UNHCR
The UN Refugee Agency

MAURITANIA

Measuring Holistic Learning Outcomes for the Forcibly Displaced

Acknowledgements

This report was prepared by Jeffrey H. Marshall, an independent consultant, and coordinated by Emiko Naka and Cirenia Chávez of the Education section, Division of Resilience and Solutions (DRS). The report benefited from many contributions, including RET International staff and UNHCR staff in the Mauritania Country Operation.

The team thanks Robson Dias, Cedic Maombi Bagalw, Prosper Icoyitungiye, Elizete Lima Zazá (RET International), and Kemegne Eric Carter and Genta Akasaki (UNHCR) for their generous support. This report would not have been possible without the support of Porticus.

Please cite the work as follows:

UNHCR. 2025. Mauritania: Measuring Holistic Learning Outcomes for the Forcibly Displaced.

Table of Contents

Executive Summary

Background

A growing body of international evidence documents the global "learning crisis" and the large number of children who are not attaining foundational learning skills, but very few studies have measured foundational skills among the forcibly displaced. The Measuring Holistic Learning Outcomes for the Forcibly Displaced project aims to contribute to establishing a baseline on the learning outcomes of refugee children while strengthening the global evidence base on learning in contexts of forced displacement. This report summarizes the results from an assessment of learning proficiency levels among forcibly displaced Malian students in the Mbera Camp in Mauritania, together with children from the host community.

This study aims to address three data gaps in the evidence base on forcibly displaced children's education in Mauritania. First, the analysis includes a detailed summary of grade 4 and 6 student performance in four foundational skill areas covering basic literacy and numeracy, socioemotional learning, and executive functioning. Second, the analysis identifies factors that facilitate or impede student learning progress. This is done with quantitative survey data comparisons of student academic and non-academic outcomes between different student and school characteristics, and with qualitative data from Focus Group Discussions (FGDs) carried out with students, school staff and parents. And third, the various outcomes and factors are compared between Mbera refugee camp schools and schools located in the local host community.

Data and methodology

The assessment used the Holistic Assessment of Development and Learning Outcomes (HALDO), developed by Save the Children, an interactive assessment tool that measures learning progress among populations impacted by crises. The assessment is designed to be administered to children ages 4-12 (in early grades through end of primary) and has an adaptive structure to account for a wide age range of skill levels in four domains for literacy, numeracy, social emotional learning (SEL) and executive functioning (EXF).

The HALDO assessment was conducted in July 2024 in all eight schools in the Mbera camp and four out of five host community schools in the commune of Fassala. Students in grades 4 and 6 were targeted as they represent the end of the primary school cycle. A total of 509 students (256 girls, 253 boys) were randomly selected from the lists of students enrolled in these grades for the 2023-2024 school year, 400 from Mbera camp and 109 from Fassala schools. A major challenge for the study is that the data collection took place during the school holidays. Sampling weights are used to correct for discrepancies between each school's share of the sample and share of the overall population, as well as the gender ratio. The samples cannot be treated as representative samples of the total population of students in the Mbera camp or host community schools.

The Focus Group Discussions (FGDs) were organized to gather additional qualitative insights. These discussions took place between July and August, after HALDO administration. The

FGDs focused on three key groups: parents of students, students, teachers, and principals. To ensure balanced representation, the discussions with students were organized with a focus on gender parity and included participants from both refugee and host communities.

Results

Key Findings from HALDO

Based on the student background questions that are included in the HALDO instrument there are significant differences between the Mbera camp and host community (Fasala) students. The host community students are significantly younger, more likely to speak Hasania, have more household possessions in the home, and are much less likely to report French language is used in their classrooms. Standing out is the age profile of the Mbera camp students, with very large shares above the expected ages of 10, for grade 4 and 12, for grade 6. Forty percent of grade 4 students in Mbera are 13 years or while in grade 6 almost half (44 percent) are 15 or older.

The HALDO assessment covers four domains and generates a large amount of data. In literacy, this includes letter identification, basic reading fluency and reading comprehension, while for mathematics the problems cover number identification, word problems and operations. For literacy and numeracy, the detailed summaries of student results on the most advanced content provide key insights into overall performance. When students are unable to successfully complete an earlier, easier task, they are automatically assigned zero scores by HALDO (i.e. all remaining difficult tasks are automatically coded as zero). The literacy results show a very large share of students who did not advance to the reading comprehension section of HALDO: roughly 80 percent of grade 4 students, and nearly 50 percent of grade 6 students. Those who could proceed to the most difficult task, i.e. reading comprehension, and answered all five reading comprehension questions correctly account for only 10 percent of grade 4 students, and 25 percent of grade 6 students.

The pattern for numeracy results is similar to that of literacy: very large shares of grade 4 (nearly 75 percent) and grade 6 (55 percent) students were unable to correctly answer easier numeracy questions, and therefore did not advance to the most difficult content (hard operations). As a result, only 17 percent of grade 4 students, and 33 percent of grade 6 students correctly answered all four hard operations questions.

Socioemotional learning (SEL) is covered in HALDO with 16 questions divided into Self-concept and Empathy domains. The results show that almost all students are able to correctly answer the initial set of Self-Concept questions, but as the questions become more demanding—for example asking them to consider factors that may impact them in the future—the scores decline substantially. A similar pattern is shown for the set of questions related to Empathy.

Finally, the HALDO questions on executive functioning (EXF) test students on a series of short-term and working memory operations, such as reciting a three number sequence. These results also show that student scores are relatively high on easier content but then drop significantly as the tasks get more difficult. When bringing together the four domains covered in HALDO to compare the scores by domain and grade level, student scores are highest in SEL content.

The main finding in this study is that student achievement levels in grades 4 and 6 are not at the expected levels. The defining feature of HALDO, in addition to its coverage across different domains, is its focus on foundational skills. This emphasis on basic content means that students in higher primary grades (such as grade 4 and grade 6) should be comfortable with the content. But the results in literacy and numeracy instead show that large shares of students did not advance to the more demanding content. Furthermore, the advanced content in HALDO is not very advanced for students in this grade range. In literacy the most difficult content requires students to answer five basic reading comprehension questions after a short passage, while in numeracy the "hard" operations are restricted to two-digit addition and subtraction problems and do not include multiplication, division, decimals, or other more challenging questions.

The results from the socioemotional learning and executive functioning components of HALDO are generally in line with the results for literacy and numeracy. Very high percentages (above 90 percent) can complete the easiest activities (i.e. state their age for SEL, repeat three numbers for EXF). But as the content gets harder the scores decline noticeably. For example, only about half of the students could identify something that may prevent them from realizing their goals in the future (SEL) or count back a four-digit number sequence in reverse (EXF).

In addition to describing student results on HALDO, one of the main activities in the report is to compare HALDO scores across different categories of students (and schools). Importantly, student scores are consistently higher in grade 6 than in grade 4, which suggests that learning is taking place in these schools. And there are other variables that are significantly associated with overall scores, such as whether or not the student reported that books are read in their homes. However, the main finding from an extensive set of comparisons—and multivariate statistical analysis—is that most of the background variables are not significant predictors of HALDO scores. This includes the comparisons of students in the Mbera camp and host community schools, although these results are somewhat different by grade. One surprising result is that the measure for socioeconomic status (SES), derived from the HALDO questions about household possessions (such as television) and home characteristics (such as type of floor), is not a significant predictor of HALDO.

It is important to remember that HALDO is designed for vulnerable populations like the students in the Mbera camp. This vulnerability should not reduce the expectations about their development of basic skills, but it does highlight the importance of taking into account the contextual features of their lives. This is why Focus Group Discussions (FGDs) data is important since it provides an additional set of information to consider a far-ranging set of topics related to life inside and outside of schools for students in the Mbera refugee camp and the host community schools.

Key findings from FGDs

The focus group discussions reveal generally positive perceptions of schooling among both camp and host community participants, with parents appreciative of school staff efforts. In Mbera camp, international partners are also recognized for their contributions. However, limited parental engagement with school staff remains a concern across communities. Child security issues are another concern raised by communities, particularly due to inadequate school fencing, which raises risks of intruders or animals entering school grounds.

A major theme emerging from the discussions is the severe lack of school resources, which affects multiple aspects of education. Infrastructure deficiencies, insufficient access to food and water, overcrowded classrooms, and a shortage of teaching and learning materials are widely cited challenges. Additionally, school staff acknowledge the need for greater inclusivity, particularly for children with disabilities, though resource constraints, lack of training, and low community awareness hinder progress. Teacher capacity issues further compound these challenges, with educators themselves recognizing the need for more training to effectively support student learning. While addressing these deficiencies is complex, tackling them is critical to improving educational outcomes and ensuring schools remain a viable and trusted option for both camp and host community families.

Recommendations

The main finding from the study is that grade 4 and 6 students are not performing at a sufficient level on basic content, and there is a convincing—albeit somewhat tentative—link between these results and the various resource, capacity and vulnerability issues that were identified in the FGDs. The recommendations are divided into policy, program, and research categories to address the learning deficiencies as well as improve the measurement and monitor of these processes in the future. These recommendations also reflect the feedback received at dissemination events at national and local levels, where the Ministry of Education officials, partner organizations, as well as Mbera camp coordinator, and local communities were present.

Policy and programme recommendations:

- **Adopt Inclusive Policies** – Collaborate with national and international partners to continue the inclusion of refugee students into the Mauritanian education system, balancing the Malian curriculum with local content.
- **Address School Resource Constraints** – Conduct a comprehensive assessment of school resource shortages and prioritize feasible items to support in collaboration with key stakeholders.
- **Enhance Parental and Community Engagement** – Establish regular meetings to foster stronger communication between school staff, parents, and community members to improve accountability and support student access and retention.
- **Improve Teacher Training and Support** – Identify specific training needs for teachers and integrate them into district-level training programs, including support for Malian teachers in Mbera camp.
- **Strengthen Student Well-being and Learning** – UNHCR should advocate with Ministry of Education for incorporating psychosocial services and social-emotional learning (SEL) into the curriculum to support forcibly displaced children's mental health, well-being, and learning.
- **Expand Access to Books and Libraries** – Strengthen library programs and encourage community initiatives to collect book donations, promote reading, and improve children's access to reading materials.

Research recommendations include:

- **Regular Learning Assessments for Forcibly Displaced Children** – Conduct learning assessments on a regular basis, ensuring they take place at the end of the school year when students have completed most of the curriculum. These assessments should be processes. The UNHCR operation could consider adding user-defined indicators for COMPASS reporting that address learning proficiency (e.g. % of forcibly displaced students with minimum skills in reading / mathematics / socio-emotional learning).

- **Expanding Early Grade Assessment** – Given the challenges faced by many grade 4 and 6 students in answering basic-level content, including earlier grades in the future assessment would provide valuable insights into early learning development.

- **Equip Teachers with Practical Assessment Tools** – Introduce simplified foundational learning assessment tools for teachers to use throughout the school year, enabling continuous monitoring of student progress.

- **Improve future assessments** by:

  - *Enhancing the HALDO Instrument for Better Usability* – Engage with program designers from Save the Children to consider simplifications, such as adjusting restrictive skipping rules (e.g. students do not need to answer all questions correctly to proceed to the next question), to improve the assessment's effectiveness and accessibility.

  - *Ensuring Meaningful Comparisons in Sampling* – To enable meaningful comparisons between Mbera camp and host community students, the sampling strategy should ensure balanced representation from both groups. Small samples may yield inconclusive results, limiting the study's usefulness and efficiency.

  - *Strengthening Data Collection and Contextual Relevance* – Expand data collection efforts to include more information on children with disabilities and tailor all questionnaires (students, FGDs, and any new instruments) to reflect key contextual factors like school feeding participation, Koranic school attendance, and school climate.

  - *Improving Data Quality and Implementation* – Allocate more time for thorough tablet programming checks and piloting, and consider adding teacher, head teacher, and parent questionnaires to complement student assessment data.

1. Introduction

A growing body of international evidence documents the global "learning crisis" and the large number of children who are not attaining foundational learning skills, with especially low learning levels among the most vulnerable populations (Moscoviz & Evans, 2021; World Bank, 2018). However, very few studies have measured foundational skills among the forcibly displaced, including forcibly displaced children who are out of school. An additional data gap is related to foundational skills in areas like socioemotional learning among forcibly displaced learners including refugees, which are increasingly recognized as necessary for success in the 21st century.

While organizations such as UNICEF, UNESCO and the OECD have recognized the importance of measuring foundational skills, undertaking assessments that include refugees and children in crisis settings remains a critical challenge. The non-inclusion of these children in national and regional assessments—and the lack of disaggregated data when they are included—masks a growing learning crisis among an especially vulnerable sub-population of young people (RET International, 2024).

The *Measuring Holistic Learning Outcomes for the Forcibly Displaced* Project aims to contribute to establishing a baseline on the learning outcomes of refugee children while strengthening the global evidence base on learning in contexts of forced displacement. **This report summarizes the results from an assessment of learning proficiency levels among forcibly displaced Malian students in the Mbera Camp in Mauritania,** and children from the host community. The assessment was conducted in July 2024 in eight refugee camp schools and four community schools, with just over 500 students in grades 4 and 6. The student assessment covered four areas of foundational skills. Additional data were provided through a student background questionnaire as well as Focus Group Discussions (FGDs) among students, teachers, and parents.

The report includes seven sections. Section 2 describes the forcibly displaced refugee context in Mauritania, including policy initiatives and challenges. Section 3 covers the aims and justification for the Mauritania report together with the key research questions that inform the data analysis. Section 4 reviews the methodology including the sample, instrumentation, data analysis methods and limitations. Section 5 summarizes the results from the various data analyses, while Section 6 provides a discussion of the implications from the main findings. Section 7 concludes with a set of recommendations.

2. Background

2.1 The forced displacement context in Mauritania

Recent figures show 149,654 Malian refugees registered in and around the Mbera camp in Mauritania's Bassikounou moughataa (department), which is located in one of the poorest regions in the country, Hodh El Chargui (Cambridge Education, 2022). The Mbera camp is located 50 kilometers from the border with Mali and was set up in January 2012 to host a large influx of Malian refugees who fled the conflict in northern Mali. **Most of the refugees who were living in the camp in 2022 had arrived with the first influx in 2012 and 2013 (65,402 refugees).** A 2021 survey conducted by UNHCR showed that 79% of the refugees interviewed did not envisage to return to Mali anytime soon (UNHCR, 2021a, 2021b). The Mbera camp is the only camp hosting refugees in Mauritania. Refugees also reside outside of the Mbera camp in two of the main economic urban areas of the country, in Nouakchott and Nouadhibou.

In early 2022, UNHCR and the Ministry of the Interior signed a Memorandum of Understanding that formalizes UNHCR's "temporary" mandate to conduct Refugee Status Determination (RSD) as well as UNHCR decisions pending the establishment of a national asylum system. The MoU also stipulates that all Malian refugees—including those in the Mbera camp who benefit from prima facie status as well as those recognized by UNHCR in urban areas—can access identification documents issued by the Ministry of the Interior and Decentralization's National Agency for the Population Register and Secure Titles (ANRPTS) (UNHCR, 2023).

According to indicators compiled by the UNHCR from various surveys, the socioeconomic situation of the Mbera refugee camp community is marked by a high level of multidimensional vulnerability. For example, the poverty rate among refugee households (59 percent) is much higher than the national average (28 percent), and the employment occupation rate of just 12 percent is substantially lower than the national average of 40 percent (UNHCR, 2023). However, there are areas where the UNHCR support has clearly had an impact, as the child malnutrition rate in the Mbera camp is marginally lower than the national average (17 percent versus 20 percent), and camp residents have much higher rates of access to drinking water and sanitation services (UNHCR, 2023).

2.2 The education context in the displaced refugee community

Based on the most recent data in December 2024, there are 7,854 students in primary and 898 students in secondary education in Mbera camp schools, of which 4,090 are girls (52 percent) (RET International, 2024). These enrolments were spread across eight primary schools in the camp, a number that increases every year.

There are also Malian refugees outside of the Mbera camp in urban areas that are enrolling in both private and public Mauritanian schools. These refugee communities are supported by UNHCR through an implementing partner. As of December 2024, there are 1,287 students in primary and 192 students in secondary education in urban areas, of which 798 are girls. However, the refugee education situation in these urban centers is difficult to monitor, and as a result less is known about their progress (Cambridge Education, 2022).

The primary and secondary schools in the Mbera camp follow the Malian curriculum in French and most of the teachers, school directors and pedagogical advisors are Malian nationals, recruited among the refugees in the camp. Salaries are paid by UNHCR (through RET International) and other education partners. Pedagogical advisors (conseillers pédagogiques) are supported and supervised by the district inspectorate (IDEN) of Bassikounou and provide support to school directors and teachers and closely monitor a range of functions and outcomes (Cambridge Education, 2022). Additional support comes from a technical working group set up by the various education partners that has monthly meetings under the leadership of the Bassikounou IDEN. The main difference between the Malian and Mauritanian education systems lies in the language of instruction: the Malian system has more hours of French, while the Mauritanian system has more hours of Arabic (RET International, 2024).

Participation rates in formal education in the Mbera camp are low. Data provided by UNHCR in April 2022 show that primary and secondary school aged children made up roughly 41 percent of the total Mbera camp population, but gross enrollment in primary in the 2021-22 school year was 37 percent, and just 6 percent at the secondary level (UNHCR, 2022). These gross enrolment rates are lower than the national Mauritanian averages for primary (56 percent) and secondary (19 percent) derived from the 2019-2021 Demographic and Health Survey (EDSM), and much lower than official figures reported through the UNESCO Institute for Statistics data portal (112 percent for primary in 2023, and 36 percent for secondary in 2020) (UIS, 2024; UNHCR, 2021a, 2023). **Amongst refugees, boys are more likely to be in school than girls (39 percent versus 35 percent) and secondary (6 percent versus 3 percent) (UNHCR, 2021).** Girls have also been observed to be more likely to be absent.

In addition to low overall rates of enrollment—which means a high share of refugee children are out of school—many children who do enter the education system are struggling.

The related challenges of enrolling—and retaining—refugee children in the Mbera camp education system are attributable to a number of factors. Forcibly displaced children are an extremely vulnerable group that can be impacted by the inherent instability of displacement, with additional psychological impacts from the conflict or displacement experience itself. There are also concerns that Malian refugee parents lack awareness of the vital importance of education, due in part to low levels of education (UNHCR, 2021a, 2023). All of these factors can contribute to late entry and grade repetition, and the resulting overage enrolment presents an additional challenge as the child's time becomes more valuable to the family (as a source of labor inside and outside of the home) and is competing with school attendance.

Another set of challenges are related to school quality. A recent evaluation of the Mbera camp schools highlighted a number of quality concerns (UNHCR, 2021a). These include issues related to infrastructure and inputs as well as management processes and capacity:

- Educational infrastructure in the camp and in the region is rudimentary (road, electricity, water, connectivity, etc.);
- Primary classrooms have high levels of crowding (upwards of 80 children per class in some cases), with insufficient benches (3 students per bench). Only 127 rooms are available for 6,500 children in primary school (average of 51 students per classroom);
- The classrooms are old, and the learning environment is not attractive;
- The use of Malian curricula and teaching staff complicates the quality assurance and assessment tasks carried out by the Mauritanian Ministry of Education district office that is responsible for these functions, and the curriculum and teacher certification processes are based on Malian standards, not Mauritanian;
- Only 23 percent of the 147 teachers and school principals are certified by the Ministry of Education of Mali;
- Only 15 percent of the Mbera camp education staff are female.

In sum, refugee children in the Mbera camp school system **face two sets of challenges: demand side factors related to the home environment** (poverty, parental attitudes) and **displacement experience, and supply side factors** in the form of school access and quality constraints.

2.3 Refugee education policy in Mauritania

The previous education sector plan for Mauritania (PNDSE II, 2011-20) did not make any reference to refugee education. This document was prepared before the Malian crisis and resulting in high numbers of refugee arrivals in 2012, but the subsequent triannual action plans also do not mention refugee education. As of June 2022, the Mauritanian government had initiated the process of developing a new PNDSE (III) as well as the Triennial Budgeted Action Plan for the education sector (PATB). With the support of international partners this process has included the framework law, the sectoral policy letter, and the diagnosis for the RESEN (Rapport d'État du Système Éducatif National in French), **which for the first time considered the presence and situation of refugees in the country** (RET International, 2024).

At the time of the 2022 review by Cambridge Education, education partners in Mauritania involved in refugee education, together with the Mauritanian and Malian Ministries of education, were exploring ways to further include Malian refugee students into the Mauritanian education system. A draft strategy for education in the Mbera camp was developed in July 2021. **This strategy describes a possible hybrid solution, based on the Malian education curriculum but gradually integrating content from the Mauritanian curriculum.** An important aspect of this integration is the exposure to Arabic which is the language of instruction for a number of subjects in the Mauritanian education system. Such an integration would also cover teacher recruitment and teacher training and continuous professional development (Cambridge Education, 2022).

Improving access to education for refugee children was mentioned as a key concern by many Ministry staff during the Key Informant Interview (KII) in May 2022 in Nouakchott and the Ministry took the engagement to include refugees in the national education system during the Global Refugee Forum 2023 that took place in from 13 to 15 December in Geneva. Reference was made especially to those refugee children living in the urban areas where education access is much less monitored by those education partners that usually intervene in support of refugee education. Another area of concern for the Mauritanian Government is the situation around the refugee camp in Mbera, a very poor area where host communities need equal support (Cambridge Education, 2022).

In summary, while significant efforts have been made to establish a functional education system in the Mbera camp, considerable challenges remain to be addressed to improve the quality and accessibility of education for all refugee children. With continued support from international organizations and the Mauritanian government, and through the implementation of new inclusive policies, it is hoped that these children will have access to quality education that will enable them to develop their potential and contribute positively to their host and home communities (RET International, 2024).

MAURITANIA. In 2023, Mauritania faced an influx of more than 55,000 arrivals, driven by the deteriorating security situation in neighboring Mali - a significant increase compared to 2022 (12,000 arrivals). ©**UNHCR/XAVIER BOURGOIS**

3. Research Framework

3.1 Aims and justifications

This study aims to address three data gaps in the evidence base on forcibly displaced children's education in Mauritania. **First, the analysis includes a detailed summary of grade 4 and 6 student performance in four foundational skill areas covering basic literacy and numeracy, socioemotional learning and executive functioning.** The existing evidence on progress indicators among Mbera camp students are focused on participation numbers and rates (enrolment, dropout, grade repetition). These are important outcomes, but it is also necessary to take stock of whether students are acquiring basic skills. This is the first time that achievement data have been collected among Malian refugee children in Mauritania.

Second, the analysis identifies factors that facilitate or impede student learning progress. This is done with quantitative survey data comparisons of student academic and non-academic outcomes between different student and school characteristics, and with qualitative data from Focus Group Discussions (FGDs) carried out with students, school staff and parents. This is not the first time that FGDs have been conducted in these schools (Cambridge Education, 2022; UNHCR, 2021a), but this data will provide an updated picture and, in combination with the quantitative survey data, can aid in providing a more holistic view of education in this context.

And third, the various outcomes and factors are compared between Mbera refugee camp schools and schools located in the local host community. These comparisons are useful for assessing progress in the Mbera camp against a reference point from the local Mauritanian community, both in terms of key student learning outcomes as well as school environments.

3.2 Research questions

The research questions follow from the three general aims of the study.
These include:

1. How do refugee students perform on an assessment of foundational skills (reading, math, socio-emotional and executive functioning)?

2. How do students' performance vary by gender, protection status (refugee vs. host), grade and locations?

3. What are the teaching and learning environment conditions like in refugee schools? What are the factors that facilitate or impede student progress?

4. Data and Methods

4.1 The HALDO instrument

The Holistic Assessment of Development and Learning Outcomes (HALDO), developed by Save the Children, is an interactive assessment tool that measures learning progress among populations impacted by crises. The instrument is applied as an interview in a one-on-one format. It provides a "holistic snapshot" of where children are in their learning across a range of foundational skill domains. The items have been sourced from validated instruments such as the International Development and Early Learning Assessment (IDELA) and International Social Emotional Learning Assessment (ISELA) (Krupar, 2019; Krupar & D'Sa, 2024).

The HALDO assessment is designed to be administered to children ages 4-12 (in early grades through end of primary) and has an adaptive structure to account for a wide age range of skill levels. Table 1 summarizes the content of the instrument in the four domains for literacy, numeracy, social emotional learning (SEL) and executive functioning (EXF). For literacy, numeracy and SEL, the initial content is classified at the *emergent* difficulty level. If learners are able to answer these questions, then they proceed to the next highest-level *intermediate*, and the adaptive process continues through *advanced* level content. Students who cannot answer emergent content are instead directed to foundational content to assess an even more basic set of skills (Krupar, 2019).

The strengths of the HALDO instrument include the ease of content adaptation to a local context, the relatively short amount of time required for application, and the "holistic" coverage across both academic (literacy, numeracy) and non-academic (SEL, EXF) skill domains. However, it is important to note that the tool is designed to provide a snapshot of basic skills rather than a detailed overview of the child's understanding of a general curriculum. This relative simplicity keeps the instrument short (about 30-40 minutes to administer) and is also appropriate for the kinds of vulnerable populations that are targeted for this kind of assessment.

MAURITANIA. Mbera camp and the hosting villages of Aghor and Lemgayes. ©**UNHCR/CAROLINE IRBY**

Table 1. Overview of HALDO content by domain

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 letter infrequent letters	Foundational
	Accuracy Reading with Comprehension	Number of words read correctly Respond to 5 comprehensions questions	Intermediate 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 double-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	0.5 - 0.75 = Intermediatel
		Ability to take the perspective of a third child in an ambiguous situation	
		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	

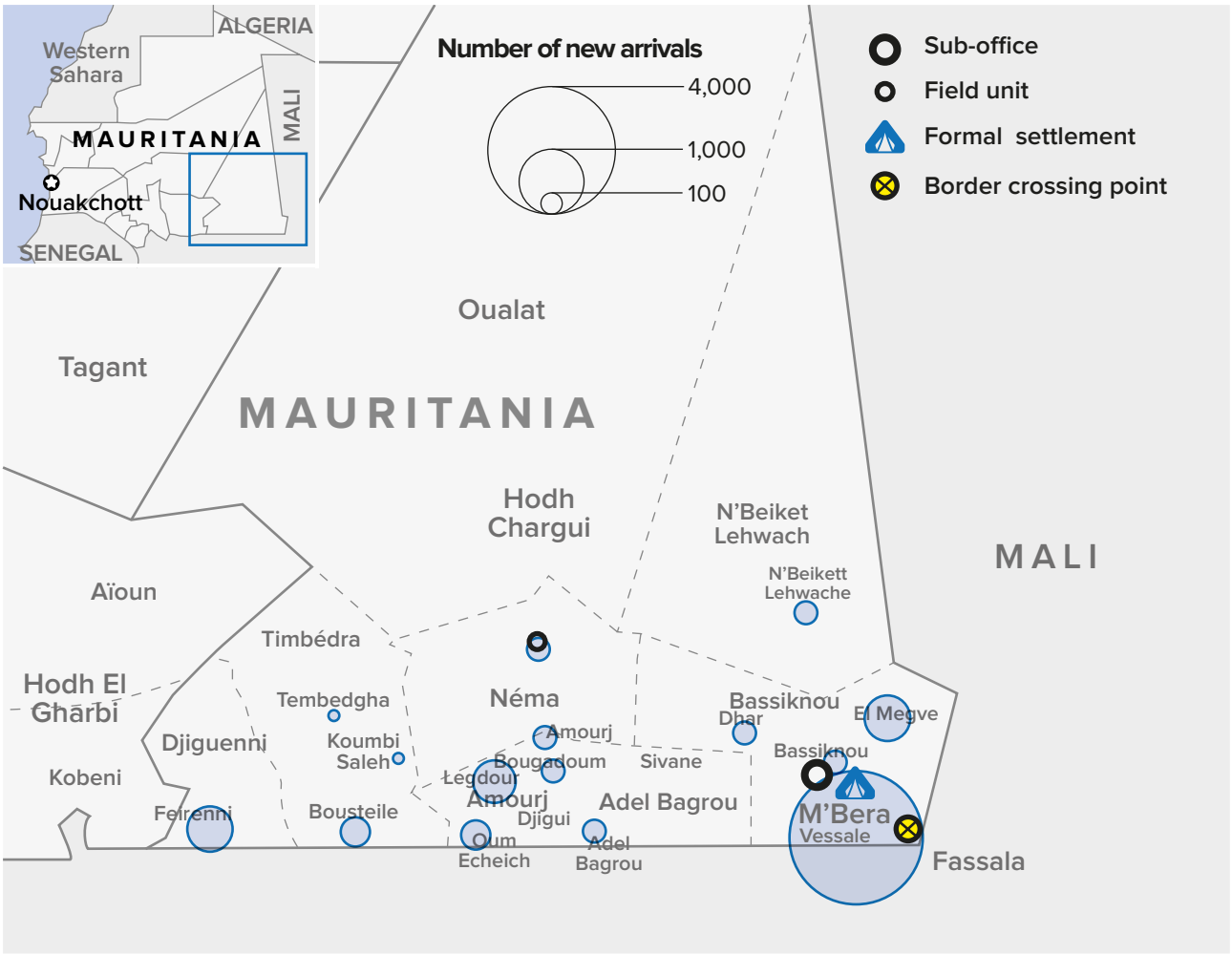
Source: Krupar (2019), Table 1

One complication with HALDO is the scoring and interpretation of results. First, the scoring is done in real time by enumerators, and for some items, the enumerator needs to decide on the scoring in an accurate and consistent manner. And second, the HALDO includes automatic recoding rules to simplify the application and calculation of overall percentage correct scores. Students that are unable to answer basic level content are automatically scored as zero (i.e. incorrect) for higher level content, while students who can answer higher level content are automatically assigned correct scores in the lowest level content. There is an underlying logic for these kinds of recodes: if a child cannot identify common letters, then they will not be able to read and understand a text. But there is the possibility that the recoding introduces some error into the calculation of overall scores. This is addressed in more detail in Annex A, which compares HALDO item and overall test performance (and characteristics) based on different scoring approaches using both classical and modern test theory tools.

4.2 Sampling and data collection

The HALDO assessment was conducted in July 2024 in all eight schools in the Mbera camp and four out of five host community schools in the commune of Fassala. The Mbera camp and Fassala district are located in the southeast corner of Mauritania along the border with Mali (see Figure 1).

Figure 1. Mbera camp and Fassala district locations



Students in the fourth and sixth grades were targeted, with a total of 509 students (256 girls, 253 boys) randomly selected from the lists of students enrolled in the fourth and sixth grades for the 2023-2024 school year: 400 in the Mbera camp and 109 in the Fassala schools (Table 2). In the host community schools, the school staff randomly chose 15 students in each grade to take the assessment, although the actual sample size is sometimes smaller or larger (Table 2). In the Mbera camp schools, the intended sample sizes were closer to 40 (grade 4) and 30 (grade 6), but again this was impacted by the number of students in attendance on the day. **The final sample included 505 students.** A team of ten trained enumerators carried out the data collection in each school, with each enumerator interviewing an average of five students per day.

A major challenge for the study is that the data collection took place during the school holidays. This required collaboration with school directors, teacher involvement, community relay interventions and the integration of recreational activities to gather students. Students were randomly selected from the full population of enrolled grade 4 and 6 students. In the host community schools, the sampled students were convened at the school, while in the Mbera camp schools all of the grade 4 and 6 classes were asked to come to school, and the actual sample was drawn from the population of students in attendance. In the camp schools, sampled students who were not present for the assessment were replaced by the next student (of same gender) on the school enrolment list.

The summary in Table 2 also provides the total enrolment by grade, gender and school type based on updated data from each school. Three things stand out. **First, in both sets of schools the grade 4 enrolment is more than twice as large as the grade 6 enrolment, which suggests high rates of grade repetition in grade 4 and/or attrition in later grades. Second, the gender ratio is fairly equal overall in grade 4 (although it does vary by school), but in grade 6 there are significantly more boys than girls in both school types.** And finally, there are two camp schools (numbers 7 and 8) where very few students showed up on the day of the assessment, and all of these students were included in the study.

Table 2. Summary of sampled school enrolments and samples by school type, grade and gender

	G4 TOTAL ENROLMENT			G4 TOTAL SAMPLED			G6 TOTAL ENROLMENT			G6 TOTAL SAMPLED			TOTAL		
	♀	♂	Total	♀	♂	Total	♀	♂	Total	♀	♂	Total	♀	♂	Total
Mbera camp schools															
Ecole 1	39	52	91	25	19	44	13	37	50	6	12	18	31	31	62
Ecole 2	52	83	135	14	23	37	17	24	41	14	10	24	28	33	61
Ecole 3	70	94	164	22	13	35	20	33	53	10	15	25	32	28	60
Ecole 4	118	94	212	17	17	34	38	37	75	7	10	17	24	27	51
Ecole 5	26	26	52	18	10	28	13	25	38	11	20	31	29	30	59
Ecole 6	31	26	57	21	16	37	8	23	31	12	4	16	33	20	53
Ecole 7	54	37	91	8	4	12	25	35	60	3	11	14	11	15	26
Ecole 8	51	31	82	7	8	15	18	20	38	4	5	9	11	13	24
Total	441	443	884	132	110	242	152	234	386	67	87	154	199	197	396
Fassala host community schools															
Fassala 1	70	65	135	3	8	11	30	45	75	8	3	11	11	11	22
Fassala 3	92	89	181	7	5	12	35	44	79	7	7	14	14	12	26
Fassala 4	22	17	39	3	3	6	7	10	17	12	12	24	15	15	30
Fassala 5	30	28	58	11	9	20	14	7	21	5	6	11	16	15	31
Total	214	199	413	24	25	49	86	106	192	32	28	60	56	53	109

Source: RET International 2024

Sampling weights are used to correct for discrepancies between each school's share of the sample and share of the overall population, as well as the gender ratio. For example, Fassala school 5 has 20 sampled students in grade 4 (out of 58 enrolled), while Fassala school 3 only has 12 grade 4 students in the sample out of an enrolment of 181. The sampling weight corrects for these imbalances and creates a weighted sample that has the same distribution of students by school and gender as is shown in Table 2 for Total Enrolment. The weighting strategy is not very consequential for the overall analysis, however, as comparisons of results with and without the weights are nearly identical; these are provided in the analysis sections below.

Finally, the samples cannot be treated as representative samples of the total population of students in the Mbera camp or host community schools. The fundamental challenge is related to selection bias since it is not possible to definitively rule out that the students who came to the school on the day of the data collection (in camp schools), or were chosen by school staff in the host community schools, are different than the students who did not come (or were not chosen). The weighting strategy does not correct for this issue, it simply adjusts the sample to ensure that students from larger schools have more weight in the sample, and the ratio of boys and girls is the same in the sample as in each school. Nevertheless, this sampling limitation **does not invalidate the study as the total number of sampled grade 4 and 6 students represents a large share of the total enrolment in these schools, about 30 percent.** The main concern is that the students who did come to the schools to participate in the assessment may be more engaged than the students who did not come, which could mean that outcomes like the HALDO assessment are higher in the sample than in the overall population.

The Focus Group Discussions (FGDs) were organized to gather additional qualitative insights. These discussions took place between July and August, immediately after the HALDO administration was completed. The FGDs focused on three key groups: students, parents, and teachers and principals. To ensure balanced representation, the discussions with students were organized with a focus on gender parity and included participants from both refugee and host communities. However, the teacher and parent samples are more weighted towards males. Table 3. provides a summary of the FGDs participants by school type.

Table 3. Summary of Focus Group Discussions (FGDs)

SCHOOL TYPE	CATEGORIES OF PARTICIPANTS	DATE OF FGDS	NUMBER OF PARTICIPANTS		TOTAL
			♀	♂	
Mbera Camp	Teachers	July 17, 2024	0	7	7
	Parents (fathers)	July 19, 2024	0	7	7
	Students	August 27, 2024	6	6	12
Fassala Commune	Students	August 24, 2024	6	6	12
	Parents (mothers)	July 28, 2024	10	0	10
	Teachers	July 28, 2024	0	8	8
Total			22	34	56

4.3 Methodology

The key methodological issues are related to the application of the HALDO instrument and Focus Group Discussions (FGDs), and survey data analysis. For the HALDO instrument, the data collection team received a two-day online training from Save the Children. Key aspects of using the HALDO tool were then reinforced through follow-up in-person and onsite refresher training. To ensure the accuracy and reliability of the assessment, the HALDO tools were pre-tested on about 30 students who were not part of the study sample prior to the actual data collection. This pre-testing phase allowed for the refinement of the tools, enhancing the quality of the data collected during the formal assessment. For example, one full day was used for two enumerators to work together as a pair to score the same child and compare the score and discuss any differences in the scoring.

The focus groups were conducted separately, and purposive sampling was used, to ensure that the discussions were tailored to the specific perspectives and experiences of the participants. The decision to group participants by gender during the FGDs was based on several cultural and social considerations, particularly in the context of the Mbera camp, where these discussions took place. Facilitators used a semi-structured interview guide that allowed for flexibility in exploring relevant topics while ensuring that key areas of interest were covered. The composition of the groups was designed to foster open dialogue within each demographic, ensuring that the views of both men and women, as well as those of refugees and host community members, were equally represented.

The qualitative data from the FGDs were reviewed by the team, and a detailed summary of the questions and responses were prepared with follow up comments for clarification. There was no coding of the FGDs data, or use of specialized software. However, the team performed meticulous **manual coding of the data, following inductively developed themes.** Selected themes were designed to reflect significant elements from the discussions, thereby ensuring precise and valuable synthesis for the final report.

The FGDs were conducted using a narrative analysis approach. This approach allowed for a focus on individual stories shared by participants to better understand their experiences, the events they have gone through, and the meanings they attribute to these events. The narratives collected were analyzed to identify contexts, main actors, and key themes with an emphasis on issues that were identified across different sets of respondents. The FGDs data are used to address research question 3 (see Section 3.2) related to teaching and learning environments and factors that impede progress.

The HALDO data analysis of learner scores and background information is based on four sets of activities. Basic **descriptive statistics** are computed using weighted data and standard errors that are corrected for the clustered sampling design. Comparisons across the main strata (gender, school type, etc.) are carried out using standard statistical tests for differences in means (i.e. t-tests). **Multivariate regression analysis** is used to identify significant predictors of student test scores (by domain and overall) from among the various student, family and school characteristics. Lastly, the HALDO assessment data are subjected to a series of detailed analyses based on both Classical Test Theory (CTT) and Modern Test Theory (e.g. IRT) tools to understand more about item performance based on different item

scoring strategies. HALDO technical analysis is summarized in Annex A. Additional details on HALDO scoring are provided in the results sections below.

The HALDO survey data analysis is used to address research questions 1 (student performance on the HALDO assessment sections) and 2 (factors associated with higher or lower scores). Research question 3 (differences between Mbera camp and host community) is cross-cutting and is addressed in both the FGDs data analysis and the HALDO quantitative data analysis.

Box 1: Ethical Considerations

RET International, who conducted the data collection for this assessment, prioritizes rigorous ethical measures to ensure that data collection activities respect the rights, dignity, and well-being of refugee children. These considerations include informed consent, confidentiality, emotional and physical safety, and alignment with international standards. The organization's robust protection policies and alignment with the Inter-Agency Network for Education in Emergencies (INEE) standards reflect a commitment to implementing interventions with integrity, inclusivity, and care. In addition, its data collection protocols comply with national laws and international standards, particularly the United Nations Convention on the Rights of the Child. All ethical measures are validated by a local ethical committee or entity to ensure contextual appropriateness. The following measures were put in place during the data collection process:

Informed Consent and Voluntary Participation: Participation in data collection processes was fully voluntary and informed. Written consent was obtained from parents or legal guardians before any activity. Additionally, children provided verbal assent following age-appropriate explanations of the evaluation's objectives and procedures. Participants were reminde*d of their right to withdraw at any time without repercussions to themselves or their families.*

Confidentiality and Anonymity: All collected data are anonymous to protect participants' identities. Databases and final reports adhere to strict confidentiality protocols, ensuring that sensitive information remains secure.

Ensuring Emotional and Physical Safety: Safe and supportive environment for children were ensured during data collection. Activities and questions were carefully designed to avoid causing emotional discomfort, stress, or trauma. Trusted adults were present during sessions to offer immediate support if necessary.

Training and Ethical Conduct of Teams: All survey teams received comprehensive training on child protection standards, ethical conduct, and sensitivity to the needs of children. This training ensures respectful, age-appropriate, and non-intrusive engagement with participants.

Protection from Sexual Exploitation and Abuse (PSEA): RET strictly adheres to its PSEA policy, ensuring all activities are conducted with respect for children's rights and protection. Rigorous supervision mechanisms were in place to uphold the highest ethical standards and prevent abuse.

Referral for Identified Needs: In cases where a child's vulnerability or specific need is identified, RET facilitates referrals to specialized services, such as psychosocial care or other forms of assistance, ensuring holistic support for participants.

4.4 Limitations

There are four main limitations of the study that affect the analysis and main findings.

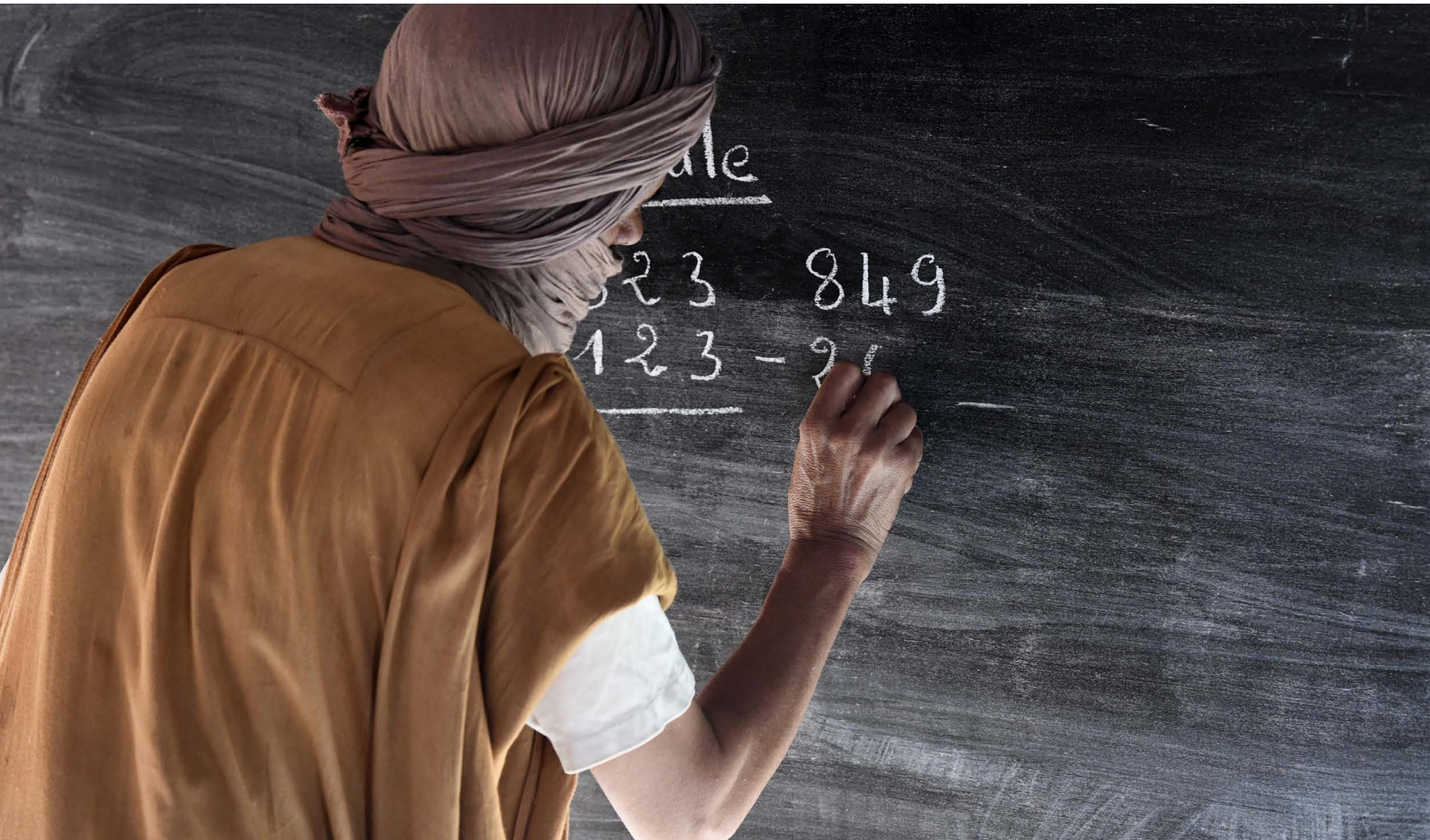
1. The samples cannot be assumed to be representative of the Mbera camp or local community schools due to the reliance on study populations made up of students who came to the schools to participate in the data collection. Also, for the Fassala commune schools the overall sample size is considerably smaller than the Mbera camp school sample.

2. Because of the high rate of overage enrolment in these schools (described below), a large share of the students that were assessed are older than the recommended HALDO target age range (4-12), although they are not outside of the intended grade range of HALDO (through the end of primary). One concern with overage is "ceiling effects", where the items are very easy and there is limited variation in the responses.

3. Due to tablet programming errors, HALDO was not implemented in exact accordance with the intended adaptive design. This means that some students incorrectly skipped some content. This is mainly a problem for constructing overall domain and HALDO averages that can be compared with HALDO results from other studies where the complete content was incorporated.

4. The quantitative data collection only draws on information from students, with very limited information on teachers or schools. This limits the scope of the quantitative analysis and the degree to which teacher and school effects can be analyzed, with a resulting emphasis on student and family background characteristics.

MAURITANIA. In 2023, Mauritania faced an influx of more than 55,000 arrivals, driven by the deteriorating security situation in neighboring Mali - a significant increase compared to 2022 (12,000 arrivals). ©**UNHCR/XAVIER BOURGOIS**



5. Findings

The data analysis findings are divided into five sections covering student background characteristics (5.1), numeracy and literacy assessment results (5.2), socioemotional learning (SEL) and executive functioning (EXF)(5.3), overall HALDO scores and zero scores (5.4), comparisons of HALDO results (5.5) and summary of focus group discussions (FGDs)(5.6).

5.1 Student background characteristics

The HALDO tool includes a short interview component to cover some basic features of students and their families. Table 4 summarizes the results based on weighted data; Table B1 in Annex B provides the same results using un-weighted data, and there are few differences between the two tables. **The main finding is that there are significant differences between the refugee and host community (Fassala) students (Table 4).** The host community students are significantly younger, more likely to speak Hasania, have more household possessions in the home (SES factor), and are much less likely to report French language is used in their classrooms.

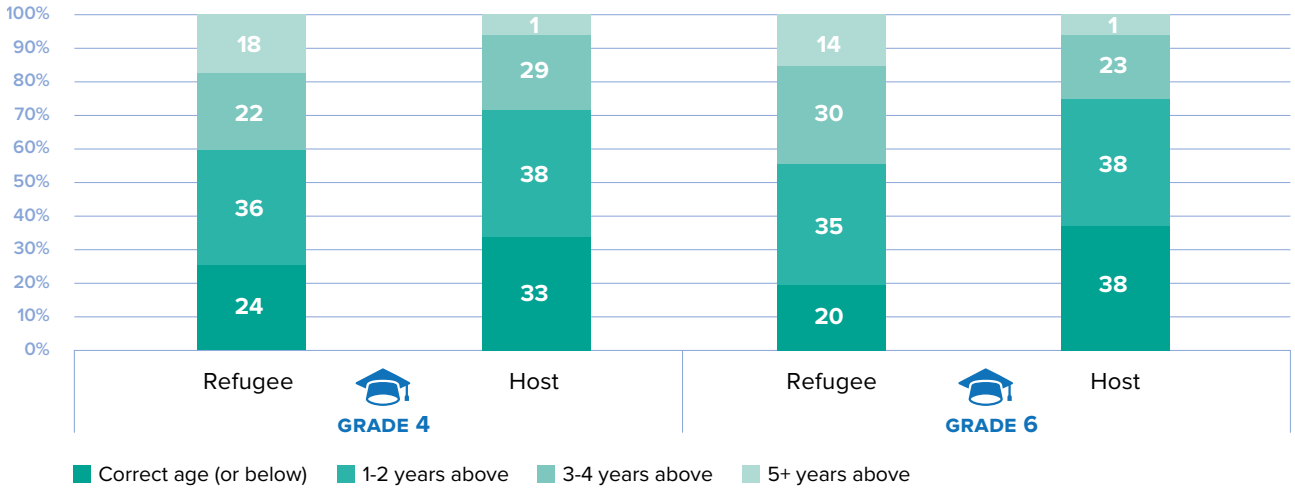
Table 4. Student background characteristics by grade and school type

CHARACTERISTICS	GRADE 4		GRADE 6	
	Refugee	Host	Refugee	Host
Female students	49.9	51.8	39.4	44.8
Student average age in years	12.4	11.6+	14.3	13.3*
Parent has phone number	41.2	28.5+	67.6	58.6
Languages spoken at home:				
Hasania	34.9	100.0*	47.4	98.1*
French	2.1	0.0	11.6	3.3+
Sonrai	2.4	0.0*	6.3	0.0*
Tamasheq	69.5	0.0*	65.9	0.0*
Arabic	3.8	0.0*	0.6	9.8*
Student lives with both parents	73.2	73.1	73.9	77.4
Student has any disability	1.4	0.7	1.8	0.6
Student has lived here:				
Less than one year	4.2	0.0*	4.1	0.3+
1-2 year	5.4	9.1	7.3	7.6
More than 2 years	88.5	90.9	84.9	91.7+
Has books in home	36.9	47.3	54.7	65.4
Someone reads books in home	27.2	43.8*	48.2	51.8
SES Factor	-0.15	1.10*	-.0.24	0.41*
Class is only in French language	36.8	3.9*	46.5	0.6*
Sample size (n)	242	49	154	60

Among refugee students there is some language variation, as sizeable shares report speaking Hasania in the home, but the most common language is Tamasheq. Also, 37 percent of grade 4 students in camp schools, and 47 percent of grade 6 students, report that their class is only taught in French (compared with less than five percent of host community students). The survey data confirm that recent arrivals are not very prevalent among refugee student families, as nearly 90 percent of the students reported that they have lived in their current residence for more than 2 years.

Figure 2 provides a more detailed summary of the age of the sample students by grade and school category using four categories for correct age (or below), 1-2 years older than correct age, 3-4 years older and 5+ years older. Standing out is the age profile of refugee students, with very large shares above the expected ages of 10, for grade 4 and 12, for grade 6. **Forty percent of refugee grade 4 students in Mbera are 13 years or older (3+ years above correct age), while in grade 6 almost half (44 percent) are 15 or older.** These age profiles are significantly older than the host community school students, although large proportions of those students are also overage.

Figure 2. Student age by grade and school category

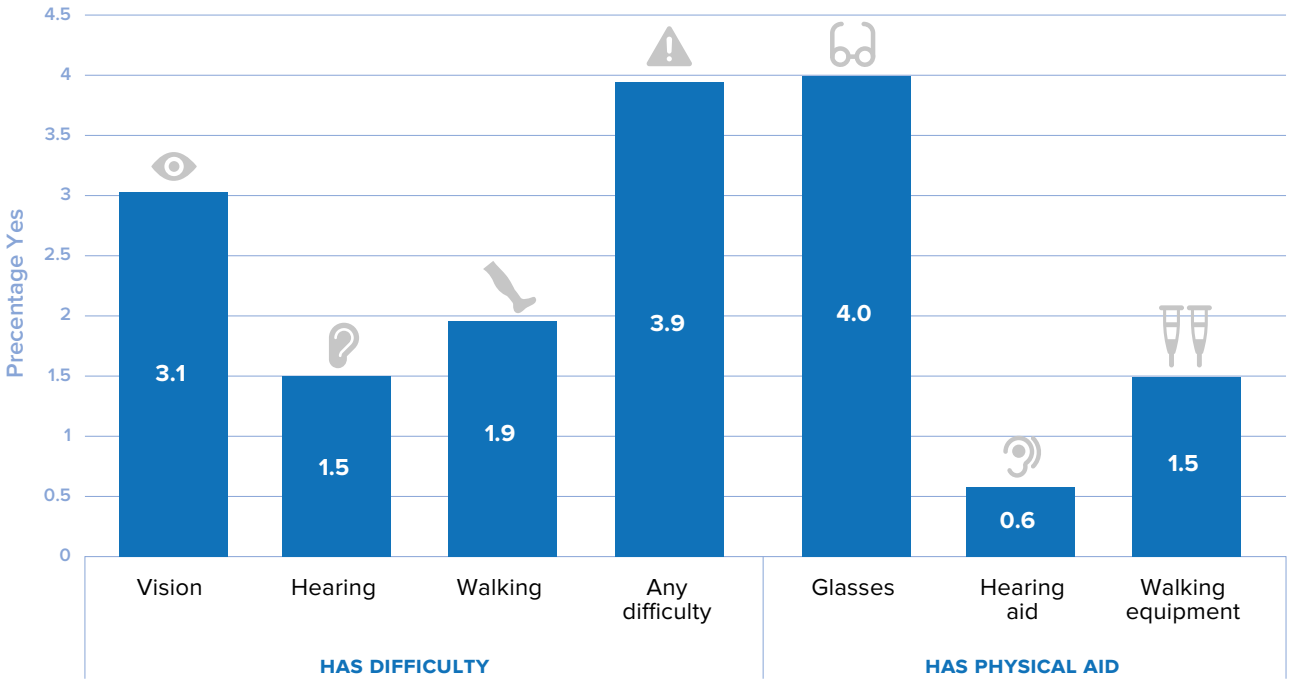


Data source: RET Mauritania HALDO assessment (2024)

Figure 3 continues with a summary of student disabilities, which is a point of emphasis in HALDO. Students were asked a series of questions about difficulties they experience related to vision, hearing and walking. **Overall, there are not a lot of reported difficulties, as just under four percent of the entire sample reported experiencing at least "some difficulty" with any of the three areas.** The most common issues are related to vision, as four percent of the sample reported using glasses and three percent (including some with glasses) reported they had at least some difficulty with vision.

The low number of students with disabilities raises some questions about access and inclusion in the camp and host community schools. This topic is addressed in the focus group discussions, but if schools are poorly prepared to diagnose disabilities, and provide support for children with different limitations, then these students may not be enrolling in school.

Figure 3. Percentage of students who report having some difficulty by physical area, and use of physical aids, whole sample

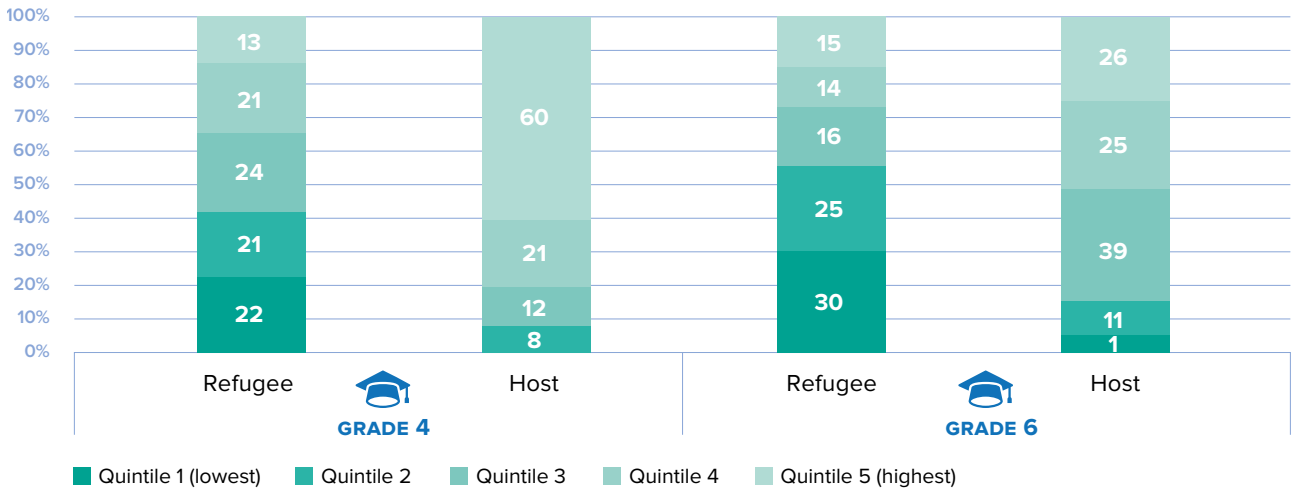


Data source: RET Mauritania HALDO assessment (2024)

Figure 4 compares socioeconomic status (SES) quintiles by grade and school category. The SES indicator is created from the HALDO questions about household possessions (e.g. television) and home characteristics (cement floor, etc.). Factor analysis was used to create an SES index, and then this was divided into five equal quintiles that are ordered from lowest (quintile 1) to highest (quintile 5). **As expected, the Mbera camp refugee students report significantly fewer household possessions (including animals like chickens and cows) than their host community counterparts.** About half of the refugee students report having electricity in their homes, compared with over 80 percent of host students. Host community school students are twice as likely to have a television, and nearly 30 percent report having a car (versus just 5 percent of Mbera camp students).

The differences between camp and host students in the individual components of the SES indicator translate into very large gaps in SES quintile. Roughly half of refugee students are in the lowest two SES quintiles (in Figure 4), compared with about 10 percent of the Host community children. By contrast, most of the host community fourth graders (and half of the grade 6 students) are in the two highest quintiles. However, it is important to note that the relatively higher SES levels among host community students does not mean that these students are affluent: **vulnerability appears to be widespread in the camp and host communities, but the host community children are clearly coming from households with more economic resources.**

Figure 4. Socioeconomic quintile summary by grade and school category



Data source: RET Mauritania HALDO assessment (2024)

Overall, the HALDO background indicators provide some useful details about the living conditions of the children in the sample. Not surprisingly, there are a number of significant differences between refugee and host community students, especially related to language, age and household resources (SES). However, there is also significant variation in many indicators within the schools and school categories, so it will be important in later stages of the analysis to consider how these background factors are associated with student scores on the HALDO instrument. Finally, one area of emphasis on the HALDO questionnaire—student disabilities—shows that relatively few students are reporting physical difficulties.

5.2 HALDO assessment results: Literacy and Numeracy

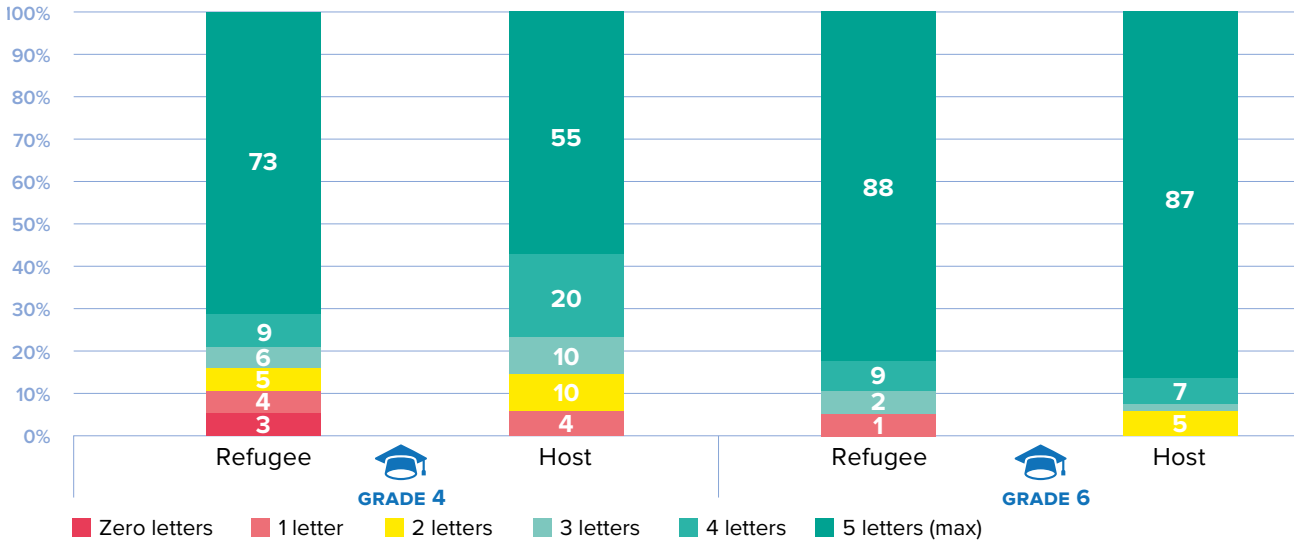
The core content of HALDO includes the literacy and numeracy assessments. In this section each component is summarized separately, with a detailed overview of how the questions were applied according to the adaptive framework, the results for each question, and a summary of the overall scores and performance levels. The data summaries are presented by grade and school category (Camp versus Host).

5.2.1 Literacy

HALDO literacy begins by asking all students to identify 5 commonly used letters ("Literacy 1"). **Nearly 75 percent of the 505 sampled students were able to identify all five common letters, and less than two percent (or 9 students) were unable to identify any letters.** Figure 5 provides the frequency-based summary for the number of correct letters from 0 to 5, by gender and school category (refugee vs. host community schools), with each bar adding up to 100 percent of the respondents in that category. Seventy three percent of grade 4 refugee students identified all five letters correctly (dark green shading), while four percent only identified one letter correctly (light red shading), and three percent were unable to identify any letters (dark red shading). In grade 6 the results are noticeably higher, as nearly

90 percent of all students identified all five letters and very few students identified fewer than four letters. **Refugee students outperformed the host community students in grade 4, but for grade 6 the scores are very similar.** Statistical comparisons of refugee versus host students are provided after the descriptive summary of each literacy item.

Figure 5. Frequency for number of correctly identified common letters (maximum=5) by grade and school category

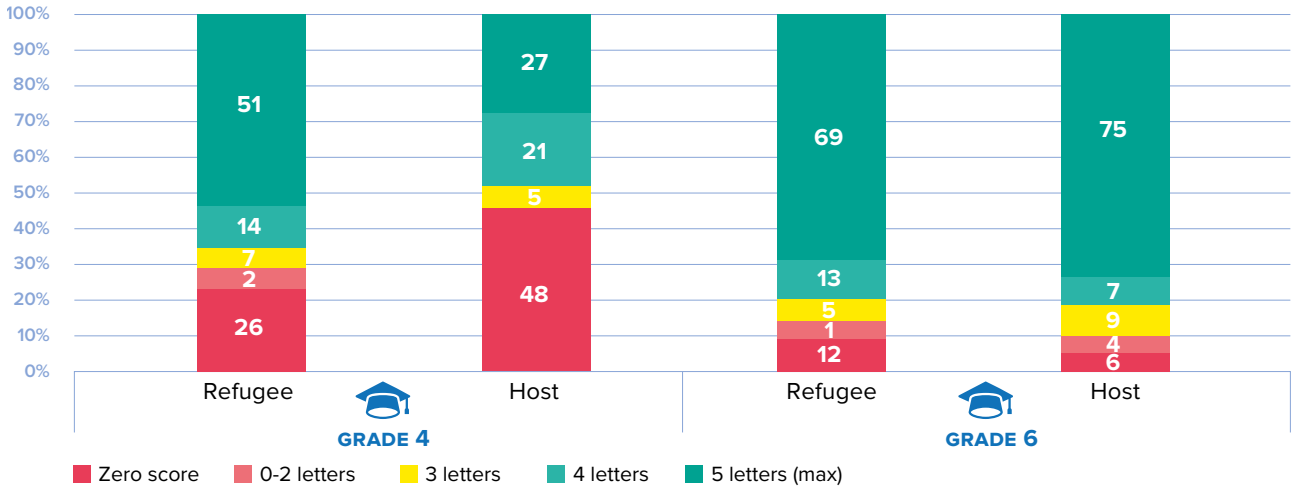


Data source: RET Mauritania HALDO assessment (2024)

Students who were able to identify all five letters skipped ahead to infrequent letter identification, while students who were unable to identify all five common letters were directed to the expressive language basic content ("Literacy 1.1"). The expressive language activity asked students to identify up to 10 animals, it did not include any kind of reading and was purely oral. The results are summarized in Figure B2 in Data Annex B. It is important to restate that 75 percent of the total sampled students (both refugee and host) were able to identify all five common letters, and therefore a much smaller sample of students answered the expressive language item: 87 grade 4 students, and just 27 grade 6 students. **The results show that for the entire sample less than 20 percent of this sub-sample were able to name 10 animals, and a sizeable share of the students—especially in the Mbera camp—were unable to name more than six animals (Figure B2).** The students who were unable to identify all five common letters in Literacy 1, and then were unable to name more than six animals in Literacy 1.1, make up the most vulnerable sector of students in terms of basic literacy skills. This group is returned to below in the summary of the overall HALDO literacy scores and performance levels.

Following the adaptive structure of HALDO, the 75 percent of the total sample (or 390/505 students) that identified all five common letters were then asked to identify five infrequent letters ("Literacy 2"). Figure 6 provides a comprehensive overview of this Literacy 2 item that combines two important aspects of student performance. First, the "Zero score" share for each sample is highlighted with dark red shading and includes the students who did not correctly answer the Literacy 1 item and are therefore automatically assigned a zero in the HALDO scoring framework. **In grade 4 this share is quite large, especially in the host community schools where nearly half of the sample did not advance past the initial basic literacy items on HALDO.** The zero scores are less common in the grade 4 Mbera camp classrooms (26 percent), and in grade 6 in both school categories.

Figure 6. Frequency for number of correctly identified infrequent letters (maximum=5) by grade and school category



Data source: RET Mauritania HALDO assessment (2024)

The second input in Figure 6 is the frequency of correct answers among the students who did complete this item by grade and by school category. These results are similar to those for the common letter identification and show that **most of the students who answered the item were able to identify at least four of the five letters.** In grade 4 refugee students were more likely than the host community students to identify all five letters, while in grade 6 the percentage of students identifying all five letters was substantially higher, but similar by school category. Almost all students were able to identify 4 or more letters, and just 11 percent of the total sample only identified fewer than four letters (yellow and light red shading in Figure 6).

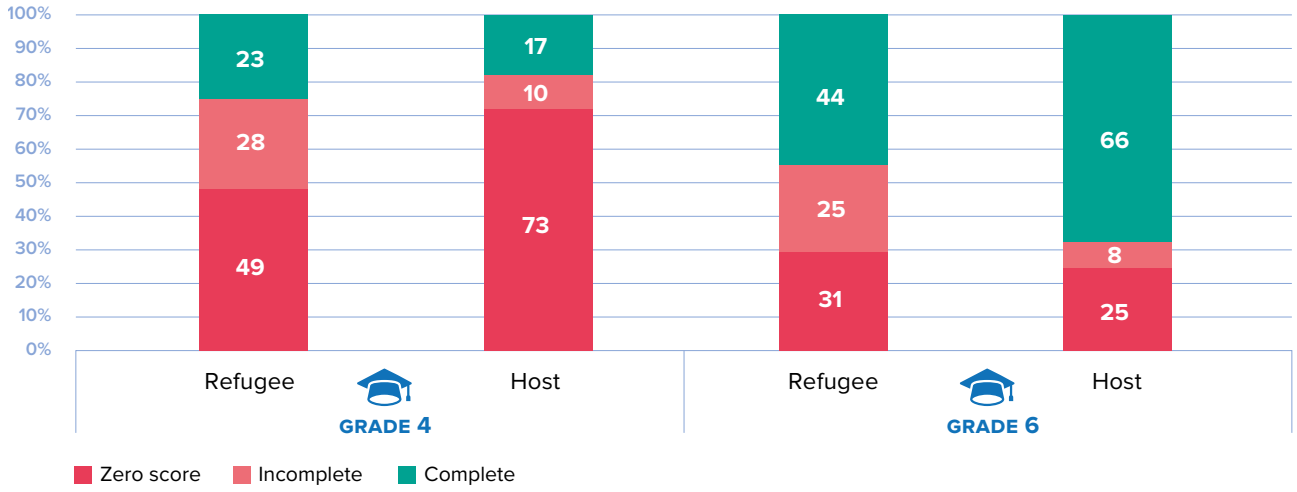
Students who were unable to identify all five infrequent letters exited the HALDO literacy assessment, while those who did identify five letters continued to the two-part reading fluency questions ("Literacy 3" and "Literacy 4"). The first part asked students to read the first five words of a written passage. Children who could read all five letters were assigned a complete score, while students who were unable to read all five words received a zero and exited the HALDO literacy assessment. The student scores were not coded in terms of the number of correct letters they were able to read, in the scoring they only receive a 0 (incorrect) or 1 (correct).

Figure 7 provides the two-part summary of student performance on the first item for reading fluency by grade and by school category. **The first result that stands out is that a sizeable share of the sample has exited the HALDO literacy assessment before answering the Literacy 3 item, and therefore they are automatically assigned the zero score.** This share includes half (49 percent) of the grade 4 refugee camp sample, 73 percent of the grade 4 host community sample, and 25-31 percent of the grade 6 samples.

Of the 279 out of the total 505 sampled students who advanced to the Literacy 3 item, 57 percent were able to read all first five words of the passage. As expected, this share is much higher in grade 6 than in grade 4. One result that stands out in Figure 7 is that relatively large shares of grade 4 refugee students were unable to read all five words ("Incomplete"),

and as a result the share who successfully completed Literacy 3 (23 percent) is not much larger than in the host schools (17 percent). This pattern continues in grade 6, where a larger percentage of grade 6 camp students was unable to read all five words (25 percent versus just eight percent in host schools).

Figure 7. Percentage of students who correctly read the first five words from reading passage by grade and school category



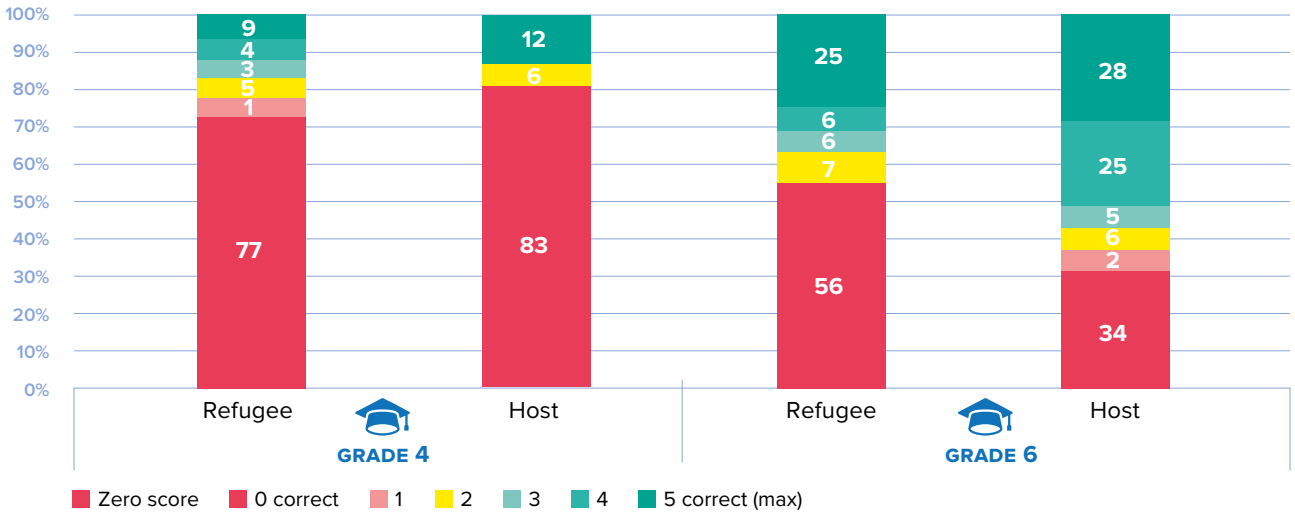
Data source: RET Mauritania HALDO assessment (2024)

The second part of reading fluency ("Literacy 4") was only completed by the 158 students who correctly read all five words in the first part (i.e. the green shaded shares in Figure 7). The students were asked to read the rest of the written passage, and the enumerator indicated the number of words the student did not read correctly. These results show that almost all of the students were able to read the passage, as more than half did not misread any words, and only 10 percent made three or more mistakes. Since the Literacy 4 item results do not vary much, they are not presented in detail.

The final element of the HALDO literacy assessment is reading comprehension, which is coded as "Advanced" content in the overall HALDO summary (see Table 1 in Section 4.1). This set of questions was also restricted to the 158 students who were able to correctly read the first five words of the reading fluency passage. They were then asked a series of questions about the passage they had just read, and the scores were coded as the number of correct answers. For the 158 students who reached this item, the overall average for the grade 4 and 6 students on reading comprehension was just under four correct answers (3.9) out of five.

The summary in Figure 8 highlights the very large share of zero score students who did not advance to the reading comprehension section of HALDO: roughly 80 percent of all grade 4 students, 56 percent of grade 6 refugee students, and 34 percent of grade 6 host community students. For the remaining students the results on the reading comprehension activity are generally positive, as nearly half of the 158 students answered all five comprehension questions correctly, and 66 percent answered at least 4 questions correctly. Only 11 students (out of 158) were unable to answer more than one question correctly.

Figure 8. Frequency for number of correct answers for reading comprehension questions (maximum=5) by grade and school category



Data source: RET Mauritania HALDO assessment (2024)

The summary of the raw scores for each of the five HALDO literacy items provides an initial overview of how grade 4 and 6 students performed on the assessment, with some comparisons between refugee and host community students. A critical element in this summary is the cumulative share of students who are not advancing in the HALDO literacy assessment. One potentially misleading result is that the raw score summaries generally show high levels of performance on each item: in all five items the majority of students received either a perfect score or missed just one answer. **However, because of the adaptive structure of HALDO, this picture is problematic since each successive item in the literacy assessment is applied to a smaller sample of students.** This is vividly demonstrated in the Literacy 5 item for reading comprehension (Figure 8), where only a small sub-set of grade 4 students answered the item, and even in grade 6 about half of the original sample did not advance this far. In other words, the sample is getting more and more selective with each item, and therefore it is not surprising that even for the most advanced content (like reading comprehension) the raw scores based on actual answers are fairly high overall.

The automatic recoding (zero scores) is a necessary component of HALDO since it is designed to move through the assessment quickly with each child and avoid asking them questions that they do not really have any chance of answering correctly. Furthermore, given the adaptive structure the only way to calculate an overall percentage correct score—which is a useful summary for the end-users of HALDO data—is to incorporate the automatic recodes. **Nevertheless, there are some potentially strong assumptions being made with these automatic recodes.** For example, if a student only identified four of the five infrequent letters in the Literacy 2 item (Figure 6) then she would automatically receive a zero score for both reading fluency items (Literacy 3 and 4) and the reading comprehension item (Literacy 5). But this is almost certainly not an accurate reflection of actual student skills: it is likely that at least some of the students who were able to identify four of the five infrequent letters can also read with adequate fluency and can perhaps correctly answer some of the reading comprehension questions. **In other words, the HALDO automatic recodes are likely to have the effect of making the student's overall reading (or numeracy) ability look lower than**

what it really is. The critical question is just how large this potential bias is. This question cannot be definitively answered. **But based on the analysis in Annex A it does not appear that the automatic recodes are a major source of bias in the HALDO analysis.** This topic is returned to in the numeracy analysis below, and the general discussion in Section 6.

Table 5 provides the global summary of the HALDO literacy assessment based on the automatic recoding scheme; Figure B3 in Annex B provides a visual summary of the results for the Mbera camp students only. For each of the five literacy items both the percent correct, and average number of correct answers are presented by grade and school category using recoded answers that automatically assign students the zero score when they did not answer the item (or for Literacy 1.1 the more advanced students are automatically assigned a perfect score of 10).

Table 5. Summary of HALDO literacy item and overall percent correct and points by grade and school category

CHARACTERISTICS	GRADE 4		GRADE 6	
	Refugee	Host	Refugee	Host
1. Letter identification (common)				
Percent correct	86.4	82.0	96.8	97.7
Number correct (max=5)	4.3	4.1	4.8	4.9
1.1. Expressive language				
Percent correct	87.7	83.1	95.6	99.1+
Number correct (max=10)	8.8	8.3	9.6	9.9+
2. Letter identification (infrequent)				
Percent correct	67.0	46.1	82.2	87.1
Number correct (max=5)	3.4	2.3	4.1	4.4
3. Reading fluency (initial)				
Can read all five words	23.3	17.2	43.7	66.4
4. Reading fluency (complete)				
Percent correct	21.9	16.7	42.1	63.1
5. Reading comprehension				
Percent correct	16.3	13.8	35.7	54.4
Number correct (max=5)	0.8	0.7	1.8	2.7
Overall summary:				
Percent correct (equal)	50.4	43.2	66.0	78.0
Percent correct (points)	65.6	58.3	78.3	85.8
Number correct (max=27)	17.7	15.7	21.2	23.2
Sample size (n)	242	49	154	60

Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to averages (means) by grade and category based on weighted data. Tests of significance for differences between Camp and Host school categories (see Host column) are based on robust standard errors that correct for student clustering.
*Statistically significant difference between Camp and Host average at p<=0.05 level
+Statistically significant difference between Camp and Host average at p<=0.10 level

At the bottom of Table 5 there are three indicators of overall performance on the HALDO literacy assessment. One takes a simple average percentage correct across the five literacy items ("equal"), and the other adds up the total number of correct answers and divides by the maximum (27, see "points") to create a modified percentage correct; this total sum of answers is also provided in Table 5 ("number correct").

There are five main results in Table 5. First, **the student literacy scores are generally high in the more basic content for common letter identification, expressive language, and infrequent letter identification.** In these three literacy activities a large share of students successfully completed all of the tasks or had just one or two incorrect answers.

Second, **there is a clear drop off in performance for the reading fluency and comprehension literacy components.** For example, only about 20 percent of the overall sample of grade 4 students received a full score for the initial reading activity (Literacy 3), and the average grade 6 student correctly answered only about 40 percent of the five reading comprehension questions. The lower scores are due primarily to the large share of students who were automatically assigned zeros for these questions, and a somewhat larger share of students who were unable to complete the first part of reading fluency (see Figure 7).

Third, **the grade 6 scores are consistently higher than the grade 4 scores, and the gaps get much larger in the more advanced literacy content (e.g. reading fluency and comprehension).** Table 5 summary does not include tests of significance to compare grade 4 and 6 averages, but in a separate analysis almost all of these differences are statistically significant (i.e. grade 6 is significantly higher than grade 4). This is important because it does suggest that there is **learning taking place in the Mbera camp and host community schools,** although the difference between grade could also be related to student attrition.

Fourth, **the differences between school categories (camp versus host) are not significantly different.** The refugee camp grade 4 students have higher averages than the host students, while in grade 6 the opposite is true. However, these differences are not statistically significant, even though in some cases they are fairly large. One limitation is the small sample size for the host community schools which makes it more difficult to establish statistical significance.

Finally, **the main finding from the summary in Table 5 is that literacy skills in the sample are below the expected levels for this age and grade range.** The overall sample percentage correct average is about 80 percent in grade 6 when based on the total points on the assessment. But this is very basic content for grade 6 students who should have no problem with letter identification, reading basic words and answering simple comprehension questions from a short passage. In grade 4 the expectations are lower, but the averages of 60 percent or lower suggest that the average student is not proficient in this basic set of literacy tasks.

Table 6 concludes the literacy analysis with a summary of literacy performance levels. These levels are created mechanically from the adaptive sequence of literacy items and are not based on a fixed set of performance (or proficiency) standards. A skills-based summary provides a somewhat more detailed overview of student performance than percentage correct.

The color-coded summary in Table 6 shows that 12.9 percent of the overall sample is categorized as Level 1, meaning they were unable to identify common letters and only able to

name five or fewer animals in the expressive language item. Nearly one quarter (24.2 percent) of the host community grade 4 students are in this category. This is followed by Level 2, which is the most common performance level for the entire sample (35.2 percent). Students in this level are either able to identify all common letters but cannot identify all infrequent letters, or they are unable to identify common letters but perform well on the expressive language item.

Only 17.1 percent of the total sample is classified as Level 5 in literacy. These students have complete reading fluency and are able to answer four or five of the reading comprehension questions. Only about 10 percent of the total grade 4 sample (both refugee and host) reached this level, compared with 29.1 percent of Mbera camp grade 6 students, and 42.6 percent of host community grade 6 students.

Table 6. Summary of literacy performance levels by grade and school category

DESCRIPTION OF LEVEL	OVERALL	GRADE 4		GRADE 6	
		Refugee	Host	Refugee	Host
1. Cannot identify all common letters, can name less than 5 animals	12.9	13.9	24.2	4.8	0.7
2. Cannot identify all common letters but can name more than 5 animals, or can identify common letters but cannot identify all infrequent letters	35.2	35.0	49.1	26.2	24.7
3. Can identify infrequent letters but cannot read first five words, or can read first five words but have multiple errors on full reading passage	27.5	35.0	9.5	30.8	24.7
4. Have reading fluency but only answered 0-3 reading comprehension	7.3	7.2	5.6	9.1	7.3
5. Have reading fluency and correctly answered 4-5 reading comprehension	17.1	8.9	11.6	29.1	42.6
Total=100 percent	100.0	100.0	100.0	100.0	100.0
Sample size (n)	505	242	49	154	60

Data source: RET Mauritania HALDO assessment (2024)

5.2.2 Numeracy

The HALDO numeracy component has a nearly identical design as literacy. All students begin with basic level content and then proceed to progressively harder content or, if they are unable to correctly answer the basic level item, they are given an easier task. Items 1, 2, and 3 are adaptive and the students only proceed to the more difficult content if they get a complete score on the item. One difference for numeracy is that the final set of items (word problems) are answered by all students, regardless of performance on previous items.

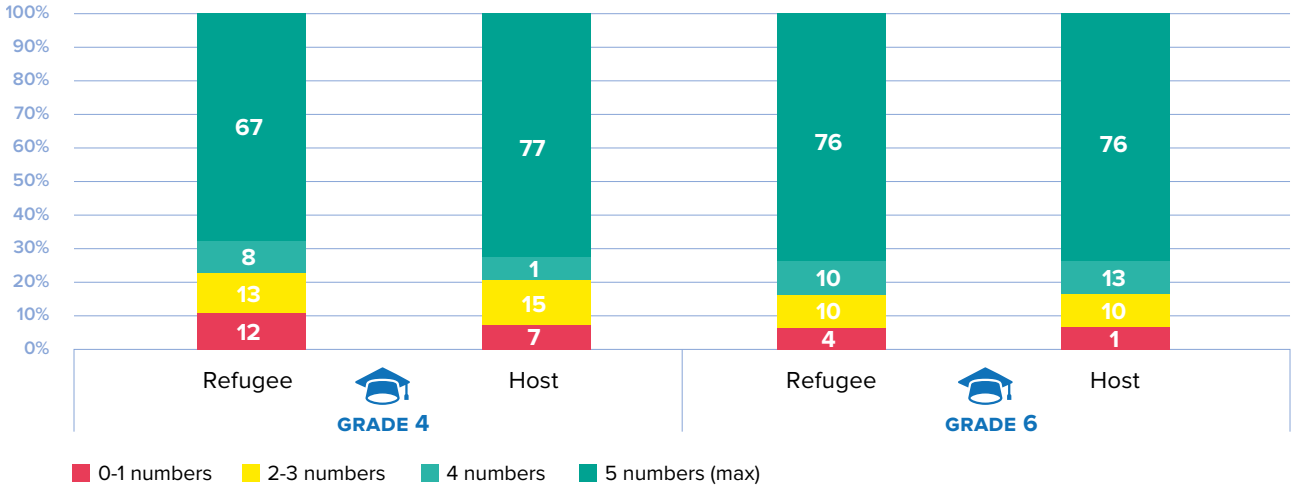
One complication with the numeracy analysis is that there were some errors in the tablet programming, and as a result the data are not complete.[1] This does marginally limit the scope

1 First, students who did not get full scores on item 1 were then directed to the more basic content in items 1.1-1.3, and then they were supposed to skip to the end of the assessment to complete the last two-word problems. This skipping was not implemented correctly for all students, so the only option is to assign these students zero scores for the word problems since they were unable to complete the first numeracy item. But since some of these students did actually answer the word problems—and some of them did get correct answers—

of the numeracy analysis and complicates calculation of an overall numeracy (and HALDO) score because it is not always possible to implement the automatic recoding. However these problems affect a small share of the students, so the analysis can be completed as designed for almost all students.

The summary in this section is not as detailed as for literacy since it follows a very similar sequence. Figure 9 shows the student scores by grade and school category for the Numeracy 1 item, which is identifying simple numbers. **Large shares of students were able to identify all five numbers, although there is a group of grade 4 students (about 10 percent) that is performing quite poorly (red shading).** One surprising result in Figure 9 is that the grade 6 average is not much higher than the grade 4 average on this content.

Figure 9. Frequency for number of correct answers for simple number identification (maximum=5) by grade and school category



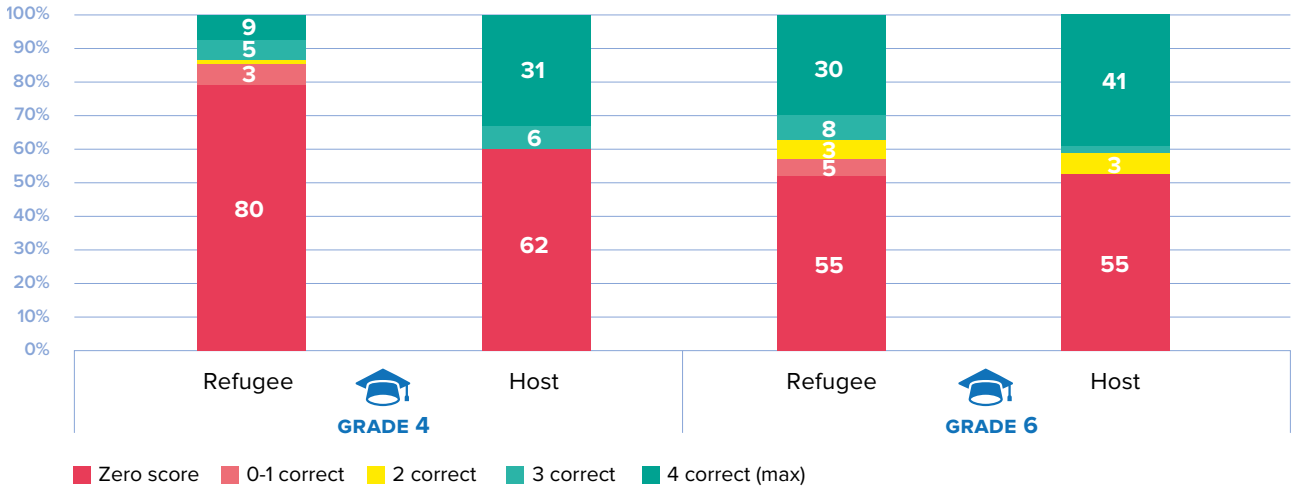
Data source: RET Mauritania HALDO assessment (2024)

The next three items in the HALDO numeracy sequence are summarized in Annex B. This includes the very basic one-to-one correspondence content (Numeracy 1.1) completed by students who could not identify all five common numbers, identification of five "hard" numbers (Numeracy 2) and then five "simple" addition and subtraction operations (Numeracy 3). These results follow a similar pattern to what was shown in literacy. Almost all of the students who were diverted to the basic one-to-one correspondence activity (students are asked to select a number of things from a group of beans or sticks) were able to complete all tasks (Figure B4). But with each successive item a larger and larger share of students are falling into the zero-score category. **For example, only about half of the grade 4 sample successfully advanced to the Numeracy 3 item for simple operations, and even in grade 6 a sizeable share of students (about 35 percent) did not reach this content.** Figure 10 skips to the hard operations (addition and subtraction with larger numbers), which is the second-to-last content on HALDO numeracy (Numeracy 4) and is the most difficult (Advanced in Table 1).

kind of recoding is problematic; this is returned to in the discussion in Section 6. Second, numeracy item 3 was incorrectly coded as 5 for some students (the maximum is 4), and these students did not continue to item 4. This means that item 4 data is missing for this group, as there is no way to automatically recode item 4 for this group. Third, students who answered any of the questions for item 3 were supposed to continue on to item 4, but in reality, the only students who continued to item 4 were those who received a full score on item 3. The solution for this problem is to automatically recode item 4 as zero for students who did not continue past item 3. The HALDO adaptive progression rules are also discussed in more detail in Section 6.

The results are again generally consistent with the literacy results at this level of difficulty: a much smaller share of grade 4 students (refugee and host combined) is proficient (about 17 percent overall), while in grade 6 the share is larger but still only about 34 percent of the total. Almost all of the other students are in the zero-score category, meaning they have not successfully completed the content leading up to the Numeracy 4 item. About 5 percent of the total sample includes students who were able to answer three of the four hard operations questions (light green shading).

Figure 10. Frequency for number of correct answers for hard operations (maximum=4) by grade and school category



Data source: RET Mauritania HALDO assessment (2024)

The final numeracy task in HALDO (word problems, Numeracy 5) is summarized in Figure B7. The word problems in numeracy are not categorized in terms of difficulty (see Table 1 in Section 4.1) as they are read out loud to the students, and do not require actual reading. All of the students were supposed to answer the first word problem and, if successful, continue with the second problem. In reality not all of the students were given this content (see footnote 2 for the summary of why), which required some automatic recoding to complete the data summary in Figure 10. **Nearly 60 percent of the total sample were able to answer both (orally applied) word problems, while roughly 25 percent scored a zero.** This is another content area where there is not much difference between grades 4 and 6, and in fact the host community grade 4 students are scoring as well as the grade 6 students.

Table 7 provides the overall summary for numeracy using the same format as the Table 5 summary in literacy. The main findings for numeracy are very similar to literacy. **First, student numeracy skills are clearly not at the expected levels for these grades, as evidenced by the fact that even in grade 6 the average student can answer less than half of the harder addition and subtraction operations.** By this grade level, the students should be familiar with addition and subtraction through large numbers, as well as multiplication and division (that are not included on HALDO). Additionally, significant numbers of students in grade 4 (especially) but even in grade 6 are unable to correctly identify double-digit numbers.

The overall percentage correct across all questions is about 60 percent, or roughly 14 correct answers out of a maximum of 23.[2] Nearly 25 percent was able to complete all tasks. But the HALDO is made up of basic skills content, so again it is important to be clear that grade 4 and 6 students should be very comfortable with these questions.

Table 7. Summary of HALDO numeracy item and overall percent correct and points by grade and school category

CHARACTERISTICS	GRADE 4		GRADE 6	
	Refugee	Host	Refugee	Host
1. Number identification (simple)				
Percent correct	80.6	87.8+	89.8	90.5
Number correct (max=5)	4.0	4.4+	4.5	4.5
1.1. One-to-one correspondence				
Percent correct	96.5	97.1	99.8	100.0
Number correct (max=3)	2.9	2.9	3.0	3.0
2. Number identification (harder)				
Percent correct	58.4	59.5	74.3	70.3
Number correct (max=5)	2.9	3.0	3.7	3.5
3. Simple operation				
Percent correct	42.0	41.6	59.7	57.0
Number correct (max=4)	1.7	1.7	2.4	2.3
4. Harder operations				
Percent correct	14.4	36.0*	37.5	43.3
Number correct (max=4)	0.6	1.4*	1.5	1.7
5. Word problems				
Percent correct	51.3	68.7*	63.8	68.4
Number correct (max=2)	1.0	1.4*	1.3	1.4
Overall summary:				
Percent correct (equal)	51.9	64.9*	67.2	69.5
Percent correct (points)	51.5	63.3+	67.6	69.3
Number correct (max=27)	11.9	14.6+	15.5	15.9
Sample size (n)	242	49	154	60

Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to averages (means) by grade and category based on weighted data. Tests of significance for differences between Camp and Host school categories (see Host column) are based on robust standard errors that correct for student clustering.
*Statistically significant difference between Camp and Host average at p<=0.05 level
+Statistically significant difference between Camp and Host average at p<=0.10 level

The second main finding in Table 7 is also consistent with literacy: grade 6 students outperform grade 4 students in almost all measures. For the overall sample (refugee and host combined) the grade 6 average is significantly higher than the grade 4 average (at the p<=0.05 level) in all areas of numeracy, with the largest gaps in Number identification (harder), Simple operations and Harder operations. However, the grade comparison pattern

2 The overall average in Table 7 is only computed for 433 out of the 504 students with maths scores. This is because of the tablet programming error where students were assigned a 5 on Numeracy 3 and then did not advance to Numeracy 4, and it is not possible to automatically recode their missing Numeracy 4 data.

does vary by school type. For the Mbera refugee students the grade 4-grade 6 gap is consistentlysignificant, which as noted for literacy does suggest that learning is taking place in these later grades in primary. For the host community schools the grade 4 student numeracy average is nearly as high as the grade 6 average, which is an unexpected result and again highlights some inherent uncertainty given the small sample for host school students. The grade 4 host averages are significantly higher than grade 4 refugee student averages in three areas of numeracy, and in the overall numeracy average, which is quite different from the literacy component where host grade 4 students actually scored lower than the Mbera camp students (Table 8).

The numeracy performance levels by grade and school category are summarized in Table 8. **Overall, just under 30 percent (28.8 percent) of the overall sample are in the lowest performance level, meaning they could not identify all five simple numbers and could not answer either of the two-word problems.** A similar share (24.3 percent) of the sample were unable to identify all five simple numbers but were able to answer at least one word problem. This leaves approximately half of the sample in the three highest performance categories where students are at least able to identify numbers (simple and hard) but vary by their ability to correctly answer operations problems. Twenty-two percent of overall sample —including 17 percent of grade 4 students and 34 percent of grade 6 students, both refugee and host students combined—are classified in the highest performance level (Level 5). These students essentially scored 100 percent correct on the HALDO numeracy assessment since they successfully completed the emergent, foundational, and intermediate level content (see Table 1) and then were able to answer all four of the harder operations.

Table 8. Summary of numeracy proficiency levels by grade and school category

DESCRIPTION OF LEVEL	OVERALL	GRADE 4		GRADE 6	
		Refugee	Host	Refugee	Host
1. Cannot identify all simple numbers, cannot answer any word problem	28.8	36.7	15.9	28.2	27.1
2. Cannot identify all simple numbers but can answer at least one word problem, or can identify simple numbers but cannot identify all harder numbers	24.3	23.9	38.6	10.1	18.2
3. Can identify all harder numbers but cannot answer all simple operations	15.0	19.6	7.9	16.2	9.7
4. Can answer all simple operations but cannot answer all harder operations	10.0	11.0	6.3	15.6	4.2
5. Can answer all harder operations	21.9	8.8	31.3	29.9	40.8
Total=100 percent	100.0	100.0	100.0	100.0	100.0
Sample size (n)	433	198	48	132	55

Data source: RET Mauritania HALDO assessment (2024)

5.3 HALDO assessment results: Socioemotional learning (SEL) and Executive functioning (EXF)

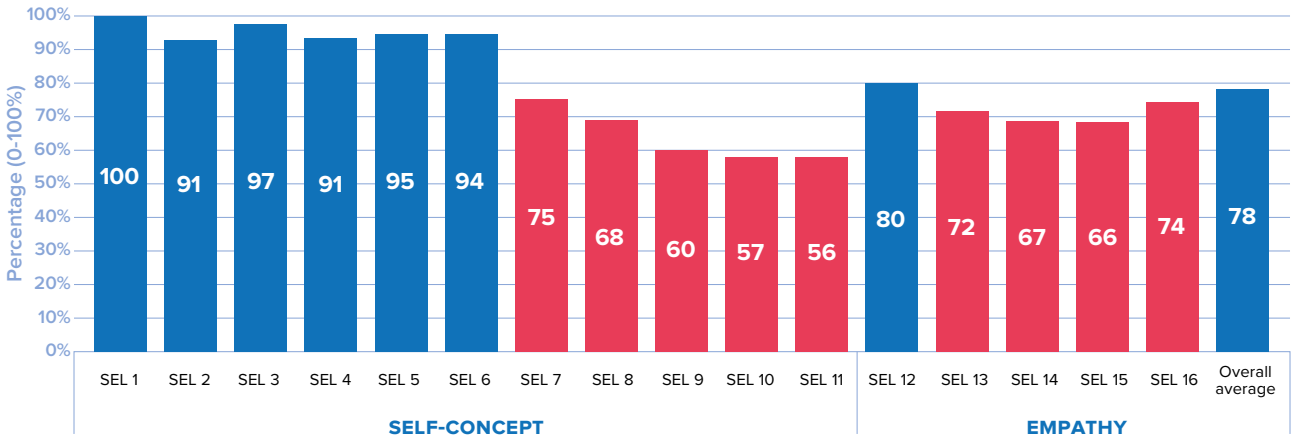
One of the innovative features of HALDO is the coverage of a broad range of foundational skills that go beyond literacy and numeracy, including socioemotional learning and executive functioning. These are summarized separately in this section, although not with the same detail as the literacy and numeracy components.

5.3.1 Socioemotional learning (SEL)

The HALDO instrument includes a brief summary of the SEL content, which "refers to a child's growing capacity for independence and confidence in a range of routine activities." The questions cover topics such as **awareness of self, understanding of his/her environment, whether they can imagine a hopeful future and how that may be impacted by realistic supports and barriers.** These are divided into two sections: Self-concept (11 items) and Empathy (5 items).

Figure 11 provides a global summary of the 16 SEL items. The SEL section begins with five basic questions about the child's self and environment, including full name (SEL 1), age (SEL 2), name of one person who takes care of them (SEL 3) and the name of their neighborhood/ community and province where they live (SEL 4-5). **All students answered these first five questions, and the scores in Figure 11 are all above 90 percent.** This is shaded in blue based on the Table 1 summary of performance levels from Section 4.1, which considers scores of 75 percent or higher on SEL to be in the "Advanced" range. The high scores on this content are not surprising for students in this age range.

Figure 11. Percentage correct on socioemotional learning (SEL) items, overall sample



Data source: RET Mauritania HALDO assessment (2024)

Questions 6-11 of the SEL assessment become more subjective and ask students to consider the future. This requires consistent scoring by the enumerators given the open format of the question. This is also where the adaptive structure in HALDO begins, as students who do not provide acceptable (or correct) answers on questions 6-10 do not advance to the next question and are automatically assigned zero scores on the remaining items in the Self-Concept section.

The student scores on the SEL 6-11 content are lower than the SEL 1-5 and following the automatic recodes get progressively lower. The sequence begins with SEL 6 where the students are asked to name one thing that they "hope/wish will happen in your life in the future". Ninety-four percent of the sample provided an acceptable answer to this question, which is in the "Advanced" range in Table 1. **The lowest averages in this section are for the questions that ask about things that could stop the child from realizing this future goal (SEL 10), or who/what could aid in making this happen (SEL 11).** These averages suggest that nearly half of the overall sample have difficulty in contemplating the future, what may happen to them and what kinds of factors may impact their lives.

The second part of the HALDO SEL assessment (Empathy) begins with SEL 12, where all students are shown a picture of a girl crying and asked to identify how she feels. The follow-up questions then ask the student about ways that they could help the girl feel better (SEL 13-14) and consider the underlying reason for why another child made her cry (SEL 15), and how that child felt afterwards (SEL 16). Questions 12 and 13 are adaptive, meaning students must provide an acceptable answer to continue to the next question. But all students then answer SEL 15 and 16.

The scores on the Empathy question block are lower than the Self-Concept scores, and in most cases are in the "Intermediate" performance range (red shading). 80 percent of the sample was able to describe how the girl feels based on the picture. The lowest score (66 percent) is for the question about why another child pushed the girl to make her cry (SEL 15).

Additional SEL comparisons are provided in Figure B8 by grade (a) and school category (b). **Not surprisingly, the SEL scores are higher in grade 6 than in grade 4, although the overall average is only about seven percent higher in grade 6 (Figure B8a).** Significant differences by grade are found for SEL 7, 8, 12, 13 and 16 (grade 6 higher). But the overall SEL average is not significantly different by grade.

For the camp versus host community comparison the results are quite similar, and the overall average across all sample participants is nearly identical in the two school categories (Figure B8b). The host community students do tend to have higher scores on the items, but the differences are generally not statistically significant. The main exception is for SEL 4 that asks the student the name of neighborhood/community/village where they reside, which was more difficult for the host community students to answer.

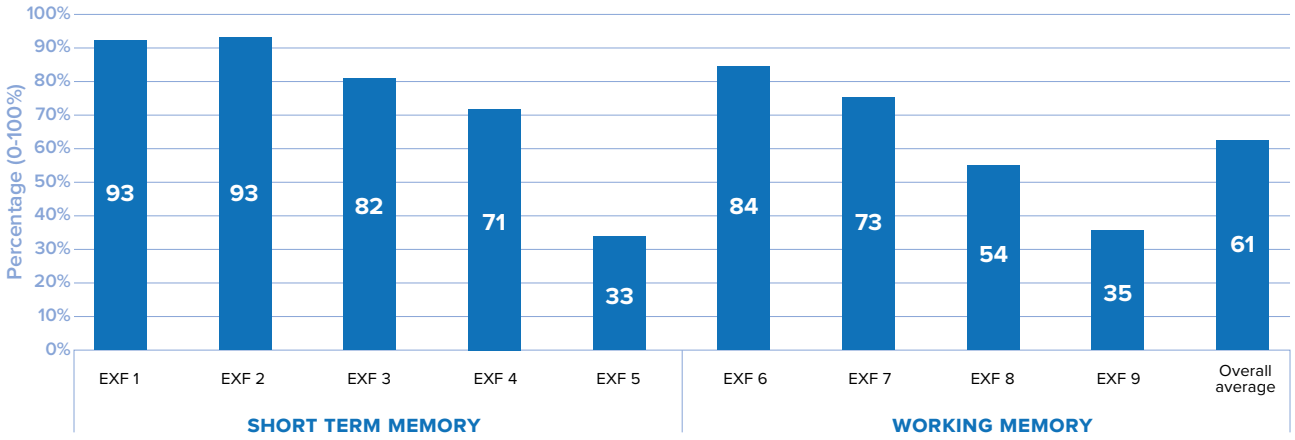
5.3.2 Executive functioning (EXF)

There are nine executive functioning items in HALDO. This begins with five questions about short term memory where students are asked to repeat a sequence of numbers that get progressively longer (from 3 numbers to 6 numbers). Unfortunately, because of an error in

the tablet programming only 114 students have data for the short-term memory questions, and these are students who performed relatively poorly on other areas of HALDO. This non-random assignment of the EXF 1-5 questions complicates the interpretation of the results, and the small sample (114) makes it quite difficult to make comparisons by grade and school category. The results in Figure 12 show high scores on the easier aspects of short-term memory (repeating short sequences of numbers), but with a substantial decline when asked to repeat 6 numbers.

The second part of executive functioning ("Working memory") asks students to repeat number sequences in the reverse order they were read to them. All 505 students answered these four items (EXF 6-9). **The summary in Figure 12 shows a steep decline in the percentage correct as the number sequence gets longer.** For example, 84 percent of the sample could correctly repeat "1, 5, 6" back as "6, 5, 1", but only 35 percent could do the same for "2, 5, 7, 1, 6". Figure 12 also provides the overall sample average for the four Working memory items only (i.e. short-term memory is not included in the overall average).

Figure 12. Percentage correct on executive functioning (EXF) items, overall sample



Data source: RET Mauritania HALDO assessment (2024)

Figure B9 provides the additional EXF analysis with comparisons by grade (a) and school category (b). As noted above, the small sample size for the Short-term memory content makes it very hard to interpret the comparisons by grade and school category. Focusing on the Working memory content, the results show that grade 6 students perform marginally better than grade 4 students on this content (about six percent higher), but the difference is not statistically significant. The comparisons by camp are even more similar, as the host community students averaged about two percent better than the Mbera camp students.

5.4 Overall HALDO scores

One of the defining features of the HALDO is its coverage of multiple academic and non-academic domains. The detailed summaries in Sections 5.2-5.3 included reviews of individual items as well as sub-domains (like "Working memory"). This detail helps to provide a complete picture of what students can and cannot do based on the HALDO tool.

Figure 13 brings together the four domains covered in HALDO to compare the scores (percentage of correct answers) by domain and grade level. **Student scores are highest in the socioemotional learning (SEL) content (about 80 percent), whereas for the other three domains the average percentage correct is close to 60 percent.** The domains do diverge in terms of the gaps between grade levels: grade 6 averages are much higher than grade 4 averages in literacy and numeracy, but less so for SEL and executive functioning (EXF). These differences then translate into a significant difference in the overall HALDO average (72 percent for grade 6 versus 61 percent for grade 4).

Figure 13. Percentage correct for HALDO domains and overall score by grade

Data source: RET Mauritania HALDO assessment (2024)

Table 9 provides a more detailed summary of the overall HALDO score based on three measures. These include the overall average percentage correct across the four domains, the overall z-score (standardized) percentage correct measure, and the main factor from a factor analysis of all of the items included together. Annex A includes more detail on the factor analysis.

Overall, the students averaged about 65 percent correct on the HALDO, with higher scores in grade 6 than grade 4. It is again important to clarify that a 65 percent overall average may give a misleading picture of performance since the content on HALDO is very basic for grades 4 and 6. The somewhat high overall average is due mainly to the inclusion of a number of very basic activities on HALDO where almost all students received high scores. But the more detailed summaries in Sections 5.2 and 5.3 clearly identify individual areas of literacy, numeracy, SEL and EXF where even grade 6 students often struggled to answer half of the questions correctly. This is also why the performance level summaries for literacy, numeracy and SEL are useful, as they make it clear that significant shares of students are unable to complete the more difficult tasks.

Table 9. Summary of overall HALDO scores by grade and school category

CHARACTERISTICS	OVERALL	GRADE 4		GRADE 6	
		Camp	Host	Camp	Host
Overall percent correct	64.2	60.5	61.0	70.2	75.7
Overall standardized (z-score)	0.02	-0.13	-0.19	0.29	0.62
Overall factor	0.01	-0.15	-0.23	0.30	0.64+
Sample size (n)	504	242	49	154	60

Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to averages (means) by grade and category based on weighted data. Tests of significance for differences between Camp and Host school categories (see Host column) are based on robust standard errors that correct for student clustering.
*Statistically significant difference between Camp and Host average at p<=0.05 level
+Statistically significant difference between Camp and Host average at p<=0.10 level

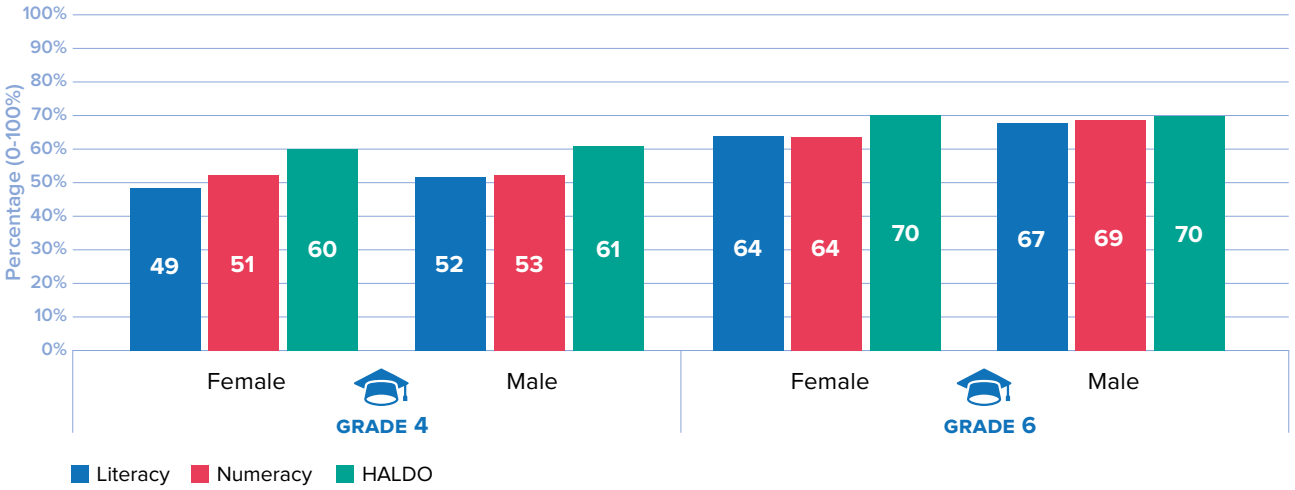
One limitation with percentage correct is that it is difficult to assess the size of the difference between different categories. The z-score and factor score both have an overall average close to 0.0, and a standard deviation of 1.0. This makes it easier to consider how large the differences are between different groups of students. **For example, the Mbera camp grade 4 students have an average z-score of -0.13, compared with a 0.29 for the camp grade 6 students; the difference between these two averages equals nearly one half of a standard deviation (or 0.42 SD), which is a substantial gap.** In grade 4 the standardized differences between camp and host students are quite small (0.06-0.08 SD). But in grade 6 the host community average is nearly one third of a standard deviation higher than the camp students.

5.5 Comparisons of HALDO scores by background characteristics

The previous sections have included comparisons of HALDO results by grade and school category. This section continues the comparisons with other student background characteristics. The purpose is to understand more about why some students perform better on HALDO than others. All of the results in this section are restricted to refugee students in order to focus the discussion on this particular context.

For gender there are no significant differences in the HALDO literacy, numeracy and overall HALDO score (Figure 14). Boys have higher scores than girls in literacy and numeracy in both grades, but the differences are generally in the 2-3 percentage point range. The overall HALDO score is the nearly identical for boys and girls in grade 4 (60-61 percent) and grade 6 (70 percent).

Figure 14. Comparison of HALDO results by student gender and grade, Mbera camp

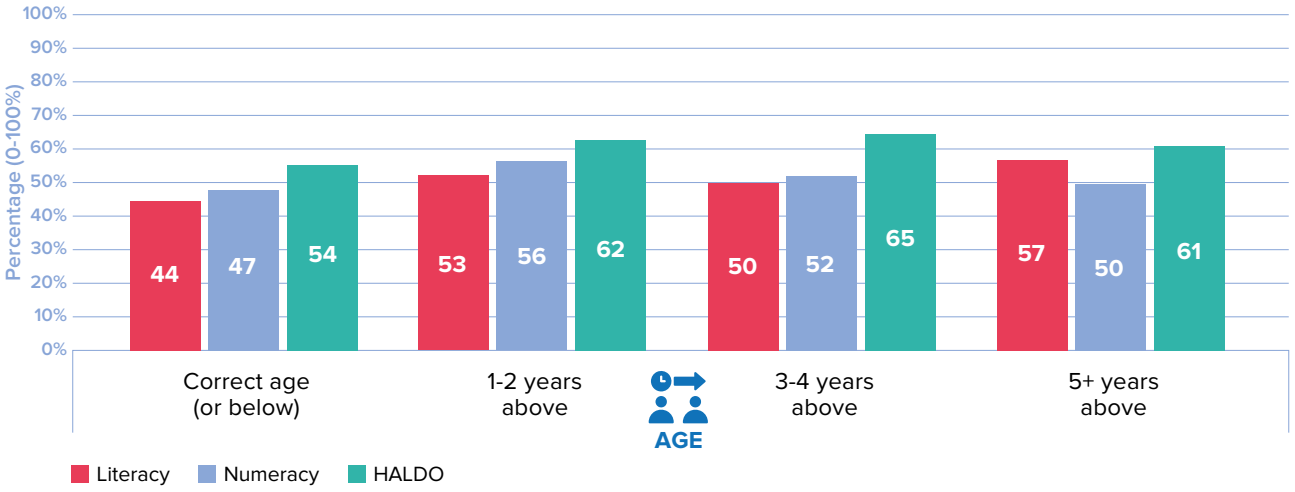


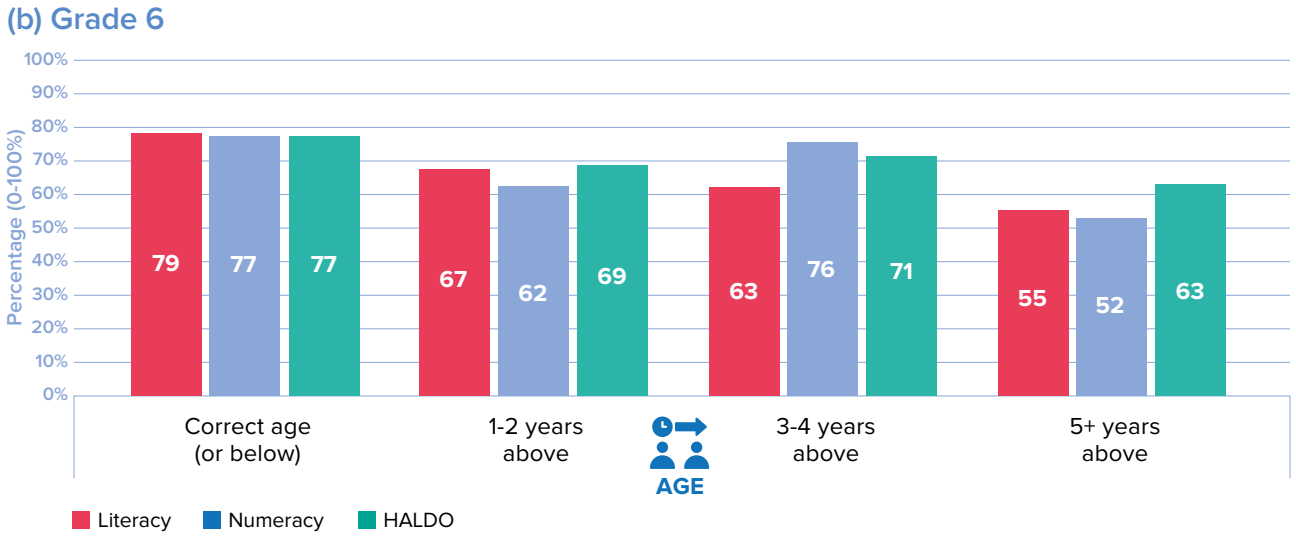
Data source: RET Mauritania HALDO assessment (2024)

The HALDO scores do vary by age, although the pattern is different by subject and grade. In grade 4 the averages for numeracy and overall HALDO are highest among students who are 1-2 or 3-4 years above the expected age (i.e. 11-15 years old), and lowest among the youngest and oldest cohorts; for grade 4 literacy the scores go up and down by age (Figure 15). In grade 6 there is a more pronounced downward trend as the youngest students (12 and under) have the highest scores on all three measures, and the oldest students (5 years older than correct age) have lower scores. **In grade 6 there are some large differences by age: correct age (or younger) students (20 percent of the sample) scored 79 percent correct on the literacy component, while the oldest (17 and above) students (who make up 15 percent of the sample) averaged just 55 percent correct.**

Figure 15. Comparison of HALDO results by student age and grade, Mbera camp

(a) Grade 4





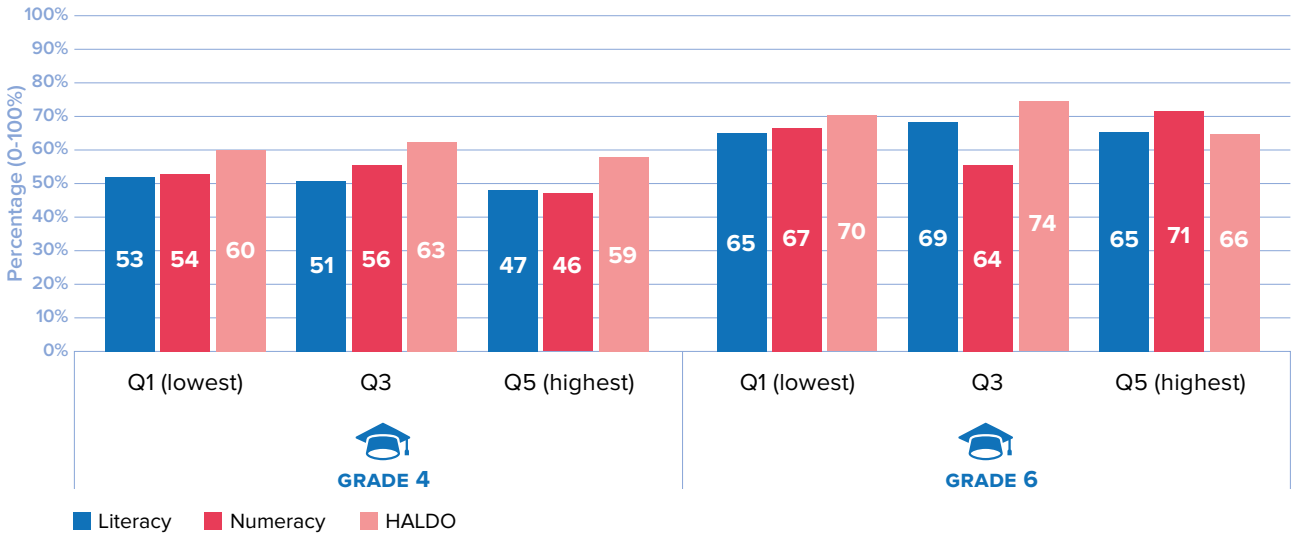
Data source: RET Mauritania HALDO assessment (2024)

Socioeconomic status (SES) is measured in HALDO with a series of questions about possessions in the home (television, electricity, etc.). These indicators were analyzed using factor analysis to produce an overall SES index, and then that index was divided into five equal groups of students (called quintiles). Students in quintile 1 have the lowest SES in the sample, and on average only report having about 10 percent of the home possessions. By contrast, the quintile 5 students reporting have over 60 percent of the home possessions. For example, 29 percent of the students from Q5 households report their family has a car, compared with less than one percent of Q1 households.

The relationship between the SES measure and student performance on HALDO is very weak (Figure 16). For example, grade 4 Mbera camp students in quintile 1 (lowest SES) scored 53 percent correct (overall) on the literacy assessment, while quintile 5 (highest SES) students answered 47 percent of the literacy items correctly. The results for other subjects (and grade 6) are similar, and clearly do not show the expected upward trend for HALDO results by SES quintile.

The lack of a significant relationship between SES and student scores is a surprising result given the very large differences in student-reported possessions in the home. One possibility is that the students did not correctly report what is in their home, in which case the SES analysis will not be valid. But the result also raises some concerns about the sample and the kinds of students that were in attendance on the day the assessment was conducted. If the most motivated and engaged students came to the Mbera camp to participate then that could have the effect of reducing the influence of SES on the results since the poorer students—who would normally be expected to have lower scores—may be relatively high performing and engaged. As was already discussed in the sampling section, it is not possible to verify if the average *sampled* student is the same as the average enrolled student in the Mbera camp (or host community schools). This again highlights the inherent problem with conducting assessments outside of the normal school calendar.

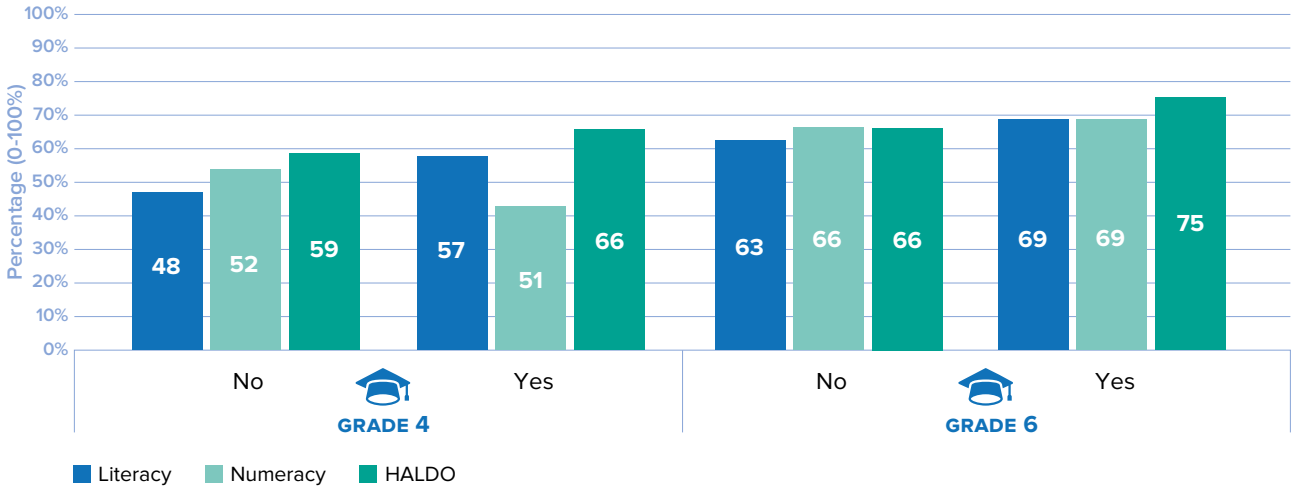
Figure 16. Comparison of HALDO results by socioeconomic (SES) quintile and grade, Mbera camp



Data source: RET Mauritania HALDO assessment (2024)

Student performance on HALDO is generally higher when books are read in their home, and this correlation is higher for literacy than numeracy (Figure 17). For example, grade 4 students who report seeing books ready in their household scored 57 percent correct on the literacy component, compared with just 48 percent for students who do not report book reading in their households. Similar magnitudes are shown in grade 6 and for the overall HALDO score.

Figure 17. Comparison of HALDO results by student has seen books read in home and grade, Mbera camp

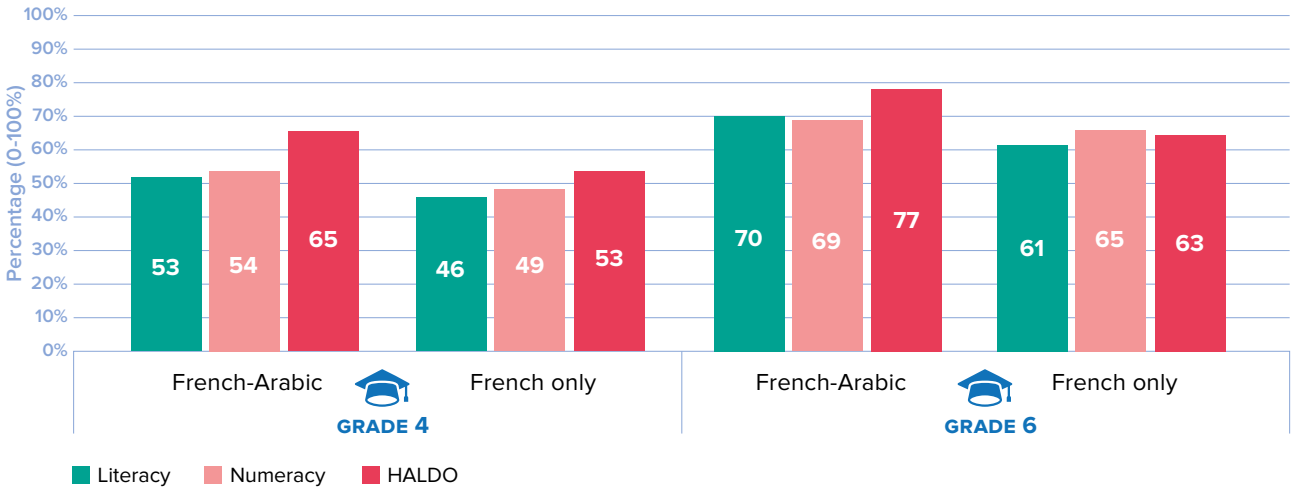


Data source: RET Mauritania HALDO assessment (2024)

Another unexpected result is shown in Figure 18 for the student-reported language of instruction. Students were asked if they received instruction in French only or in French and Arabic combined. **The results show that students who report French only instruction have consistently lower HALDO scores than students who report receiving instruction in French and Arabic.** This is surprising since the HALDO was conducted in French, which would

presumably provide an advantage to students who receive more instruction in this language. However, it should be noted that the student responses to this question were not consistent within each school, especially in grade 4 where some students would report French only instruction, but their classmates would report differently.

Figure 18. Comparison of HALDO results by language of instruction in class and grade, Mbera camp



Data source: RET Mauritania HALDO assessment (2024)

Additional indicators from the HALDO interview were incorporated into the analysis of how student and family background predict performance, including language spoken in the home and disability/difficulty status. These results do not show any significant relationships with HALDO scores, so they are not presented.

The background analysis concludes with multivariate statistical modeling, which is presented in Table 10. Each of the four HALDO components, and the overall HALDO score, were modeled using a block of independent variables. The main feature of multivariate analysis is that it can consider how multiple independent variables simultaneously predict an individual outcome. This is like correlation analysis between a dependent and independent variable, but the difference is that each individual independent variable is interpreted as a "net effect" when controlling for other variables in the model. The main goal with this type of analysis is to identify significant predictors of the outcome among a group of variables, and in this way shed light on what are the most important (or significant) predictors of the outcome. Significant variables in table B1 are flagged with an asterisk (p<=0.05) or + (p<=0.10).

Very few predictors among the set of independent variables are consistently significant in the estimations in Table 10. The control for Mbera camp schools is insignificant in all estimations, which means that when controlling for other factors the HALDO scores in the camp and host community schools are not significantly different. Grade 6 students have much higher scores than grade 4 students in literacy and numeracy, but the grade control is not significant for the SEL or EXF outcomes. Student age and gender are also not very significant, although girls do have a marginally significant advantage over boys in socioemotional learning (SEL).

Table 10. Covariates of HALDO outcomes, full sample

INDEPENDENT VARIABLE	HALDO COMPONENT				OVERALL
	Literacy	Numeracy	SEL	EXF	HALDO
Mbera camp school	-0.11 (-0.48)	-0.26 (-1.32)	0.06 (0.26)	0.14 (0.58)	0.02 (0.08)
Student in grade 6	0.52* (2.18)	0.39* (2.43)	0.12 (1.10)	0.14 (0.98)	0.36+ (2.05)
Student is female	-0.14 (-1.63)	-0.09 (-1.01)	0.12+ (2.03)	-0.11 (-1.43)	-0.01 (-0.01)
Student age in years	0.01 (0.06)	-0.01 (-0.17)	0.06* (3.48)	-0.02 (-0.77)	0.03 (1.24)
Parent has phone number	0.40* (3.15)	0.28* (2.38)	0.08 (0.72)	0.20+ (1.81)	0.30* (2.63)
Languages spoken at home (reference=Hasania only):					
Tamasheq only	0.05 (0.26)	-0.01 (-0.04)	-0.07 (-0.45)	-0.14 (-0.68)	-0.07 (-0.36)
Hasania and Tamasheq	-0.14 (-0.65)	0.10 (0.57)	0.04 (0.19)	-0.71* (-4.68)	-0.17 (-1.03)
Other combination	0.18 (1.17)	-0.16 (-0.61)	0.06 (0.63)	0.05 (0.34)	0.05 (0.44)
Student lives with both parents	0.16* (2.18)	-0.09 (-0.57)	-0.18 (-1.39)	-0.04 (-0.32)	-0.08 (-0.69)
Student has any disability/difficulty	0.33+ (1.88)	-0.46+ (-1.91)	-0.03 (-0.14)	-0.03 (-0.25)	-0.09 (-0.91)
Someone reads books in home	0.21* (2.42)	-0.07 (-0.55)	0.28* (2.45)	0.08 (0.46)	0.24* (2.17)
SES quintile (reference=Q1)					
Quintile 2	-0.01 (-0.20)	-0.37+ (-1.88)	0.27* (2.71)	-0.03 (-0.32)	0.16+ (1.87)
Quintile 3	0.16 (0.84)	0.03 (0.15)	0.33* (2.67)	0.06 (0.43)	0.29 (1.68)
Quintile 4	-0.39* (-233)	-0.21 (-0.76)	0.05 (0.25)	-0.22 (-1.15)	-0.18 (-0.71)
Quintile 5 (highest)	-0.37* (-2.70)	0.09 (0.39)	-0.03 (-0.17)	0.19 (-0.86)	-0.13 (-0.66)
Class is only in French language	-0.13+ (-1.96)	-0.02 (-0.08)	-0.40* (-2.42)	-0.47* (-3.13)	-0.42* (-2.68)
Sample size (n)	499	428	499	499	499
Explained variance (R^2)	0.22	0.11	0.13	0.10	0.17

Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to averages (means) by grade and category based on weighted data. Tests of significance for differences between Camp and Host school categories (see Host column) are based on robust standard errors that correct for student clustering.
*Statistically significant difference between Camp and Host average at p<=0.05 level
+Statistically significant difference between Camp and Host average at p<=0.10 level

The variable for "parent has phone number" is consistently positive and significant in the multivariate analysis. This variable was captured by the enumerators, and roughly half of the entire sample of students have parents with a phone number. However, it is not clear whether this is measuring phone access for parents or whether or not the school has the phone number of the parents. It may be an indicator of parental engagement with the school, meaning that when the school has the phone number the parents are more engaged. This could explain why these children have consistently better results.

Student scores are also significantly higher when they indicate that books are read at home. This is especially true for literacy but also for SEL and the overall HALDO score. This indicator may be indicative of a "reading culture" in the home, or it may even be capturing students who are themselves reading at home.

Finally, the indicator for French language use is consistently negative and significant. However, it is interesting that the effect sizes are largest (and most significant) in the SEL and EXF components of the HALDO. This is another result that may require some further follow-up to better understand the underlying explanation.

5.6 Summary of focus group discussions (FGDs)

The focus group discussions were conducted with students, parents and teachers in the Mbera camp and Fassala commune (see Table 3 summary). The interviews were translated and summarized in text form. With some exceptions, the FGDs data are generally consistent across the two communities (camp and host), and a number of key issues were identified by all three groups (students, parents and teachers).

The camp and host community participants are positive about the schooling experience. Children are excited to go to school to learn and be with their friends, and they are aware of the importance of education and the impact it can have on their futures.

"I love school because my dream requires me to go through school."

-Fassala student

"The school allows me to progress little by little towards my goal of becoming a doctor one day."

-Fassala student

Parents are also appreciative of the school staff's efforts, and in the case of the Mbera camp they are also thankful to the international partners who are active in the camps. **Parental perceptions about the relevance of public education are important in this context given the availability of religious education options.** Two fathers in the Mbera camp parental FGDs noted that many families in the past considered formal (public) education to be a threat to religious beliefs, but that this has changed as parents are more likely to recognize the schools' value to strengthen cultural and religious values. Many children in this context still attend Koranic religious classes in the early morning before coming to public school, but references to the declining share of parents who are withholding children from these schools does suggest that camp (and host) schools have earned more trust from the local population.

A concern shared by parents in both communities is the limited parental engagement with school staff. Fassala parents appear to give this issue more weight, as they acknowledge that some teachers do reach out to them individually, but they do not feel like they are involved in the schools' decision-making process. The Mbera parents reference the heavy workload of school staff as a reason for limited engagement, and they also indicated that parents are more likely to deal with school-related issues through the parent association than directly with school staff. Mbera camp parents also indicated that they are actively involved in education through their interaction with other camp families.

Child security issues are a concern among camp and host community students, parents, and staff, although the underlying problems do vary. Instances of bullying and violence between students do not stand out in the discussions, as the references seem to indicate somewhat isolated events. However, a common point of concern is the lack of adequate fencing around the school and the potential dangers this brings for students during regular school hours (such as strangers or animals coming on school grounds). Potentially dangerous commutes to school were also cited by multiple participants.

The Fassala students reported use of corporal punishment in their schools, which is a serious issue and is officially prohibited. At least some of the students made a connection with the use of punishment in their homes, indicating that their teachers are treating them in a similar way as their parents. It also should be restated that Fassala students were positive about their teachers (overall) and the schooling experience. There was no reference to this kind of punishment in the Mbera camp, as one girl noted that "our teachers love and protect us."

School staff are aware that more needs to be done to make schools more inclusive, especially for children with disabilities. This is related in part to resources (infrastructure and class sizes, see below) which make it difficult for school staff to adequately engage with vulnerable children. A lack of staff training was also cited (also returned to below), as well as the need to improve awareness about this issue, make disabled children more visible in the community, and make schools more receptive to their inclusion.

References to gender as an inclusion issue were limited in the FGDs. All of the participant teachers were male in both school categories, which is reflective of the gender imbalance among staff in the Mbera camp schools. All participants in the Mbera camp FGDs with parents were fathers, while in Fassala the parent FGDs participants were all mothers. The main point is simply that females were not represented in at least some of the focus group discussions, which limits the insights about gender from their perspective.

The topic that stands out from the focus group discussions is the lack of sufficient school resources. This is a cross-cutting issue that is mentioned directly and indirectly in all of the interviews, with numerous references to a range of resource-related deficiencies in schools.

These include:

- Infrastructure components such as deteriorated classrooms, not enough classrooms and space, insufficient play areas, lack of fences around school buildings, lack of electricity and insufficient/inadequate toilets. Students are concerned that their schools lack adequate infrastructure, and are unattractive.
- Canteen, food and water access. Another commonly referenced issue that is related to larger problems of food (and water) insecurity. This is a very large topic given the presence of a school feeding program (in Mbera), the targeting of Mbera camp families for program participation which in turn can act as an incentive to send children to school, and also the participation of children in Koranic schools in the early morning that increases the nutritional demands given the very long day attending school.
- Classroom resources such as tables and chairs.
- Teaching and learning materials (kits, textbooks, teacher materials).
- More teachers to reduce overcrowding.

"Classrooms need to be more attractive",

-Fassala student

"Some classrooms are abandoned and in very poor condition, giving the impression of devastated schools, like those seen in conflict zones, such as in Mali"

-Fassala student

It is important to recognize the potentially far-reaching impacts of these kinds of deficiencies in schools, while also recognizing the challenges of addressing so many things (see discussion in Section 6). Insufficient infrastructure, hardware and learning materials can impact the teaching and learning environment, which not only affects student learning progress but can also undermine confidence in the education system, and lead children (or parents) to discontinue school attendance. Food and water security issues impact the child's ability to physically cope with the demands of school attendance. Also, the resource constraints have direct and indirect impacts on other problems that have been identified by participants, such as school security (fences) and teacher engagement with the community (due to heavy workload).

A critical question is how the problems identified by FGDs participants help explain the inadequate overall (average) levels of HALDO basic knowledge among grade 4 and 6 students in both sets of schools. **The resource constraints identified above are certainly relevant, but school staff (and some parents) also identified a number of teacher capacity issues that provide additional insight into the challenges faced by these schools.** Teachers in Mbera and Fassala schools cited the need for additional teacher training, but there are multiple layers to be addressed. First, the teachers reference the importance of better monitoring and evaluation tools to identify specific training needs for individual teachers given the difficulty for teachers to pinpoint exactly what it is they need. The goal of this professional support is to better equip teachers to deal with a diverse set of challenges and student needs in the classroom related to student motivation, special needs and other vulnerable children, and classroom diversity. This kind of capacity building program requires a long-term commitment and cannot be quickly accomplished with a handful of "fixed" trainings.

Teacher awareness about the need for more training to effectively instruct their students is a very clear signal that the teaching and learning environment is not sufficiently effective. The FGDs data provide additional glimpses that support this contention. A parent in Fassala noted that the difficulties that students are encountering in secondary education are attributable to a large degree to deficiencies in primary schools. There are also references to problems with frequent student absences and eventual dropout, and Mbera staff referenced a troubling dynamic where students enter the school system with high levels of motivation, but "slowly become unmotivated" as they spend more time in these schools. These kinds of concerns—and the results from HALDO—make sense given the resource constraints, teacher concerns about capacity and the vulnerability of the student populations.

Finally, the FGDs data are consistent with a previous round of discussions and observations carried out in the Mbera camp by Cambridge Education (2022). This includes inadequate educational infrastructure, classroom crowding, unattractive learning environments, a lack of certified teachers and extremely uneven gender balance among teachers.

6. Discussion

The discussion focuses on the implications of the main findings from the analysis, divided into two general categories: summative results (mainly from HALDO) about student learning and school quality, and formative results about the study design and implementation.

The main finding in this study is that student achievement levels in grade 4 and 6 are not at the expected levels. The defining feature of HALDO, in addition to its coverage across different domains, is its focus on foundational skills. This emphasis on basic content means that students in higher primary grades (such as grade 4 and grade 6) should be comfortable with all of the content. But the results in literacy and numeracy do not show this. For example, the most advanced content in literacy includes five reading comprehension questions following a short text. Only 10 percent of grade 4 students, and 25 percent of grade 6 students, were able to correctly answer all five questions. Similar shares of students were only able to answer four or fewer of the five questions. This means that the largest blocks of students—80 percent in grade 4 and 50 percent in grade 6—did not even answer these questions because they were unable to successfully complete an easier task on the literacy assessment. These large shares of students who did not advance to more advanced content—who were then automatically assigned "zero scores" in the advanced content—are the main reason why the overall performance on HALDO is considered to be low for these grades. For numeracy the student performance on the four "hard" operations items follows a similar pattern: 17 percent of grade 4 students, and 34 percent of grade 6 students, answered all four questions correctly, while 75 percent of grade 4 and 57 percent of grade 6 students received automatic zero scores. And again, what is identified as "advanced" content on HALDO is not very demanding for students in this grade range: the "hard" numeracy operations on HALDO are restricted to two-digit addition and subtraction problems (e.g. "22+40"), and do not include multiplication, division, decimals, or other more challenging questions.

Given these concerns about the level of difficulty of HALDO versus the actual grade level (and age) of students in the sample, it is not realistic to assess the number of students who have attained "minimum" level skills. The main point again is that for grade 4 and (especially) grade 6 students *all of the content* on HALDO is at a basic level, and the student performance should reflect this. That is why the overall percentage correct averages (as high as 80 percent in grade 6, depending on the measure that is used) are misleading, and should not be interpreted in the same way as the results from a more comprehensive assessment that includes a wider range of difficult or advanced content. The more realistic scale for evaluating the HALDO results in this sample is more like what a teacher would use in his or her classroom, meaning an overall average of 65 percent is indicative of a very low score (e.g. a "D" or even an "F").

The results from the socioemotional learning (SEL) and executive functioning (EXF) components of HALDO are generally in line with the results for literacy and numeracy. Very high percentages (above 90 percent) can complete the easiest activities (state their age for SEL, repeat three numbers for EXF). But as the content gets harder the scores decline noticeably. For example, only about half of the students could identify something that may prevent them from realizing their goals in the future (SEL) or count back a four-digit number sequence in reverse (EXF).

The relatively low scores on HALDO (for this grade level) are consistent with the kinds of challenges facing these students, schools, and communities. Refugee students are an extremely vulnerable group, but the host community students also appear to be relatively poor. These students often have limited resources in the home to support their education (materials, homework help, reading books, etc.), which in turn puts extra pressure on the school to develop skills. But these schools suffer from resource and capacity constraints, and the end result is that learning is not sufficient.

There are other notable findings from the HALDO analysis, although these are given less attention than the inadequate performance on the literacy and numeracy components. First, student scores are consistently higher in grade 6 than in grade 4. This does suggest that learning is taking place in these schools, although this comparison could be affected by student attrition between these grades (i.e. better students remain enrolled). Second, there are very few significant differences between the Mbera camp and host community student averages, although these comparisons are complicated by the small sample size for host schools. And finally, there are only a handful of student and family background variables that are significant predictors of HALDO performance. This is a concern since background variables like SES are usually quite robust predictors, and the general absence of significant raises concerns about the sample composition since the students were contacted during the holiday period.

The second set of study findings serve a formative feedback function on the HALDO instrument and overall study design and implementation. First, the HALDO instrument was subjected to a very detailed analysis, both as part of the review of the student results (Section 5) as well as in a separate analysis focusing on the performance and psychometric properties of the HALDO items under different scoring scenarios (see Annex A). **Overall, the HALDO instrument has acceptable performance characteristics based on things like reliability and item discrimination (in this sample of schools).** This is important because HALDO is designed to provide a snapshot of student performance across four distinct domains of knowledge using a simple tool that can be implemented quickly (with each student).

However, there are tradeoffs with using an instrument like HALDO instead of more comprehensive assessments, especially in higher grades (like grade 4 and grade 6). First there is the basic level of content that is covered, although in earlier primary grades this is less of a problem. Second, the automatic recoding functions in HALDO—where students are assigned a zero on more difficult items if they are unable to correctly answer easier content—may be too restrictive. For example, at least some of the students who are able to identify four out of five simple letters are also likely to be able to identify at least some of the harder letters; but based on HALDO scoring rules the only students who advance to the hard letter identification are those with perfect scores on the simple letter identification activity.

For the overall study design and data collection, there are additional formative findings. **Standing out is the difficulty of carrying out student assessments during the holiday period when students are not in regular school attendance.** For the present study there is an inherent degree of uncertainty about the representation of the sample because it cannot be assumed that the students who came to the schools on the day of the assessment are a representative sample of the entire population of students in these schools. **Another lesson learned is that comparisons between Mbera camp and host community students should rely on relatively**

similar sample sizes for each group. The small number of host community students in this sample (109)—especially in combination with the uncertainties about the sampling—make it very difficult to compare camp and host community outcomes. Given the multiple challenges in delivering educational services in refugee camps, the comparison against an external "reference point" is potentially quite important, but this requires sufficient study power.

Finally, there were several tablet programming errors for HALDO content that resulted in missing data and less complete overall information on HALDO performance. This is partly related to HALDO recoding strategies that require extensive skipping, which does complicate the programming work in comparison with traditional assessments where students answer all questions.

MAURITANIA. Children from Mbera refugee camp are educated on environmental conservation as part of World Environment Day
©**UNHCR/XAVIER BOURGOIS**

7. Recommendations

The recommendations follow the main findings that were discussed in Section 6, again divided into summative and formative sections. However, it is important to note that the bulk of the data analysis comes from a HALDO tool that is, by design, quite streamlined. There is no survey data from schools, teachers, or parents to provide additional information and context, which leaves a small set of focus group discussions as the only other source of information.

The main finding from the study is that grade 4 and 6 students are not performing at a sufficient level on basic content, and there is a convincing—albeit somewhat tentative—link between these results and the various resource, capacity and vulnerability issues that were identified in the FGDs. Unfortunately, all three of these problem areas—resource constraints in schools, teacher capacity limitations, and student (and family) vulnerability—are multi-layered and not amenable to easy policy fixes. The following recommendations are mainly intended to stimulate additional discussions on these topics to facilitate actual actions.

These recommendations also reflect the feedback received at dissemination events at national and local levels, participated by the Ministry of Education officials, partner organizations, as well as Mbera camp coordinator, and local communities.

7.1 Policy and Programme Recommendations

- **Adopting Inclusive Policies** – Collaborate with national and international partners to continue the inclusion of refugee students into the Mauritanian education system, balancing the Malian curriculum with local content.

- **Address School Resource Constraints** – Comprehensively document school resource constraints and discuss these among a range of stakeholders to prioritize feasible items to support. These constraints could address infrastructure, hardware, teaching and learning materials. Resource constraints can become overwhelming, and it is simply not possible to address every deficiency, highlighting the need to identify the most critical areas.

- **Enhance Parental and Community Engagement** – Stimulate more contact between school staff, parents, and members of the community. Establishing regular meetings to share concerns and provide updates on school functions can serve many purposes, including identifying ways for parents and the community to support education, alert school officials to specific concerns of the community, and bring together different ideas to identify problems and discuss proposed solutions. If possible, school-based solutions should be considered to address some of the priority resource constraints that have been identified. Possible examples include making school spaces more attractive (and green) and enhancing the security of the school grounds. Compared with issues related to resources this is a relatively low-cost intervention. The finding in Section 5 that student HALDO scores are higher when books are read in the home is a reminder of how parental engagement can affect key student outcomes.

- **Improve Teacher Training and Support** – Teachers need to be supported by school principals, district staff and other stakeholders to help identify specific training needs and look for ways to address those needs. This is another challenging area to address in detail given the overall demands on teacher training resources, and the difficulty of delivering quality in-service training on the topics that are most relevant for teaching staff.

  The largely Malian teaching force in the Mbera camp schools need to be better integrated into existing district training and support functions (see Cambridge Education, 2022). To avoid complete reliance on district support processes, local solutions can be considered, including the use of experienced staff to provide support and mentoring, as well as cluster-based training mechanisms where schools work with each other to address needs and share experiences.

- **Strengthen Student Well-being and Learning** – UNHCR should advocate with Ministry of Education for incorporating psychosocial services and social-emotional learning (SEL) into the curriculum to address the trauma experienced by forcibly displaced children and support their well-being and learning.

- **Expand Access to Books and Libraries** – Strengthen library programs and community initiatives to collect donations of books, promote reading and access to books among children.

7.2 Research Recommendations

In terms of future research using HALDO in these communities, one challenge is that it is difficult to provide detailed guidance without knowing the overall research goals. For example, a follow-up study in the same grades could require a different design than a new study targeting earlier grades. Regardless, the following recommendations are largely developed based on the lessons learned from this study (see Section 6):

- **Regular Learning Assessments for Forcibly Displaced Children** – Collect HALDO data on a regular basis and during the regular school calendar, preferably towards the end of the school year when students have completed most of the curriculum for that year. These assessments should be integrated into refugee education monitoring and evaluation processes. The UNHCR operation could consider adding user-defined indicators for COMPASS reporting that address learning proficiency (e.g. % of forcibly displaced students with minimum skills in reading / mathematics / socio-emotional learning).

- **Include Earlier grades for future data collections** – The results for this study provide very relevant information given the difficulties experienced by many grade 4 and 6 students to answer even basic level content. But by design HALDO is intended to measure curriculum in early primary grades, so it would be useful to assess these schools during this phase of primary. Future assessments should also include younger grade levels to capture early learning development.

- **Equip Teachers with Practical Assessment Tools** – Introduce simplified foundational learning assessment tools for teachers to use throughout the school year, enabling continuous monitoring of student progress.

- Improve future assessments by:

  - **Enhancing the HALDO Instrument for Better Usability** – Discussions should be conducted with program support resources (including designers of HALDO from Save the Children) to consider simplifications to the HALDO instrument. For example, less restrictive skipping rules could be used (e.g. students do not need to answer all questions correctly to proceed to the next question). This will depend on numerous factors, including the grade level that is being tested and the desirability of making comparisons with other HALDO results.

    - **The skipping instructions that are provided in the HALDO instrument that was incorporated in this study need to be clarified**. In several places one set of instructions indicated that if the student answered any of the questions correctly, they could proceed to the next question, but additional instructions then indicated that only those students who answered all questions in the block could proceed.

  - **Ensuring Meaningful Comparisons in Sampling** – If comparisons between Mbera camp and host community student results are a priority, then the sampling needs to reflect this and include roughly equal numbers in each category. Small samples can be a waste of resources since the results are not likely to be very informative.

  - **Strengthening Data Collection and Contextual Relevance** – More information is required on students and children with disabilities. This is not an easy aspect of student background to measure via surveys like HALDO, which based on a widely used metric (Washington Group/UNICEF). The low rate of reported disabilities raises concerns about inclusion in these schools, which in turn highlights the importance of understanding more about the prevalence of disabilities in the community, the schools' efforts to include these children, and the challenges they face.

    In addition, All questionnaires—student, FGDs and any other new instruments—should be tailored to address more specific contextual features related to education in these contexts. Examples include school feeding participation, Koranic school attendance, nutrition, and school climates.

  - **Improving Data Quality and Implementation** – More time should be spent on tablet programming double checks and piloting. Ideally, if time permits, an initial set of trial data could be analyzed by a data specialist to review the skipping procedures based on actual data (as opposed to just reviewing tablet programming). This would help to identify incorrect skips and other errors.

    Additional questionnaires (or interviews) for teachers, head teachers and parents should also be considered for inclusion to complement the student assessment data generated by HALDO. This study is largely dependent on FGDs for additional contextual data, but this kind of information can be collected through the surveys as well (although in more quantitative format). The mixed study design should be retained however, if possible, since it is important to have a data collection venue where students, staff and parents can discuss a wide range of issues.

References

Cambridge Education. (2022). *Strengthening Data for Quality Education for Refugee Children and Youth*. https://www.oxfordmeasured.co.uk/projects/strengthening-data-management-for-quality-education-for-refugee-children-and-youth

Krupar, A. (2019). *Return to Learning Midterm Report: Bekaa, Lebanon. Save the Children.* https://resource-centre.savethechildren.net/pdf/return_to_learning_midterm_lebanon_2019.pdf/

Krupar, A., & D'Sa, N. (2024). *Measuring learning during crises: Developing and validating the Holistic Assessment of Learning and Development Outcomes (HALDO).* International Journal of Educational Research Open, 6, 100320. https://doi.org/10.1016/j.ijedro.2024.100320

Moscoviz, L., & Evans, D. (2021). *Learning Loss and Student Dropouts during the COVID-19 Pandemic: A Review of the Evidence Two Years after Schools Shut Down. Center for Global Development.* https://www.cgdev.org/publication/learning-loss-and-student-dropouts-during-covid-19-pandemic-review-evidence-two-years

RET International. (2024). *Inception Report: Measuring Holistic Learning Outcomes for Forcibly Displaced Children.*

UIS. (2024). *UIS Statistics.* https://data.uis.unesco.org/

UNHCR. (2021a). *Esquisse d'une stratégie pour l'Education dans le camp de Mbera.*

UNHCR. (2021b). *Statistiques de l'éducation à Mbera.*

UNHCR. (2022). *Briefing Note: Education Working Group, Bassikounou.*

UNHCR. (2023). *UNHCR Mauritania—Country strategy evaluation 2020-2022: Report.* https://www.unhcr.org/media/unhcr-mauritania-country-strategy-evaluation-2020-2022-report

World Bank. (2018). *World Development Report—Learning: To Realize Education's Promise* (p. https://www.worldbank.org/en/publication/wdr2018). World Bank.

Annex A: HALDO assessment background analysis

This section provides a more detailed analysis of the student assessment data from the HALDO instrument. For each of the four HALDO test sections (literacy, numeracy, SEL and EXF) the various items are analyzed using both Classical Test Theory (CTT) and Modern Test Theory tools to assess item difficulty, discrimination, reliability, and dimensionality. The overarching purpose is to compare the item performance characteristics of the "scored" items (using CTT tools) against the characteristics of the raw data using modern tools like item response theory, (IRT). The analysis concludes with "global" analysis that assesses the degree to which the four HALDO section items perform as a group, and how feasible it is to combine them into a single overall HALDO measure of student performance.

The results in this section are not critical for the student results summary in Section 5 of the report since that discussion is based entirely on the "official" percentage correct scoring scale for presenting HALDO results. However, the HALDO is a relatively new tool that has not been used in many countries, so additional details about the underlying measurement properties of the assessment provide additional context for understanding the results in the various analyses—including report section 5.5 where multivariate analysis is used to analyze variation between students on the various assessment measures—and potentially add to the formative evidence base on HALDO performance.

A1. Literacy

Table A1 begins with a summary of the six items from the Literacy section. Section 5.2.1 in the main report already reviewed the HALDO literacy items together with the scoring basics and skipping rules. All students answered the initial letter identification (common) question, and the remaining items were applied "adaptively" based on student performance on the previous item. The literacy items are presented in Table A1 in raw as well as scored form. The raw summary is only for those students who answered the item, while the scored summary applies automatic recodes so that all students have data for all six items. The results for number correct are very different by raw and scored category, which is to be expected given the large numbers of student data that is recoded according to the HALDO scoring rules (see Section 5.2.1 for more details).

The summary in Table A1 relies on the basic tools of Classical Test Theory (CTT) for difficulty (number correct), discrimination and reliability (captured by factor analysis and Chronbach's Alpha). The CTT parameters are calculated using the scored (complete) data, as it is not possible to carry out the alpha reliability or factor analysis with missing data. The results in Table A1 show acceptable levels of item discrimination[3] and reliability for the scored literacy

3 The discrimination parameters in Table A1 were obtained using the alpha command in Stata (version 17) and are based on the "item-rest" correlation that measures the correlation between the individual item and the rest of the test items. This is a version of the item discrimination index, but it is calculated differently from the standard discrimination index measure. 0.30 and above is generally considered to be an acceptable level for the item-rest correlation.

items. The six (scored) literacy items also load onto a common factor with an Eigenvalue of 3.6, which easily clears the minimum threshold of 1.0 for identifying meaningful dimensions. And the overall reliability indicator (Chronbach's Alpha) is also at an acceptable level (0.88).

Table A1. Basic summary of HALDO literacy items

TEST QUESTION	N (RAW)	NUMBER/PERCENT CORRECT		DISCR.	FACTOR LOADING
		Raw	Scored		
Letter identification (common) (max=5)	505	4.5 (1.1)	4.1 (1.1)	0.58	0.60
Expressive language (max=10)	114	5.9 (2.6)	9.1 (2.1)	0.57	0.58
Letter identification (infrequent) (max=5)	390	4.5 (0.9)	3.5 (2.1)	0.75	0.75
Reading fluency (initial, 0-1)	279	0.6	0.3	0.76	0.91
Reading fluency (complete, max=30)	158	29.2 (1.4)	9.1 (13.6)	0.76	0.91
Reading comprehension (max=5)	158	3.8 (1.4)	1.2 (1.9)	0.71	0.83
Eigenvalue	----	----	----	----	3.6
Chronbach's Alpha (α)	-----	----	----	----	0.88

Data source: RET Mauritania HALDO assessment (2024)
Notes: Standard deviations in parentheses. Number (or percent) correct in the Scored column is based on all 505 students in the sample. Eigenvalue, Discrimination and Chronbach's Alpha are only calculated for the entire sample (505 students). Factor loadings refer to the first factor from Principal Component Factor (PCF) analysis.

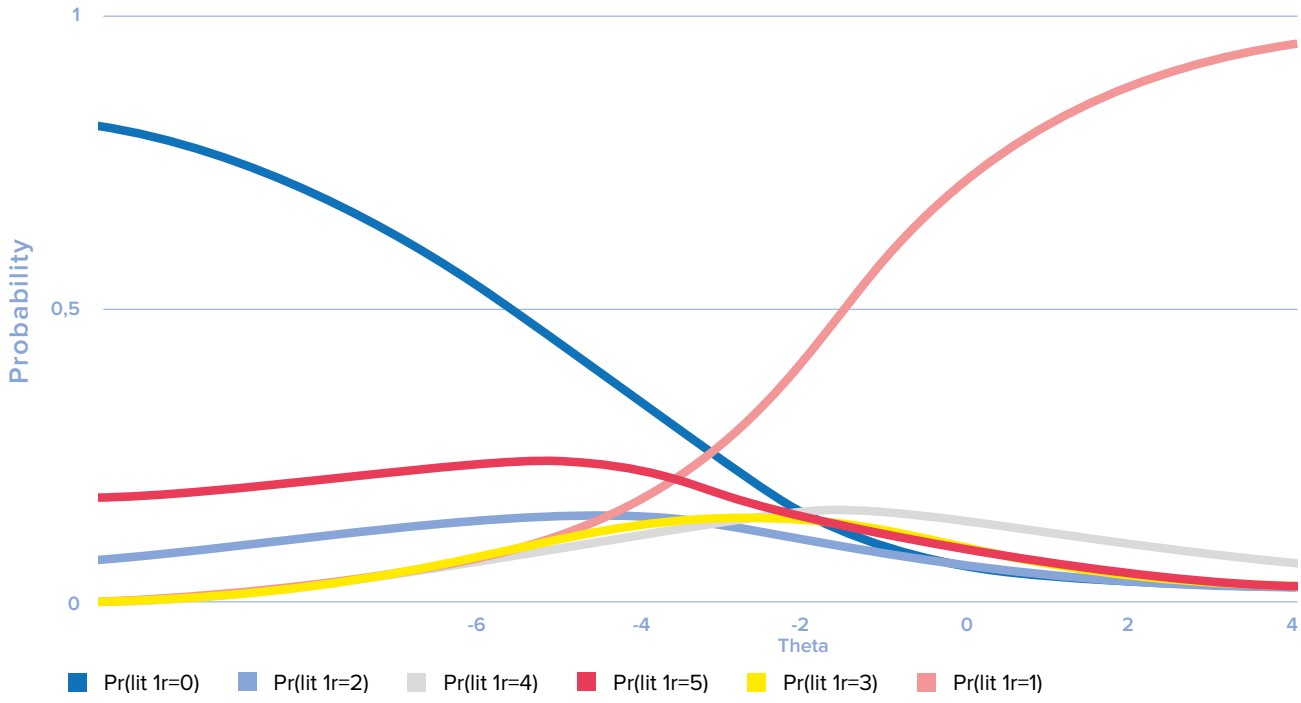
The HALDO test relies on automatic scoring rules to deal with missing data. For example, if a student cannot identify any common letters then it is safe to assume that the student cannot complete the reading comprehension tasks (and is therefore assigned zeros on these questions). As discussed in Section 5, these automatic recode rules simplify the application of the HALDO for both the enumerator and the student, but they do potentially introduce some error. Standing out is the possibility that students are being assigned zeros (or perfect scores) on questions where their actual score would be different (if tested); a specific example of this was discussed in Section 5.2.2 for numeracy.

Modern Test Theory incorporates item response theory (IRT) tools to assess item and test performance. This is only possible with the raw data, as the automatic scoring rules in HALDO preclude the use of IRT with the scored data.[4] One of the defining features of IRT analysis is that it does not require complete data to assess item performance or compare student achievement levels. In the case of HALDO this is important because the students do not take the same tests: some students answer only the basic content, others answer all of the questions except the most basic content and other students are in the middle (i.e. they answer some of the basic and some of the more advanced content). IRT analysis effectively gives students more points for correctly answering harder items, and fewer points for correctly answering easier items. Using powerful statistical processing it is able to evaluate thedifficulty levels of all items as well as their effectiveness based on their ability to predict student performance on the overall construct (discrimination).

4 IRT estimations fail to converge with the HALDO data when the scored items are used instead of the raw items.

The results for the literacy items are somewhat more mixed when applying IRT. Figure A1 provides a visual example of the IRT analysis output for literacy item 1. Most of the literacy items on HALDO are "partial credit" items since the students receive scores on a scale (item 3 is the only question that is coded as 0-1 for incorrect-correct). The upward sloping (orange) curve in Figure A1 for students who identified all five common letters means that students with a higher overall performance on the literacy domain (measured along the x-axis with "Theta") have a higher probability of identifying all five common letters (probability is on the y axis). The negative sloping (blue) curve for students who scored zero on item one simply shows that lower scoring students (on overall literacy) have a high probability of getting zero on this item.

Figure A1. Item characteristic curve (ICC) for literacy item 1



Since the other item 1 response categories (1 correct, 2 correct, etc.) have fairly flat curves, the visual summary in Figure A1 suggests that item 1 is somewhat dichotomous: the meaningful outcomes are for students who identified all letters or could not identify any letters. However, the interpretation of overall performance for the students who did not identify all five letters is somewhat difficult given the HALDO scoring formula directs these students to answer an even more basic literacy item (expressive language), and then terminates the assessment. As noted above, IRT is designed to handle situations where students do not answer the same set of items, but the fact remains that low scoring students do not have a lot of information in HALDO.

The item characteristic curves for the remaining partial credit literacy items generally follow the same pattern in Figure A1, with a dichotomous curve structure that is sharply upwards for students who received the maximum score on the item, sharply downwards for students who scored zero, and fairly flat for students in the middle ranges. The main exception to this pattern is for item 3, which is coded as 0 if the student could not read all five initial letters in a passage, and 1 if the student was able to read all five. This dichotomous item has a downward sloping ICC that is very hard to explain since it suggests that students who were unable to complete

this task have higher literacy ability than those that were able to complete it.[5] One possibility is that the automatic skip rules in HALDO are in some way complicating the calculation of ability for these students, although this problem is less pronounced when the second part of the reading fluency (literacy 4 item) is not included in the IRT analysis.

The final element of the IRT analysis is to assign an equated score to each student based on their raw score performance (i.e. no automatic recoding) on the HALDO literacy items. This is another standard feature of IRT that makes it possible to compare student scores on tests even when they did not answer the same set of questions. In theory the IRT-generated measure of student performance should provide a more accurate picture of student performance on HALDO than the overall percentage correct measure that relies on automatic scoring rules. The equated IRT score is added to the Section 5. statistical analysis of HALDO scores, while the "official" HALDO scoring formulas based on overall percentage correct are used for the summary of student results in report Sections 5.2-5.5. However, in the case of literacy the equated IRT score is compromised by the negative discrimination result for the literacy 3 item, so some caution is required with the interpretation of this indicator.

5 The downward sloping curve for item 3 (or "negative discrimination") is shown using both the IRT module in Stata as well as a specialized IRT software called Item and Test Analysis (IATA)(Cartwright, 2013).

A2. Numeracy

Table A2 provides the Classical Test Theory summary statistics for the six sets of items on the HALDO numeracy assessment. The overall results are quite similar to the literacy results. The scores are generally very high on all of the items based on raw scores, but the more difficult content scores decline substantially when lower performing students are automatically assigned zeros; for example, the 133 students who answered the Hard Operation questions averaged 3.3 correct answers out of 4, but this average is just 1.0 for the entire (Scored) sample. The discrimination and factor loading parameters are relatively low for the very easy items, especially the three one-to-one correspondence questions. But once again the scored items have acceptable levels of discrimination, factor loadings and overall reliability (Alpha).

One complication with numeracy is that item 4 (Hard Operations) was incorrectly skipped by 71 students due to a tablet programming error (this was discussed in Section 5.2.2). To address this issue the factor analysis (and Alpha reliability) statistics were computed with and without the numeracy 4 item. The results are not very different.

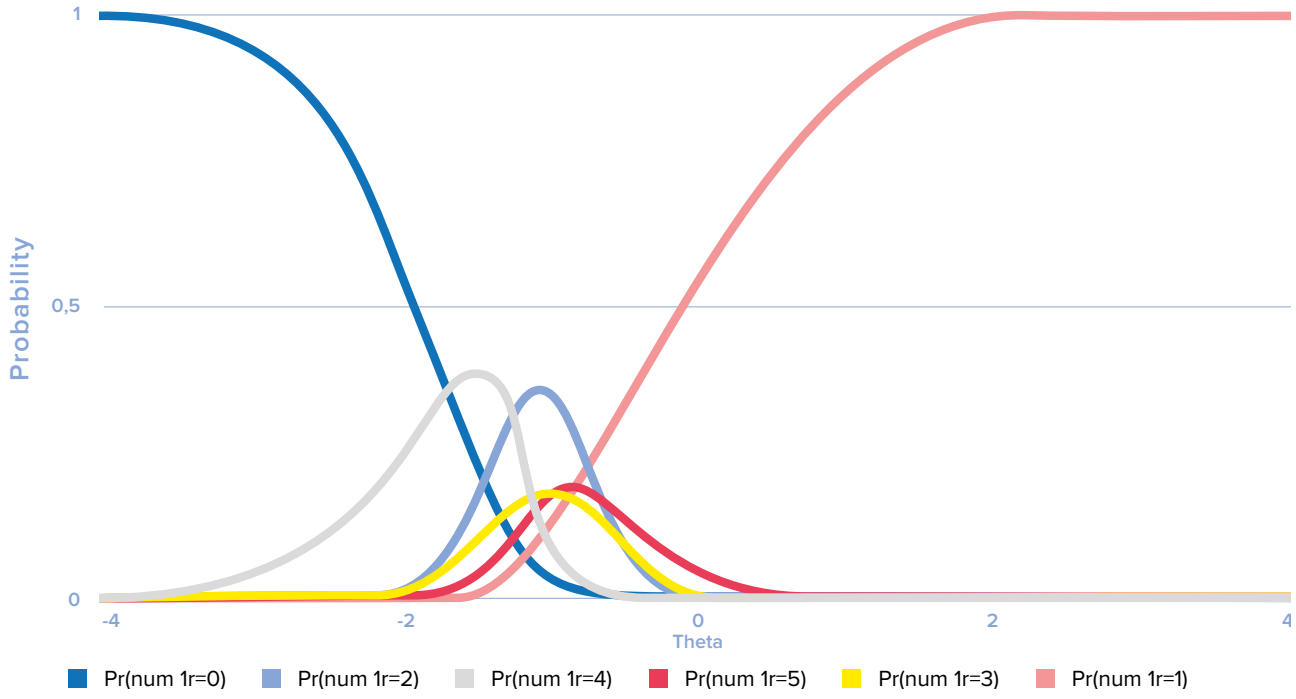
Table A2. Basic summary of HALDO numeracy items

TEST QUESTION	N (RAW)	NUMBER/PERCENT CORRECT		DISCR.	FACTOR LOADING	
		Raw	Scored		(1)	(2)
Simple number identification (max=5)	504	4.2 (1.4)	4.2 (1.4)	0.70	0.74	0.79
One to one correspondence:						
Item 1	143	0.98	0.99	0.35	0.25	0.30
Item 2	143	0.97	0.99	0.40	0.28	0.34
Item 3	143	0.88	0.97	0.44	0.35	0.40
Harder number identification (max=5)	361	4.5 (1.2)	3.2 (2.2)	0.76	0.90	0.92
Simple operations (max=4)	273	3.5 (1.0)	1.9 (1.9)	0.62	0.85	0.78
Harder operations (max=4)	133	3.3 (1.1)	1.0 (1.6)	----	0.72	----
Word problems:						
Item 1	378	0.69	0.51	0.61	0.75	0.73
Item 2	378	0.82	0.60	0.65	0.78	0.78
Eigenvalue	----	----	----	----	4.1	3.6
Chronbach's Alpha (α)	-----	----	----	----	0.85	0.84

Data source: RET Mauritania HALDO assessment (2024)
Notes: Standard deviations in parentheses. Number (or percent) correct in the Scored column is based on all 504 students in the sample. Eigenvalue, Discrimination and Chronbach's Alpha are calculated for the entire set of items (but with only 433 students, see Factor Loadings column 1), and then without the Harder operations item (Numeracy 4) for all 504 students (Factor loadings column 2). Factor loadings refer to the first factor from Principal Component Factor (PCF) analysis.

The IRT analysis of the raw response data from the HALDO numeracy section shows that the numeracy items performed somewhat better than the literacy items. This is mainly related to item discrimination, which was substantially higher for numeracy and there were no items with negative discrimination (like item 3 in literacy). Figure A2 provides a visual example with the Item Characteristics Curves for the different scores on numeracy item 1. The basic pattern is the same as most of the literacy items, with a sharp negative slope for students who were unable to identify any of the five simple numbers, and sharply positive slope for the students who correctly identified all five. One difference with the literacy items is that the other scores on numeracy item 1 have more defined curves that suggest more differentiation between the different levels of performance on the item. But these are not very large differences and in general the numeracy and literacy items have similar IRT characteristics.

Figure A2. Item characteristic curve (ICC) for numeracy item 1



A3. Socioemotional learning (SEL)

Table A3 continues with a summary of the 16 socioemotional learning (SEL) items. The SEL content is different from the literacy and numeracy content areas in two ways. First, all of the SEL items are dichotomous where the student either completed the task (coded as 1) or did not (coded as 0). And second, the SEL content is divided into two general constructs for Self-Concept (items 1-11) and Empathy (12-16). All students were asked to complete the first six SEL tasks, and the results in Table A3 show that almost all students in the sample were able to do so (overall average close to 95 percent). This is another example of HALDO content that is quite easy for the study sample, which from an item analysis standpoint means the item provides very little useful information. This is not the same thing as saying that the item is not useful: very high (or very low) scores on items provide useful information for a range of stakeholders (program staff, government counterparts, teachers, etc.), but in the narrow scope of item analysis these items do not provide much information because there is very little variation between students. The limited variation on the first six SEL items results in very low discrimination index scores, and low factor loadings, for SEL 1-6. This is again confirmation that these items are not really generating any useful information for comparing levels of student socio-emotional learning.

Table A3. Basic summary of HALDO socioemotional learning (SEL) items

TEST QUESTION	N (RAW)	NUMBER/PERCENT CORRECT		DISCR.	FACTOR LOADING		
		Raw	Scored		(1)	(2)	(3)
One to one correspondence:							
Item 1	505	0.99	0.99	0.04	0.02	----	0.04
Item 2	505	0.93	0.93	0.14	0.15	----	0.14
Item 3	505	0.97	0.97	0.18	0.16	----	0.16
Item 4	505	0.94	0.94	0.09	0.03	----	0.07
Item 5	505	0.94	0.94	0.21	0.16	----	0.20
Item 6	505	0.94	0.94	0.33	0.36	----	0.34
Item 7	475	0.79	0.74	0.68	0.80	----	0.75
Item 8	373	0.91	0.67	0.75	0.89	----	0.83
Item 9	338	0.87	0.58	0.78	0.95	----	0.87
Item 10	295	0.95	0.55	0.79	0.95	----	0.88
Item 11	280	0.97	0.54	0.79	0.93	----	0.88
Word problems:							
Item 12	505	0.80	0.80	0.56	----	0.78	0.59
Item 13	401	0.90	0.71	0.68	----	0.95	0.72
Item 14	358	0.96	0.68	0.67	----	0.91	0.71
Item 15	505	0.64	0.64	0.41	----	0.49	0.41
Item 16	505	0.71	0.71	0.55	----	0.62	0.55
Eigenvalue	----	----	----	----	4.3	3.0	5.6
Chronbach's Alpha (α)	-----	----	----	----	0.84	0.86	0.88

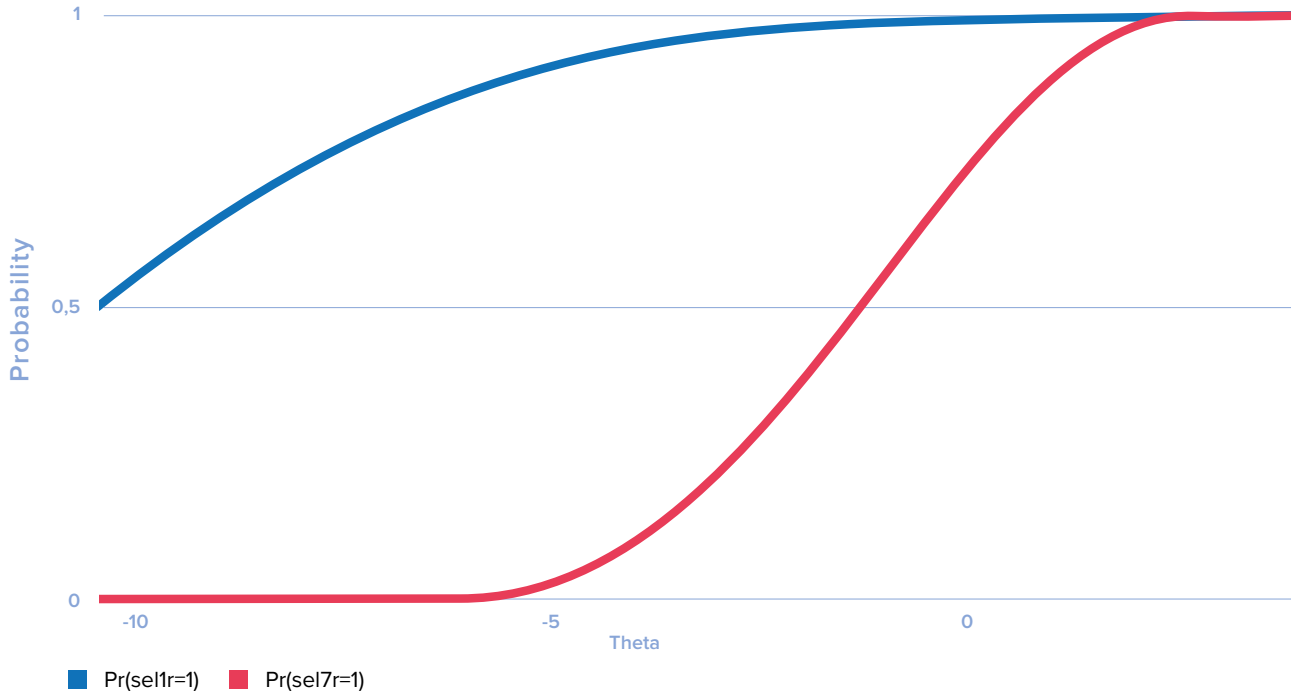
Data source: RET Mauritania HALDO assessment (2024)
Notes: Percent correct in the Scored column is based on all 504 students in the sample. Eigenvalue, Discrimination and Chronbach's Alpha are calculated for the entire set of items (but with only 433 students, see Factor Loadings column 1), and then without the Harder operations item (Numeracy 4) for all 504 students (Factor loadings column 2). Factor loadings refer to the first factor from Principal Component Factor (PCF) analysis.

The SEL items 7-11 were applied adaptively as the student's response on the individual item then determined if they continued. The SEL assessment was then reset at item 12 as all students answered the initial question related to Empathy, and then the adaptive structure was applied to items 13 and 14; SEL items 15 and 16 were answered by every student. Since there was some skipping in the SEL instrument the average scores in Table A3 are quite different between Raw and Scored on a number of items.

Table A3 summarizes three different factor analysis results. The first estimation (1) only included the 11 items for Self-Concept, the second estimation only included the five items for Empathy, and the third estimation included all 16 items. There is not much evidence that the Self Concept and Empathy questions are loading onto distinct dimensions of SEL. This is shown by the generally large factor loadings for items 7-16 on the overall factor (3).[6]

The IRT analysis for SEL was also somewhat different than for literacy and numeracy. This is mostly due to the total reliance in SEL on dichotomous items, with no partial credit questions with different values. The use of dichotomous items facilitates the IRT analysis, and the results for discrimination are generally similar to the results in Table A3 based on scored data. Figure A3 provides a visual comparison of two SEL items with very different levels of discrimination. The curve for SEL 1 (blue line) is positively sloped, but it is not very steep (meaning low discrimination) and it is pushed to the left side of the overall ability scale (which means it was an easy item). By contrast, SEL 7 has a very steep slope (high discrimination) near the middle range of the SEL ability scale (theta=0), which means that the information provided by this one item does a good job of predicting how students will do on the remaining SEL items.

Figure A3. Item characteristic curves (ICC) for SEL items 1 and 7



6 Another way to assess the degree of "multi-dimensionality" in the SEL content (i.e. whether or not the responses are very different between Self Concept and Empathy questions) is to examine the factor loadings on different factors from the same analysis. The results in table C3 are for the first factor created by the factor analysis, but there was a second factor produced in the analysis with an Eigenvalue above 1.0 (factor 2, eigen=1.72). But on closer review of the loadings for this factor there is no clear evidence that the Self Concept and Empathy domains are distinct.

A4. Executive Functioning (EXF)

The final component of the HALDO is executive functioning, which like SEL was divided into two general sub-dimensions (Short-term memory and Working memory). Table A4 provides the basic summary of the nine EXF items. Due to a programming error in the tablets most of the students skipped the first five EXF items, and answered items 6-9. This complicates the classical test theory analysis for discrimination and reliability because these tools rely on complete data (i.e. no missings), so in Table A4 they are only presented for items 6-9. There is also no automatic recodes for the EXF items because all students answered items 6-9, and the skipping pattern for items 1-5 was not systematic so the recode rules cannot be applied.

There is a lot of variation in the EXF results, with high scores on two of the items (EXF 1 and 2) and lower scores on EXF 5 and 9. Even when restricted to just four questions the discrimination and reliability indicators are at acceptable levels.

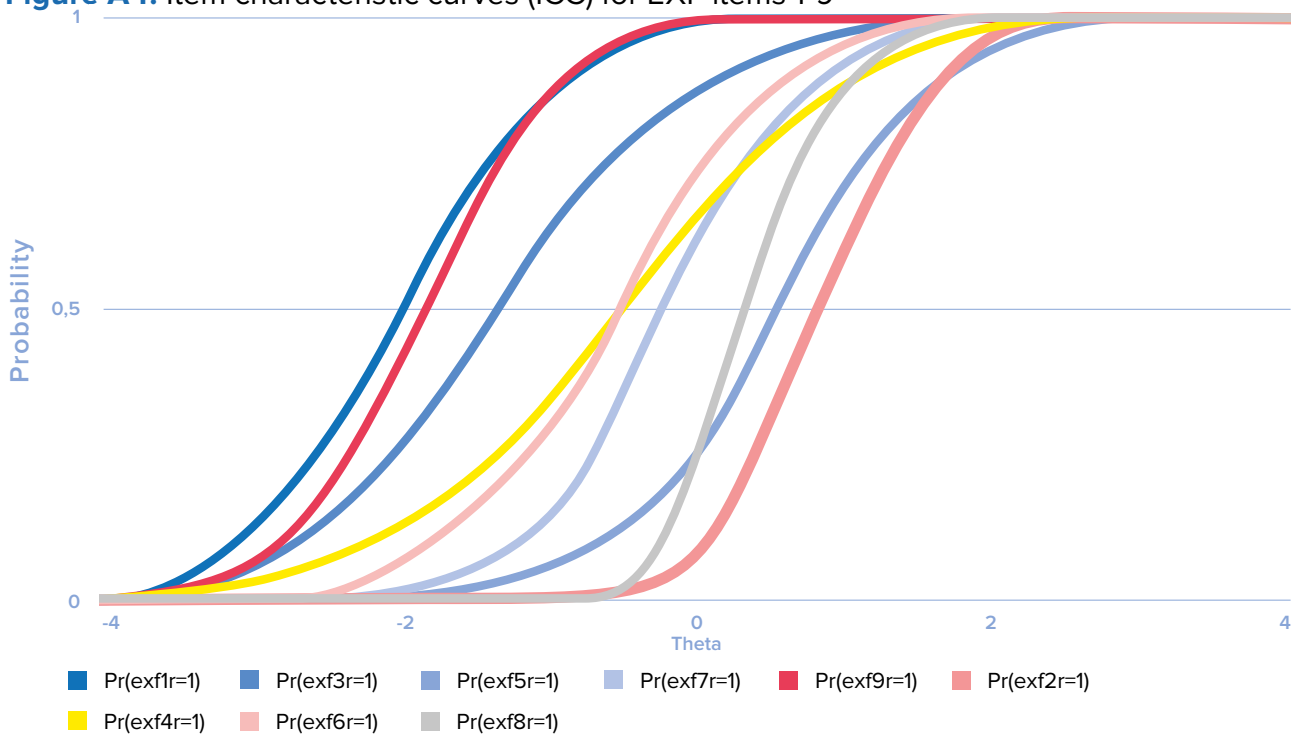
Table A4. Basic summary of HALDO executive functioning (EXF) items

TEST QUESTION	N (RAW)	PERCENT CORRECT	DISCR.	FACTOR LOADING
Short-term memory:				
EXF 1	114	0.93	----	----
EXF 2	114	0.92	----	----
EXF 3	114	0.78	----	----
EXF 4	114	0.62	----	----
EXF 5	114	0.29	----	----
Working memory:				
EXF 6	505	0.81	0.49	0.56
EXF 7	505	0.72	0.61	0.69
EXF 8	505	0.52	0.70	0.79
EXF 9	505	0.32	0.54	0.64
Eigenvalue	----	----	----	1.9
Chronbach's Alpha (α)	-----	----	----	0.78

Data source: RET Mauritania HALDO assessment (2024)
Notes: Percent correct is based only on the students who answered the items (i.e. raw scores). Eigenvalue, Discrimination and Chronbach's Alpha are only calculated for the entire sample (505 students). Factor loadings refer to the first factor from Principal Component Factor (PCF) analysis.

The IRT analysis is again useful given the missing data for EXF 1-5 and the inability to consider things like discrimination for these items using classical tools. The IRT analysis shows that the nine items have generally good discrimination ability, especially items 7 and 8 (which is consistent with the results in Table A4). Figure A4 presents the visual summary of the item characteristic curves (ICCs) for all nine EXF items. Since the items are all dichotomous it is possible to put them on a single graph. The results show that the items have generally steep ICCs that cover a range of difficulty levels.

Figure A4. Item characteristic curves (ICC) for EXF items 1-9



A5. Global analysis

The final activity combines the HALDO content from all four sections into a single ("global") analysis. The purpose is to assess the degree to which the different HALDO sections are "multi-dimensional", meaning that each one measures a distinct set of knowledge (or learning domain). Given the importance of summarizing the HALDO content in specific areas (and items, see Section 5), the global analysis is not critical for the overall report analysis. But it does provide another area where further exploration can provide some additional background on HALDO assessment performance. And it also considers the validity of the single, overall measure of HALDO performance that is incorporated in the Section 5.4 summary.

The global analysis begins with the classical test theory tools for discrimination and reliability, which are presented in Table A5. This was done by incorporating all of the literacy, numeracy, socioemotional learning and executive functioning items simultaneously into the discrimination and factor analysis routines. Since both tools require complete data (for all students), it was not possible to include numeracy item 4 (Hard Operations) or the first five EXF items since not all students have data on these questions (i.e. not automatic scoring is possible).

The results for the global analysis of all items are generally consistent with the results from the individual section-specific analyses. The discrimination parameters are generally above 0.30 (or 0.40), with the exception of relatively easy content in numeracy (especially one-to-one correspondence) and the first six SEL items (which were also quite easy for the sampled students). The factor loadings follow the same pattern: somewhat low for the easier numeracy and SEL content, but otherwise in an acceptable range. The overall reliability indicator (alpha) is 0.88.

The factor analysis did identify other significant factors (i.e. factors with an Eigenvalue greater than 1.0). But these additional factors were not particularly robust, and there was no clear pattern from the factor loadings about the dimension of knowledge that was identified. This is not an unusual result in factor analysis with a large number of items included. The important point is simply that the factor (and discrimination) analysis results do not suggest that the individual HALDO sections are measuring clearly distinct domains of knowledge. This in turn provides some justification for combining the different content into an overall score, but again as noted above this is not really a priority in the analysis given the need to understand student performance on specific areas of HALDO content.

Table A5. Discrimination index and factor loadings for all HALDO items combined ("global analysis")

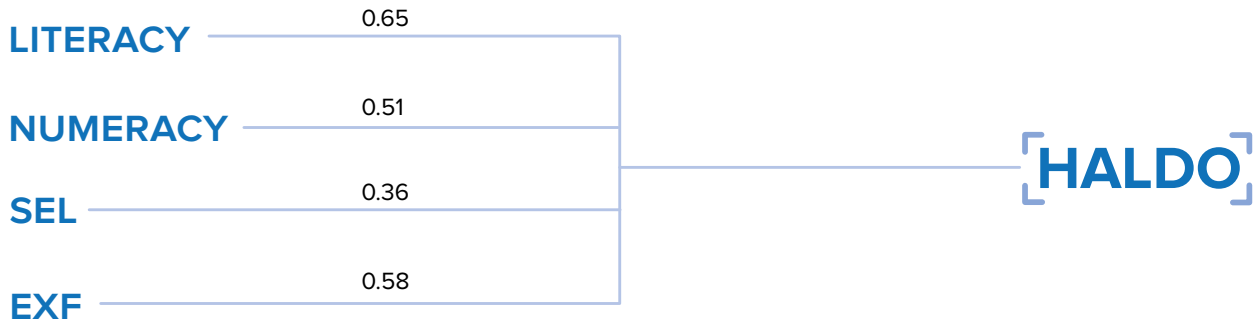
TEST QUESTION	DISCR. INDEX	FACTOR LOADING
Letter identification (common) (max=5)	0.52	0.52
Expressive language (max=10)	0.44	0.45
Letter identification (infrequent) (max=5)	0.50	0.53
Reading fluency (initial, 0-1)	0.53	0.63
Reading fluency (complete, max=30)	0.53	0.63
Reading comprehension (max=5)	0.53	0.62
Numeracy:		
Simple number identification (max=5)	0.33	0.32
One to one correspondence:		
Item 1	0.26	0.23
Item 2	0.29	0.26
Item 3	0.35	0.32
Harder number identification (max=5)	0.41	0.41
Simple operations (max=4)	0.52	0.54
Harder operations (max=4)	----	----
Word problems:		
Item 1	0.42	0.43
Item 2	0.39	0.38
SEL (Self Concept):		
SEL 1	0.01	-0.01
SEL 2	0.23	0.23
SEL 3	0.14	0.15
SEL 4	0.19	0.17
SEL 5	0.17	0.17
SEL 6	0.28	0.29
SEL 7	0.48	0.58
SEL 8	0.54	0.65
SEL 9	0.55	0.68
SEL 10	0.56	0.69
SEL 11	0.54	0.68
SEL Empathy:		
SEL 12	0.36	0.42
SEL 13	0.44	0.52
SEL 14	0.44	0.52
SEL 15	0.38	0.40
SEL 16	0.40	0.44

TEST QUESTION	DISCR. INDEX	FACTOR LOADING
EXF (Short-term memory):		
EXF 1	----	----
EXF 2	----	----
EXF 3	----	----
EXF 4	----	----
EXF 5	----	----
EXF (Working memory):		
EXF 6	0.44	0.44
EXF 7	0.46	0.48
EXF 8	0.48	0.51
EXF 9	0.45	0.50
Eigenvalue	----	7.4
Chronbach's Alpha (α)	----	0.88

Data source: RET Mauritania HALDO assessment (2024)
Notes: Standard deviations in parentheses. Number (or percent) correct in the Scored column is based on all 505 students in the sample. Eigenvalue, Discrimination index and Chronbach's Alpha are only calculated for the entire sample (505 students). Factor loadings refer to the first factor from Principal Component Factor (PCF) analysis.

The full set of items were also included in a global IRT analysis. These findings are not as easy to summarize, but the main takeaways are similar to the global analysis using CTT tools. The relatively easy items do not add much discrimination power, but most of the other items do have predictive power for an overall measure of knowledge. In general the discrimination power is lower than in the content area-specific analyses, but this kind of "dilution" in predictive power is not surprising when combining four different areas of content.

Figure A5 provides a visual summary of how the four HALDO components load onto a single factor based on a simple principal component factor analysis specification. The largest loadings are for Literacy and Executive Functioning, followed by Numeracy. The socio-emotional learning (SEL) block of questions has the smallest loading onto the overall HALDO factor.



Additional correlational and factor analyses were carried out using different measures of HALDO performance on the four domains (e.g. percentage correct, IRT-generated averages, etc.). These results are mixed but this is not surprising given the diverse content that is included in HALDO. One takeaway from this analysis is that the overall HALDO measure—while serving a purpose by providing a basic summary of student foundational skills—is a difficult construct to interpret in terms of what it is capturing.

The automatic recoding rules in HALDO are necessary to streamline the assessment and derive easily interpretable measures of overall scores based on percentage correct. But these recodes could potentially introduce error into the measurement of student performance by assuming that low scoring students cannot complete more advanced tasks, or assigning more advanced students perfect scores on lower-order content. This section generated individual item performance characteristics using classical test theory (CTT) tools for the recoded (scored) data, and modern (IRT) test theory tools for the raw data answers (no automatic scoring). The overarching goal of these comparisons is to assess "item stability" across the different scoring strategies. The main finding is that the individual item performance characteristics from the four areas of the HALDO assessment do not vary much between the two scoring strategies.

This analysis is not intended as a validation of the HALDO assessment tool. The data are specific to a single sample of older children (grades 4 and 6) from one country, so it cannot be assumed that the same results will be found in other contexts. The purpose is simply to understand more about how the instrument is performing under different scoring scenarios in order to provide background context for the data analysis in Section 5, and identify any serious issues with the instrument that could impact interpretation of the main findings (none were found).

The consistency of the CTT- and IRT-derived parameters provides support for the HALDO tool. But there are still concerns about the use of the automatic recodes in the scoring, which are addressed in more detail in Section 5 and again in the concluding sections of the report. Furthermore, just because the instrument has acceptable levels of reliability does not mean that the assessment is measuring meaningful content and skills, which is a much harder question to address and again may vary substantially across different contexts (and age groups).

MAURITANIA. In 2023, Mauritania faced an influx of more than 55,000 arrivals, driven by the deteriorating security situation in neighboring Mali - a significant increase compared to 2022 (12,000 arrivals). ©**UNHCR/XAVIER BOURGOIS**

Annex B: Additional results

Table B1. Student background characteristic comparisons by school type and grade (unweighted)

CHARACTERISTICS	GRADE 4		GRADE 6	
	Camp	Host	Camp	Host
Student is female	54.6	49.0	43.5	53.3
Student age in years	12.6	11.9	14.4	13.6+
Parent has phone number	41.7	24.5+	61.7	63.3
Languages spoken at home:				
Hasania	42.7	100.0*	46.7	98.3*
French	2.5	0.0	12.3	1.7*
Sonrai	2.5	0.0*	7.1	0.0*
Tamasheq	62.7	0.0*	64.9	0.0*
Arabic	3.7	0.0*	0.6	6.7
Student lives with both parents	70.7	71.4	70.8	76.7
Student has any disability	1.2	2.0+	1.3	1.7
Student has lived here:				
Less than one year	5.0	0.0*	2.6	1.7
1-2 year	5.4	8.2	9.7	6.7
More than 2 years	87.2	91.8	83.8	90.0
Has books in home	36.5	53.1*	54.6	68.3
Someone reads books in home	27.4	44.9*	45.5	53.3
SES Factor	-0.20	0.90*	-0.20	0.56*
Class is only in French language	38.4	4.1*	50.6	1.7*
Sample size (n)	242	49	154	60

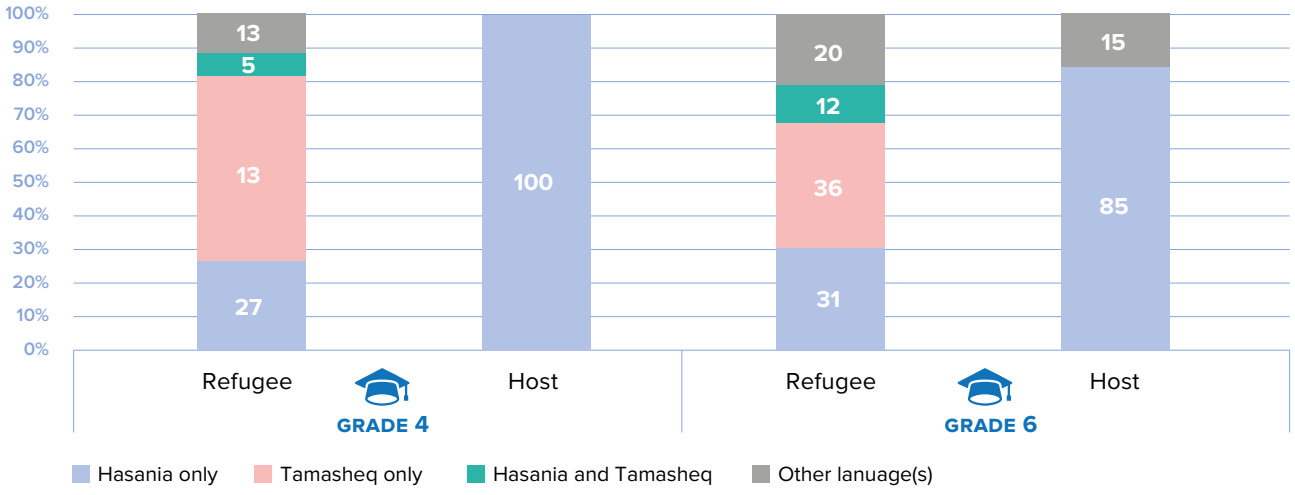
Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to averages (means) by grade and category based on un-weighted data. Tests of significance for differences between Camp and Host school category averages are based on robust standard errors that correct for student clustering.
*Statistically significant difference at p<=0.05 level
+Statistically significant difference at p<=0.10 level

Table B2. Comparison of overall grade 4 and 6 HALDO scores with and without sampling weights

CHARACTERISTICS	GRADE 4		GRADE 6	
	No weight	Weight	No weight	Weight
Literacy percent (equal)	48.8	48.1	65.7	70.0
Literacy percent (points)	64.0	63.3	77.9	80.8
Numeracy percent (equal)	67.2	67.9	75.6	77.8
Numeracy percent (points)	60.2	60.7	70.0	72.8
SEL percent	74.6	75.5	80.3	82.7
EXF percent	58.8	59.4	60.2	65.1
Global percent (equal)	66.4	66.9	74.2	77.2
Global percent (points)	66.3	66.6	76.6	79.5
Sample size (n)	290	290	214	214

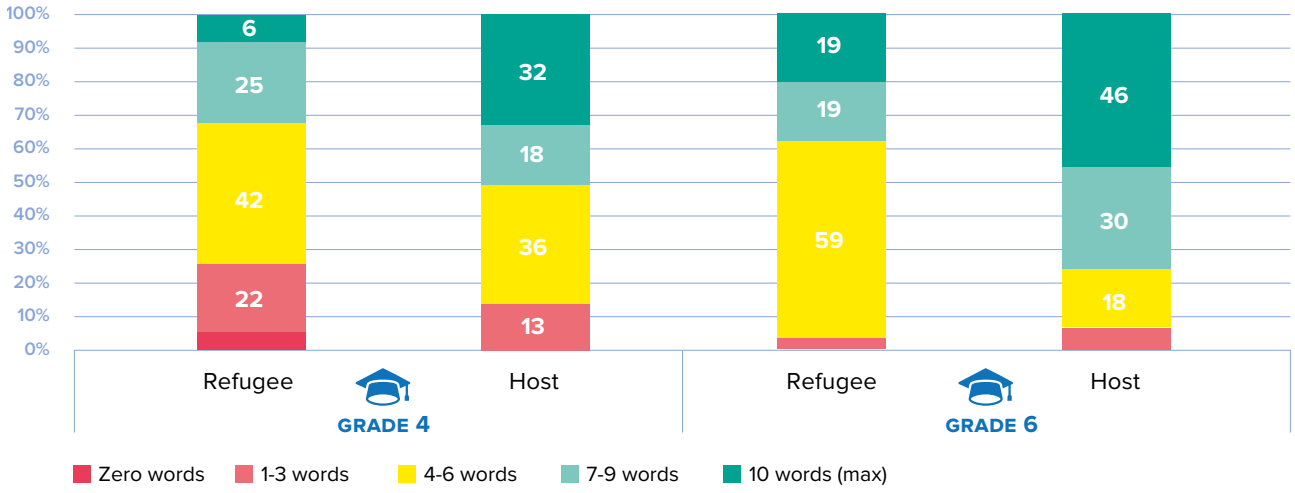
Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers refer to overall averages (means) by grade (Camp and Host schools combined) with and without the sampling weight.

Figure B1. Student language use in the home by grade and school category



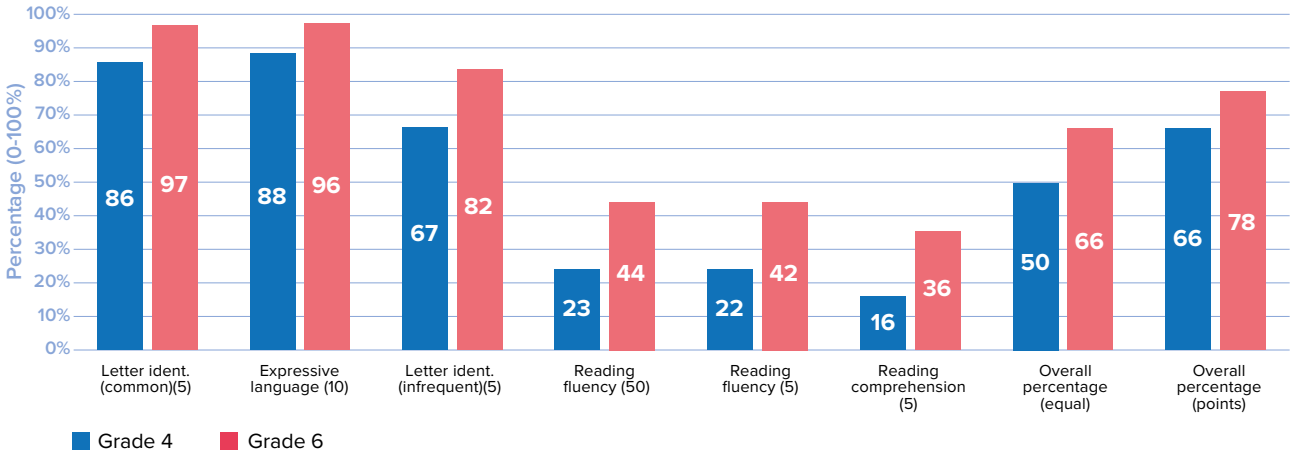
Data source: RET Mauritania HALDO assessment (2024)

Figure B2. Summary of expressive language (Literacy 1.1) by grade and school category



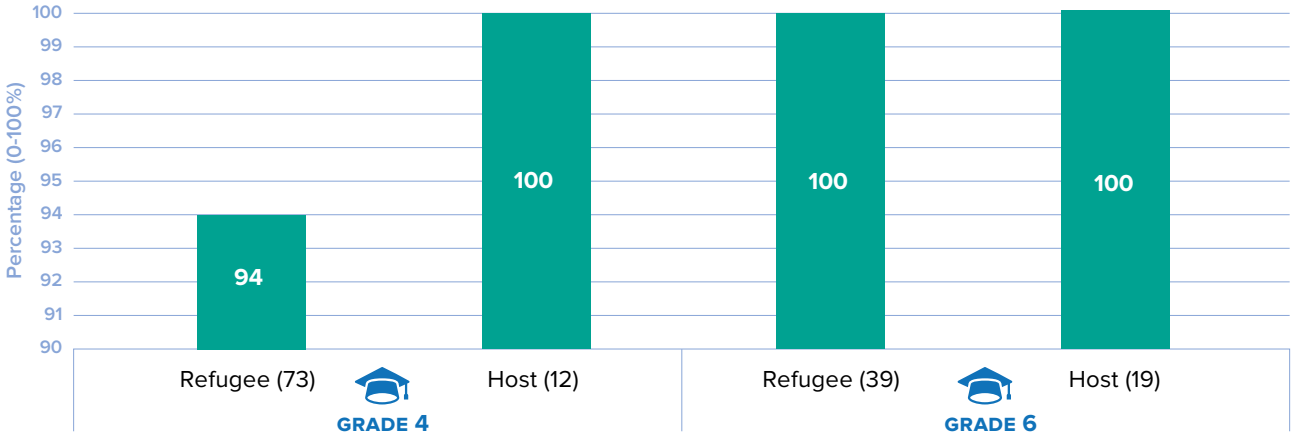
Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers of students who answered this question are indicated in parentheses.

Figure B3. Percentage correct by literacy item and grade, Mbera refugee camp students



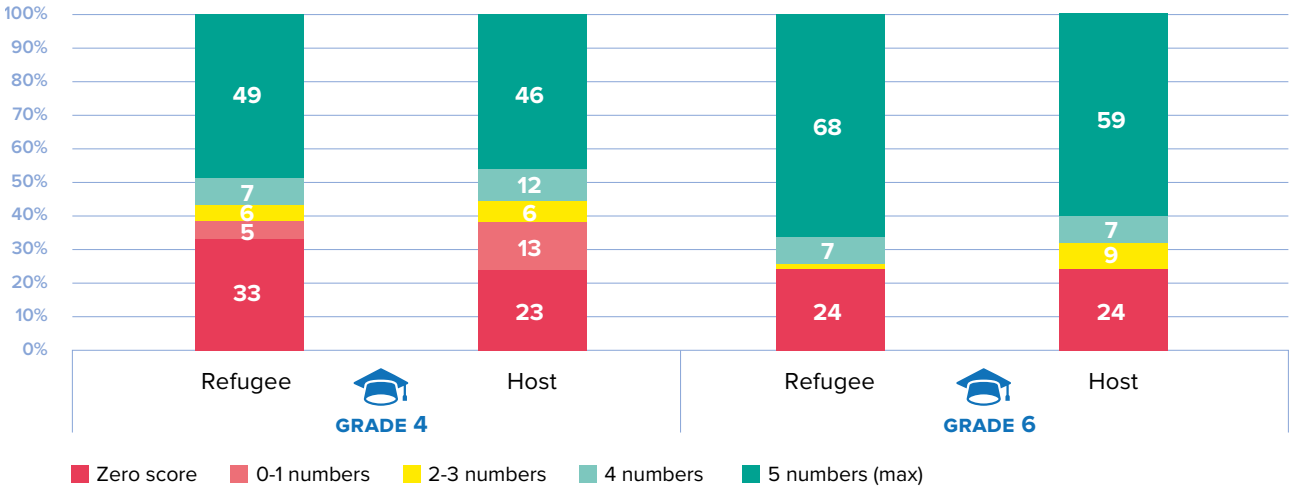
Data source: RET Mauritania HALDO assessment (2024)

Figure B4. Percentage of students who completed one-to-one correspondence tasks (Numeracy 1.1) by grade and school category



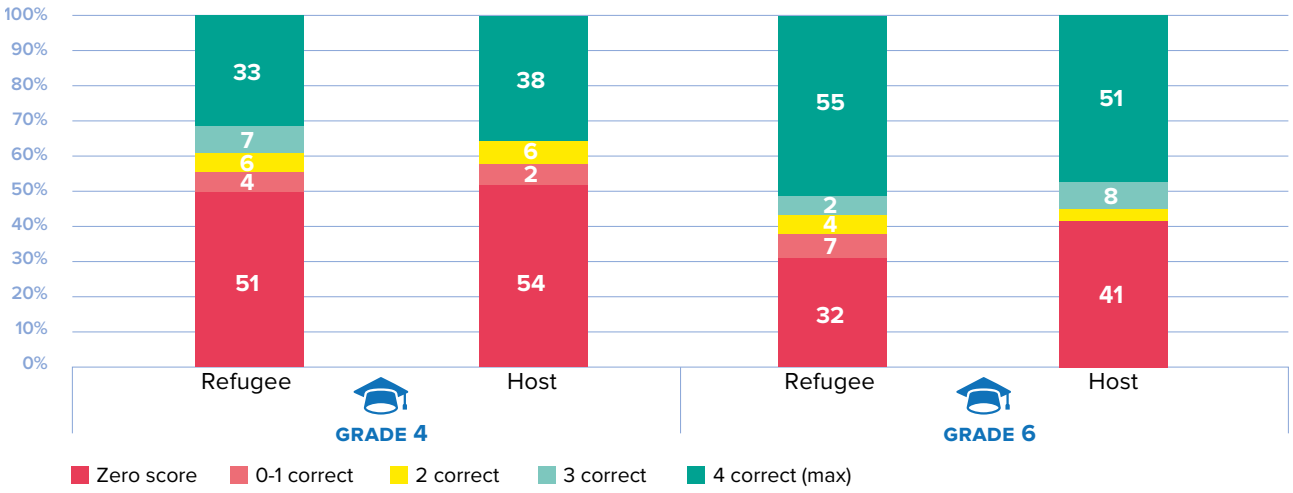
Data source: RET Mauritania HALDO assessment (2024)
Notes: Numbers of students who answered this question are indicated in parentheses.

Figure B5. Summary of hard number identification (Numeracy 2) by grade and school category



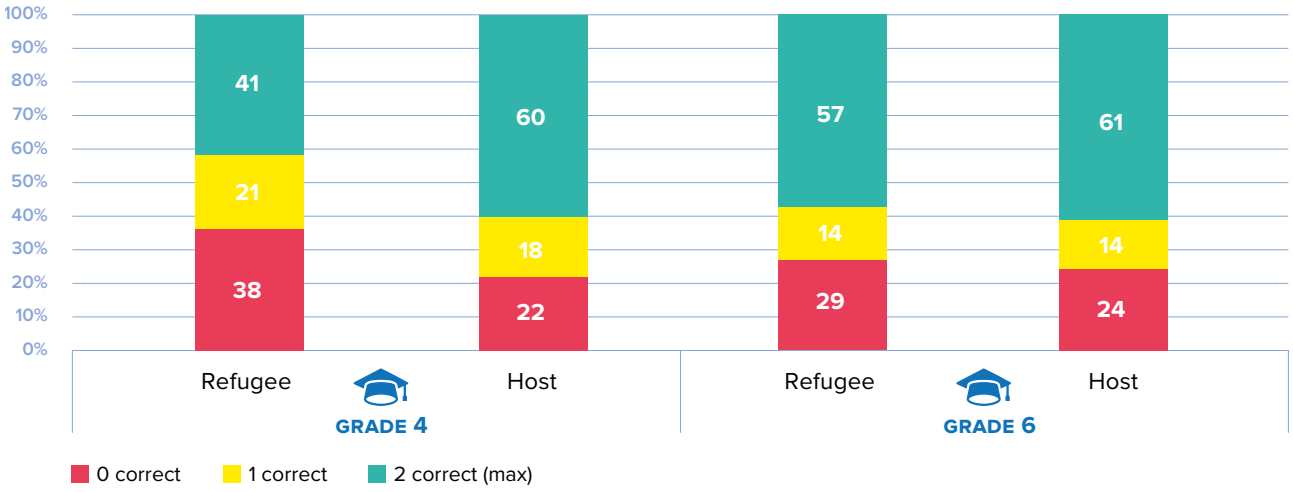
Data source: RET Mauritania HALDO assessment (2024)

Figure B6. Summary of simple operations (Numeracy 3) by grade and school category



Data source: RET Mauritania HALDO assessment (2024)

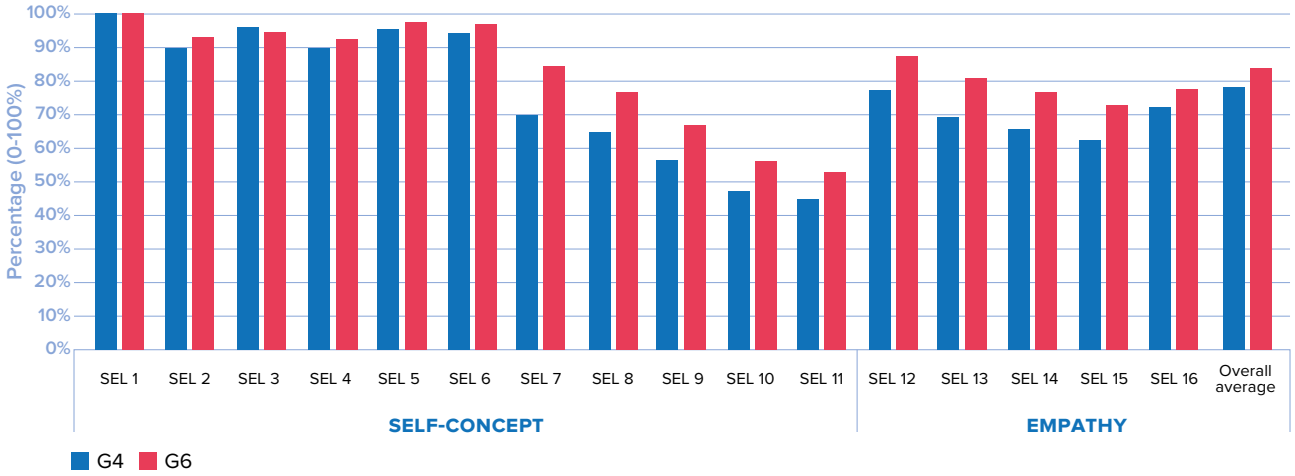
Figure B7. Frequency for number of correct answers for word problems (maximum=2) by grade and school category



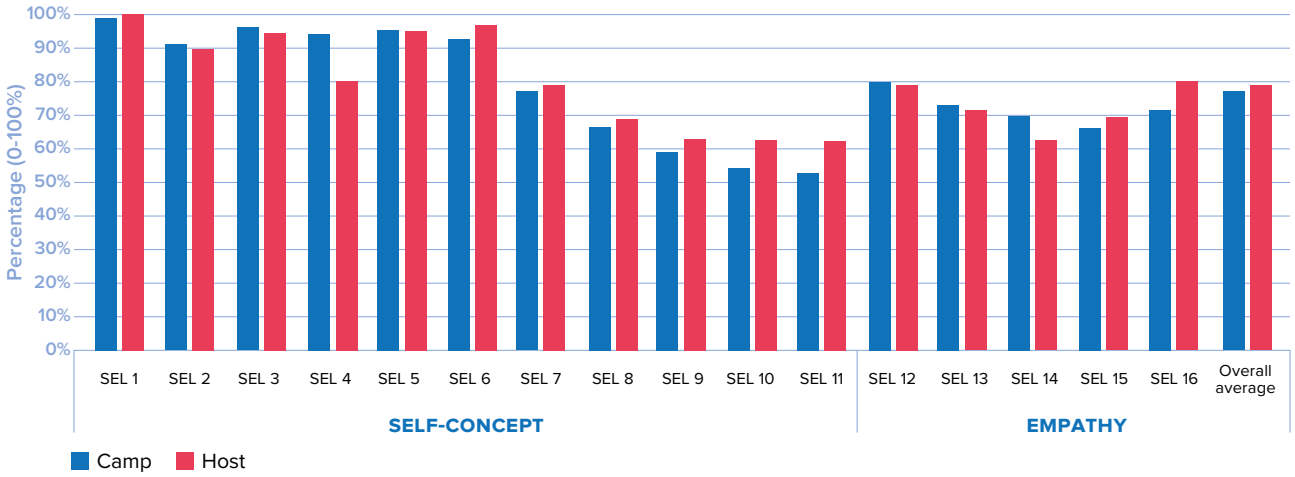
Data source: RET Mauritania HALDO assessment (2024)

Figure B8. Socioemotional learning (SEL): percentage correct by item, grade level and school category

(a) By grade



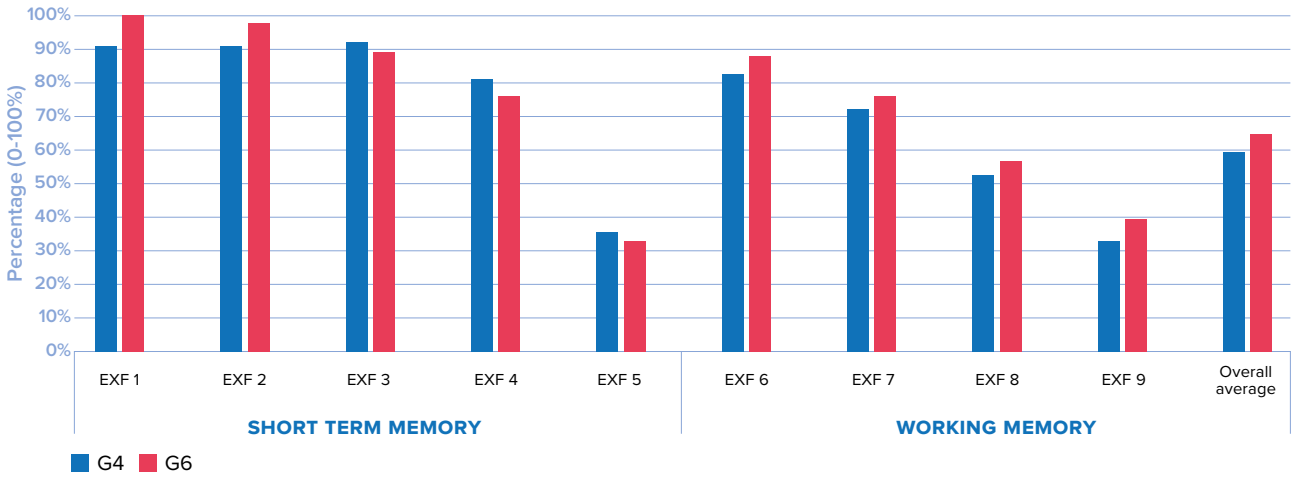
(b) By school category



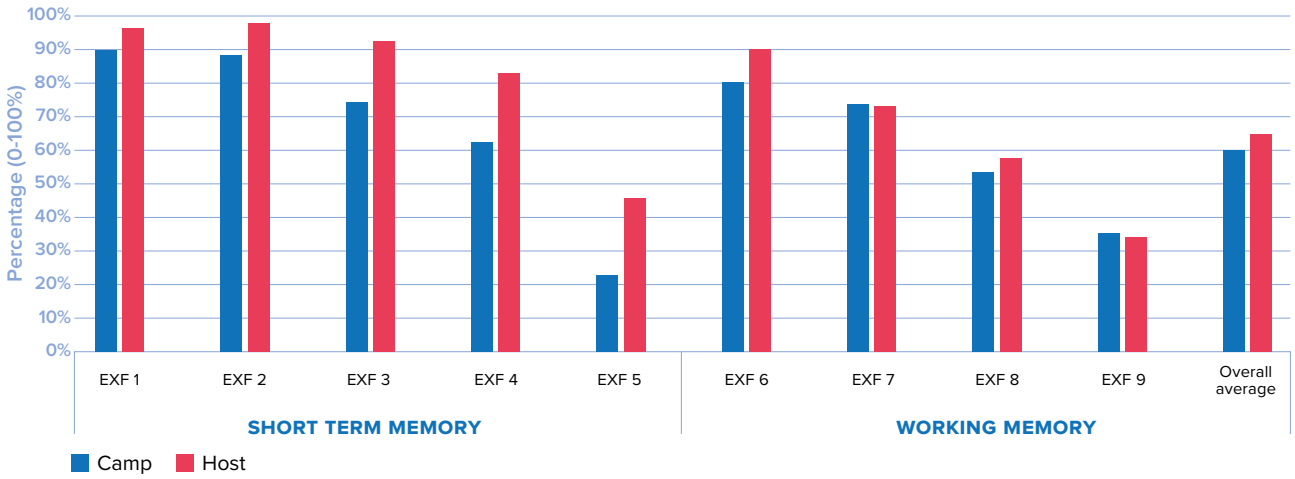
Data source: RET Mauritania HALDO assessment (2024)

Figure B9. Executive functioning (EXF): percentage correct by item, grade level and school category

(a) By grade



(b) By school category



Data source: RET Mauritania HALDO assessment (2024)

MAURITANIA

Measuring Holistic Learning Outcomes for the Forcibly Displaced

PRODUCED BY UNHCR
January 2025

FRONT COVER AND BACK COVER
MAURITANIA
©UNHCR/RET International in Mauritania

UNHCR
The UN Refugee Agency

This document along with further information is available on UNHCR's education website: www.unhcr.org