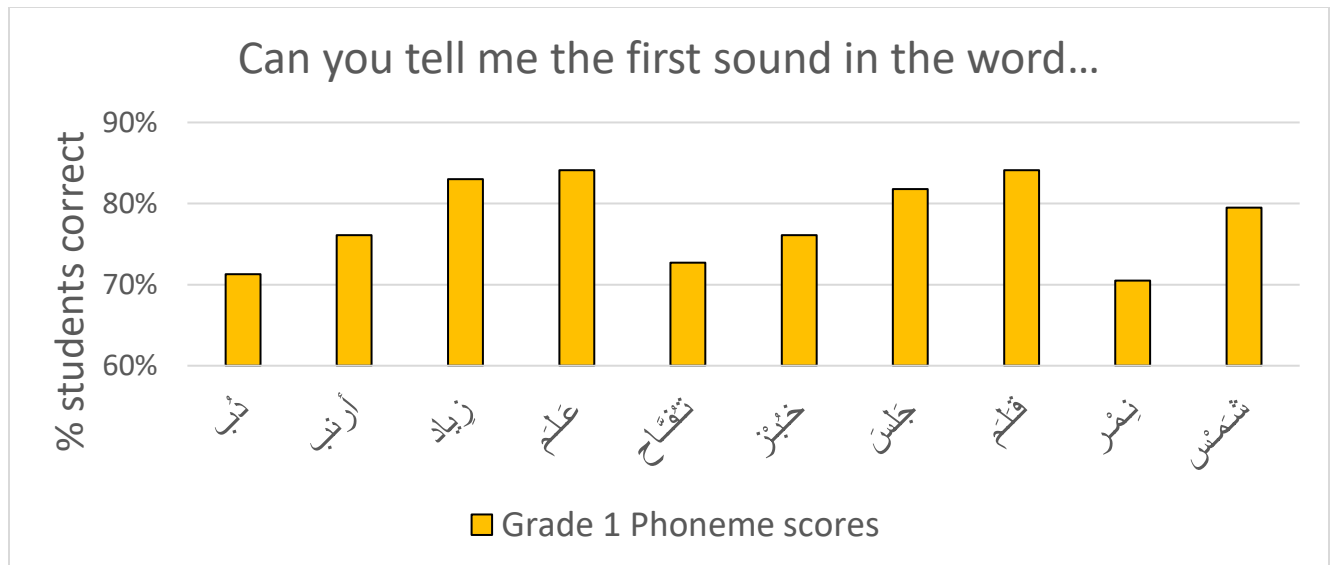


Almost 20% of learners named fewer than five letters and 2% could not name a single letter. We note that the relatively poor performance on letter names may reflect that this is not emphasized in the current curriculum, or that students are able to name letters but only slowly. Learners performed better on the untimed phonemic awareness task.

6.2.1.2 Task 2: Phonemic Awareness

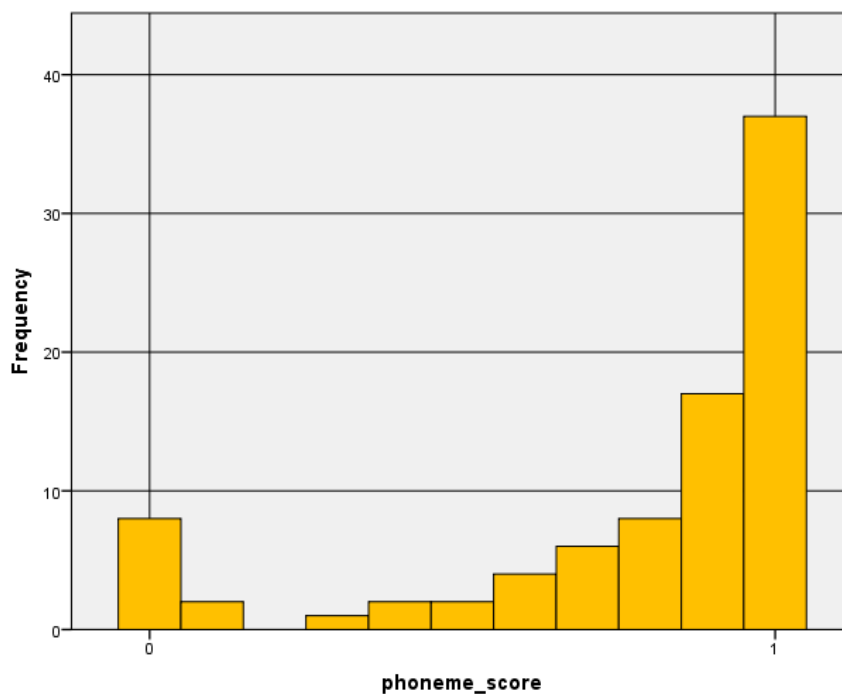
The table below shows the distribution of EGRA scores for Task 2: Phonemic Awareness. This task was untimed and administered without stimulus to Grade 1 only. The items are presented in the order they were asked. Accuracy is 78% across all tasks, ranging from 71% - 84%, and there is little or no evidence of respondent fatigue towards the final items.

Figure 15: Task 2: Phonemic Awareness (Grade 1 only)



Again, the average masks two inflection points. Around 10% of respondents scored zero percent on the phoneme activity, while over 40% of respondents scored 100% on phonemes.

Figure 16: Average Grade 1 scores on Phonemic Awareness

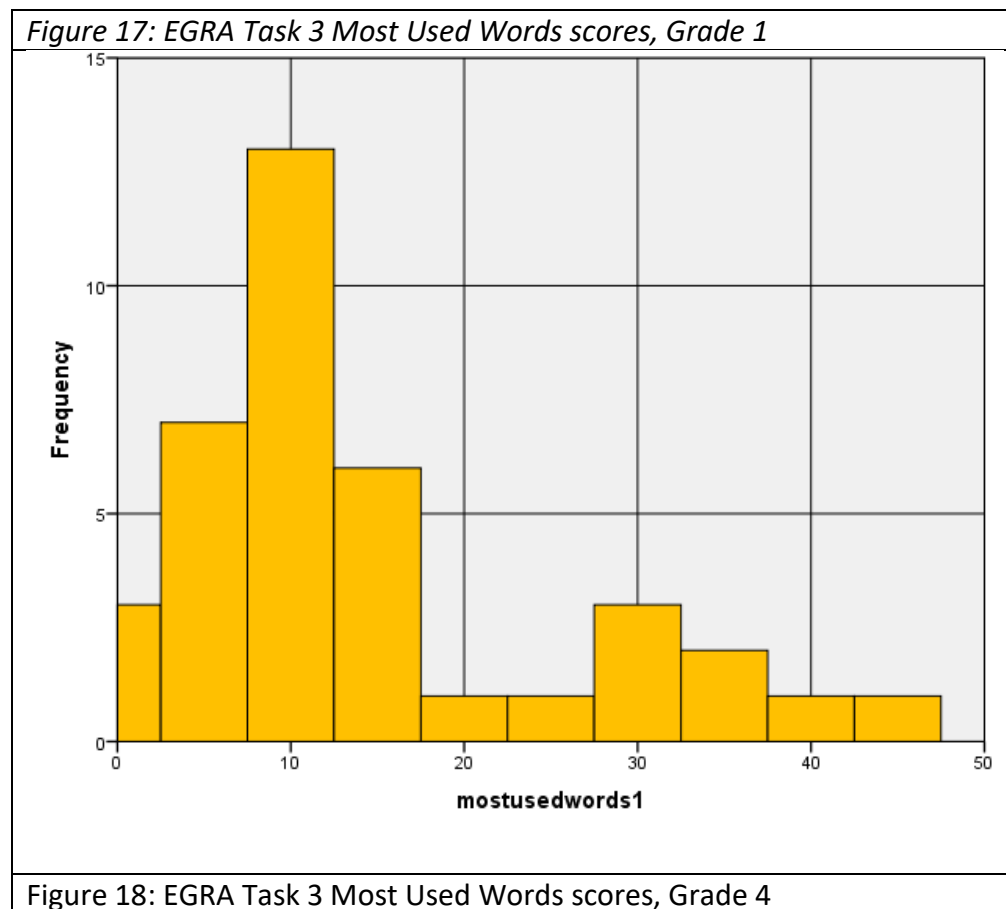


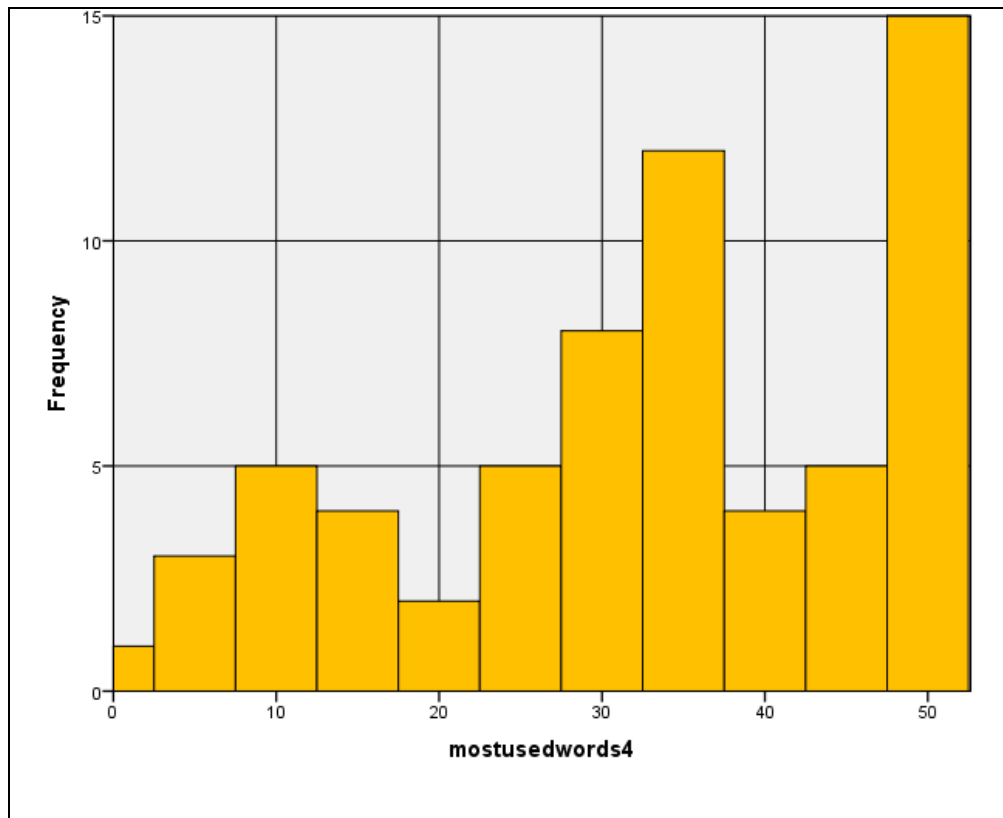
We recommend extending the exercise to 20 phonemes at midline and endline.

6.2.1.3 Task 3: Most Used Words

The table below shows the distribution of EGRA scores for Task 3: Most Used Words for Grade 1 and Grade 4. This was a timed task (1 minute) administered with stimulus, but different word lists were used for Grade 1 and Grade 4 learners. Scores were not adjusted for time remaining.

For Grade 1 learners, although the average was 14.6 words read correctly per minute, the main normal curve is centered around the median of 11 correct words. There is another curve of higher performers centred around 30 words correct.





For Grade 4 learners, the average of 33.3 words read correctly masks three groups in the sample. Low performers are centered around 10 words correct (20% of learners scored less than 20 words correct), medium performers are centered around 35 words correct (34% of learners scored between 20 and 40 words), and a quarter of the sample got 47 or more words correct out of 50.

Given that 12% of the sample got the maximum number of words correct, we recommend that RTP consider increasing the number of words on the task to 100 at midline and endline to account for a potential ceiling effect.

6.2.1.4 Task 4a: Reading Passage

This task was administered to both Grade 1s and Grade 4s. It was composed of a reading passage of 31 words for Grade 1s and of 50 words for Grade 4s along with five reading comprehension questions. The reading passage was timed (1 minute) but the questions were not. The survey was scripted so that each reading comprehension question was only asked if the student had reached the

appropriate point in the passage. Questions which were not asked or skipped were counted as incorrect.

For Grade 1s, the average score of 9.0 words read correctly in a minute is close to the median score of 7 words per minute. Around 30% of Grade 1 learners read 10 or more words per minute, and only 3.6% of the sample were unable to read a single word. These scores are around 60% of the scores for most used words, showing that there are challenges in fluency and reading without diacritical marks in this exercise.

For Grade 4s, the average score of 36.5 words read correctly in a minute masks that 30% of those with a valid reading score read 47 or more words correctly, including 18% who achieved the maximum score of 50. Around 8% of learners could not read the first five words and had the task discontinued; these were excluded from the graph below.

The relatively short length of the reading passage may have accounted for a ceiling effect to this task as well, even though the passage was made more challenging by the absence of diacritic marks showing learners the short vowel sounds.

To account for potential ceiling effects and allow for additional improvement, we recommend lengthening the passage to 100 or more words at midline and endline.

Using linear regression modelling, the best and only significant predictor of a child's reading score was whether the child reported having story books in the home, with a p value of $<.05$ and a standardized beta of 0.33. Grade was excluded from the analysis as the reading passage was only administered to Grade 4 students. No other variable was significantly correlated with the outcome.

Figure 19: Distribution of Grade 1 reading scores

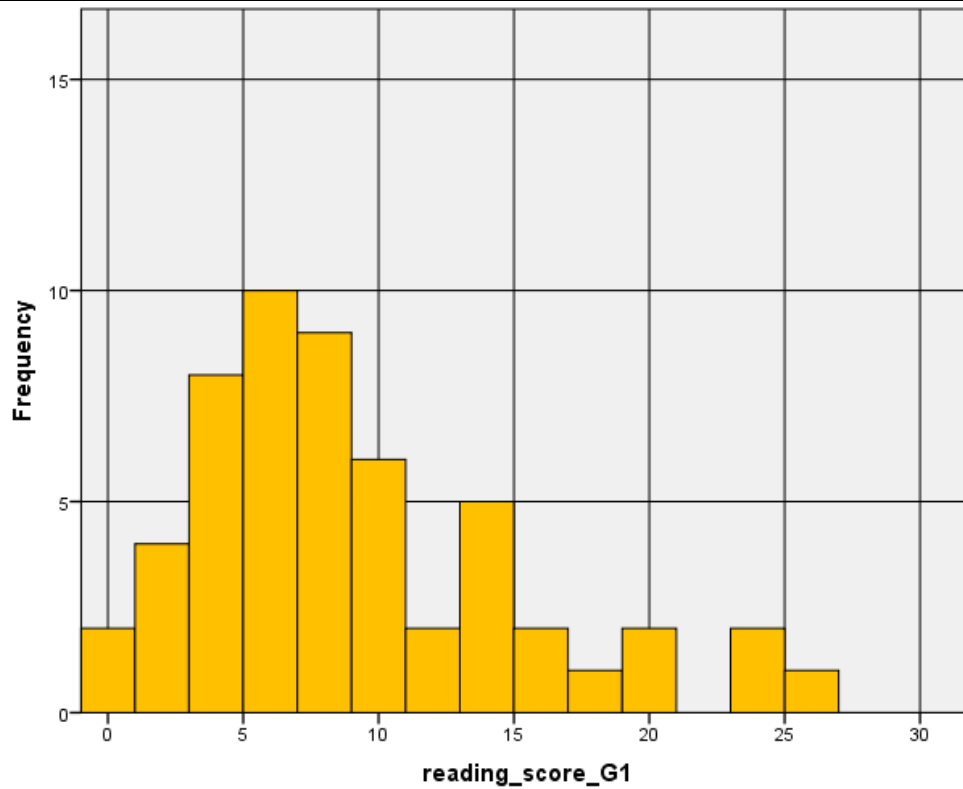
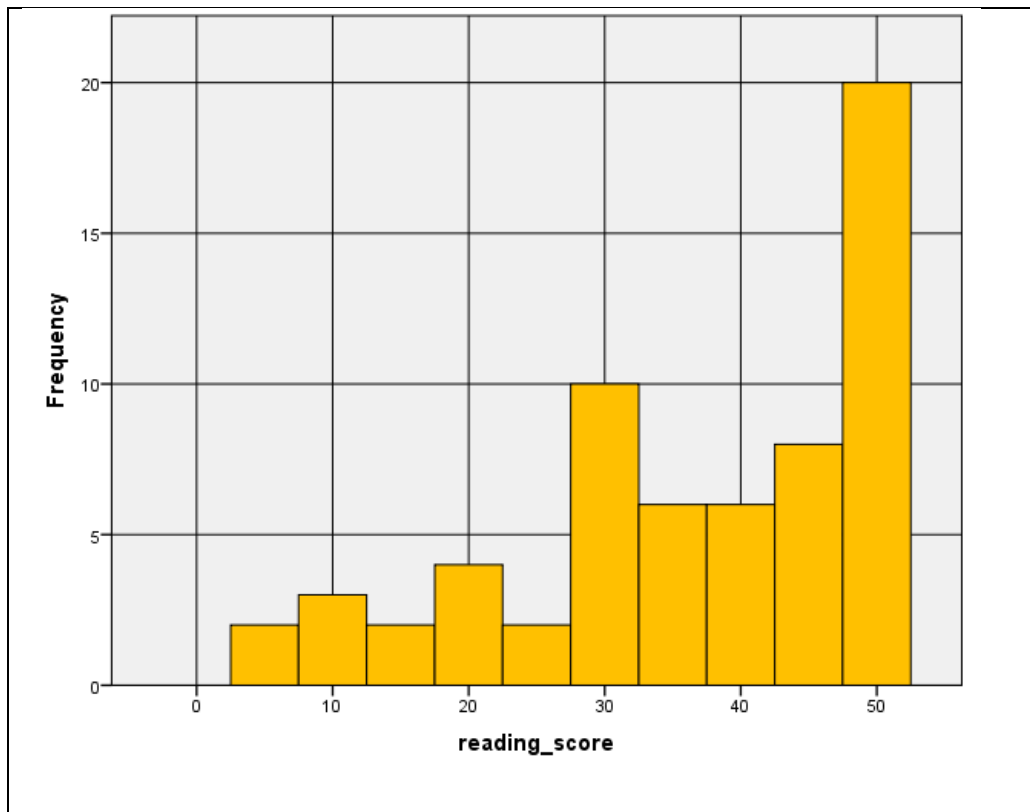


Figure 20: Distribution of reading scores (readers only)



6.2.1.5 Task 4b: Reading Comprehension

The tables below show the distribution of reading comprehension scores for Grade 1 and Grade 4 learners. In Grade 1, almost half of students were non-readers, answering 0-1 comprehension questions correctly, while only 18% of students met the target of 80% comprehension.

By Grade 4, nearly two-thirds of students (58%) demonstrated reading with comprehension, and only 5% were non-readers.

Figure 21: Distribution of Grade 1 Reading Comprehension

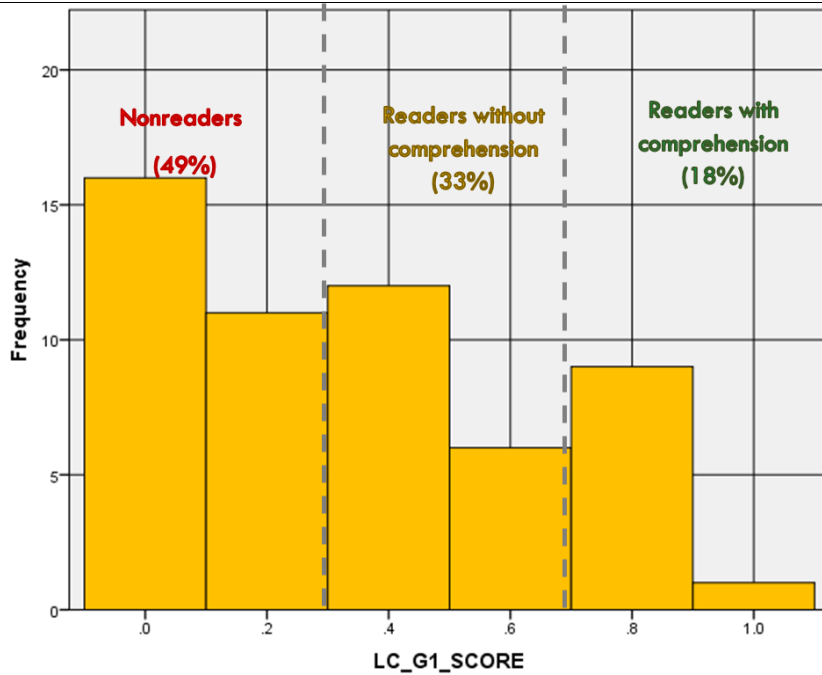
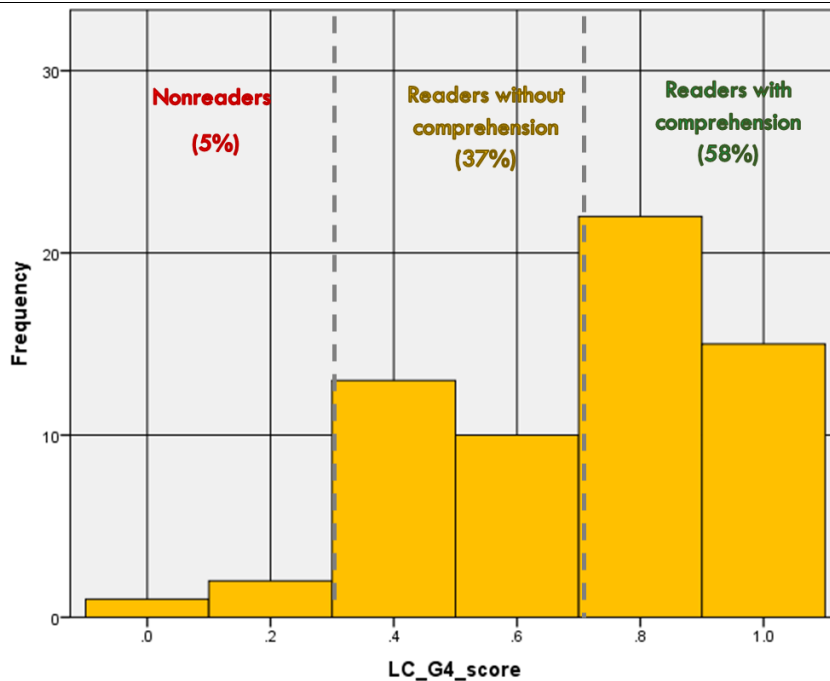


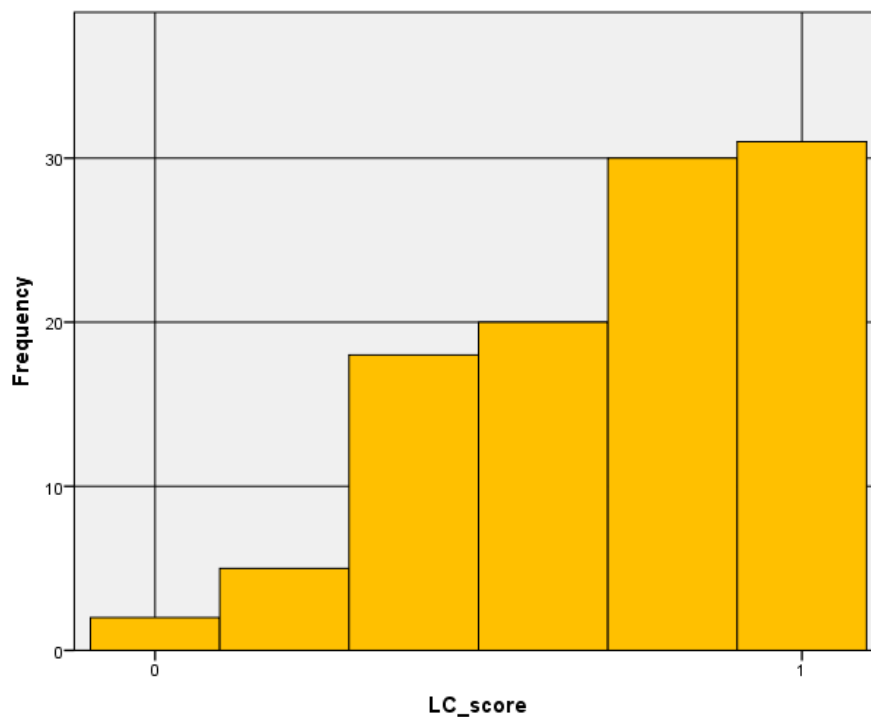
Figure 22: Distribution of Grade 4 Reading Comprehension



6.2.1.6 Task 5: Listening Comprehension

Enumerators read a short passage twice out loud to Grade 4 learners only and then asked them a series of comprehension questions. Questions were mostly factual but the last was inferential i.e. the answer was not directly present in the passage. Learners' responses were not timed.

Figure 23: Distribution of Listening Comprehension scores (Grade 4 only)



Over half of learners got 80% or more of the listening comprehension questions correct and 30% got all five questions correct. This task, the only one not requiring some reading skill, underscores the strength of learners in understanding oral Arabic as the language of instruction.

6.2.1.7 EGRA results Summary

A deeper dive into the five EGRA task results shows that it is helpful to look past average scores. At baseline, these findings suggest:

- Tasks are too easy for some Grade 4 learners, with ceiling effects demonstrate on Task 3 Most Used Words and Task 4 Reading Passage. These tasks should be extended so that further

improvement over and above baseline can be captured. Midline and endline analysis will need to be adjusted to account for ceiling effects at baseline if these tasks are extended.

- Tasks 1 and 2 seem to indicate different basic literacy levels among Grade 1 students, but these can be reconciled. It may be that Grade 1 students are accurate but slow at identifying letters, or it may be that current curriculum emphasizes letter sounds (phonemes) over letter names. In any case, relatively good performance by Grade 4s suggests that most students already seem to have a solid foundation in literacy in Arabic.
- Learners in Lebanon are at an advantage because the language of instruction, Arabic, is also the main language used at home for virtually all learners. In general, learners in Arabic speaking countries have relatively strong literary skills boosted by Qur’anic education, but an approach to rote learning may come at a cost to learning comprehension.
- From these positive take-aways, Right to Play should not overlook around 8-10% of learners who scored very poorly on the tasks (five letters or less on Task 1, five words or less on Task 4). Surely it is the same bottom 10% of learners who score poorly on basic tasks who then show little progress on most used words or the reading passage.

6.2.2 How do these reading skills differ by equity indicators?

6.2.2.1 Gender

Although the program assumed that girls outperformed boys in the classroom, the baseline EGRA assessment did not provide evidence of this. In Grade 1, there was no clear pattern of girls outperforming boys, and by Grade 4, although girls’ scores were higher than boys’ scores across all tasks, none of the differences were statistically significant. The lack of statistical significance may be a result of small sample size and overperformance of 3-8%.

Figure 24: Performance on Grade 1 EGRA Tasks, girls vs. boys

		Intervention centers						
	Variable	Girls	N	Boys	N	Total	N	Sig.
Task 1: Letter recognition	Average # correct letters	17.9	32	17.0	11	17.7	43	-
	Average % letters correct	65%	32	67%	11	66%	43	-
Task 2: Phonemes	Average % PA correct	78%	58	78%	29	78%	87	-
	Average # MUW correct	13.9	24	15.0	13	14.6	36	-

<i>Task 3: Most Used Words</i>	<i>Average % MUW correct</i>	70%	24	69%	13	70%	37	-
--------------------------------	------------------------------	-----	----	-----	----	-----	----	---

Figure 25: Performance on Grade 4 EGRA Tasks, girls vs. boys

		<i>Intervention centers</i>						
	Variable	Girls	N	Boys	N	Total	N	Sig.
<i>Task 3: Most Used Words</i>	<i>Average # MUW correct</i>	34.1	38	32.3	26	33.3	64	-
	<i>Average % MUW correct</i>	89%	38	83%	26	87%	64	-
<i>Task 4a: Reading</i>	<i>% of students qualifying as readers^A</i>	95%	38	88%	25	92%	63	-
	<i>Average % reading passage correct (accuracy)</i>	89%	38	86%	25	87%	63	-
	<i>Average words correct per minute (fluency)</i>	37.7	38	34.6	25	36.5	63	-
<i>Task 4b: Reading Comprehension</i>	<i>Average % reading comprehension</i>	73%	38	66%	25	70%	63	-
	<i>% of students qualifying as readers with comprehension^B</i>	61%	38	56%	25	58%	63	-
<i>Task 5: Listening Comprehension</i>	<i>Average % listening comprehension</i>	60%	62	55%	44	58%	106	-

6.2.2.2 Other equity dimensions

There were no significant differences in any EGRA scores by CWD status, ECD attendance, language spoken at home, or origin (refugee, internally displaced person, or neither).

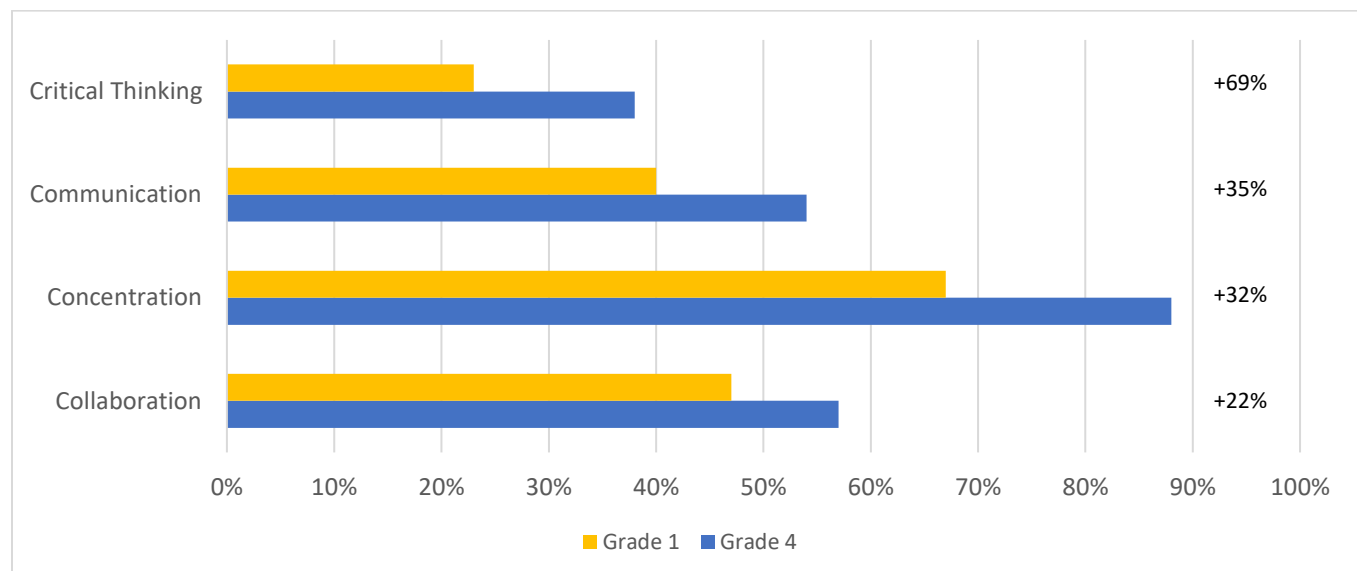
6.3 Life skills scores in intervention learning centers

6.3.1 What can the baseline tell us about learners' life skills?

The learner assessment adopted questions from the International Social and Emotional Learning Assessment (ISELA) developed by Save the Children. The in-country team chose the 4Cs life skills that were used for this assessment.

The progression of life skills scores between the Grade 1 and Grade 4 students is shown in the table below. Life skills scores are standardized and presented as a percentage of the maximum possible score, although each index score is constructed differently. See the 'ISELA' section above for details on each of the indexes presented below.

Figure 26. Change in Life Skills between Grade 1 and Grade 4



The baseline report shows that life skills already increase from Grade 1 to Grade 4 even in the absence of the Right to Play intervention. In Lebanon, the life skill that naturally increases the most over time is critical thinking, increases by over 2/3 (69%). The life skill that naturally increases the least is collaboration, which increases by less than a quarter (22%) from Grade 1 to Grade 4.

There is a potential ceiling effect for concentration/perseverance for both grades, with over half of Grade 1 students (52%) and nearly half of Grade 4 students scoring the maximum score. The communication and collaboration scores follow more of a normal curve.

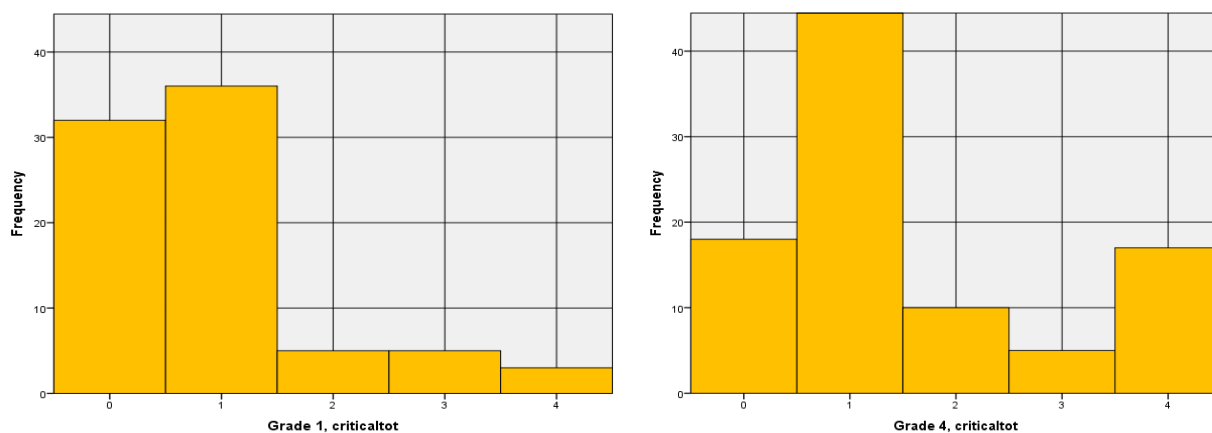
Grade 4 Life Skills scores are mostly above 50%, and it is hence expected that making a difference on these scores for the students already scoring above the average, or in the fourth quartile, will be more

difficult to improve at midline and endline than looking to boost higher scores for the lowest quartiles. This evolution of scores, provided that no intervention aimed at developing life skills has been provided to Grade 4 students in the past years, are a good proxy to show what the normal trajectory of improvement will look like without RTP support.

Disaggregated results for each Life Skill score, for Grade 1 and Grade 4 are presented below. To help visualizing, results are grouped by band (0-1, 1-2, etc.) but here are left unstandardized, so some indexes have a higher maximum score than others.

6.3.1.1 Critical thinking

Figure 27. Critical thinking scores (out of 4)



Using a multiple linear regression model, the significant predictors of high critical thinking scores are grade, student reporting newspapers in the home, student reporting playing games with the family after school, and student correctly answering the child protection question.

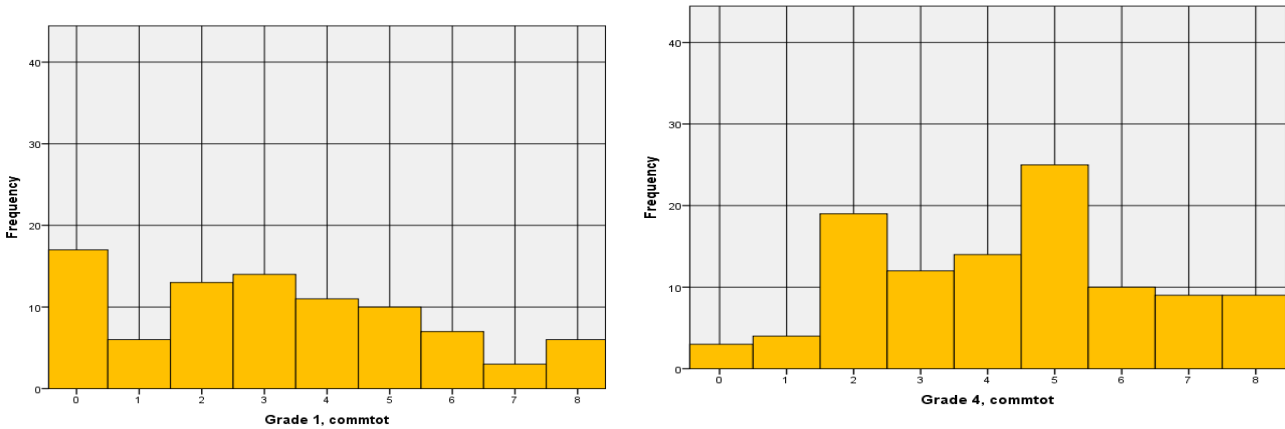
Figure 28: Multiple regression model for critical thinking scores

	<i>b</i>	SE B	β	<i>p</i>
Constant	.047	.061		.440
Grade	.049	.015	.234	.002
At home, do you have (the following types of reading materials): Newspaper	.165	.064	.187	.011
What do you usually do in the evening/after school? Play games with my family	.095	.047	.146	.044
If you or a friend were being harmed by someone else, what would you do? Student answered correctly	.104	.046	.168	.023

Adjusted R-square value = .111.

6.3.1.2 Communication

Figure 29: Communication scores (out of 8)



The linear best fit model for predicting communication scores has grade and students reporting a family member reading to them and/or playing literacy games with them in the last week. Student reported watching television in the morning is negatively correlated with the communication score.

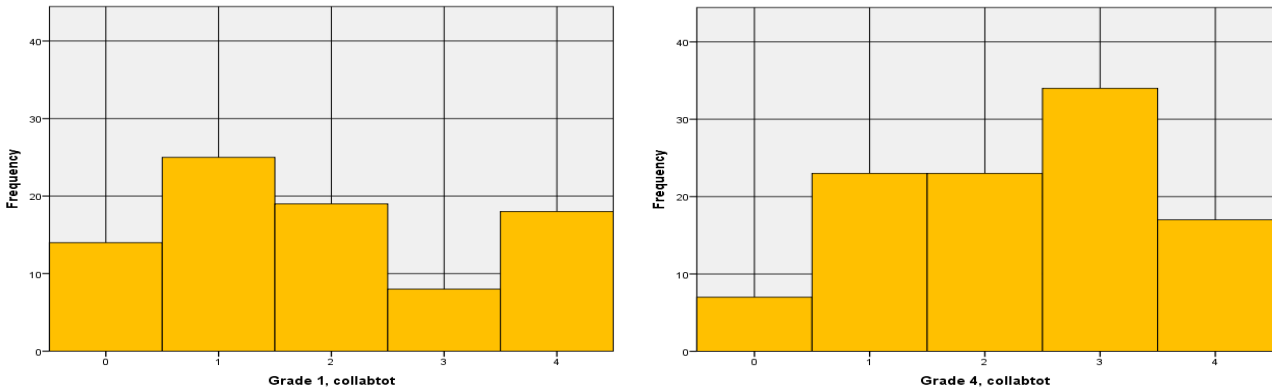
Figure 30: Multiple regression model for communication scores

	<i>b</i>	SE B	β	<i>p</i>
Constant	.331	.044		.000
Grade	.032	.013	.166	.014
In the past week, has a family member read to you?	.146	.038	.257	.000
What do you usually do in the morning / before school? Watch TV	-.143	.038	-.248	.000
In the past week, has a family member played literacy games with you?	.106	.039	.186	.007

Adjusted R-square value = .223.

6.3.1.3 Collaboration

Figure 31. Collaboration scores (out of 4)



Using a multiple linear regression model, the significant predictors of high collaboration scores are student reporting religious materials in the home, student reporting participation in virtual literacy extracurricular activities, and student correctly answering the child protection question.

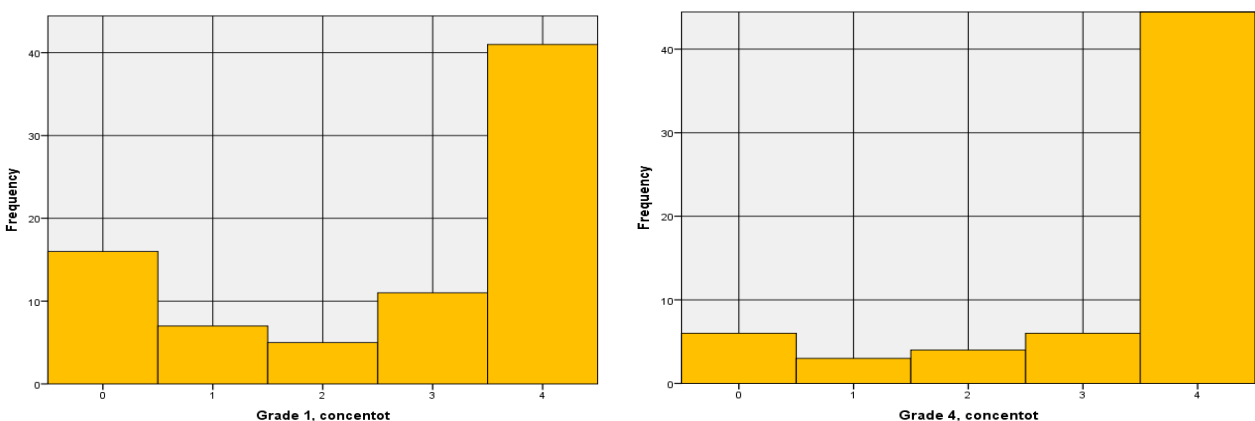
Figure 32: Multiple regression model for collaboration scores

	<i>b</i>	SE B	β	<i>p</i>
Constant	.233	.081		.004
At home, do you have (the following types of reading materials): Religious materials	.201	.079	.181	.012
Do you participate in out-of-school reading activities, like those available through radio, TV, or mail?	.164	.049	.239	.001
If you or a friend were being harmed by someone else, what would you do? Student answered correctly	.124	.046	.194	.007

Adjusted R-square value = .114.

6.3.1.4 Concentration

Figure 33. Concentration scores (out of 4)



Using a multiple linear regression model, the only significant predictor of high concentration scores is grade.

Figure 34: Linear regression model for concentration scores

	<i>b</i>	SE B	β	<i>p</i>
Constant	.597	.052		.000
Grade	.072	.017	.300	.000

Adjusted R-square value = .085.

Collaboration and communication scores seem to be already above the halfway mark by Grade 1, and by Grade 4 they have naturally improved further. For collaboration, the number of zero scores have halved by Grade 4, while the number of students scoring 3 or 4 have tripled. For communication, the number of zero scores drops by 80% in Grade 4, while the median score shifts from 3 in Grade 1 to 5 in Grade 4.

However, critical thinking and concentration scores seem to follow a different logic, with most of the improvement from Grade 1 to Grade 4 coming from more learners achieving the maximum possible score. Critical thinking thus seems to be either more difficult or slower to develop naturally in children, while concentration strongly improves throughout grades and communication and collaboration follow a more normal distribution. There might be a case where RTP intervention could benefit critical thinking scores as these are more complicated to increase, potentially because critical thinking as a concept is multidimensional and assumes a certain level of cognitive development.

Moreover, although an average score is reported in the PMF and used to set objectives for the upcoming programming year, we recommend using the distribution of scores to set objectives in increasing life skills scores; this would already give a first idea of how much can the scores increase above their natural evolution, but also allow RTP to target the sub-groups of students who need the most support. Instead of looking at further increasing scores for children already scoring high in the last quartile, we recommend that RTP focus on improving the scores of the lowest-scoring students. RTP will be able to show more improvement by targeting, for instance, 0 to 3 scores on communication than for 5 to 8 scores.

6.3.2 Life Skills and gender

There is no significant difference on life skills scores for boys and girls overall.

For Grade 1, girls have better communication scores*(44%) than boys (30%). By Grade 4, these differences are not significant anymore, and the only significant difference is on conflict resolution, for which girls score 10% higher than boys*. Conflict resolution was not selected to be reported upon in the PMF, however this competency was assessed as part of the Learner assessment.

Figure 35: Disaggregated scores for Grade 1&4 boys and girls

Grade 1				Grade 4			
Measure	<i>n</i>	Score	%	Measure	<i>N</i>	Score	%
Concentration: Overall	80	2.68	67%	Concentration: Overall	101	3.53	88%
Female	55	2.75	69%	Female	60	3.68	92%
Male	28	2.52	63%	Male	41	3.32	83%
Communication: Overall	87	3.21	40%	Communication: Overall	105	4.32	54%
Female	59	3.58	45%	Female	62	4.56	57%
Male	28	2.43	30%	Male	43	3.98	50%
Collaboration: Overall	84	1.89	47%	Collaboration: Overall	104	2.3	57%
Female	57	1.84	46%	Female	62	2.35	59%
Male	27	2	50%	Male	42	2.21	55%
Critical Thinking: Overall	81	0.9	23%	Critical Thinking: Overall	101	1.52	38%
Female	57	0.96	24%	Female	59	1.66	42%
Male	24	0.75	19%	Male	42	1.33	33%

6.3.3 Life Skills and poverty levels

Socio economic scale (SES) items were selected as questions for the students and parents' interview, as to proxy the level of poverty of the household and measure whether it is correlated to lower achievement scores for Life Skills and EGRA.

Linear regression was run on four Life Skills scores variables and each socio-economic status proxy variable was tested independently on each life skills scores, however these are not significant. It was noted earlier in section "Marginalization and Equity indicators" that socio-economic scores were normally distributed.

6.3.4 Life Skills and disability

Similar to what was noticed for EGRA scores, there is no significant difference on Life Skills scores for children with disabilities. However, the numbers overall are very low and even more so at Grade 4, and likely too small to allow to detect significant differences.

Figure 36. Disaggregation of Life Skills scores by children with or without disability

Grade 1				Grade 4			
Measure	<i>n</i>	Score	%	Measure	<i>n</i>	Score	%
Grade 1 Concentration: Overall	80	2.68	67%	Grade 4 Concentration: Overall	101	3.53	88%
Grade 1 Children with Disability	17	2.18	54%	Children with Disability	8	3.88	97%
Grade 1 Children without Disability	63	2.81	70%	Children without Disability	92	3.54	89%
Grade 1 Communication: Overall	87	3.21	40%	Grade 4 Communication: Overall	105	4.32	54%
Grade 1 Children with Disability	19	3.11	39%	Children with Disability	9	2.89	36%
Grade 1 Children without Disability	68	3.24	40%	Children without Disability	95	4.44	56%
Grade 1 Collaboration: Overall	84	1.89	47%	Grade 4 Collaboration: Overall	104	2.3	57%
Grade 1 Children with Disability	18	1.67	42%	Children with Disability	9	1.44	36%
Grade 1 Children without Disability	66	1.95	49%	Children without Disability	94	2.38	60%
Grade 1 Critical Thinking: Overall	81	0.9	23%	Grade 4 Critical Thinking: Overall	101	1.52	38%
Grade 1 Children with Disability	17	0.47	12%	Children with Disability	8	1.25	31%
Grade 1 Children without Disability	64	1.02	25%	Children without Disability	92	1.57	39%

6.3.5 Life Skills and daily activity patterns

Attending pre-school, repeating Grade 1 or Grade 4 makes no significant difference on Life Skills scores overall. It is however worth noticing that repeat frequencies are low for Lebanon and thus the subsample is very small.

When disaggregating by grades, Girls in Grade 1 are less likely to play with friends in the morning or after school, and even less so by Grade 4 where only 19% of girls play with friends in the morning versus 40% of boys. Although playing with friends in the morning and after school does not significantly impact Life Scores at Grade 4, it might be a problem in terms of programming if

girls do not access outdoor play-based activities to the same extent as boys, as it might exclude them from activities.

6.3.6 Reading activities with significant correlation to Life Skills scores

The two following tables present the correlations between Life Skills scores and two of the PMF indicators, Supplemental reading (index variable comprised of variables “virtual_camp” and “read_camp”) and home reading activities, which is an index variable built upon the following questions:

In the past week, have you seen anyone in your home reading?

In the past week, has someone in your family helped you with your homework?

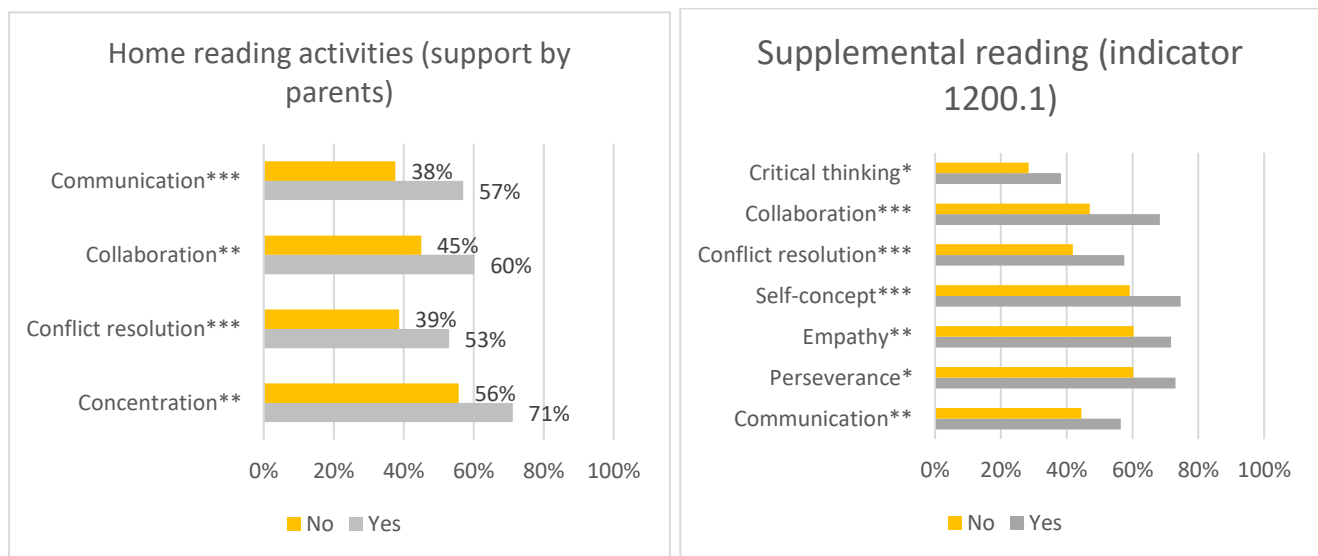
In the past week, has a family member read to you?

In the past week, has a family member told you a story?

In the past week, has a family member played literacy games with you?

Correlations in the tables below confirm that these indicators are relevant to follow for RTP as there is a big and significant difference between students benefitting from these support reading activities and others. The variables were tested on the “Four Cs” but also on additional Life Skills tested through the Learner Assessment, such as stress management, self-concept, perseverance, and conflict resolution.

Figures 37 and 38: Reading support with significant correlation with Life Skills Scores



These two indicators are important as they impact two Life Skills which are less influenced than others by most of the variables controlled in the Learner Assessment: Self-concept and Critical thinking. While several contextual variables influence collaboration and communication quite often, variables requiring a certain level of cognitive complexity seem more difficult to influence.

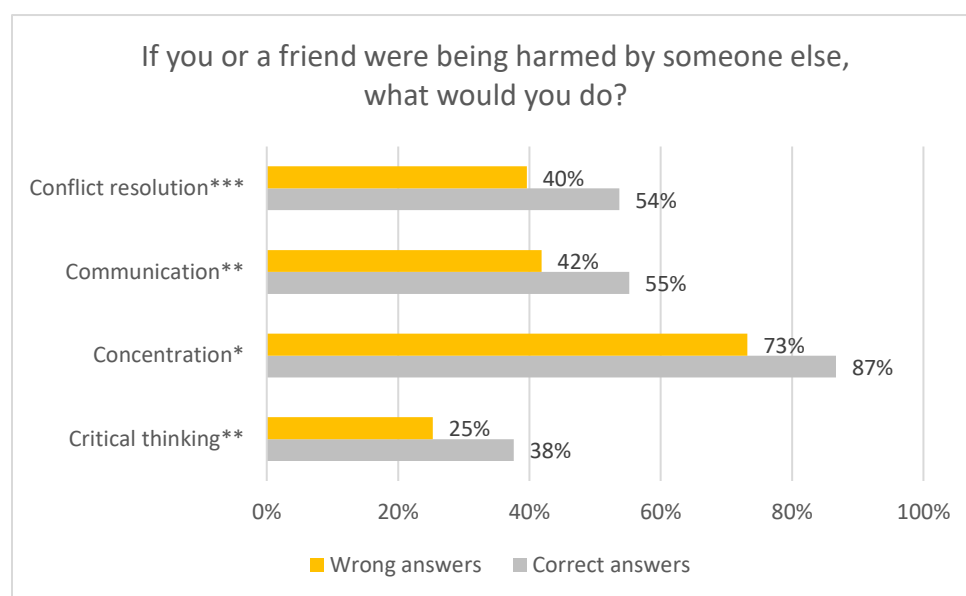
6.3.7 Risk variables with significant correlation to Life Skills scores

Eight questions were asked to students to proxy the level and type of risk they were faced with and whether this correlated with some of their Life skills scores, as a proxy for their development.

There is a correlation between students having had to earn money to support their family and some of the Life Skills scores. As was the case in Mozambique, here too in Lebanon, the answer to the “child protection” variable, where students are asked what they would do if themselves or a friend were harmed and to which they can either answer to “tell a trusted adult” (which stands for the “correct answer”), “don’t know” (which stands for the “wrong answer”) or chose not to respond, significantly correlated with Life Skills scores, as children selecting the “don’t know” answer have significantly lower scores on Life Skills across the board.

36% of Grade 1 students and 60% of Grade 4 students gave the right answer, with the remainder giving the wrong answer or refusing to answer. Considering that this variable correlates significantly with most scores, this variable could be used to disaggregate and identify, in the students’ population, the children who need the most support to increase their life skills.

Figure 39: Correlation of child protection variable with 4Cs



* $p < .05$, ** $p < .01$, *** $p < .001$

6.3.8 Positive answers to “Relate” variables significantly influences Life Skills scores

Students were asked a set of 18 questions under the section “relate”, aimed to assess the strength of their relationships but also the behavioral patterns dictating their relationships to others. Correlations were tested on life skills which were not constructed as index variables including some of the relationship questions, as it was done for Mozambique.

Overall, answering positively to the “Relate” variables significantly impacts all the Life Skills scores compared to answering negatively.

The three “Relate” variables which significantly impact most of the Life Skills are Relate 6, 10, 11 and 16; three of these refer to seeking help or support when needed. It goes in the same direction as the results presented above, showing that children who correctly answer the question “If you or a friend were being harmed by someone else, what would you do?” have increased Life Skills Scores compared to those who don’t. The correct answers refer to seeking support or telling an adult. More analysis is needed to determine whether the behaviour of seeking help enhances Life Skills or else whether children with good scores on Life Skills questions are more likely to seek support when needed.

Figure 40: Correlations between “relate” variables and Life Skills scores

		Stress management	Critical thinking	Concentration	Collaboration	Communication	Self-concept	Perseverance	Conflict	Empathy
relate2	Do you talk to anyone at home when you are sad?	+36%***	X	X	+10%*	+37%***	X	+14%**	+14%*	+8%*
relate3	Do you talk to anyone else at home when you are sad?	+19%***	X	X	X	+21%***	X	X	X	X
relate4	Do you ask anyone at home for help when you have a problem with a friend?	+15%***	X	X	+13%**	+34%***	X	+21%***	X	+7%*
relate5	Do you ask anyone else at home for help when you have a problem with a friend?	+14%*	X	X	X	+24%***	X	+14%*	X	X
relate6	If you are working on something difficult, do you ask anyone at home for help?	+27%***	X	+16%*	+21%***	+27%***	+17%**	+57%***	X	+12%**
relate7	If you are working on something that takes a lot of time and it is difficult to finish, do you ask anyone else at home for help?	X	X	X	X	X	X	+34%***	X	X
relate8	Do you know when someone in your home is feeling sad?	+11%*	X	X	+17%**	+11%*	+19%**	+17%**	+13%*	+35%***
relate9	Can you tell me the names of your good friends?	X	X	+64%*	X	X	X	X	X	X
relate10	Do you talk to any of your friends when you are sad?	+42%***	+10%***	X	+30%***	+28%***	+14%**	+19%***	+13%*	+18%***
relate11	When you have a problem with another child, do you ask any of your friends for help?	+27%***	+13%**	X	+24%***	+31%***	+19%**	+26%***	+10%*	+14%***
relate12	If you are working on something difficult, do you ask any of your friends for help?	+18%***	X	X	+24%***	+20%***	X	+56%***	X	+13%***
relate13	Do you know when your best friend is feeling sad?	+14%**	X	X	+18%**	+13%**	*	+20%***	X	+35%***

relate14	Now, I want you to think about adults who you talk to almost every day in your school (teacher, counsellor, principal, etc.). Can you tell me which adults you talk to almost every day?	X	X	X	+12%**	X	X	X	X	X
relate15	Do you talk to any adult in your school when you are sad?	+40%***	X	X	X	+36%***	X	+18%**	X	X
relate16	Do you ask any adult in your school for help when you have a problem with a friend?	+24%***	X	X	+21%***	+36%***	X	+29%***	X	+12%**
relate17	If you are working on something difficult, do you ask an adult in your school for help?	X	X	X	+15%*	+17%**	X	+54%***	X	+10%*
relate18	Do you know when an adult in your school (teacher, counsellor, principal, etc.) is feeling sad?	+14%**	X	X	+16%**	+13%**	X	X	X	+26%***

Critical thinking, concentration and self-concept are the Life Skills for which the answers to the Relate variables make the least difference and on the least Life Skills scores.

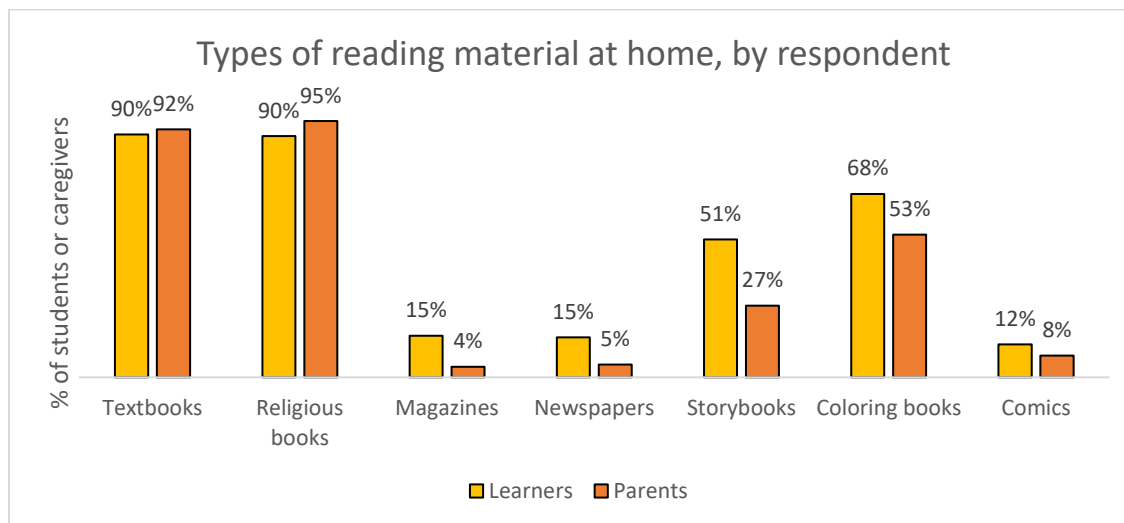
5.4 READING ACTIVITIES IN THE COMMUNITY: PARENT SURVEY

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

6.4.2 How do reported reading activities differ by parents and children?

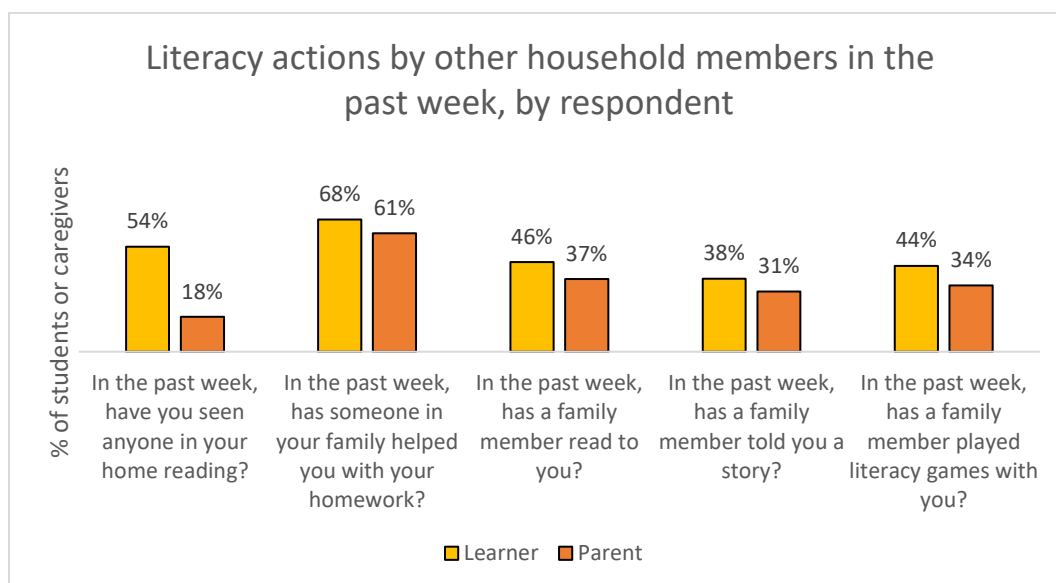
Both learners and parents were asked what types of reading materials were available at home. The sample of parents was a subset of households of children, and 75% of parents were linked to learners. Both learners and parents report similar prevalence of reading materials in the home, with nearly all reporting that their homes had textbooks and religious books. Learners reported magazines and newspapers in the home at about three times the rate of parents, and learners reported more storybooks and coloring books at home than parents.

Figure 41: Types of reading material at home, by respondent



Learners and parents were asked whether certain literary activities had happened in the household. Learners reported somewhat higher rates of these activities happening in the last week. The greatest disparity was in the response to ‘have you seen anyone in your home reading?’ where learners reported triple the rate of the activity in the past week compared to parents.

Figure 42: Literary actions by other household members in the past week, by respondent



6.4.3 Of those activities, which do children take part in? How frequently?

Learners were asked whether they had attended both in-person extracurricular literacy activities (reading camps, reading centers, or reading days out) and media-based extracurricular literacy activities (radio, TV, mail, etc.), and if so, whether they attended such activities in the past week. Attendance was similar for both in-person and media-based extracurricular activities. Grade 4 students were significantly more likely to attend media-based extracurricular activities compared to Grade 1 students, although a higher proportion of Grade 1 students reported attending both in-person and media-based extracurricular activities in the past week.

Figure 43: Self-reported participation in literacy extracurricular activities

Variable	Grade 1	Grade 4	N
Student attends in-person extracurricular literacy activities	25%	37%	180
Student attended in-person extracurricular literacy activities in the past week	63%	38%	56
Student attends media-based extracurricular literacy activities	23%	39%*	181
Student attended media-based extracurricular literacy activities in the past week	89%	58%*	56

6.4.4 How does participation differ by equity indicators?

There was no statistically significant difference in in-person or media-based attendance in extracurricular activities, either overall or in the past week, by sex, disability status, ECD attendance, or SES score.

6.4.5 How does participation in at-home reading activities differ by presence of reading materials in the home?

6.4.5.1 Extracurricular literary activities

The reported presence of certain reading materials in the home was significantly ($p < .05$) or near-significantly ($p < .08$) correlated with participation in extracurricular literacy activities.

- Students reporting storybooks in their household were nearly twice as likely to report having attended in-person literacy extracurricular activities in the last week as those without, and 1.5x as likely to have attended a virtual literacy extracurricular activity as those without.
- Students reporting comic books in household were 1.6x more likely to have attended an in-person literacy extracurricular activity as those without.
- Students reporting magazines in their household were 2.3x more likely to have attended an in-person literacy extracurricular activity in the last week as those without.

6.4.5.2 EGRA scores

Students who reported having storybooks in the home had higher EGRA scores than those who did not.

- Grade 1 students who reported having coloring books in the home had higher most used scores than those who did not (15 words vs. 13 words, $p < .05$).
- Grade 4 students who reported having storybooks in the home had higher most used words accuracy (92%) compared to those who did not (81%), and a higher reading score (41 words in a minute) than those who did not (32 words per minute). Both of these were significant at $p < .05$.

5.4.5.3 Life skills

- Grade 4 students who reported having storybooks in the home had significantly or borderline significantly higher scores on two life skills areas – collaboration (2.6 vs. 2.0, $p < .05$) and conflict (1.7 vs. 1.3, $p < .07$).
- Grade 4 Students who reported having coloring books in the home had borderline significantly higher scores on conflict (1.7 vs 1.2, $p < .06$).

It is not clear from the survey materials whether there is a causal effect between the two sets of variables, i.e. providing children's reading materials will encourage further extracurricular literacy activity, boost EGRA and life skills scores, or whether there is a third intervening factor, i.e. households that are more likely to have reading materials are also more likely to have children who attend literary extracurricular activities. We recommend that Right to Play explore this causal link to determine whether it is effective to provide children's reading materials in the household to encourage participation in extracurricular literacy activities.

5.5 TEACHER PROFESSIONAL DEVELOPMENT

6.5.1 How does teaching practice correlate with teacher well-being?

Palestine baseline report found the observed teacher practices index did not correlated well with self-reported teacher practices index. However, at the time of the baseline, teaching was virtual and in-person classroom observations could not be conducted.

6.5.2 How does teachers' perception of their self-efficacy affect their ability to adapt to new teaching practices?

At the time of the baseline, in-person classroom observations could not be conducted to triangulate these variables. We recommend that Right to Play conduct a follow-up piece of analysis to answer these evaluation questions after conducting in-person classroom observations.

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

At baseline, teachers reported using most play-based learning techniques between 50-75% of the time in classrooms. The two techniques used less frequently were giving students time for free play and using games to teach or review letters.

Figure 44: In the most recent full week of school, how often did you...

Variable	Usage†	N
give children time to ask each other questions and discuss the lesson with each other?	75%	10
read to students or ask them to read aloud?	73%	10
use subject-specific games and play activities that are linked to the lesson?	70%	10
ask students to answer questions, summarize, or act/draw a story read in class?	65%	10
engage children in arts? (role play/theatre/drama, painting/drawing, etc.)	63%	10
engage children in a Reflect, Connect and Apply discussion? ‡	60%	5
engage children in singing, rhymes, or chants?	58%	10
use a game to teach or review sounds, rhymes, or syllables?	55%	10
encourage the children to take leadership roles in class?	53%	10
use a game to teach or review new words?	50%	10
give the students time for free play or to structure their own activities?	38%	10
use a game to teach or review letters?	38%	10

†Every day = 100%, Most days = 75%, Some days = 50%, One day = 25%, Not at all = 0%. ‡Only asked of teachers completing RTP training.

Of the five teachers who received RTP training on play-based learning, all said they found it ‘very useful’.

See also Section below.

6.5.4 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.

The number of teacher respondents at baseline (n=10) was insufficient to meaningfully analyze these differences statistically.

6.5.5 How do teachers adapt teaching practices for CWD and gender-responsiveness?

All teachers reported having a choice in selecting reading materials for their classroom. Teachers were not prompted on selection criteria, and all teachers reported considering criteria which considered children's reading level, relevance, and equity considerations for both girls and children with disabilities.

Figure 45: Criteria teachers report using selecting reading materials for lessons

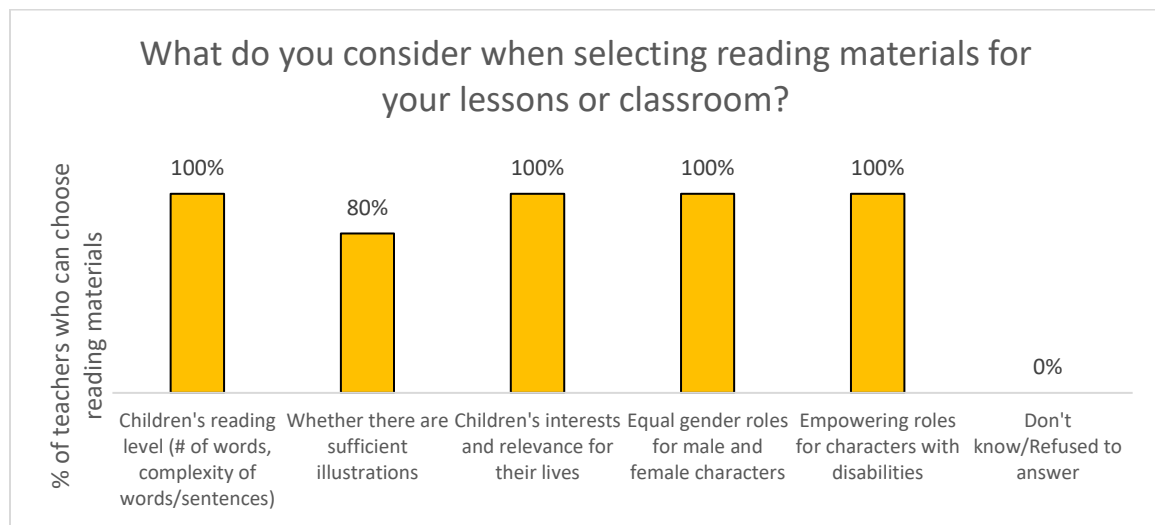
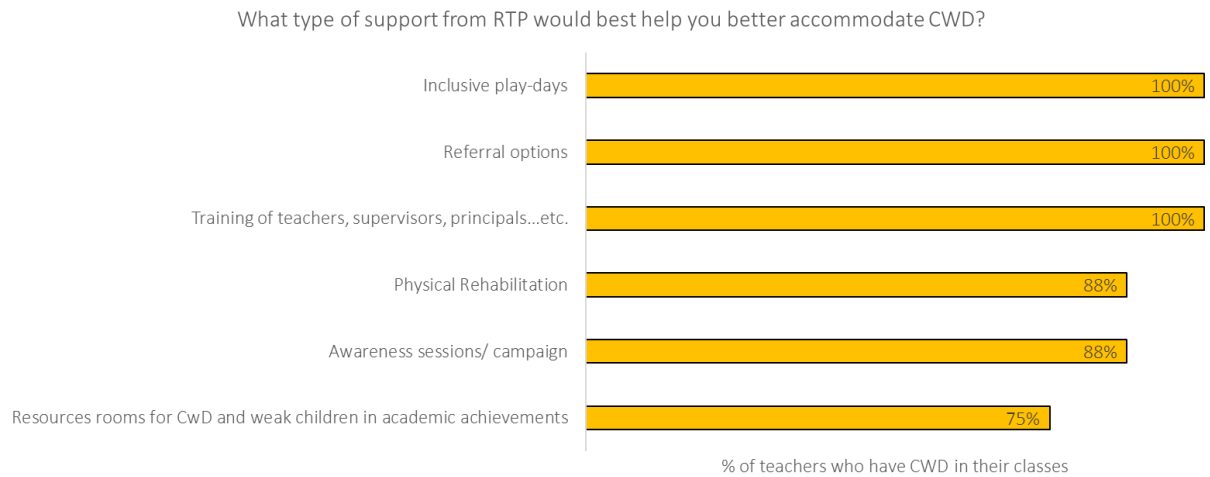


Figure 46: Support teachers request for CWD from RTP



Responses to these questions were strongly positive, with all teachers agreeing with most items. However, there was no way to triangulate these responses and so they should be treated with caution.

2.5 To what extent does RTP training influence teacher knowledge, attitudes, and practices regarding inclusive PBL/literacy instruction?

The number of teacher respondents at baseline (n=10) was insufficient to analyze these differences meaningfully statistically.

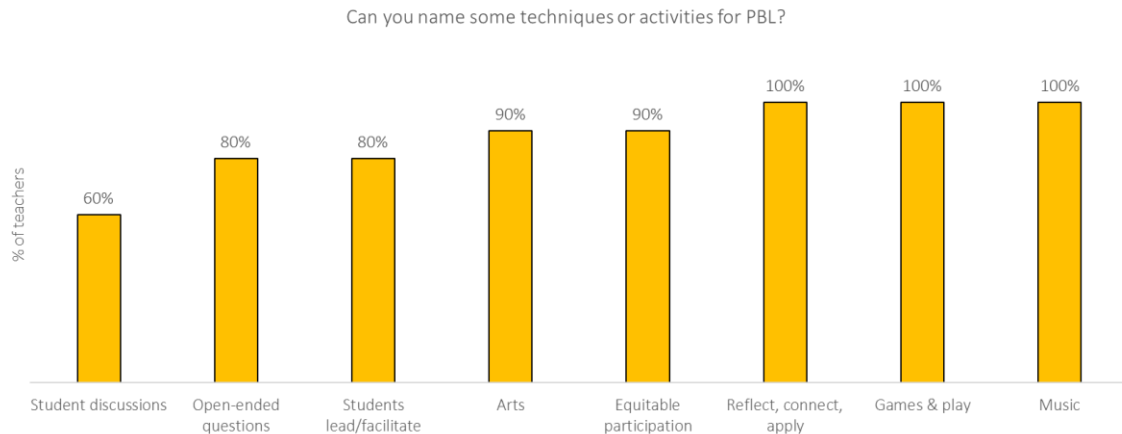
6.6 BUILDING A CULTURE OF READING

6.6.1 Knowledge and practice of PBL and inclusive education in the learning centers (Lebanon)

Teachers were asked to name techniques or activities for play-based learning. They were not prompted about these categories. The table below lists their responses. At the time of the baseline, five teacher respondents had received training from RTP, and five had not.

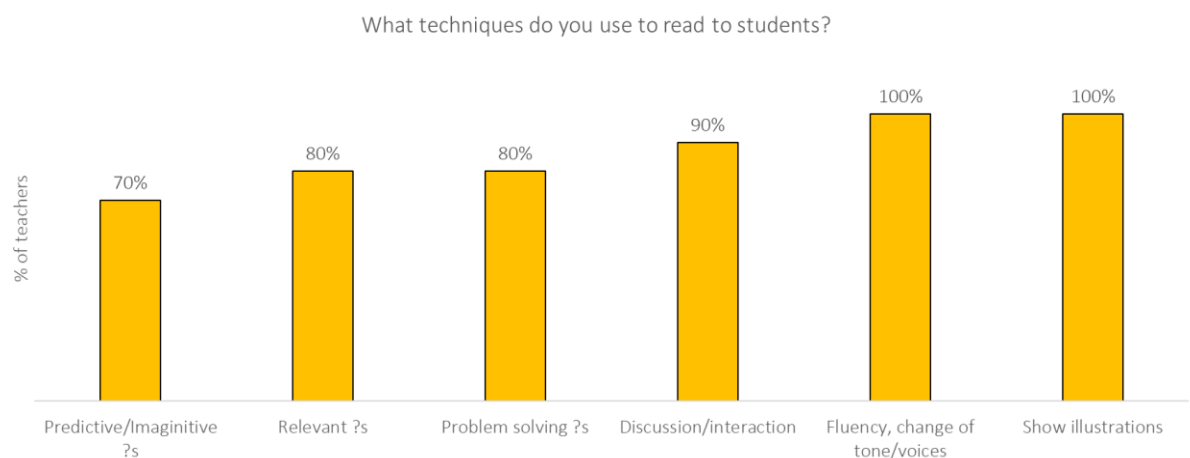
Knowledge of play-based learning approaches is generally high, with 80% of teachers able to name all categories of play-based learning except for student discussions. Some teachers who had not received training from RTP were not aware of arts-based activities, giving students leadership/facilitation opportunities, or open-ended questions, and only 60% of both RTP-trained and untrained teachers mentioned student discussions. Due to small sample sizes, differences between trained and untrained teachers were not statistically significant.

Figure 47: Playful learning techniques, as reported by teachers



Teachers were asked what techniques they use when reading to students. They were not prompted about these categories. Responses were nearly identical between RTP-trained and untrained teachers and results were not statistically significant. As the table shows, the majority of teachers mentioned all playful reading techniques, but the techniques least mentioned were around asking predictive or imaginative questions, questions that relate to the children’s own lives, and problem-solving questions about the story.

Figure 48: Self-reported playful techniques teachers use in the classroom



The table below shows the frequency teachers reported using different methodologies over the most recent full week of school. Totals have been weighted according to the values below the table and reordered in descending order. Teachers on average reported using most of the methodologies in

the last week 50% of the time; two methodologies, giving students free time and using a game to teach letters, were used less than 50%. Results were similar between RTP-trained and untrained teachers and due to small sample size differences were not statistically significant.

Figure 49: In the most recent full week of school, how often did you...

Variable	Overall	N
give children time to ask each other questions and discuss the lesson with each other?	75%	10
read to students or ask them to read aloud?	73%	10
use subject-specific games and play activities that are linked to the lesson?	70%	10
ask students to answer questions, summarize, or act/draw a story read in class?	65%	10
engage children in arts? (role play/theatre/drama, painting/drawing, etc.)	63%	10
engage children in a Reflect, Connect and Apply discussion?	60%	5
engage children in singing, rhymes, or chants?	58%	10
use a game to teach or review sounds, rhymes, or syllables?	55%	10
encourage the children to take leadership roles in class?	53%	10
use a game to teach or review new words?	50%	10
give the students time for free play or to structure their own activities?	38%	10
use a game to teach or review letters?	38%	10

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

6.6.2 Knowledge and practice of PBL and inclusive education in the community

Parents were asked their opinions on students and learning.

Statements 3 and 5 are typical of a 'direct instruction' type of approach to teaching, as opposed to statements 1, 2 and 4 which are considered 'constructivist'. The desired opinion for statements 3 and 5 would be for respondents to disagree with the statement.

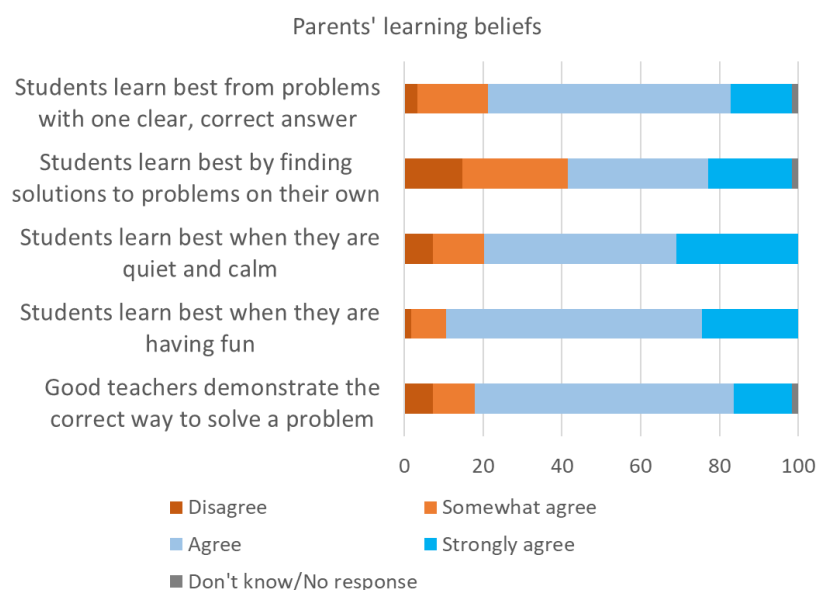
The results show that parents have substantial support for a ‘direct instruction’ approach to teaching. Nearly half (41%) of respondents disagreed or strongly disagreed that students learned best by finding solutions to problems on their own, with the strongest disagreement (15%) of any item, whereas only 20% disagreed or strongly disagreed with the statement ‘students learn best when they are quiet and calm’ and 18% disagreed or strongly agreed that ‘good teachers demonstrate the correct way to solve a problem’.

We believe that parents’ responses to these questions and the ones presented below are evidence of parental support for a traditional, teacher-centered, directive approach to learning, rooted in Arabic and Palestinian culture. Right to Play’s constructivist approach presents a potential challenge to this approach and we suspect that a significant proportion of parents may resist this approach if Right to Play does not work to change traditional attitudes and beliefs around learning.

6.6.3 How do parents’ beliefs about teaching support or conflict with RTP’s approach and teacher’s approach to learning?

At baseline, survey data on parent’s beliefs and opinions show that they broadly support a ‘direct instruction’ type of teaching more than a constructivist type. Nearly 80% of parents agreed or somewhat agreed with three statements supporting a direct instruction approach to learning. This, read in conjunction with gendered attitudes of girls being quiet and boys needing discipline in the classroom (See Section 5.1), illustrates conservative/traditional attitudes towards education which may be a challenge for RTP to overcome.

Figure 50: Parents’ attitudes towards inclusive and collaborative learning methods



We recommend that Right to Play reassess the potential cultural barriers to implementing its model in a relatively conservative Arabic target population, and consider the extent to which it will need to challenge and change existing norms - with the resistance that comes with it – rather than accommodate them.

We also recommend that Right to Play focus on measuring parents’ and teachers’ attitudes towards directive versus constructivist teaching methods more closely going forwards. We believe that acceptance or resistance of the Right to Play model by both parents and teachers will be an important factor determining the uptake of the model at midline and endline.

6.7 SAFE AND INCLUSIVE LEARNING ENVIRONMENTS

6.7.1 Emotional and physical safety in the learning spaces

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

There is a mismatch between what objectives teachers and directors say are included in the school lesson plan. However, this may be because of Directors’ lack of visibility of school lesson plans in the individual learning centers.

Figure 51: What objectives are included in the school lesson plan?

Objective	Teachers	Directors
Inclusion of children with disabilities	100%	33%
Bullying	100%	33%
Violence (general)	100%	33%
Exploitation	100%	33%
Hygiene	100%	33%
Gender-based violence	90%	33%
Gender	80%	0%
Teacher behavior	60%	0%
Other	0%	0%

Students in Lebanon were not asked questions at baseline about emotional safety or wellbeing concerns such as bullying or violence at learning centers. This is because learning centers were not yet back in session at the time of the baseline.

Do children and teachers know how to report and get support with safety issues?

Responses from directors and teachers point to different perceptions of students reporting safety concerns at learning centers. While all directors replied that students came to them or a member of staff to address center safety concerns, and two of three directors replied they had a system for doing so, half of teachers responded that students ‘never’ or ‘rarely’ come to them to report safety concerns inside the center. 40% responded that they ‘always’ do.

All 10 teachers said they knew what to do, or where to go for support, when they were concerned about a child’s safety.

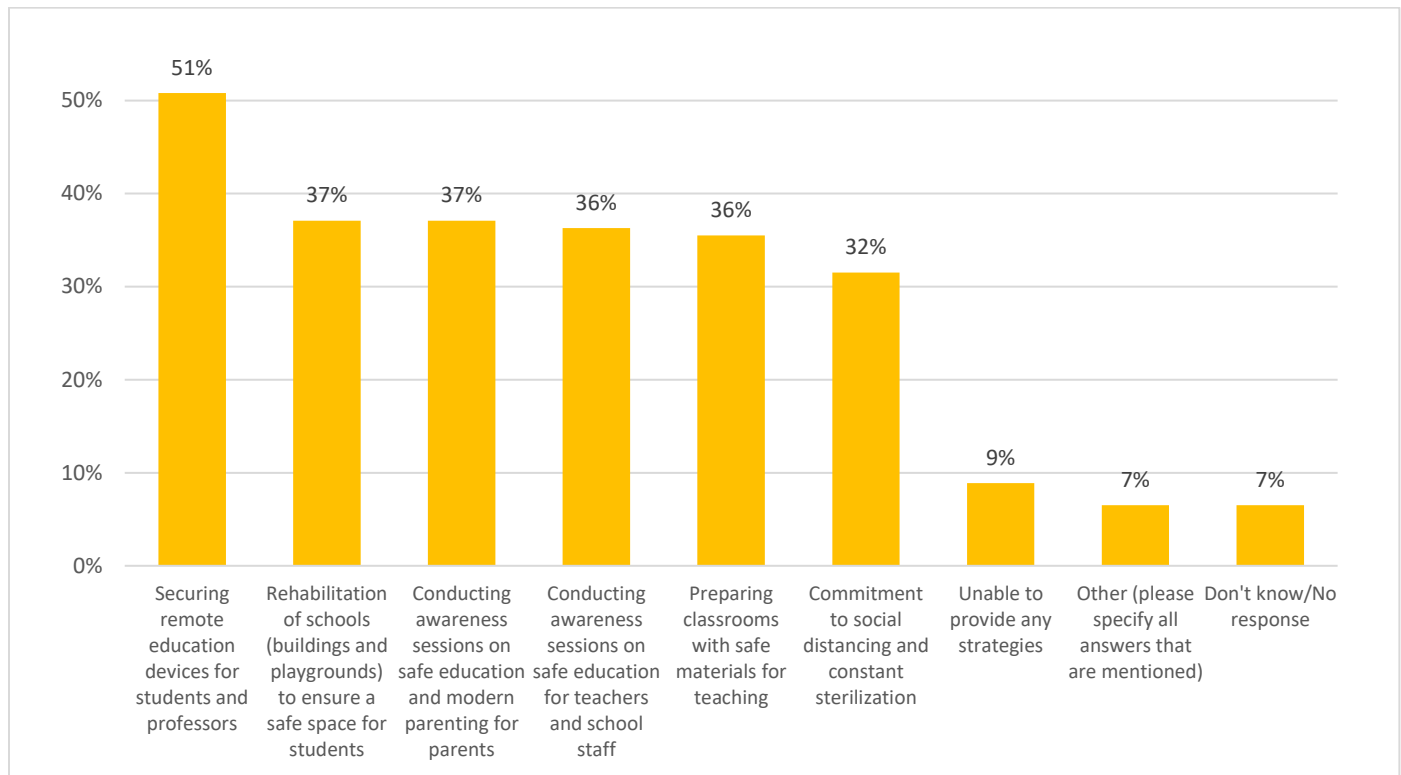
6.7.2 What is the baseline capacity (measured by knowledge, attitudes, and practice) of community members in creation and maintenance of safe learning environments?

These questions were not asked to parents in Lebanon at baseline.

6.7.3 What barriers do community members face in creation and maintenance of safe learning environments, despite knowledge and attitudes?

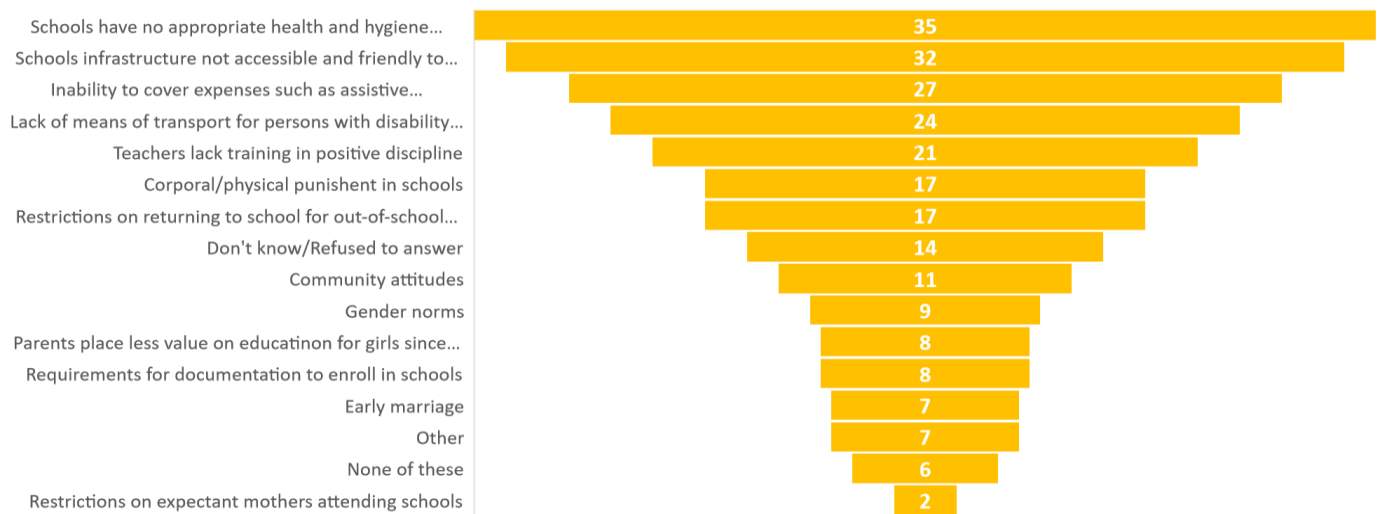
Parents were asked what strategies learning centers could use so that students are not in danger of physical violence or verbal abuse, where they feel safe, happy, and relaxed, so that they participate and learn better. More than one response could be recorded. Results are presented below.

Figure 52: Parents' self-reported strategies learning centers can use to make the environment safer for children



Parents were also asked about barriers to Right to Play's mission to build more inclusive learning, with equal participation for all genders and disabilities, in their community. More than one response could be recorded. Results are presented below.

Figure 53: Parents' self-reported barriers to Right to Play's mission



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

These questions were not asked to parents in Lebanon at baseline.

6.8 LEAVING NO ONE BEHIND

6.8.1 What are the beliefs held by parents, teachers and principals regarding gender and disability and access to education?

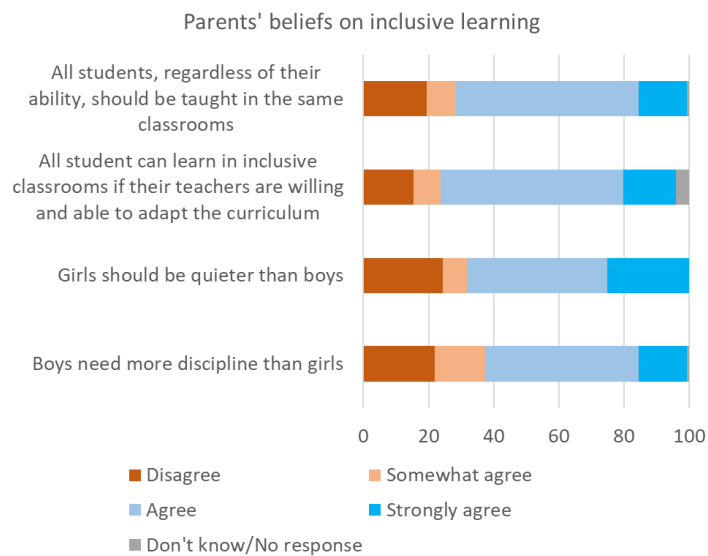
There is broad support for parents for supporting inclusive teaching at baseline. 88% of parents stated that it was equally important for boys and girls to have good reading skills, while 11% said it was more important for girls.

Parents did reveal gendered attitudes towards classroom behaviors of girls and boys. Nearly two-thirds of parents (62%) agreed or strongly agreed that boys needed more discipline than girls, while more than two-thirds (70%) agreed or strongly agreed that girls should be quieter than boys. There is further evidence of parents' attitudes towards gender roles and behaviors in the widening gap between girls' and boys' before and after school activities from Grade 1 to Grade 4, with girls increasingly likely to do chores and boys more likely to work outside the home.

6.8.2 How much do all the actors support inclusive teaching versus differentiated teaching for CWD?

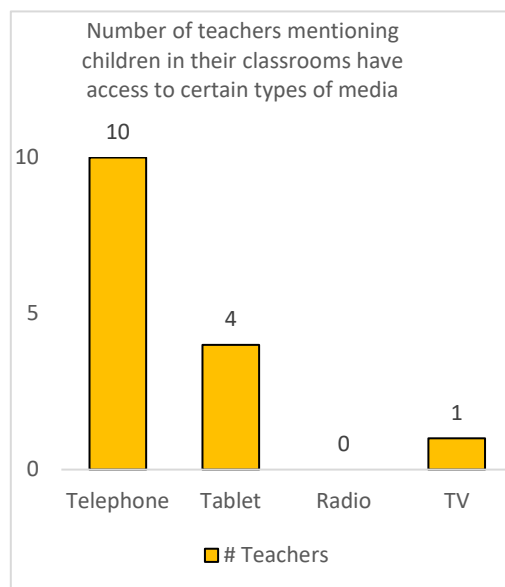
Over 70% of parents agreed or strongly agreed that all students should be taught in the same classrooms regardless of their ability, and 72% agreed or strongly agreed that all students can learn in inclusive classrooms if their teachers are willing and able to adapt the curriculum. However, 65-70% of parents also agreed with statements confirming traditional attitudes towards gender roles in the classroom.

Figure 54: Parents' attitudes and beliefs towards inclusive learning



6.9 INNOVATION LABS AND ICT USAGE

Figure 55: Teachers' reporting of selected media types in the classroom



At baseline, 7 of 10 teachers had never heard of Innovation Labs and the associated multilingual access platforms. Two teachers had heard of these and said some students were participating. Based on these responses, there is not enough information to answer the remaining evaluation questions about Innovation Lab usage at baseline.

At baseline, all 10 teachers said that children use phones, tablets, television, or radio at home to access educational resources. Teachers estimated that 98% of their class, on average, had access to at least one of these media. However, only phones were mentioned by all 10 teachers. All 10 teachers agreed that students with access to these materials performed better than students without.

6. KEY FINDINGS AND RECOMMENDATIONS

7.1 EGRA Arabic scores are relatively high at baseline, but averages mask that 10% are struggling

For learning outcomes, Lebanon learners have relatively high levels of Arabic and by Grade 4 some are achieving the maximum score on some EGRA tasks which will make it harder to show improvement at midline and endline (ceiling effect).

We recommend lengthening the Task 4 reading passage to 100 or more words at midline and endline so learners can show improvement beyond the ceiling of the current assessment. Analysis at midline and endline would need to draw a fair comparison accounting for any adjustments to these tasks.

This being said, the baseline survey identified that around 10% of learners are struggling with the basics of literacy, which may hold them back on more complex tasks. The country office chose not to administer more basic Task 1 and Task 2 to Grade 4 learners in the interest of saving time on the survey, but this may have hidden that some Grade 4 learners still have not mastered the fundamentals of letter names or sounds which are holding them back on the more advanced tasks.

We recommend that RTP consider whether they can deliver remedial content targeted at the bottom 10% of learners in Lebanon, as this is a way of boosting non-zero EGRA scores at midline and endline.

We recommend that RTP identify and focus on underperforming students –the 10% or so who are struggling with Arabic literacy basics.

7.2 Children who do not correctly report harmful behavior also have the lowest Life Skills scores

We found that one question, where students are asked what they would do if themselves or a friend were harmed and to which they can either answer to “tell a trusted adult” (which stands for the “correct answer”), “don’t know” (which stands for the “wrong answer”) or chose not to respond, was the best marker of low life skills scores. This question was significantly positively correlated with all four life skills measures, $p < .01$, with students answering correctly scoring 12-16% higher on the life skills items than those who did not. 64% of Grade 1 and 40% of Grade 4 students did not correctly report what to do when they or a friend were harmed in learning centers, and these children are lower performers across all Life Skills scores.

It is not clear whether this is a marker of overall childhood development or there is a causal connection, but we recommend using it as a quick proxy to identify which students are the most likely to have the lowest Life Skills scores. It can be used as part of ongoing monitoring efforts and whenever some discussions are led with students to automatically flag and target the likely lowest performers.

7.3 Life Skills levels can benefit from specific approaches for high and low-scoring students

At baseline, before the intervention has started, Grade 4 Life Skills scores are already mostly above 50%. Although the PMF has set a target of an average score, this focus masks different needs of high and low scoring students.

We found several other demographic markers for lower life skills scores within grades:

- In Grade 1, boys had significantly lower communication skills (2.4 points) than girls (3.6 points), $p < .05$, and children with an identified or self-reported disability had significantly lower conflict skills (0.5 points) than those without (1.0 points), $p < .05$.
- In terms of activities, Grade 1 children who reported doing chores in the morning or afternoon had significantly lower concentration scores (1.8 to 2.1 points) than those who did not (3.0), $p < .05$, while children who reported playing with friends in the morning had significantly lower collaboration (1.3 vs. 2.0) and concentration scores, (1.8 vs. 2.9), $p < .05$.
- In Grade 4, children with disabilities had significantly lower scores versus those without on collaboration (1.4 vs. 2.4) and communication (2.9 vs. 4.4), $p < .05$.

- In Grade 4, children who reported living in a home where people used physical aggression towards each other had higher scores versus those who did not on collaboration (2.9 vs. 2.2) and communication (5.8 vs. 4.1), $p < .05$, while children who reported someone being away for six months or more had significantly lower scores on conflict compared to those who did not (0.8 vs. 1.7), $p < .01$.
- There were no reliable socio-economic scale (SES) score predictors for life skills for either grade.

Moreover, we found that life skills scores were not always normally distributed. Some distributions had high and low scoring clusters (See Section 6.3, e.g. Grade 4 critical thinking, Grade 1 collaboration, Grade 1 Concentration).

We therefore recommend that RTP consider the idea of cooperation, where peer work and “mentorship” could be organized between higher- and lower-scoring students to support development. This approach will also promote collaboration, which is one of the life skills supported by RTP in Lebanon.

7.4 Availability of selected reading materials at home is significantly correlated with attendance of extracurricular literacy activities

Although there is some discrepancy between learners and parents in the reported presence of these materials in the household, it seems that most homes do not have magazines, newspapers, storybooks, coloring books, or comics in the household.

The baseline found evidence of significant positive correlations between some home reading materials and attendance at literacy extracurricular activities. There was some evidence that students reporting storybooks, comic books, or magazines in their home were more likely to have attended literacy extracurricular activities, and that students who reported having storybooks or coloring books had higher EGRA and life skills scores than those who did not (Section 6.4.5).

We recommend that RTP explore whether there is a causal relationship between providing child-friendly reading materials in the home and increased attendance at extracurricular literacy events, increased life skills and EGRA outcomes. If shown, providing students with child-friendly reading materials may have a significant impact on their life skills and reading scores at midline and endline.

7.5 Cultural attitudes towards girls may hinder RTP’s implementation and further gender stereotypes

The baseline survey highlighted traditional attitudes among the target population, particularly along gender lines. The survey showed a clear shift in expectations around girls’ and boys’ daily patterns of activities outside of school. Whereas in Grade 1 there were no significant differences between behaviors of girls and boys, by Grade 4 girls were around one-third as likely to report they played

with friends in the morning, almost twice as likely to report doing housework in the afternoon, and never reported working outside. Validation activities confirmed that due to increasing concerns about safety, parents are less likely to let girls out of the house. With good intentions, these parental attitudes may risk retrenching behavioral differences between boys and girls, which may among other things affect outside play.

In Section 6.8.1, we found that 60-70% of parents hold traditional gendered attitudes towards boys and girls in the classroom. These attitudes may present a challenge to Right to Play's model of inclusive play, especially outside the home.

We recommend that Right to Play reassess the potential cultural barriers to implementing its model in a relatively conservative Arabic target population, and consider the extent to which it will need to challenge and change existing norms - with the resistance that comes with it – rather than accommodate them.

7.6 Parents' attitudes are key to inclusive classrooms, and an opportunity to seize for RTP

Parents mostly display inclusive attitudes towards gender and disability, but also show agreement with traditional directive attitudes towards learning as opposed to RTP's playful constructivist approach.. We recommend involving parents as much as possible when advocating for change, as they are ultimate decision makers when it comes to children dropout and inclusive access to education.

Indeed, research shows that reading to children 1) makes them feel safer and more secure (which ties in with the previous finding on child protection), 2) improves future reading skills, given listening is a precursor for reading, developmentally, and 3) improves interaction, which, like cognitive priming techniques (asking what will happen next, etc.) are great predictors of ability scores.

We also recommend that Right to Play focus on measuring parents' and teachers' attitudes towards directive versus constructivist teaching methods (Section 7.4) and attitudes to boys' and girls' roles both inside and outside the classroom. We believe that acceptance or resistance of the Right to Play model by both parents and teachers will be an important factor determining the uptake of the model at midline and endline.

7.7 Teachers self-reporting usage of playful methodologies should be triangulated using classroom observations

Classroom observation data were not available inclusion in the baseline report, as due to COVID and ongoing security issues in Lebanon classes were being held mostly virtually. However, InformEd's

previous evaluation of the Palestine program called the validity of teachers' self-reported measures of using playful methodologies into question. Data from the baseline teacher surveys suggest that already nearly all teachers consider multiple integrated learning criteria for reading materials, display strong knowledge and adherence to play-based learning techniques, and report all safety concerns. If so, then this undermines the case that further support is needed. However, these responses may be subject to a strong positive response bias as they were aware enumerators were acting on behalf of the RTP program. In practice RTP's playful and constructivist methods may go against a hierarchical and traditional structure within the established educational system. Teachers from the teachers' survey were relatively young (average age 29 years), female (80%), and less experienced (average 8.5 years as a teacher and 4.8 years at the center).

We recommend that Right to Play conduct in-person classroom observations of all teachers included in the teachers' survey when it is feasible to do so and match the data to see if observed playful practice matches self-reported practice.

7.8 There is evidence that around half of children with disabilities may drop out from Grade 1 to Grade 4 and that teachers may need help identifying children with disabilities in their classroom

The percentage of self-reported CWD in random samples of Grade 1 and Grade 4 classrooms suggests that around half of CWD may drop out from Grade 1 to Grade 4. This should be confirmed through other methods and RTP should then consider how the program can address reasons for dropout.

The number of children who self-reported having "a lot of difficulty" or "cannot do at all" any of the six Washington Group questions is five or six times the number of learners identified by teachers. This means that teachers may need support identifying children with disabilities in their classroom, and especially recognizing 'invisible' disabilities such as speech or hearing impairments, which were not identified by any teacher.

Right to Play has focused on children with disabilities and whether they are disadvantaged on learning outcomes or life skills at baseline. However, the limited number of children with disabilities in the sample means that quantitative analysis may not be the best method for this. In fact, baseline survey data seemed to suggest that children with disabilities may perform better and show more resilience than their peers, at least in Grade 1.

We recommend that Right to Play explore drivers for children with disabilities dropping out of school through qualitative and participatory methods, so that RTP can understand what the program can do to address their needs.

Annex 1 – PMF Baseline indicators – Lebanon



Annex 1 - EQIE
Baseline PMF_Final - L

Annex 2 – Terms of Reference



Annex 2 - Terms of
Reference.pdf

Annex 3 – Analysis framework



Annex 3 - Analysis
framework.docx

Annex 4 – ISELA competencies mapping

Competencies mapped by the International Social and Emotional Learning Assessment (ISELA), developed by Save the Children.

Collaborative for Academic, Social, and Emotional Learning (CASEL), 2017		Skill included in ISELA (name of ISELA task)
SEL Competency Area	SEL Skill	
Self-Awareness	- Identifying emotions - Accurate self-perception - Recognizing strengths - Self-confidence - Self-efficacy	- Accurate self-perception (Self-concept) - Self-efficacy (Self-concept)
Self-Management	- Impulse control - Stress management - Self-discipline - Self-motivation - Goal setting - Organizational skills	- Stress management (Stress management) - Self-motivation (Perseverance)
Social Awareness	- Perspective-taking - Empathy - Appreciating diversity - Respect for others	- Perspective-taking (Empathy) - Empathy (Empathy)
Relationship Skills	- Communication - Social engagement - Relationship building - Teamwork	Relationship building (Relationships)
Responsible Decision-Making	- Identifying problems - Analyzing situations - Solving problems - Evaluating - Reflecting - Ethical responsibility	Solving problems (Solving conflict)

Annex 5 – Baseline survey research instruments

				
EQIE Lebanon Learner Assessment	EQIE Lebanon Teacher Interview.xlsx	EQIE Lebanon Principal Assessment:	EQIE Lebanon Parent Survey.xlsx	EQIE Lebanon Learner Assessment.xl

Annex 6 – Inception report



Annex 6 - EQIE
5-country baseline inc