Return to Learning Evaluation Bekaa, Lebanon

Save the Children

Endline Report

September 2019

Allyson Krupar

Contents

ACKNOWLEDGEMENTS

We would like to thank staff at Save the Children US and Save the Children International, Lebanon for their coordination and support in developing and implementing the endline analysis described here.

We would especially like to thank Bouchra Hayek, Rita Mahrouk, Joyce Massoud, Khaled El Hajj Ismail, Tarek Daoud, Miled About Jaode, Claudia Pasotti, Martha Wilkes, and Ayman Ali Hassan for their support prior to and during the enumerator training in March 2019, midterm data collection in May 2019, and endline data collection August-September 2019.

Finally, we would like to thank the assessors for their contributions and dedication to working with children that helped shape the baseline described below: Anita Garabette Apartian, Ayat Abdel Hadi Khaldieh, Melissa Elias Moaakar, Fatima Mousallem Darwish, Jouhaina Hasan Al Khatib, Nadine Mahmoud Shmoury, Walid Ahmad Dandach, Amer Hisham Reda, Alaa Sami Abdel Rahman, Milda Jean Haber, Yousef Haidar Nseir, Georgette Tony Fren Jabour, Jiana Georges Nehme, Mohammad Ali Jarah, Sami Ahmad Kamar, Layal Elias Bader, Rania Jamil Al Mais, and Hussein Hasan Chames.

EXECUTIVE SUMMARY

The *Return to Learning* (RtL) program includes six weeks of content, two days a week, in non-formal learning environments. RtL aims to support continuous learning for those affected by violence and conflict and decrease the gap in time between displacement and access to learning opportunities in contexts of forced displacement. Working with facilitators who may have limited experience teaching children, RtL provides facilitators with support and tailored content to help children develop skills to return to non-formal and formal education opportunities. In June 2019, facilitators were trained on how to use an online dashboard which showed the average scores of children in their centers and of the ages with whom they worked for differentiated instruction. This report analyses how RtL supports learners through this tailored approach and what potential outcomes we can measure in the two instances of the program implemented in Lebanon in 2019.

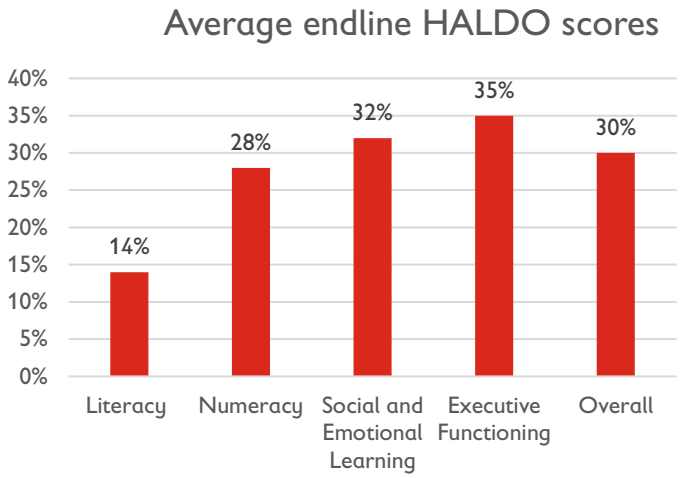
We evaluated the implementation and outcomes of RtL using the Holistic Assessment of Learning and Development Outcomes (HALDO), facilitator observations, and focus group discussions with parents, facilitators, and children participating in the project. HALDO is a rapid response tool for use in contexts of emergency and displacement. HALDO was developed through the RtL program as a means to profile learner populations and evaluate program outcomes. HALDO provides program staff with a profile of the literacy, numeracy, social-emotional learning (SEL), and executive functioning (EF) skills of children 4-12 years of age. In emergencies, many children are arriving with diverse learning needs and experiences. The results of HALDO can be used to inform decision-making processes by generating evidence upon which to base education response priorities, policies, investments, and discussions with host communities and donors in the first phase of a response. Directly, HALDO informs facilitators in RtL classrooms about what skill levels children are achieving in aggregate by age group and center so that they can more accurately select learning content based on the assessment results.

RtL's evaluation is a quasi-experimental design with two intervention groups. This endline report outlines results that compares HALDO scores between two cohorts: Cohort 1, who received the RtL intervention from March to May, and Cohort 2, who received a more tailored the intervention from June to July. Baseline data was collected with both cohorts in March 2019 before the intervention. The midterm data was collected with both cohorts in May 2019 after the end of cohort 1. We conducted a final evaluation in August to compare any gains or losses for these two groups. The data presented here includes results from HALDO, findings from qualitative focus group discussions with parents, facilitators, and children, and observational notes from the field (see Protocols in Appendix 2). Qualitative findings are integrated throughout HALDO findings to better contextualize results from the assessment. Quantitative findings outline the effect of the program using a difference-in-difference estimation technique to identify effect sizes. The mixed methods approach allows for a deep dive into the ecology and implementation of RtL.

FINDINGS

Endline data was collected in August – September 2019 to compare the results from children in the first cohort, who received the Return to Learning content in sequence with those in the second who received the content organized based on their outcomes from the HALDO assessment from June to July.

The endline HALDO sample included 368 children, 53% female and 47% male, ages 4 to 16. Focus group discussions at endline were conducted with 53 children (45% female), 34 parents (50% female), and 16 facilitators (75% female).



AVERAGE HALDO SCORE

The average child responded to 30% of total HALDO questions correctly at endline. The literacy subtask measured reading skills in Arabic with children answering an average 14% of literacy items correctly at endline. The average child answered correctly 28% of numeracy items, 32% of SEL items, and 35% of executive functioning items.

HALDO AND EQUITY

HALDO measures equity factors in four dimensions: sex, socioeconomic status, home learning environment, and inclusion/disability. Boys were more likely to have higher literacy scores than girls. Children with higher socioeconomic status (SES) had higher scores in literacy, numeracy and overall. The home learning environment was significantly associated with higher overall and SEL scores at endline. Specifically, seeing someone reading at home was highly correlated with higher overall HALDO outcomes, suggesting that programming could focus on home learning environments and literacy in the home to improve learning outcomes.

OUTCOME ANALYSIS

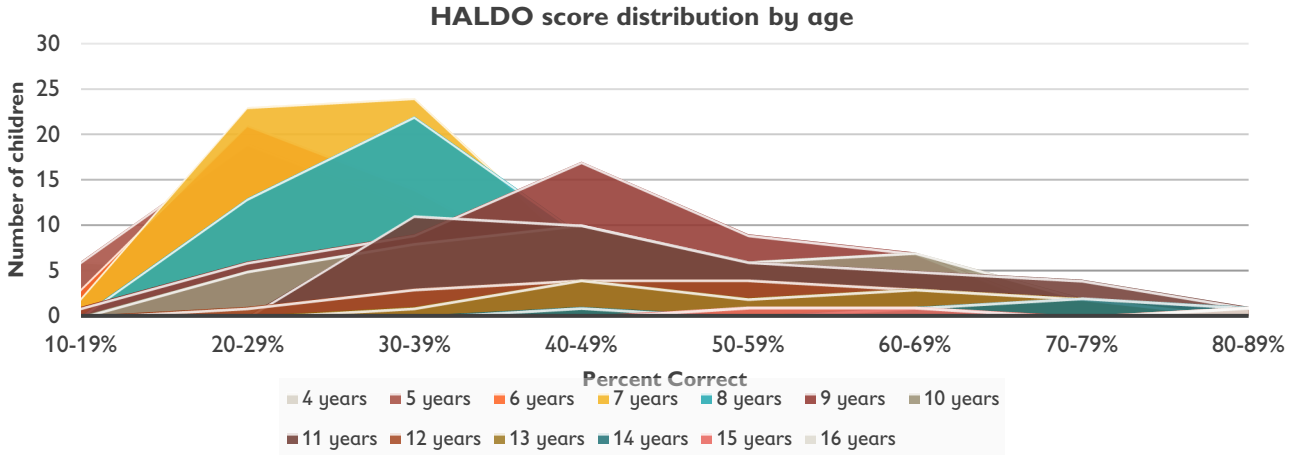
The main findings presented here identify the outcomes of the RtL treatment on HALDO outcomes. When we compare the difference between cohort 1's scores at midterm and cohort 2's scores at endline, we hypothesize that the intervention would increase children's scores in cohort 2 more than in cohort 1 related to project implementation and differentiated instruction. We see small gains across HALDO domains, with statistically significant increases in all domains and overall for children who participated in the 2nd cohort. However, when comparing to the gains we measured in the midterm for cohort 1, the revised content in cohort 2 was only related to increases in SEL. This signifies that RtL overall may be supporting children's learning but that the differentiated instruction did not show significant measurable learning gains in HALDO results.

RELIABILITY OF HALDO DOMAINS

We assessed two types of reliability: interrater reliability, or whether assessors agreed on responses when working in pairs, and internal consistency reliability, or whether items within each domain (literacy, numeracy, SEL, and EF) measured similar topics. Between assessors, we found strong levels of agreement for all HALDO domains. 32 children were assessed in pairs at endline (60 at midterm and 67 at baseline), presenting robust interrater reliability measures. For internal consistency, we found very strong reliability for the overall HALDO score as well as for the domain composites that we created. This suggests that the individual items in each domain reliably measure similar topics.

VALIDITY OF HALDO DOMAINS

Because we did not have other measures of children's development, we used the child's age as a proxy for predictive validity. We assumed that older children would have a higher HALDO score. HALDO clearly shows progression in average scores by age with 4 year olds scoring the lowest and scores gradually improving with age at endline.



LEARNING GAPS

The literacy outcomes in the sample are the largest learning gaps in the refugee response. One potential avenue of exploration from the data is looking specifically at literacy in the home learning environment, since seeing someone reading at home was highly correlated with increased scores overall. Literacy and executive functioning scores plateau for older learners, unlike numeracy and SEL, which increase more noticeably with age. Most importantly, we see limited gains between midterm and endline for children in cohort 2, suggesting that the participants in the Return to Learning content developed more literacy and executive functioning skills than the counterparts who did not participate in the program.

RECOMMENDATIONS

These findings can inform programming and implementation through focusing on variation and learning of children in specific ages. For ECD-age children (3-6 years), gaps identified in HALDO results suggest that programming could focus on:

- Literacy through exposure to reading materials, phonemic awareness, simple vocabulary, and alphabet games. Children under 10 emphasized the need to learn more letters in the Return to Learning content.
- Numeracy through awareness of numbers, simple operations, spatial concepts, patterns, amounts, and basic math vocabulary. Children under 10 in focus group discussions emphasized that they learned numbers 1-10 but did not mention any other aspects of math.
- SEL through understanding emotions in others, using role-play games and stories to give children an opportunity to start interpreting emotions in others. Children under 10 emphasized learning not to hit each other, to control one's volume, and to wait before speaking.
- Executive functioning through more memory-related games that are related to classroom routines and activities.

For early-grade children (7-10 years), programming could focus on:

- Literacy through letter identification, vocabulary, phonemic awareness, letter knowledge, and basic reading skills Children under 10 emphasized the need to learn more letters in the RtL content during focus group discussions. Children over 10 emphasized storytelling and writing.

- Numeracy through basic and advanced numeric operations and application of these operations to real life activities. Children under 8 in focus group discussions emphasized that they learned numbers 1-10 but did not mention any other aspects of math. Children over 10 emphasized learning measurements and using numeracy skills in the market to read expiration dates, add and subtract. Activities could focus more on real-life activities to support HALDO findings.
- SEL through understanding emotions in others, using role-play games and stories to give children an opportunity to start interpreting emotions in others, and ensuring games address interpreting ambiguous or violent behaviours in others. Children under 10 emphasized learning not to hit each other, to control one's volume, and to wait before speaking. Children over 10 also mentioned the need for more structure in the classroom to keep fights from breaking out and to keep each other from littering.
- Executive functioning through more memory related games that are related to classroom routines and activities.

For older-age children (11-16 years), programming could focus on:

- Literacy through reading with comprehension at a grade 2 level. Focus on supporting children reading skills and their comprehension ability. Children over 10 emphasized storytelling and writing during focus group discussions about RtL activities
- Numeracy through multi-digit operations and multi-step word problems related to real-life activities for the children. Children over 10 emphasized learning measurements and using numeracy skills in the market to read expiration dates, add and subtract. Activities could focus more on real-life activities to support HALDO findings.
- SEL through understanding and interpreting emotions of others, conflict resolution and relationship management, reviewing self-concept items. Children over 10 emphasized relaxation techniques and conflict resolutions skills during focus group discussions.
- Executive functioning through mindfulness activities that will increase child's concentration

Related to the feasibility of HALDO, recommendations include refining how training is focused and using peer learning tools to improve time utilization during training and piloting a paper-based version of the assessment. Validity tests show that the tool is valid as a measure of learning by age, but nuances between age outcomes are described in the overall findings.

BACKGROUND

WHAT IS THE PURPOSE OF THIS REPORT?

342,560[1] Syrian refugees live in and around Bekaa, Lebanon, a valley known for agriculture and situated on the Syrian border. Historically, this area has hosted Syrian migrants who were involved in seasonal agricultural work prior to the war.[2] Lebanon hosts around 1.5 million Syrian refugees, roughly a quarter of the national population, but national policies can make it difficult for refugees to legally work. The World Food Programme's 2017 vulnerability assessment found that 75% of Syrian refugee households have no access to basic food and shelter, and 58% are living in extreme poverty.[3] This vulnerability is compounded by lack of education opportunities, with more than half[4] of Syrian refugee children in Lebanon out of school.

To respond to this growing crisis, in 2013, Save the Children in Lebanon scaled up programming. Together with UN agencies and other INGOs, the country office supports education, child protection, shelter, water, sanitation and hygiene, and food security and livelihoods programming, particularly in Bekaa and other refugee and host communities. The Save the Children response in Bekaa has identified that out-of-school children have learning gaps that programming ought to accommodate. Providing RtL to these children generates evidence through HALDO and qualitative finds that are critical to understanding these gaps and informing future programming.

This report presents the findings from the endline evaluation of RtL, focusing on the results from HALDO and qualitative data collected with field staff in August to September 2019. The purpose of this report is to identify what the next steps are for the Return to Learning programming in 2020. In the next few sections, we present the study methods and describe the overall profile of learners. We summarize the impact of RtL, focusing on HALDO results, learning gains in the two cohorts, and implementation feasibility. We also briefly summarize the validity and reliability of HALDO to verify it's robustness as a tool monitoring literacy, numeracy and SEL. The final section outlines the next steps for Return to Learning in Lebanon.

METHODS

WHAT WERE OUR RESEARCH QUESTIONS?

The purpose of this research is to answer the following research questions.

1. How has the sample of children changed over time?
    a. Are the children in cohort 1 who were able to be found at midterm and cohort 2 who were able to be found at endline different than those who were not able to be found? If so, how?
    b. Did the attrition rate differ between cohort 1 and cohort 2 children?
2. Did RtL participation exhibit impact on children's learning and development skills for cohort 1 in comparison to cohort 2?
    a. Does this impact result in more equitable outcomes for traditionally disadvantaged groups?
3. How do changes in children's development and learning vary by home learning environment and/or classroom learning environment/center?

[1] https://data2.unhcr.org/en/situations/syria/location/71
[2] https://gh.bmj.com/content/4/1/e001122#ref-6
[3] https://www.wfp.org/content/2017-vulnerability-assessment-syrian-refugees
[4] https://www.unhcr.org/lb/education

   a. What skill areas are identifiable in each center to inform facilitation?
4. Have facilitators advanced active, child-centered teaching practices in RtL?
5. What are the psychometric properties of the HALDO tool?

As the first attempt to identify the utility of HALDO as it measures learning over time, this report explores how children score on specific measures and focuses on children's performance by age.

WHAT WAS THE SAMPLE?

The Lebanon endline sample included 368 (53% female) children from 5 centers serving 5 communities:

- NABAD Association Center, Ali Nahri community, 25 children (44% female)
- SAWA Association Center, Mansoura community, 156 children (52% female)
- El Khiara Municipality Center, Khiara community, 67 children (46% female)
- Cultural and Social Council Center, Bar Elias community, 76 children (62% female)
- Qabelias Center, Qabelias community, 44 children (55% female)

Of the 600 out-of-school children that comprised the sample, 300 were randomly assigned to cohort 1 (to receive the intervention first) and the remaining 300 were assigned to cohort 2 (as a wait listed comparison group scheduled to receive the intervention after the midterm). Children in the same family were identified by their caregivers' phone number and placed in the same cohort based on feedback from the field staff. Prior to finalizing the sample, we evaluated if age was significantly different between each cohort and found that there were no significant differences by cohort and age, suggesting that although the random samples were not restricted by age, the age distribution in each cohort is comparable. The table below presents the sample tracked between baseline to midterm and midterm to endline.

	Baseline	Midterm	Endline
Cohort 1	300	262	97
Cohort 2	300	254	271
Total	600	516	368

97 children were tracked from midterm to endline in cohort 1 and 271 from cohort 2.

HOW WERE THE DATA COLLECTED AND ANALYSED?

9 enumerators were maintained from the originally 18 trained in March 2019. Enumerators assessed children one-on-one and in pairs using the HALDO tool programmed on tablets on the KoboToolbox platform. Paired assessments were used to identify interrater reliability, discussed in more depth below. 32 children were assessed in pairs (9% of the sample).

During data collection, data were uploaded to KoboToolbox and reviewed by staff in Bekaa and Washington, DC. Enumerators were provided ongoing feedback to address any discrepancies in interrater reliability during data collection and clarify remaining questions from the training. After data collection was complete, the dataset was analysed in Stata to identify the psychometric qualities of the tool and compare children's outcomes by cohort, age, sex, camp residence, SES, and home learning environment. Data analysis focused on identifying frequencies and distribution by camp, age, and sex then conducting multivariate regressions to analyse key equity indicators. The psychometric qualities of the tool were explored through analysis of intra-class coefficient (ICC) and kappa statistics to understand the difference between enumerators and internal consistency reliability was analysed using Cronbach's alpha. Without a secondary validated measure of children's learning, HALDO's validity was established by using children's ages as a proxy for predictive validity.

WHAT DID THE HALDO TOOL INCLUDE?

HALDO assesses a range of learners' skills, including literacy, numeracy, social and emotional learning (SEL), and executive functioning (EF) to identify potential diversified learning approaches and programs in emergencies. Assessors administered HALDO one-to-one with a child between 4-16 years of age. In Table 1, we present the domains measured in HALDO as well as the individual items in each domain. To identify overall skill levels by center and to target programming in cohort 2, we created four skill level categories, Emergent, Foundational, Intermediate and Advanced. The below table includes which questions fall within each skill level. In SEL, the skill levels were determined by overall outcome scores rather than individual questions.

Table 1. Domains, skills, and item descriptions from HALDO

Domain	Skill	Items	Skill Level
Literacy	Letter Identification	Identify 5 letters common letters	Emergent
	Expressive language	Name 10 animals (only asked if child cannot identify any common letters)	
	Letter Identification	Identify 5 letters infrequent letters	Foundational
	Accuracy	Number of words read correctly	Intermediate
	Reading with comprehension	Respond to 5 comprehension questions	Advanced
Numeracy	Number identification	Identify 5 single-digit numbers presented visually	Emergent
	One-to-one correspondence	Understand concept of different numbers related to objects (3 items) (only asked if child cannot identify any single-digit numbers)	
	Number identification	Identify 5 double-digit numbers presented visually	Foundational
	Simple Operations	Complete 5 simple numerical operations with single-digit numbers	Intermediate
	Hard Operations	Complete 5 harder numerical operations with two-digit numbers	Advanced
	Word Problems	Complete 2 numerical problems from a verbal word problem	
Social Emotional Learning	Self-Concept	Knowledge of name, age, sex, community name, country name	< 0.25 = Emergent
		Ability to identify positive hope(s) for future, what could support and stop this future	0.25 - 0.5 = Foundational
	Empathy	Ability to identify how someone else might be feeling	
		Ability to show empathy	
		Ability to take the perspective of a third child in an ambiguous situation	0.5 - 0.75 = Intermediate
		Tendency to not attribute hostility to ambiguous provocation	0.75 – 1 = Advanced
Executive Functioning	Short Term Memory	Ability to remember 4 number sequences	
	Working Memory	Ability to remember and reverse 4 number sequences	

Besides the core developmental domains, HALDO allows us to collect information on some key demographic and home characteristics. This allows us to understand how children's learning and development is affected by the following equity factors: sex, socioeconomic status (SES, index of household wealth assets like refrigerator

and mobile phone), home learning environment (HLE, two questions about the presence of reading materials in the home and seeing someone read in the home), and disability status (three questions about challenges with vision, hearing, and physical movement[5].

FINDINGS

The findings below present the profile of learners, focusing on general demographic information by center, overall scores, and distribution of overall scores. The findings suggest that there are some differences between learners in each center, but the variations in sample size make it difficult to compare. We focus particular program recommendations by center without comparing centers and the feasibility of program implementation overall. We focus the findings on the impact of RtL on cohort 2 in comparison to cohort 1. We then present equity, reliability, validity, and feasibility analyses for overall scores and each subsequent domain (literacy, numeracy, SEL, and EF).

WHAT WERE THE DEMOGRAPHIC PROFILES OF LEARNERS?

HALDO allows us to collect demographic and household data on children so that program staff get a clear profile of the children with whom they are working. Because the children in each center have slightly different demographic backgrounds and socioeconomic statuses, focusing on programming alternatives in each center will allow the Bekaa response team to make strategic decisions about next steps in programming. There are some characteristics of the sample populations in each center, which may affect learning outcomes.

Table 2. Demographic characteristics of sample from the five centers in Bekaa

Variable	NABAD Association Center, Ali Nahri	SAWA Association Center, Mansoura	El Khiara Center, Khiara	Cultural and Social Council Center, Bar Elias	Qabelias Center, Qabelias	Overall
Sex (% female)	44%	52%	46%	62%~	55%	53%
Average age (years)	7***	8	9	9***	7	8**
Attended school before	52%*	21%***	46%*	46%*	23%	33%
Language						
Arabic (%)	100%	100%	100%	100%	100%	100%
Kurdish (%)	0%	0%	0%	0%	0%	0%
English (%)	0%	1%	1%	1%	0%	1%
Other (%)	0%	1%	0%	1%	1%	1%
Home learning environment						
Books present in home	36%	13%*	10%*	34%	39%	22%**
See someone read at home	28%	13%	13%	32%	20%	19%
Inclusion/disability	12%	14%	7%	7%~	15%	15%
Sight	4%	6%	4%	1%	11%	6%

[5] WGDS. (2017). The Washington Group Short Set on Functioning. Retrieved from http://www.washingtongroup-disability.com/wp-content/uploads/2016/12/WG-Document-2-The-Washington-Group-Short-Set-on-Functioning.pdf.

Hearing	12%	2%	3%	0%	5%	3%
Physical	12%	11%	1%	4%	5%	8%
Socioeconomic status[1]	38%	34%***	36%	48%***	36%	38%
n	*25*	*156*	*67*	*76*	*44*	*516*

Differences significant at p<0.001(***), p<0.01 (**), p<0.05 (*) and p<0.10 (~) – No significant differences found.

Overall, there were many more children sampled in Mansoura and Bar Elias. When comparing by center, there are significant differences between the sample of children that may be relevant for programming and to better understand each center's results. For example, children in Bar Elias were significantly older than children in the other centers. Children in Ali Nahri were more likely to have attended school before especially compared with Mansoura who were less likely.

6 and 7 year olds are a bit overrepresented in the sample (see Figure 1).[6]

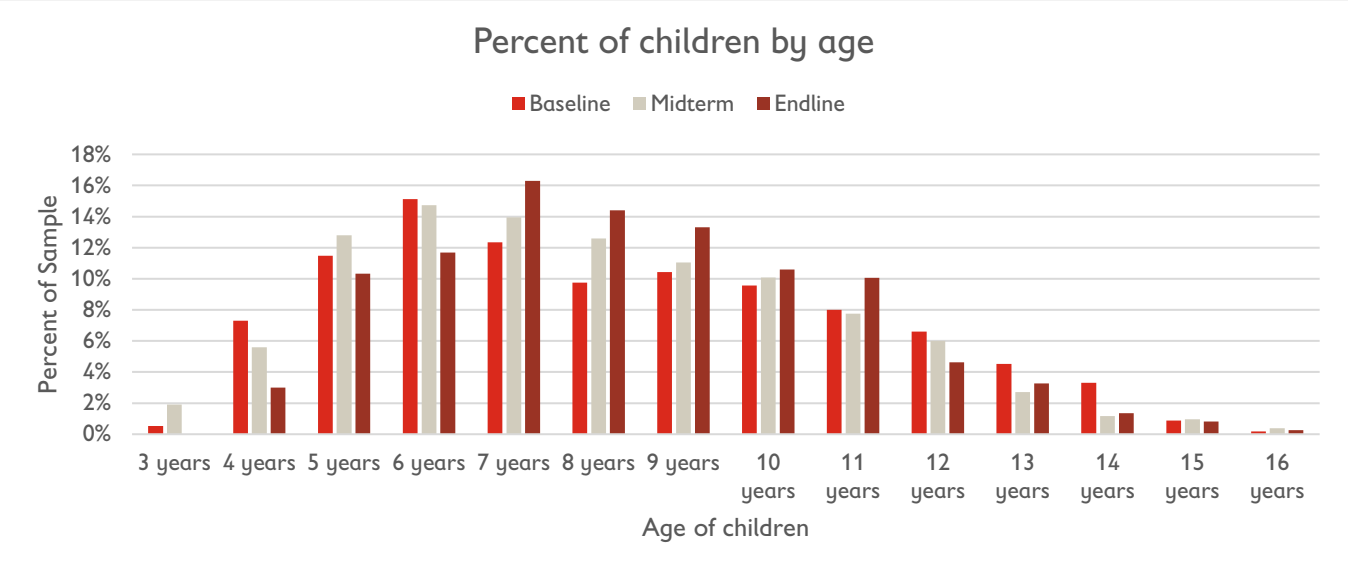


Figure 1. Distribution of children by age in Lebanon baseline sample (n=575, baseline. N=516, midterm. N=368, endline.)

WHAT WAS THE DISTRIBUTION OF HALDO SCORES?

For each HALDO domain—literacy, numeracy, SEL, and EF—we created a composite score that reflected the number of items that a child answered correctly in that domain. We present the distribution of these domain scores versus the age of the child in Figure 2.

We observed a floor effect for the literacy and numeracy scores: a large proportion of children could not respond to any of the literacy or numeracy questions. The SEL and EF skills were more evenly distributed, but the distribution was still skewed to the left, as visible in Figure 2.

[6] In future analysis, to compare children in this sample to the larger population in Save the Children ECD, primary schools, and AEPs, each age range could be assigned a weight reflecting the proportion of the larger population represented. These weights would allow for an analysis that can be generalized to the overall population of children in Save the Children programming. This representation can inform programming, but the findings section below focuses on the unweighted population to identify specifically the strengths and gaps in the HALDO tool's assessment.

While the floor effects are important to consider in any predictive models, the overall distribution of scores in the four domains illustrates that HALDO captured a very large range of skills for children in the refugee response in Lebanon.

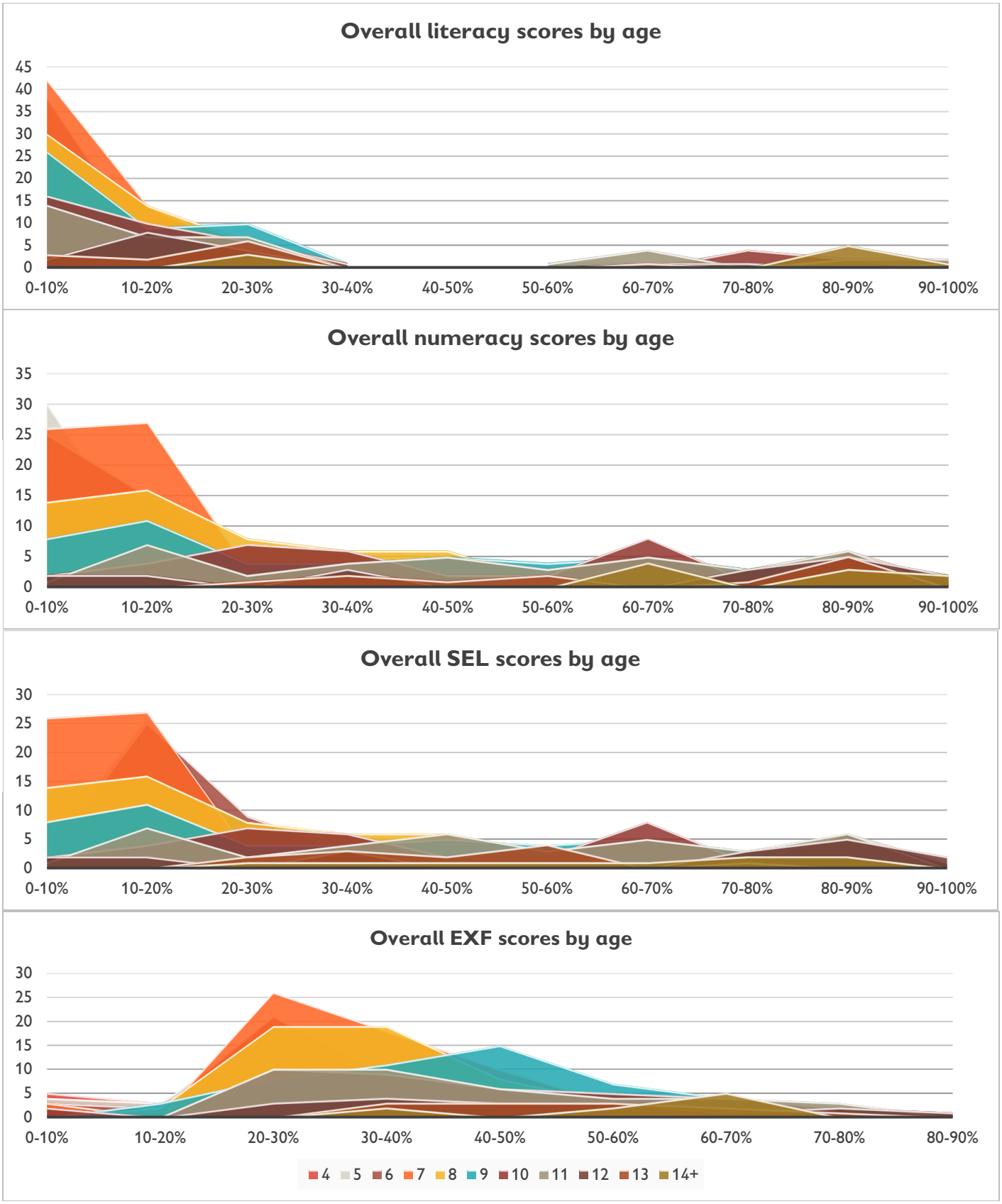


Figure 2. Distribution of HALDO literacy, numeracy, SEL, and EF composite scores versus child age

In Figure 3, we present the percent of questions the average child answered correctly in each HALDO domain as well as for the total assessment. On average, children responded correctly to 30% of the total number of HALDO questions. SEL and EF had the highest average correct at 32% and 35% respectively and literacy was the lowest at 14%.

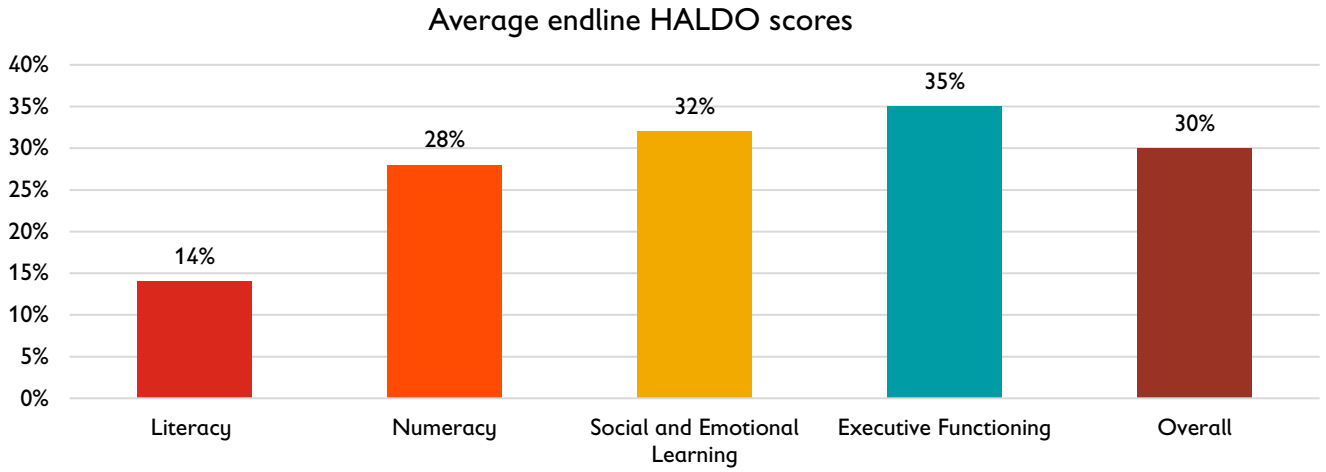


Figure 3. Percent of items average child answered correctly, overall and domain-specific

WHAT WERE THE LEARNING AND DEVELOPMENT PROFILES OF LEARNERS?

One of the main aims of HALDO is to provide program designers and front-line emergency staff with a quick picture or profile of the learning and development needs of the children with whom they are working. This allows them to make more informed programmatic decisions. Since the programming in the Lebanon refugee response is focused on out-of-school children and the program is being implemented in two cohorts, we disaggregated the domain scores for children into the four skill categories described above to easily identify where children may be struggling in cohorts. In Tables 3-5, we present a profile of the average child who has emergent skills, foundational skills, intermediate skills, and advanced skills with their average age. Each table also includes top-line programming recommendations for the country team to consider in program design conversations with their Education Technical Experts. These tables provide additional information to the Dashboard available here.

Table 3. Domain-specific skills for average ECD-age (3-6 years) child in Bekaa, Lebanon refugee response, and recommendations for future programming

Domain	Skill	Max	3-6	Programming Recommendations
Literacy	# of common letters identified	5	.35	Child has very few emergent literacy skills. Only 1 child in this age group could identify all 5 common letters and no child in this age group was able to read five words. Focus on emergent literacy skills for young children, like exposure to reading materials, phonemic awareness, alphabet games, and simple vocabulary
	# of animal names child can express	10	1.8	
	# of difficult letters identified	5	.1	
	Ability to read 5 words	1	0	
	Accuracy (% words read correctly by number of readers)	100%	0	
	# of comprehension questions answered	5	0	

Numeracy	# of single-digits identified	5	.5	Child has weak emergent numeracy skills. Only two children in this age group could do any operations. Focus on emergent math skills for young children like awareness of numbers, number identification, simple operations, spatial concepts, patterns, amounts, and basic math vocabulary.
	# of amounts identified	3	.8	
	# of double-digits identified	5	.5	
	# of simple operations solved	5	.01	
	# of harder operations solved	5	0	
	# of word problems	2	0	
SEL	# of self-concept items answered appropriately	12	3.1	Child can identify some basic self-concept items but is still struggling with understanding emotions in others. Use role-play games and stories to give children an opportunity to start interpreting emotions in others.
	# of empathy items	5	.24	
Executive Functioning	# of working memory items answered correctly	9	2.2	Child has some working memory skills. To boost their working memory play memory-related games and emphasize classroom routines.

Table 4. Domain-specific skills for average early grades (7-10 years) child in Bekaa, Lebanon refugee response, and recommendations for future programming

Domain	Skill	Max	7-10	Programming Recommendations
Literacy	# of common letters identified	5	1.7	Child has some emergent literacy skills but struggling with reading. Only 9 children in this age group could read the first five words of the passage. Focus on letter knowledge and basic reading skills.
	# of animal names child can express	10	3.6	
	# of difficult letters identified	5	1	
	Ability to read 5 words	1	.05	
	Accuracy (% words read correctly by number of readers)	100%	3%	
	# of comprehension questions answered	5	.19	
Numeracy	# of single-digits identified	5	2.5	Child has weak emergent numeracy skills. Focus on emergent math skills like double-digit number awareness, simple operations, spatial concepts, patterns, amounts, and basic math vocabulary.
	# of amounts identified	3	1.5	
	# of double-digits identified	5	3	
	# of simple operations solved	5	.9	
	# of harder operations solved	5	.2	
	# of word problems	2	.36	
SEL	# of self-concept items answered appropriately	12	4.7	Child can identify almost 50% of basic self-concept items but is still struggling with understanding emotions in others. Use role-play games and stories to give children an opportunity to start interpreting emotions in others.
	# of empathy items	5	1	
Executive Functioning	# of working memory items answered correctly	9	3.2	Child has some working memory skills. To boost their working memory play memory-related games and focus on classroom routines.

Table 5. Domain-specific skills for average 11-16 year-old child Bekaa, Lebanon refugee response, and recommendations for future programming

Domain	Skill	Max	11-16	Programming Recommendations
Literacy	# of common letters identified	5	3.3	Child has strong emergent literacy skills but struggling with reading with comprehension at a grade 2 level. Only 23 children in this age range were able to read the full passage. Focus on
	# of animal names child can express	10	5.2	
	# of difficult letters identified	5	2.5	

	Ability to read 5 words	1	.3	supporting children's reading skills and their comprehension ability.
	Accuracy (% words read correctly by number of readers)	100%	11%	
	# of comprehension questions answered	5	1	
Numeracy	# of single-digits identified	5	4.4	Child has strong emergent numeracy skills but struggling with double digit operations and word problems. Focus on number identification and using operations in real-world problems.
	# of amounts identified	3	2.7	
	# of double-digits identified	5	4.4	
	# of simple operations solved	5	2.3	
	# of harder operations solved	5	1	
	# of word problems	2	1	
SEL	# of self-concept items answered appropriately	12	6.5	Child can identify over half of basic self-concept items but is still struggling with understanding emotions in others. Use role-play games and stories to give children an opportunity to start interpreting emotions in others.
	# of empathy items	5	1	
Executive Functioning	# of working memory items answered correctly	9	4.2	Child has some working memory skills. To boost their working memory play more memory-related games based on to classroom routines.

WHAT WERE THE EQUITY CONSIDERATIONS IN THE ENDLINE?

When considering what programs would be most applicable for the refugee children in Bekaa, it is important to consider how the literacy, numeracy, SEL, and EF skills of children differ by key equity factors. HALDO allows us to disaggregate the domain and overall scores by equity factors to understand if some children are struggling and need additional attention.

In this endline, we looked at four dimensions of equity: sex, socioeconomic status, home learning environment, and inclusion/disability. We found significant differences (see Table 6) in the average child's overall HALDO and domain-specific scores across two of the four different equity factors.

One important finding was that seeing someone reading at home, which signifies literate family members, had a statistically and practically meaningful positive relationship with overall HALDO scores. These findings suggest that the home learning environment could be an important space for the Lebanon program to support children's learning and development. Encouraging home reading through family outreach, reading festivals, reading camps, and book banks in addition to more educational programming, could help children develop strong literacy and numeracy skills in the refugee response.

Table 6. Equity findings in HALDO Lebanon Endline[7]

	Sex	SES	Home learning environment		Vision, hearing, or physical disability
			Books in the home	Reading at home	
Overall	No relationship	Child with higher SES responded to 9% more questions correctly than their peers with lower SES	No relationship	Child who sees reading at home was able to correctly answer 4% more questions compared to peers	No relationship
Literacy	Girls had 3% higher scores on average on literacy questions	Child with higher SES responded to 22% more SEL questions than their peers	No relationship	No relationship	No relationship
Numeracy	No relationship	Child with higher SES responded to 23% more SEL questions than their peers	No relationship	No relationship	No relationship
SEL	No relationship	No relationship	No relationship	Child who sees reading at home was able to correctly answer 6% more SEL questions compared to peers	No relationship
EF	No relationship	No relationship	Child with books in the home was able to correctly answer 4% more EF questions compared to their peers	No relationship	No relationship

[7] Fitted estimates from regression models predicting overall and domain-specific HALDO scores by equity factors and age are reported in table 10 in the appendix.

HOW FEASIBLE WAS RETURN TO LEARNING TO ADMINISTER?

The main findings from the endline about feasibility related to implementing the Return to Learning content and using HALDO data to inform programming. This section focuses on data collected in focus group discussions (n=53 children, 34 parents, and 16 facilitators), facilitator observation (n=16), and field notes.

16 FGDs were conducted with children in four groups, males and females (separately) below 10 and above 10, with parents (mothers and fathers separated) and with male and female facilitators.

- **Children above 10** frequently brought up challenges related to balancing their work obligations and schooling. In cohort 1, several boys reporting that if they had to decide between work and school, they would finish their work first before going to school. Boys in cohort 2 similarly did not see school as a priority and several reported that they do not use reading or math in any ways outsides of school. In cohort 1, this age group named specific activities as their favorite aspects of the program, including the "Giraffe and lion stories", "Keeping secrets", "Who do we love", lessons and relaxation techniques, conflict resolution activities, how to respect each other, and how to apply math and literacy outside the classroom. In cohort 2, the older children wanted to focus more on literacy skills, including in French, and drawing.
- For **children under 10**, challenges in cohort 1 focused more on the environment within the classroom, where facilitators had trouble with classroom management. In cohort 2, facilitators participated in further training on classroom management but some girls still found boys to be bossy and bullying towards them. Outside the learning environment, children mentioned that the bus ride was long and they had work to do at home. Many of these children in cohort 1 and 2 learned the letters A and B and numbers 1 and 2 in Arabic, but wanted to learn more letters and numbers. During conversations with field staff, this was a point that the education assistants and facilitators raised as well – that the curriculum did not include the whole alphabet. Children specifically liked playing with clay and thought of how their parents help them to learn to pray and support them in school
- Many of the **parents** who participated in the focus groups were illiterate and had no suggestions on how to support their children's learning at home other than getting them enrolled in schools. Parents in cohort 1 also recognized that their children start to work around the ages 7 or 8 which affects their schooling. Fathers in both cohorts were not convinced that RtL was satisfactory in helping their children improve their literacy and numeracy skills both due to length and the game-based nature of activities. Overall, parents in both cohorts emphasized that they wanted their children to have opportunities beyond agricultural work.

Facilitators both participated in FGDs and classroom observations in August and September. Classroom observations address whether facilitators are using active, child-centered pedagogy in the classroom. 16 observations were conducted over the course of 9 days from July 29 to August 7. Observations identified that most children were on-task and the facilitators were using a diverse set of pedagogical tools (group work, presentations, individual work, etc.) during their facilitation. This signifies that facilitators were attempting to put children at the center for their teaching. After participating in classroom management training between cohort 1 and cohort 2, facilitators had less classroom management issues, with more orderly classrooms observed. Overall, facilitator observations noted that cohort 2 facilitation was more child-centered but that facilitators needed further active learning, inclusive learning, and SEL training to better work with children in the program.

Program implementation trends evident in the facilitator observations can also identify other variables that may affect learning outcomes and instruction. For instance, overall, 81% of children enrolled were in attendance during facilitation observation, 83% of boys were on average in attendance and 80% of girls. Attendance was not correlated with HALDO outcomes but could explain variations in the data.

In FGDs, facilitators reflected on their work and emphasized how they had learned how to cope with classroom management challenges from cohort 1 to cohort 2. The familiarity with the content and classroom environment supported there comfort in the program. Although the curriculum was beneficial on the social and emotional learning, facilitators felt it lacked enough time and focus on literacy and numeracy to effectively increase children's learning. They also found emergent and foundational activities to be age-inappropriate for older learners who scored at these skill levels on the HALDO assessment.

The below recommendations were identified with staff and facilitators for future RtL implementation in Lebanon.

Programmatic recommendations from the field:

- Given that facilitators were not trained teachers, RtL should ensure that **facilitators have additional training opportunities,** including classroom management, active learning, inclusive education, and SEL training (such as HEART).
- During the facilitation and during assessment, **education assistants acted as support at each center** for facilitators and enumerators. These staff are crucial to project success and should be further trained on how to use the RtL tools.
- Final results can be used for **advocacy** purposes to address more funding, programming and policy to increasing educational access for refugees in Bekaa.
- **Further mobilization** and sensitization with the community before the project begins would support child recruiting and ensure that any siblings or other children understand why they can or cannot participate.

Content recommendations from the field:

- Save the Children staff found the RtL content to be too short to build skills in this context where many children have been out of school for a long time or never in school. Children were lacking emergent literacy and numeracy skills, so more skill focused activities were recommended. HALDO results supported tailoring the learning content, but emergent and foundational activities were not inclusive of older learners. Content should be further developed to include more skill development and focus more on inclusion of overage children.
- Facilitators also highlighted that the content could be more in-depth and the program longer than 2 hours a day for 6 weeks. Parents, for example, did not understand why the program was so short and had difficulty envisioning what their children could learn in that time. These recommendations could lead to **more activity development within RtL** as well as increased focus on **more community outreach** to support parents' engagement with the project. One proposed solution to this was to engage CO staff in developing content themselves, as well as drawing on the materials already developed.
- Staff suggested connecting RtL content with **PSS and PFA** programming so RtL can support other learning outcomes.

MEAL recommendations from the field:

- MEAL capacity building support should include **step-by-step instructions** on how to run the data analysis, program management, enumerator training, and dashboard generation connecting Kobo to PowerBi. The RtL MEAL steps should be written out as methods so that the CO can use these steps for future project management. SCUS has begun developing this tool for further review and refinement with the CO.

- The **Dashboard could link directly to the learning content** in high resource environments so the facilitator has both electronic and paper copies of the content.
- Future programming could **track lesson use** by creating an app with content and results. This app would need to be able to work off and online to sync results. This is a possible avenue for future development in high resource settings.

Next Steps:
To conclude the lessons learned, there are three main programmatic next steps that need to be pursued:

1. **ECCD tracking**: 100 of the cohort 1 children who are between 3-6 years old will be placed in an ECCD program beginning June 24 and ending in late September. The ECCD program will use the CB ECE curriculum and associated national school readiness pre and post tool to qualify for an MEHE certificate. Data from these children will need to be tracked systematically and compared to other children who did not participate in RtL.
2. **BLN programming**: As of writing, funding for further BLN programming in Bekaa is unknown. Children over the age of 6 have no further educational programming to attend after completing RtL in cohort 1 or 2. Future programming should consider this particular population.

IMPACT ANALYSIS

Student & Household Characteristics

Table 3 displays the proportions of the sample by student background characteristics, condition, and phase. At endline, the average age of RtL students (cohort 2) was 8, which was significantly different than the children from cohort 1 who were assessed at endline. At midterm, the average age of RtL students (cohort 1) was 8.26 years of age, while for the comparison group (cohort 2) it was 8.06. At baseline, the average student age was 8.21. The sex composition of both condition groups sampled at midterm and endline are statistically the same.

The impact analysis only includes children who could be identified at both midterm and endline. Overall, there was 39% attrition in the sample from midterm to endline due to limited sampling from cohort 1. Cohort 2 attrition was only 7%. Children with lower SES were significantly more likely to leave the program. Broken down by age, 7, 8, 9, 12, 13, and 15 year olds were more likely to leave the program and 4 year olds were less likely. This confirms the qualitative responses from the focus group discussions where older children discussed the difficulty in deciding whether to go to work or to school.

Table 7: Student background characteristics at endline

	Intervention Group	Comparison Group	Total (endline)	Total (midterm)	Total (baseline)
Female (%)	53%	53%	53%	50%	47%
Age (years)	8***	9.25***	8.33	8.04	8.21
Attended School Before	31%	40%	33%	15%	10%
Disability	15%	16%	15%	13%	20%
SES: sum of 10 common household items	3.81	3.63	3.76	3.68	4.04
Total observations	271	97	368	516	575

Differences significant at p<0.001(***), p<0.01 (**), p<0.05 (*) and p<0.10 (~)

The previous analyses explored relative differences between condition groups and phases while employing t-tests to test for statistical significance of those differences. While this provides an understanding of how groups differ between phases and any changes over time, it does not allow us to derive conclusions about the impact of the RtL intervention. In this section, we fit multivariate regression models controlling for key background characteristics clustered at the center level to estimate the impact of the program on HALDO scores.

Table 4 outlines the effect sizes (or estimated impact of the intervention) in both standardized and absolute values. This report will review the results from the standardized effect size column. The analysis shows impact of the program for an emergent numeracy skill (one-to-one correspondence), self-identification skills which include identifying a hopeful future, and overall SEL and HALDO scores. At 0.029 to 0.088 standard deviations for all statistically significant outcomes in bold in table 4, these are small effect sizes, but suggest positive impact of RtL on overall learning outcomes for the average student.

In addition to overall impact, we explored whether there might be differential treatment effects for boys relative to girls and students from households with higher SES versus those with lower SES. We found that boys are outperforming girls on numeracy skills but girls are outperforming boys on literacy skills. Children with reportedly lower SES had lower overall scores but SES was not a significant factor for any of the domains. Disability was also not related to children's outcome scores.

Table 8: Return to Learning outcomes by standardized effect size and absolute effect size

	Skill Outomes	Return to Learning Effect Size (standardized)
Literacy	Common letter identification average	**0.070~**
		[0.186]
	Expressive language	-0.014
		[-0.056]
	Hard letter identification	**0.088****
		[0.263]
	Reader	**0.075~**
		[0.297]
	Oral reading accuracy	0.003
		[0.019]
	Reading comprehension	0.052
		[0.261]
Numeracy	Easy Number identification	0.023
		[0.049]
	One-to-one correspondence	0.014
		[0.041]
	Hard Number identification	0.023
		[0.049]
	Operations	0.034
		[0.154]
	Word problems	**0.069*****

		[0.202]
SEL	Self-identification score	**0.044****
		[0.242]
	Empathy score	0.033
		[0.136]
EX	Executive Functioning	**0.029***
		[0.165]
Domains	Literacy	**0.046***
		[0.239]
	Numeracy	**0.040~**
		[0.151]
	SEL	**0.041~**
		[0.240]
	HALDO overall	**0.038***
		[0.251]

Differences significant at p<0.001(***), p<0.01 (**), p<0.05 (*) and p<0.10 (~), standardized errors in parentheses and significance reported in the narrative.

WHAT WAS THE VALIDITY OF HALDO?

Validity refers to how well HALDO actually measures the domains in question. To understand this we would have to use a secondary validated measure and compare the HALDO results to that. However, we did not have the opportunity to use a separate validated measure. Hence, we used the range in children's age as a proxy for predictive validity. We expect that older children would score higher on most of the HALDO domains. If we were actually measuring these domains then we would expect older children to have a higher overall and domain specific HALDO score. In the equity models we fit above, we looked at the age variable to see if, in fact, this hypothesis was supported (see Table 10).

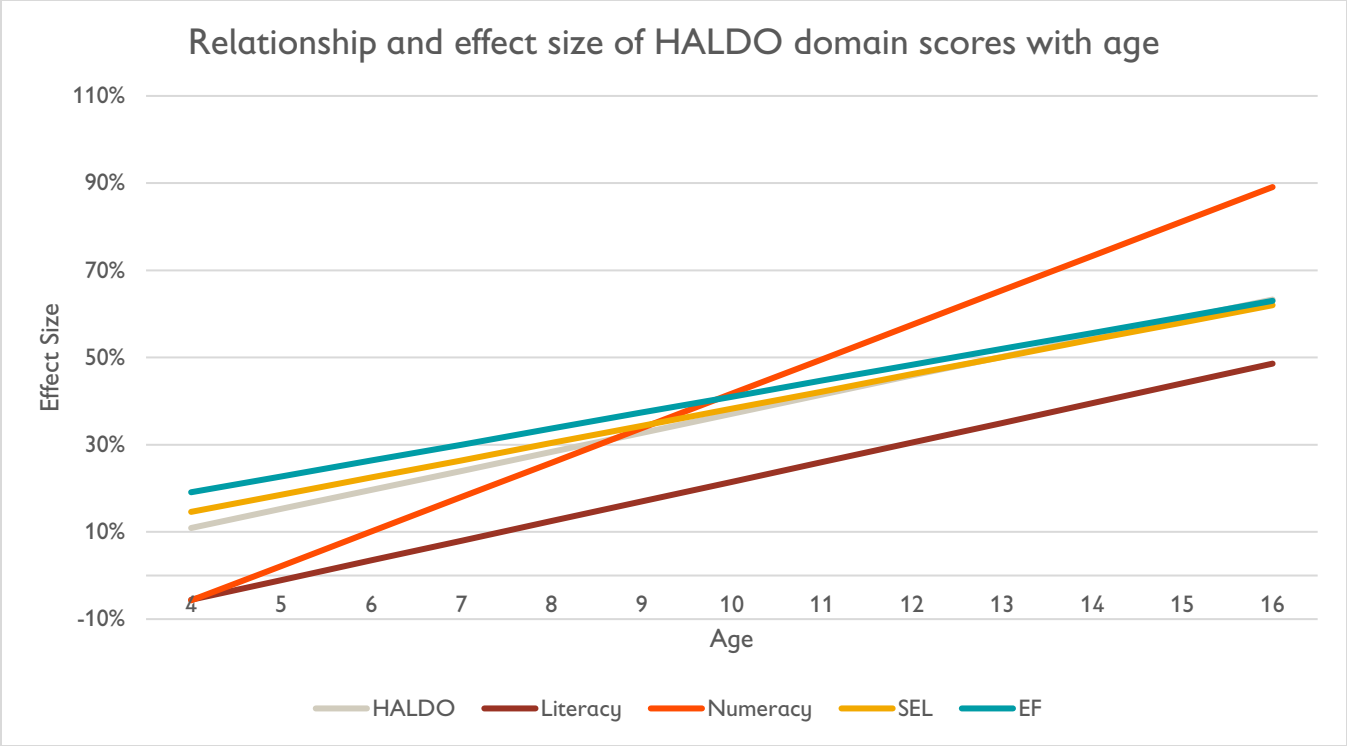


Figure 5. Predicted relationship and effect size of HALDO total and domain scores with child age (controlling for sex, SES, disability status, and home learning environment)

The relationship between age and overall skills measured in HALDO are illustrated in Figure 5. HALDO clearly shows progression in average scores by age. A one year difference in age was associated with a 3-7% point difference in the percent of items a child answered correctly. This difference represented a moderate to large effect size (omega-squared) of between 0.24-0.51 across the four domains.

These findings suggest that, in the context of the refugee response in Lebanon, HALDO measured a developmental progression such that older children scored higher than younger children. HALDO was able to validly measure the developmental difference in learning and wellbeing outcomes between 4-16 year olds. Thus, we can trust the tools' findings.

The programmatic conclusions from this chart emphasize what has been discussed above, that literacy is currently the lowest performing and literacy programming should be emphasized across all age groups.

WHAT WAS THE RELIABILITY OF HALDO?

We assessed two types of reliability: interrater reliability and internal consistency reliability.

1. **Interrater reliability** has to do with whether each assessor would come to the same conclusion about how to record a response while assessing a child. We measure the level of agreement between assessors using a kappa statistic or the intra-class coefficient (ICC). The kappa measures whether assessors agree on the scores measured for nominal and ordinal variables. The ICC measures the same agreement for continuous variables and is reported for reading accuracy. Interrater reliability scores range from 0-1 with 1 representing complete agreement between assessors. Kappa and ICC measures were assessed as .90 - 1 having almost perfect levels of agreement, .80 - .89 with strong agreement levels, .60-.79 showing moderate agreement, and .40 - .59 showing weak agreement.

2. **Internal consistency reliability** has to do with how well the individual items hang together to measure the domain in question. We use a Cronbach's alpha statistic to understand the extent that each domain is measuring a uniform concept, to identify internal consistency within HALDO. The alpha ranges from 0-1, and values above 0.7 are considered good.

INTERRATER RELIABILITY

We looked at the level of agreement for all HALDO items between assessors who assessed the same child in pairs. The data presented here from assessors (n=36) presents strong levels of agreement (kappa between 0.73-1) for all items (see Table 7).

Table 7. Interrater reliability findings in HALDO Lebanon

Domain	Skill: Items	Baseline Kappa[8]/ ICC	Midterm Kappa/ ICC	Endline Kapp/ ICC	Agreement
Literacy	Letter Identification: Identify 5 common letters	1	1	1	High
	Expressive language: Name 10 animals	1	1	1	High
	Letter Identification: Identify 5 rare letters	.98	1	1	High
	Reading with comprehension: Ability to read first five words.	1	1	1	High
	Respond to 5 comprehension questions	1	1	1	High
	Accuracy: Percent of words read correctly	.80	.93	.73	Strong
Numeracy	Number identification: Identify 5 single digit numbers presented visually	1	1	1	High
	One-to-one correspondence: Understand concept of different numbers related to objects (3 items)	.95	.97	.90	High
	Number identification: Identify 5 double digit numbers presented visually	1	1	1	High
	Simple Operations: Complete 5 simple numerical operations with single-digit numbers	1	1	1	High
	Hard Operations: Complete 5 harder numerical operations with two-digit numbers	1	1	.89	High
	Word Problems: Complete 2 numerical problems from a verbal word problem	1	.94	.87	High
SEL	Self-Concept: Knowledge of name,	1	1	1	High
	Age,	.94	.89	.94	High
	Caregiver name	1	.79	**	High
	Community name,	1	.96	1	High
	Country name.	.94	1	1	High
	Ability to identify positive hope for future,	.91	.96	.84	High
	What could stop this future	.95	1	1	High
	What could support this future	.94	1	1	High
	Ability to identify second positive hope for future,	1	1	**	High
	What could stop this future	1	1	**	High
	What could support this future	1	.91	**	High
		1	.91	**	High
	Empathy: Ability to identify how someone else might be feeling.	1	1	1	High
	Ability to show empathy.	1	.94	1	High
	Ability to take the perspective of a third child in an ambiguous situation.	.84	1	1	Strong
		.97	1	.78	High

[8] All results presented are Kappa statistics, except for accuracy, which is a continuous variable and assessed using ICC.

	Tendency to not attribute hostility to ambiguous provocation.				
EF	Executive Functioning Average of 9 number sequences	.94	1	.85	High

INTERNAL CONSISTENCY RELIABILITY

We found good to very strong internal consistency reliabilities (see Table 8) for the overall HALDO score as well as for the domain composites that we created. This suggests that the individual items in each domain hang together well and do a good job in reliably measuring the domains of HALDO.

Table 8. Cronbach's alpha for HALDO overall and domain composites

Domain	Baseline (n=575) Alpha	Midterm (n=516) Alpha	Endline (n=368) Alpha
Literacy	0.83	0.84	0.86
Numeracy	0.81	0.77	0.83
SEL	0.84	0.83	0.79
Executive Functioning	0.70	0.82	0.71
Overall	0.92	0.86	0.82

CONCLUSION

In the HALDO endine from Bekaa, Lebanon, we found that the assessment has strong reliability and validity when it comes to assessing the literacy, numeracy, social and emotional learning, and executive functioning of children between 4 and 16 years old. The endline data collection showed modest positive learning outcomes associated with RtL participation in multiple skills. Further study is needed to understand the relationship between the HALDO results and data-driven programming in the classroom.

The main programmatic findings from the endline results are as follows:

- RtL Content: Given the short duration of the project and minimal attention to literacy and numeracy skills, it is difficult to relate the learning outcomes measured here as impact measure from the program implementation. Children, facilitators, and staff all commented on wanting additional programming and training content, with suggestions to focus more on literacy and numeracy outcomes to better support children to re-enter non-formal or formal education.

Overall, the HALDO endline in Lebanon presents the learners' profile in the Return to Learning program. The endline results show modest learning gains for children who participated in cohort 2. Although the sample of cohort 1 participants was smaller, we can compare the two groups to see that cohort 2 has higher gains than in the absence of the implementation. Further programming, particularly focusing on children's literacy and numeracy skills, could strengthen these modest gains and further support children's reintegration in school.

APPENDICES

APPENDIX 1: FITTED ESTIMATES

Table 9. Fitted estimates from regression models predicting demographic factors by community, compared with dummy variables for each community.

	Age	Sex	Educational background	Books in the home	Reading at home	Disability	SES composite
Khiara	.0350	-.1950	.7387*	-1.2361*	.3542	-.5250	-.5997
Qabelias	-.2604**	-.1175	-.6032	1.3572**	-.5992	.1896	-1.4163
Bar Elias	.2625***	.6847*	.7581*	.3161	.5992	-1.093*	7.008***
Mansoura	.0256	-.0693	-1.050***	-.6006	-.3351	.3952	-3.499***
Ali Alnahri	-.5003***	-.4915	1.187**	.3566	.7823	.705	-.1933

p<0.05 (*), p<0.01 (**), and p<0.001(***).

Table 10. Fitted estimates from regression models predicting overall and domain-specific HALDO Endline scores by equity factors and age

	HALDO	Literacy	Numeracy	SEL	EF
Sex	0.000151	0.0213	-0.0251	-0.0112	0.0189
	[0.01]	[0.02]	[0.02]	[0.01]	[0.02]
Treatment	**0.0460*** **	**0.0639** **	**0.0532***	**0.0463** **	**0.0344~**
	[0.01]	[0.02]	[0.02]	[0.02]	[0.02]
Age	**0.0437*** **	**0.0447*** **	**0.0791*** **	**0.0396*** **	**0.0365*** **
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
SES	**0.0919***	**0.219** **	**0.229** **	0.0545	0.0496
	[0.04]	[0.07]	[0.07]	[0.05]	[0.06]
Disability	-0.00423	0.0142	-0.0171	0.000989	-0.0153
	[0.01]	[0.02]	[0.02]	[0.02]	[0.02]
Books at home	0.0185	0.0113	0.00461	0.0116	**0.0403~**
	[0.02]	[0.03]	[0.03]	[0.02]	[0.02]
Reading at home	**0.0352~**	0.0178	0.0298	**0.0538***	0.00814
	[0.02]	[0.03]	[0.03]	[0.02]	[0.03]
Community	0.00461	0.00288	0.0131	0.00953	-0.00741
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	**-0.153*** **	**-0.387*** **	**-0.515*** **	**-0.0928***	-0.00216
	[0.03]	[0.05]	[0.05]	[0.04]	[0.04]
df_m	8	8	8	8	8
Observations	368	368	368	368	368
Standard errors in brackets					

="~ p<0.10	* p<0.05	** p<0.01	*** p<0.001"	

Table 11: Literacy Impact scores with effect sizes

	Common letter identification	Expressive language	Hard letter identification	Reader	Oral reading accuracy	Reading comprehension
Sex	**0.055~**	**0.129****	0.027	**-0.006***	0.005	0.003
	[0.261]	[0.557]	[0.124]	[-0.037]	[0.037]	[0.022]
Age	-0.005	0.012	-0.002	0.001	**0.004~**	0.000
	[-0.021]	[0.050]	[-0.007]	[0.004]	[0.029]	[0.002]
Has attended school before	0.001	0.055	0.025	0.025	-0.000	**0.034~**
	[0.005]	[0.236]	[0.116]	[0.147]	[-0.000]	[0.232]
SES	0.007	0.066	-0.009	-0.076	-0.081	-0.086
	[0.035]	[0.286]	[-0.041]	[-0.436]	[-0.623]	[-0.581]
Disability	-0.009	0.050	-0.011	-0.000	0.002	-0.011
	[-0.042]	[0.217]	[-0.051]	[-0.002]	[0.012]	[-0.074]
Recentness of arrival	-0.013	-0.007	**-0.009~**	-0.007	0.007	0.003
	[-0.063]	[-0.031]	[-0.042]	[-0.042]	[0.051]	[0.018]
Reading at Home	0.022	0.011	0.060	0.020	**0.018~**	0.009
	[0.104]	[0.046]	[0.274]	[0.116]	[0.137]	[0.064]
Books at Home	-0.009	-0.032	**-0.104~**	0.008	**0.012***	0.018
	[-0.042]	[-0.139]	[-0.474]	[0.048]	[0.090]	[0.121]
Treatment	0.029	0.069	0.031	0.005	-0.027	0.001
	[0.138]	[0.298]	[0.141]	[0.030]	[-0.205]	[0.005]
_cons	0.040	-0.165	0.031	0.038	-0.019	-0.002
	[0.190]	[-0.710]	[0.142]	[0.222]	[-0.150]	[-0.012]
R-sq	0.027	0.130	0.036	0.015	0.029	0.021
N	335	103	335	335	335	335

Table 12: Numeracy Impact scores with effect sizes

	Easy Number identification	One-to-one correspondence	Hard Number identification	Operations	Word problem
Sex	-0.006	0.050	-0.006	0.018~	0.049
	[-0.022]	[0.169]	[-0.022]	[0.130]	[0.183]
Age	**-0.008***	0.016	**-0.008***	-0.002	-0.003
	[-0.029]	[0.055]	[-0.029]	[-0.018]	[-0.009]
Has attended school before	0.018	-0.044	0.018	0.003	0.037
	[0.068]	[-0.149]	[0.068]	[0.019]	[0.138]
SES	0.129	0.136	0.129	0.045	-0.129
	[0.486]	[0.459]	[0.486]	[0.328]	[-0.483]
Disability	**0.059~**	-0.084	**0.059~**	0.019	**-0.023***
	[0.223]	[-0.283]	[0.223]	[0.135]	[-0.088]
Recentness of arrival	0.028	-0.034	0.028	0.001	0.000
	[0.104]	[-0.114]	[0.104]	[0.005]	[0.002]
Reading at Home	0.011	-0.072	0.011	0.035	**-0.051~**
	[0.040]	[-0.244]	[0.040]	[0.258]	[-0.191]
Books at Home	0.024	0.025	0.024	-0.036	0.008
	[0.091]	[0.084]	[0.091]	[-0.260]	[0.030]
Treatment	-0.055	-0.066	-0.055	-0.019	-0.032
	[-0.207]	[-0.221]	[-0.207]	[-0.141]	[-0.120]
_cons	-0.064	0.135	-0.064	0.002	**0.074****
	[-0.241]	[0.456]	[-0.241]	[0.017]	[0.276]
R-sq	0.039	0.049	0.039	0.023	0.027
N	335	109	335	335	335

Table 13: SEL Impact scores with effect sizes

	Self-identification score	Empathy score
Sex	-0.013	0.001
	[-0.079]	[0.003]
Age	-0.004	-0.012
	[-0.027]	[-0.047]
Has attended school before	-0.014	0.026
	[-0.084]	[0.101]
SES	-0.052	0.030
	[-0.317]	[0.116]
Disability	0.028	-0.021
	[0.170]	[-0.082]
Recentness of arrival	-0.001	-0.006
	[-0.003]	[-0.023]
Reading at Home	0.035	0.042
	[0.212]	[0.162]
Books at Home	-0.008	0.033
	[-0.047]	[0.130]
Treatment	**0.027~**	0.035
	[0.162]	[0.137]
_cons	0.047	0.062
	[0.288]	[0.242]
R-sq	0.028	0.034
N	335	335

Table 14: Domain Impact scores with effect sizes

	Literacy	Numeracy	SEL	Executive Functioning	HALDO overall
Sex	0.016	**0.028***	-0.009	**0.035***	0.011
	[0.152]	[0.207]	[-0.059]	[0.237]	[0.109]
Age	0.001	**-0.004***	**-0.007***	**-0.011***	**-0.007***
	[0.010]	[-0.031]	[-0.047]	[-0.073]	[-0.070]
Has attended school before	**0.022***	0.003	-0.001	**0.038***	**0.013~**
	[0.210]	[0.020]	[-0.009]	[0.260]	[0.137]
SES	**-0.065***	0.004	-0.027	0.014	-0.018
	[-0.615]	[0.027]	[-0.181]	[0.094]	[-0.183]
Disability	-0.005	-0.004	0.013	-0.010	0.002
	[-0.046]	[-0.026]	[0.086]	[-0.068]	[0.021]
Recentness of arrival	-0.001	-0.000	-0.002	-0.002	-0.002
	[-0.007]	[-0.001]	[-0.015]	[-0.013]	[-0.019]
Reading at Home	0.019	-0.013	0.037	0.018	0.025
	[0.180]	[-0.093]	[0.251]	[0.122]	[0.265]
Books at Home	-0.014	-0.006	0.005	-0.005	-0.002
	[-0.130]	[-0.047]	[0.035]	[-0.031]	[-0.018]
Treatment	-0.000	**-0.039****	**0.029***	-0.006	0.009
	[-0.004]	[-0.286]	[0.199]	[-0.044]	[0.092]
_cons	0.002	**0.066***	0.052	0.084	0.056
	[0.022]	[0.486]	[0.353]	[0.571]	[0.588]
R-sq	0.023	0.033	0.038	0.058	0.047
N	335	335	335	335	335

APPENDIX 2: RETURN TO LEARNING FOCUS GROUP DISCUSSION PROTOCOL

INTRODUCTION

The purpose of this protocol is to guide focus group discussions (FGD) with children and adults for the Return to Learning program in Lebanon.

➔ FGDs conducted with children require a specific set of skills and questions. **Only facilitators with experience working with this age group should conduct FGDs.**

FGD PREPARATION

PARTICIPANTS

- Should constitute a reflection of the different groups in the respective community, e.g. different types of people/ professions/ background within the community/ people living with disabilities/ elderly people/ etc. At the same time, the group should be as homogenous as possible with regards to social status of participants, given that this has proved more successful in past FGDs.
    - ➔ Consider issues of inclusion such disabilities
- The focus groups should ideally be between 7 to 12 people maximum and last between 45 and 70 minutes.
- The focus groups must be separated between male and female and age (separated FGDs for girls, boys, women and men). If these requirements are not met, the FGD will have to be interpreted separately/differently and weight of findings adjusted accordingly.

NOTES ON FACILITATION

- Facilitators should strictly follow the FGD guidelines, be familiar with the tool before conducting the interviews and receive appropriate training. They must not provide their opinion, influence the conversation or argue a point with participants, even if they feel that the participant is wrong.
- While guiding the discussion, facilitators should first of all be good listeners. They should ensure that all participants are heard, without pressurizing those who prefer not to talk. Facilitators should also ensure that the opinions and views of all participants are respected.
- The facilitator should try to always get a sense of who the participants are talking about (if it is women, men, girls or boys).
- The facilitator should also be careful to pay attention to any non-verbal communication, including tone of voice, facial expression (use encouraging nods and smiles) and eye contact.
- It is preferable to arrange participants in a circle for a friendly and interactive setting. Discussion will take place in a safe, comfortable and confidential location.
- The facilitator should try to ensure a relaxing and comfortable environment; controlling his/her voice, body language and choosing the culturally appropriate language.

NOTES ON NOTE-TAKING

- The facilitator should be accompanied by a person – of the same sex of the group - who takes notes during the discussion and also supports the facilitator in compiling the FGD report.
- It will be of importance to write up detailed notes of the discussions, not summaries or interpretations.
- When possible and not causing harm – recording the discussion should be considered, with the consent of participants, as this usually leads to more accurate note-taking.
- Notes should not contain any names of participants. Confidentiality has to be ensured.

KEY TIPS FOR COMMUNICATING WITH CHILDREN

- Stop an activity or discussion if a child feels upset.
- Be sensitive to identify when a child might need additional support and attention.
- Be sensitive to the mood and energy of the group. Quick breaks or energizers can be added in between themes, if needed.

CAREGIVER/PARENTAL CONSENT

- It is important to ask caregivers/parents for agreement for their children to participate in the FGDs. This must take place before the FGDs are conducted.
- Parents/caregivers should clearly understand the purpose of the FGDs, voluntary nature of participation and issues around confidentiality.
- Consent may be verbal and signed depending on your context.

TOPIC

Understanding facilitators' and learners' experiences in the RtL program and parents perspectives on education.

PURPOSE

The Return to Learning focus group data collection supports the overall project evaluation and ongoing process evaluation. These focus groups occur in May 2019, between the first and second cohort of children who receive the Return to Learning content. The data will inform the second cohort implementation and serve as a benchmark to compare endline focus group data with, to be collected in July-August 2019.

The focus group data supports additional data collection in the program, including a facilitator observation to assess and monitor facilitator's teaching practices and the Holistic Assessment of Learning and Development Outcomes (HALDO), to assess children's knowledge and potential learning gains. The qualitative data from the focus groups will tell us *how* and *why* any changes observed in facilitator observation data or HALDO assessment data may occur.

RESEARCH QUESTIONS

The following research questions frame this project:

<u>Baseline</u>

1. How comparable are children's outcomes measured by HALDO between the intervention group (Cohort 1) and control group (Cohort 2) in terms of learning and development skills, background characteristics, and home learning environment?
2. What can HALDO tell us about children's emergent learning and development skills? What does this mean for programming?
3. Do children's learning and development skills vary by student background characteristics like sex, poverty, or home learning environment? If so, what does this mean for effectively targeting our learning and development program?
4. Do facilitators display active, child centered teaching practices after RtL training?

<u>Endline</u>

6. How has the sample of children changed over time?
    a. Are the children in Cohort 1 who were able to be found at endline different than those who were not able to be found? If so, how?
    b. Did the attrition rate differ between Cohort 1 and Cohort 2 children?
7. Did RtL participation exhibit impact on children's learning and development skills?
    a. Does this impact result in more equitable outcomes for traditionally disadvantaged groups?
8. How do changes in children's development and learning vary by home learning environment and/or classroom learning environment?
9. Have facilitators advanced active, child centered teaching practices in RtL?

Focus group data will help us understand what changes are occurring in the children who participate in the program and why those changes may be occurring. We will also use data to inform what is working in the project as part of a process evaluation and to support project improvements in the second cohort.

STUDY DESIGN

The qualitative data collection is not seeking representativeness, but rather we are using an exploratory approach to understand what is happening in the program and what could be changed to improve programming for various stakeholders in the second cohort.

Table: FGD sample size per RtL target groups to conduct qualitative tracer study

Target communities	Number of facilitators		Number of parents/ caregivers		Number of learners < 10		> 10	
	Male	Female	Male	Female	Male	Female	Male	Female
Khiara	1	1			1	1	1	1
Qabelias	2	1	7	7	2	2	2	2
Bar Elias	0	3	7	7	2	2	2	2
Mansora	0	4	7	7	1	1	1	1
Ali Alnahri	0	3			1	1	1	1
Total	5	11	21	21	7	7	7	7

GUIDANCE

1. 2 focus groups with facilitators (male and female)
2. 2 focus groups with parents/ caregivers (male and female)
3. 4 focus groups with children (under 10, male and female, and over 10, male and female).
   a. All children who participate in the focus group should be from **cohort 1.**
4. Two FGDs can be completed a day. Four days are necessary to complete all focus groups
5. The FGD facilitator and note-taker should take 15-20 minutes after completing one FGD to write field notes.
6. After completing data collection, the FGD facilitator and note-taker will take one day for preliminary data analysis
7. All FGDs will be audio-recorded
8. All recordings should be transcribed and all notes and transcriptions shared with Beirut and Washington, DC staff in English.
9. Analysis will be conducted using Dedoose and results shared with field and country office staff.
10. It is recommended that field office staff share results with participants and facilitators as appropriate.

FGD TOOLS AND IMPLEMENTATION

FACILITATOR FGD PROTOCOL:

Welcome participants: *Good morning/afternoon. Thank you for taking the time to join us for this discussion today.*

Introduce yourself and your role: *My name is <insert FGD facilitator name> and this is my colleague <insert names of note taker>. We're here on behalf of Save the Children. We would like to ask you some questions about your experiences as facilitators in the Return to Learning program.*

Explain purpose: *The aim of this discussion is to learn from your experiences and improve programming in cohort 2 which will begin next month.*

Explain confidentiality: *Participation in the discussion is completely voluntary and you do not have to answer any questions that you do not want to answer. You may leave the discussion at any time or ask for a short break. This focus group should last about one to two hours.*

We will not be writing your names down or use them in any way after this discussion. We will treat everything that you say today with respect, and we will only share the answers you give as general answers combined with those from all the people who speak to us. We ask that you keep everything confidential, too. Please do not tell others what was said today and by whom.

To be sure that we hear you accurately, and with your permission, we would like to audiotape (record) the focus group discussion. Is this acceptable to you? Yes or No (If a participants replies with "no", he/she should leave the discussion at this point. The FGD facilitator is responsible for following up on this.)

While we hope that the whole group can stay for the entire discussion, participants can also choose not to respond or leave at any time if they are not comfortable with the questions asked.

Do you have any questions before we begin? Should we start?

After all the facilitators have given their consent, please ask them to introduce themselves to the group.

General Information

Guidance:

- Fully complete this section after you receive permission from participants to take notes.
- Every section below needs to be filled for the FGD data to be used in analysis.

Date (day / month / year): _____

Number of FGD participants: _____

Area: _____

FGD Questions		
1.	Conduct FGDs with 7 facilitators in each FGD. Discussion should focus on program implementation, efficiency, and efficacy.	**Guiding Questions:** a. Do you think this program fills a need for out of school children? Why or why not? b. What worked well during facilitation? c. What was challenging during facilitation? d. What program implementation challenges did you face? e. How did you solve those challenges? f. What do you believe children gained from participating in the lessons?

g. What changes should be made to improve the quality of the program before we roll out the second cohort?

INTRODUCTION FOR PARENTS/CAREGIVERS FGDS (ABOVE 18 YEARS):

Welcome participants: *Good morning/afternoon. Thank you for taking the time to join us for this discussion today.*

Introduce yourself and your role: *My name is <insert FGD facilitator name> and this is my colleague <insert names of note taker>. We're here on behalf of Save the Children. We would like to ask you some questions about education issues that may affect your family or community so that we can better understand your needs and concerns.*

Explain purpose: *The aim of this discussion is to understand how we can support parents and their children's education.*

Explain confidentiality: *Participation in the discussion is completely voluntary and you do not have to answer any questions that you do not want to answer. You may leave the discussion at any time or ask for a short break. This focus group should last about one to two hours.*

We will not be writing your names down or use them in any way after this discussion. We will treat everything that you say today with respect, and we will only share the answers you give as general answers combined with those from all the people who speak to us.

We ask that you keep everything confidential, too. Please do not tell others what was said today and by whom.

To be sure that we hear you accurately, and with your permission, we would like to audiotape (record) the focus group discussion. Is this acceptable to you? Yes or No (If a participants replies with "no", he/she should leave the discussion at this point. The facilitator is responsible for following up on this.)

While we hope that the whole group can stay for the entire discussion, participants can also choose not to respond or leave at any time if they are not comfortable with the questions asked.

Do you have any questions before we begin? Should we start?

After all the parents/ caregivers have given their consent, please ask them to introduce themselves to the group.

General Information

Guidance:

- Fully complete this section after you receive permission from participants to take notes.
- Every section below needs to be filled for the FGD data to be used in analysis.

Date (day / month / year): _____

Number of FGD participants: _____

Area: _____

FGD Questions	
1. Desired Information: Reenrollment attitudes, obstacles to education, and how parents can be more involved in learning (home learning environment, reading at home, etc.) FGDs	**Guiding Questions:** a. From your perspective, how can we best serve out of school children in future programs? b. As parents and caregivers, what are our responsibilities in our children's education? c. What do you hope for your children's education?

should occur with at least 5 parents, one from each center area.	d. How long has your child/children been out of school? a. Why? b. What do you hope will change, if anything, about this situation? e. Refer to Activity 1 f. What can parents do to support education and learning at home? g. Refer to Activity 2

Activity 1: Ranking obstacles to education

On a large piece of paper, make a grid with at least 9 boxes, like the example below. Ask the participants what obstacles they know are the way of their children's education and write each unique obstacle in a box. You can use the examples in the boxes below as ideas, but focus on what the participants think are obstacles to their children's education, not only the examples here.

Then ask the participants to vote on each item by reading the item aloud and asking the participants to raise their hands, or to play a sticky note on the box if they think this is the **biggest obstacle to education** and tally the responses in each box.

Obstacle 1: Not enough schools	Obstacle 4: Early marriage	Obstacle 7: Not safe to travel to school	Obstacle 10: Education is not a priority/ need
Obstacle 2: No money for school supplies	Obstacle 5: Language barriers	Obstacle 8: Poor quality education	Obstacle 11: Education content is not relevant
Obstacle 3: Child needs to work to support family	Obstacle 6: No teachers	Obstacle 9: Family is not settled in one place	Obstacle 12: Don't trust education staff/ teachers

After the participants vote, discuss which obstacles got the most votes and why.

Take a picture of the paper before ending the FGD.

Activity 2: Home learning environment

On sticky notes, write down the ideas that parents come up with about how they can support education and learning at home. Then read the sticky notes aloud to the parents and ask them which would be most likely/ easiest to least likely/ hardest. As they discuss what order the sticky notes should be in, identify when there is disagreement and ask for more information. The parents may not all agree on the ranking, which should be noted and the FGD facilitator should ask follow up questions whenever there is disagreement.

INTRODUCTION FOR LEARNER FGDS (UNDER 18 YEARS)

Before starting the FGD you should receive verbal consent from all children to participate. Please read the following script to the children and ensure that each child provides verbal consent. If a child decides not to participate in the FGD, that is alright! The child can be excused from the FGD. Please ensure that these children still receive any benefits (e.g. transport fee/snacks) that is provided to all other children.

Welcome participants: *Good morning/afternoon. Thank you for taking the time to join us for this discussion today.*

Introduce yourself and your role: *My name is <insert FGD facilitator name> and this is my colleague <insert names of note taker>. We're here on behalf of Save the Children.*

Explain purpose: *We are here today to ask you some questions about the learning program you have been participating in. We want to know about the positive and negative things. We want to make sure that all of you can share your views honestly. Everyone's opinion is important. We welcome your responses and want to emphasize that there are no right or wrong answers. Your insights will help us better understand what we can do better in the future.*

Explain Ground Rules: *You do not have to participate or answer any questions that you do not want to answer. You may leave the discussion at any time or ask for a short break. This focus group should last about one to two hours.*

We will not be writing your names down or use them in any way after this discussion. We will treat everything that you say today with respect, and we should all be respectful when others speak. We would like to request that everyone agree to keep what they hear during this focus group private. This means that you should not go outside this focus group and start gossiping about each other. Also, If sharing examples or personal experiences, please do not use any names. In the same way, we (point to all the facilitators in the room) will keep everything you tell us private. When we talk about this conversation, we will only talk about the general things that we have learned from doing multiple focus groups like this one.

To be sure that we hear you accurately, and with your permission, we would like to audiotape (record) the focus group discussion. Is this acceptable to you? Yes or No (If a participants replies with "no", he/she should leave the discussion at this point. The facilitator is responsible for following up on this.)

Do you have any questions? Should we start?

After all the youth have given their consent please ask them to introduce themselves to the group.

General Information

Guidance:

- Fully complete this section after you receive permission from participants to take notes.
- Every section below needs to be filled for the FGD data to be used in analysis.

Date (day / month / year): _____ **Number of FGD participants:** _____

Area: _____

FGD Questions	
1. Desired Information: Questions should focus on summative conversation about how the program was and their attitudes towards education. FGDs should include 7 children in each, 4 FGDs with (1) only girls, (2) with only boys, (3) with	**Guiding Questions:** a. What did you learn in the program? b. What do you hope will happen in your life after you finish this program? c. Activity 1 d. Activity 2 e. What did you like about it?

only younger children (under 10) and (4) with only older children (over 10).	f. What did you not like about it? g. What else would you like to learn in school? h. Do you feel prepared to go to school all day, every day? If not, why not? i. From your perspective, how can we help other out of school children?

Activity 1

Divide the youth into two groups. Ask each youth to sit in the circle and give them one big flipchart paper with colored pencils/markers. During the activity the note-taker who speaks the program language should move through the group and help the youth who may struggle because of literacy. The facilitator should stand at the front of the room with the main flip chart. As the facilitator gives the youth instructions, he/she should demonstrate what is expected from the youth with examples on the main flip chart in the front of the room.

Say: *We are going to start this focus group with an activity that you will do in your groups. We have you in small groups so you can help each other. We are giving each group a piece of paper. On this paper I would like you to draw what the average day in the learning program is like. You can include pictures of activities you did, your facilitator, classmates, games you played, or anything that you remember. You can draw more than one picture, but we will ask you to talk about each picture you draw.*

When the children are finished drawing, ask the two groups to sit together and share their pictures. Probe children by asking about what they learned on that day, or with that facilitator, or during that game.

Offer the youth a short break between Activity 1 and 2. Play a game or energizer when the youth return

Activity 2

Note: Some of the information you need to get through this activity may be covered by children's responses in Activity 1. Use only the questions that you need from this activity. No need of asking youth redundant or repeated questions.

Say: We are going to play another game. I am going to read a few statements to you. For each statement I want you to think about how much you agree or disagree with the statement. If you agree with the statement fully then you should move to the right side of the room. If you disagree with the statement fully then you should move to the left side of the room. If you neither agree nor disagree with the statement, then you should stand in the center of the room.

Now some of you may not want to be in the center or against the right wall. What would be standing in this area suggest? Get some responses from youth. *It would mean that you agree with the statement but not fully.*

Where will you stand if you disagree with a statement but not fully? Point to the space in between the left wall and the center of the room.

Let's play a game to see if you understand. Okay?

Read the following statements to the youth. After reading each statement ask the youth to move to the place in the room that best represents their level of agreement with the statement. Once youth are standing in their chosen space, ask them questions about why they chose the space to see if they understand the activity?

- *You enjoy eating <enter name of local treat or sweet>*
- *You play football with my friends*
- *You watch television everyday*

Once youth have understood the activity you can proceed to ask them the following statements. For each statement use the prompts to get additional information from the youth. After each statement, ask the follow up questions to gather more information.

- *Before you started this program, you were working*
  Follow up questions:
  - *What work did you do? Why?*

- *Currently, you want to go back to school*
  Follow up questions:
  - *How does your family support you?*
  - *What did they give you?*
  - *What has changed from before this program and now?*

- *Before you joined this program, you were learning outside of school*
  Follow up questions:
  - *Were you reading? If so what?*
  - *Did you do math in the market?*

- *Currently, you study out of school*
  Follow up questions:
  - *Do you read? If so what?*
  - *Do you do math? If so where/what kind?*