Girls' Education Challenge

UKaid from the British people

USAID FROM THE AMERICAN PEOPLE

Prepared On Behalf Of:

care®

Leave No Girl Behind

Adolescent Girls' Education in Somalia (AGES)

Midline Evaluation - 2

Prepared By:

consilient

June 2023

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABE	Accelerated Basic Education
AGES	Adolescent Girls' Education in Somalia
ALP	Alternative Learning Programme
al Shabaab	Harakat al-Shabaab al-Mujahideen
AMISOM	African Union Mission to Somalia
BDR	Banadir (region)
BL	Baseline
C1 NFE	Cohort 1 Non-Formal Education
C4 NFE	Cohort 4 Non-Formal Education
C5 NFE	Cohort 5 Non-Formal Education
CEC	Community Education Committee
CPYDS	Chinese Positive Youth Development Scale
DEO	District Education Officer
EGMA	Early-Grade Math Assessment
EGRA	Early-Grade Reading Assessment
FCDO	Foreign and Commonwealth Development Office
FE	Formal Education
FGD	Focus Group Discussion
FGS	Federal Government of Somalia
FMS	Federal Member State
GEF	Girls' Empowerment Forum
IDP	Internally-Displaced Person
JSS	Jubaland State
KII	Key Informant Interview
LSI	Life Skills Index
ML	Midline
MOE	Ministry of Education
NFE	Non-Formal Education
OOS	Out-of-School
PSU	Primary Sampling Unit
REO	Regional Education Officer
SOS	Somali Shilling
SWS	South West State
USAID	United States Agency for International Development
USD	United States Dollar
VSL/VSLA	Village Savings and Loan Association
YLI	Youth Leadership Index

Executive Summary

Learning

Original Baseline Cohorts

Learning outcomes among the original baseline cohorts are mixed. The FE and ABE cohorts showed substantial learning gains at ML1 and continue to perform above their baseline levels, but have experienced significant backsliding in both literacy and numeracy since ML1. This learning loss is concerning, and is coupled with a failure, even at ML1, to meet the learning benchmarks established at baseline.[1] C1 NFE girls, on the other hand, have shown no learning improvements since baseline on either literacy or numeracy, in either the ML1 or ML2 rounds.

Learning gains among FE girls include an increase in literacy of 20.5 points since baseline, compared to an expected (benchmarked) achievement of 29.7 points. This cohort came closer to meeting the numeracy targets, gaining 15.3 points since baseline, compared to an expected improvement of 18.5 points. ABE girls have also failed to meet their benchmarks, but by larger margins, falling short of the literacy and numeracy benchmarks by 15.4 points and 11.5 points, respectively.

An important caveat when assessing learning outcomes concerns differential exposure to schooling among the cohorts. Girls who remained enrolled in FE consistently improved considerably, but the cohort's overall scores are reduced by the number of girls who dropped out over time; the same is true of ABE and NFE girls.[2] While even consistent attenders generally failed to meet improvement benchmarks, these differential gains underscore the importance of continuous enrolment to achieving the programme's learning goals.

C4 NFE Cohort

Improving learning outcomes—numeracy and Somali literacy—are at the centre of the AGES programme's goals. C4 NFE girls were enrolled before ML1 and completed the 10-month NFE programme. At the ML2 evaluation point, in the aggregate, we find that both numeracy and literacy scores improved significantly from a low base set of scores at ML1. The C4 NFE girls improved their average numeracy scores from 21.1 points to 49.6 points, and average literacy scores increased by 18.3 points to 34.9 points.

Examining changes in learning outcomes by region, C4 NFE girls in Banadir, Lower Shabelle, and Middle Shabelle improved their numeracy scores from 24.6 to 28.2 points on average while in Bay C4 NFE girls only improved by 2.3 points on average. With regard to literacy scores, C4 NFE girls in Banadir and Lower Shabelle improved their literacy scores 23.2 and 24.9 points, respectively, while scores in Bay stagnated with only a 1.8 point increase. Girls from Bay scored significantly higher than girls from all other regions at ML1, but by ML2 the numeracy and literacy gains were minimal and not statistically significant.

C5 NFE Cohort (Baseline)

Before ML2, C5 NFE girls were enrolled into the 10-month NFE programme designed to provide them with basic arithmetic and literacy skills useful in the labour market. The mean Somali literacy score of C5 NFE

[1] Note from the project: The sentence is inaccurate. The first FE cohort surpassed benchmarks at the first midline round, although the first ABE cohort failed to achieve the benchmark. No benchmarks were set for NFE.

[2] Note from the project: Girls who dropped out continue to be longitudinally tracked by the project and completed learning assessments as part of data collection.

girls was 27.4 and the mean numeracy score is 40.2. The scores were relatively low compared to all girls, but higher than C4 NFE girls when they enrolled before ML1.

Transition

Original Baseline Cohorts

Transition, as an outcome, is difficult to assess, given the lack of comparison group and the lack of national statistics on grade promotion rates and the frequency of consecutive enrolment. Despite this, a surprising share of FE girls remain enrolled 3+ years after their initial enrolment in late 2019; indeed, 60.7 percent of FE girls remain enrolled and have advanced three grade levels since that time, despite school disruptions caused by COVID-19. In total, 75.5 percent of FE girls remain enrolled.

The most compelling finding regarding the original ABE and NFE cohorts is the number of NFE girls who have successfully moved into employment since the conclusion of their learning programmes, in a context of high youth unemployment. Over half of all NFE girls (59.6 percent) are employed at ML2, and a further 8.6 percent are either self-employed (2.6 percent) or enrolled in formal school (6.0 percent). A significant share of ABE girls are also employed, though at a lower rate, which likely reflects their younger age and the fact that a higher share (18.4 percent) are now enrolled in formal school. A large share of ABE girls are, unfortunately, neither in employment, school, or a learning programme.

C4 NFE Cohort

C4 NFE girls experienced relatively higher levels of successful transition by ML2 than that of C1 NFE girls by ML1. The overall successful transition rate is 74.4 percent at ML2 for the C4 NFE girls with the most common successful transition outcomes being retention in the NFE programme (37.2 percent) and age-appropriate, non-exploitative work (28.0 percent). It is important to note that retention in an NFE programme – for this cohort – is an "easier" outcome, compared to the C1 NFE cohort mentioned above. This is because the ML2 round comes just 12 or 13 months after this cohort's enrolment into NFE; thus, remaining enrolled in the NFE programme is a more common outcome than for the C1 NFE cohort at ML1, when over two years had passed since enrolment.[3]

Nevertheless, there were a substantial proportion of unsuccessful transitions with 25.6 percent of girls now out-of-school after the NFE programme but also not employed following the NFE programme. We observed strong regional differences for transition rates with C4 NFE girls in Middle Shabelle having a transition rate 20.6 percent higher than the aggregate and Banadir with a transition rate 11.7 percent lower.

Sustainability

Sustainability in the AGES evaluation is assessed only for the original baseline cohorts, supported by FCDO. The sustainability of outcomes or improvements after the end of programme support might be challenging. While there have been improvements in all sustainability indicators from BL to ML1, at ML2 we didn't find any further

[3] Transition outcomes are self-reported by girls. Importantly, this means that girls may report current enrolment in a learning programme that has recently ended (or even one from which they dropped out). This almost certainly applies to some girls who self-report continued enrolment in ABE (since late 2019) or in the Cohort 4 NFE programme, the latter of which ended prior to the start of data collection for this ML2 evaluation round. Alternatively, girls may have shifted into other learning programmes or continuation programmes that are not part of the AGES intervention. We discuss this issue in more detail in Sections 5 and 10 of the report, which analyse transition outcomes using alternative coding schemes. Here we take respondents' self-reports uncritically, though this may result in overstating the transition rate. *Note from the project: It is possible that part of the respondents may have shifted into other programmes beyond those offered by AGES, including but not limited to the ABE course provided by Bar ama Baro, which allows for transition into levels 3-4.*

significant improvement on most areas of analysis from ML1 to ML2, with the notable exception of the adoption of new teaching practices and increased support for inclusive education.

Parents' and caregivers' attitudes towards girls' education has improved overall, with 44.6 percent of them (from 62.9 percent at BL) believing that costs of education and families' financial difficulties are reasons to withdraw a girl from school. Considering limited household savings and poor participation in VSL by both caregivers (7.9 percent) and girls themselves (25.3 percent, with only half of them active), it is not clear whether these improvements can be maintained in case of adverse economic shocks. Efforts aimed at empowering girls to be changemakers in their own communities have likewise achieved mixed results: while there has been a certain rate of continued participation and engagement in GEFs, the range of activities carried out has significantly decreased from ML1 to ML2 and is likely to indicate a progressive reduction in their effectiveness.

On the other hand, the main positive signal in terms of sustainability of programme intervention is observed among teachers, as not only a higher number (and proportion) of teachers has been trained in new teaching methodologies, but quantitative and qualitative evidence point to an uptake, although slow, of these new approaches. Increased adoption of inclusive teaching approaches is also mirrored in the overall improvement in enrolment of GWDs, which however is still challenged by the limited availability of costly assistive devices and inadequate school facilities. We can also assume a spillover effect of the adoption of new teaching methodologies practices, including on inclusivity, as their successful implementation by some teachers can be observed by others, including new and not yet trained teachers and therefore support their spread and sustainability in the longer term.

Attendance

Original Baseline Cohorts

Attendance is a crucial intermediate outcome that is hypothesized to have a positive impact on learning outcomes and transition rates. The attendance data of FE girls has been collected in BL, ML1, and ML2 and reveal several important changes to their attendance levels since BL. First, the headcounts of classrooms by our enumerators reveal that the share of female students in the classroom has risen to nearly equal that of boys since 2019. At ML2 girls make up 49.3 percent of the students in a classroom which is up from only 39 percent in 2019. This change in the gender composition of a classroom is due to the rapid enrolment of girls into school which rose 91 percent compared to 79 percent for boys over the same period. Second, the attendance rate of girls has dropped slightly from baseline from 87.2 percent in the BL to 83.9 percent at ML2. Finally, our regression analysis suggests that the fall in attendance may in part be due to the sample of girls aging over time and entering grades in which they are more predisposed to under-attend.

C5 NFE Cohort

C5 NFE girls began their programme in 2023, immediately prior to this evaluation; thus, this evaluation represents a baseline for their attendance outcomes. At ML2, this cohort of NFE girls had an average attendance rate of 64.9 percent and the dropout rate was 9.2 percent. The 9.2 percent dropout rate is not dissimilar from the 5.7 percent dropout rate of FE girls who dropped out since F1, but the C5 NFE girls are only a few months into their program, and further attrition may occur before the end of their programme.

Teaching Quality

We analyse four dimensions of teaching quality and practices: professionalism, gender equity, the use of physical punishment, and pedagogical practices. AGES' interventions to strengthen teaching quality are expected to improve learning and transition outcomes as girls are more effectively engaged by teachers.

Original Baseline Cohorts

For FE schools, we find a significant improvement since baseline in teacher absenteeism as reported by girls. Regarding gender equity, results are mixed; girls were significantly more likely to report that teachers treat boys and girls differently at ML2, but we find a significant increase in the percent of classrooms where teachers directed questions at both genders. We also find that teachers tend to use practices favouring girls in class, but that there has been little relative change in teachers' use of equitable practices for girls over time.

Looking at disciplinary practices, we find a significant and consistent decline in the use of physical punishment since baseline. However, around three-quarters of FE girls reported that their teachers punish them for wrong answers and around one-third reported that teachers use physical punishment. While the latter result is a significant decline from baseline, reported use of physical punishment increased since ML1, suggesting a need for continued focus on reducing use of this negative practice.

Lastly, we find some improvement in use of positive pedagogical practices, including formative assessments and engagement of students who are not participating in class. However, overall, the use of positive teaching practices remains somewhat limited, and many teachers still rely on traditional practices including having students copy from the board and repeat after the teacher. Continued emphasis on positive teaching practices and sensitization on the inefficacy of traditional teaching practices may help further bolster learning and transition rates by endline.

C5 NFE Cohort

We find that over 90 percent of C5 NFE students stated that their teachers made them feel welcome. However, around one-quarter of girls stated that their teachers were often absent; teacher absenteeism appears to be a widespread issue across all zones addressed by the programme. Furthermore, around half of girls reported that teachers punished them for wrong answers and 15 percent reported the use of physical punishment by teachers, showing a clear need to reduce NFE teachers' use of negative disciplinary practices. In a more positive finding, in classroom observations, C5 NFE teachers were frequently observed engaging students who were not participating, using open-ended questions, soliciting student opinions, and providing positive feedback to girls. Overall, these results suggest that teachers in C5 NFE centres across all zones are attempting to engage students by using participatory, positive teaching practices, and that many girls have positive perceptions of the quality of teaching in their NFE centres. However, there remains substantial room for improvement across all assessed pedagogical practices.

Leadership and Life Skills

Original Baseline Cohorts

The acquisition of leadership, self-esteem, and life skills is an intermediate outcome which the Theory of Change hypothesizes will lead to improved learning outcomes as well as the ability to transition successfully after leaving the programme. Life skills were measured primarily through the Youth Leadership Index (YLI), a set of 21 self-reported questions which asks girls about the sense of control they have over their lives, their confidence in class, and their ability to organize others toward a common goal. The YLI is scaled from 0 to 100. The scores of FE, ABE, and C1 NFE girls on the YLI have risen significantly, from 49.0 percent in the baseline to 66.5 percent at ML2. Furthermore, the proportion of girls who met the AGES target of 70 percent on the YLI has risen from 11.5 percent in the baseline to 42.1 percent.

The girls in the various FE, ABE, and C1 NFE cohorts experienced similar YLI improvements, however there were important regional differences. Girls in Jubaland experienced the highest increase in YLI scores – from an average score of 44.6 at BL to an average of 68.0 at ML2 – while those in Banadir saw their scores rise

from 49.7 to 68.1, and those in South West State experienced a smaller rise, from 52.7 to 62.0. The share of girls who reached the target of 70 percent on the YLI score reached 46.6 percent in Banadir, 44.3 percent in Jubaland, and 32 percent in South West State.

A noteworthy finding is the link between life/leadership skills and learning outcomes. YLI scores are among the best predictors of a girl's learning outcomes over time; for example, a 10-point increase in YLI scores from BL to ML2 is associated with a 1.7-point improvement in numeracy and 2.5-point increase in literacy scores, respectively, among FE girls. Similar findings among ABE and C4 NFE girls suggest a consistent link between leadership skills and learning outcomes. In addition, our analysis shows that participation in a Girls' Empowerment Forum (GEF) – which is intended to increase leadership skills – also impacts learning and transition outcomes positively, lending further support to the link between life/leadership skills and higher-order educational outcomes.

C4 NFE Cohort

The C4 NFE girls showed a marked improvement in their YLI scores from ML1 to ML2, increasing 16.6 points to 66.4, indicating a clear win for the programme in raising leadership skills among the C4 NFE cohort. In addition, the proportion of girls who attained a score of at least 70 on the YLI rose from 11.3 percent to 42.7 percent, another sign of the substantial progress made with the C4 NFE cohort.

C5 NFE Cohort

ML2 is the baseline study for C5 NFE girls, however the YLI scores for them are actually higher than for any other cohort of girls. Furthermore, the proportion of girls scoring at least 70 on the YLI is the highest among all cohorts with 47 percent of them attaining that mark. The higher-than-expected YLI of the C5 NFE girls cannot be attributed to age since they are not older than the other NFE cohorts. There were significant differences by region. Girls in Banadir have a significantly higher baseline YLI score (67.2) than Jubaland (56.1) or South West State (55.7).

School Management and Governance

Strong school management benefits student learning, enrolment, and retention by strengthening the quality of teaching, improving student attendance, and expanding community support for girls' education. To improve school management and governance, AGES provides support to Community Education Committees (CECs) to improve their ability to reach and assist marginalised girls, especially girls with disabilities, to enrol and remain in school.

Original Baseline Cohorts

For FE schools, at ML2, we find that 97 percent of FE schools had a CEC and that 97 percent of CECs had visited schools in the past year to monitor facilities, a substantial increase since baseline. CECs were most active in promoting the enrolment of out-of-school children and students who had dropped out, tracking student and teacher attendance, and undertaking child protection activities, including preventing the use of physical punishment by teachers. These results suggest a relatively high level of activity of CECs. However, many CECs had little involvement in addressing financial barriers, which were a key reason cited by caregivers for why children might not attend school. Indeed, CEC members frequently claimed that resource constraints were a major barrier to their efficacy.

C5 NFE Cohort

NFE girls face many barriers to education, including financial, accessibility, and cultural barriers. These barriers can be more acute for the ultra-marginalised girls targeted by AGES, including girls with disabilities,

pregnant girls and mothers. To address these barriers, C5 NFE girls reported that CECs most commonly support enrolment of out-of-school students and students who have dropped out and follow up on student absenteeism. However, NFE girls reported a somewhat limited level of activity of CECs. Overall, our analysis suggests that CECs are moderately active in supporting NFE centres and girls, but that they lack resources and that activities may not align perfectly to the barriers faced by girls and marginalised students.

Community Attitudes

Original Baseline Cohorts

Community attitudes around girls' education can play an instrumental role in helping marginalised girls overcome the barriers preventing them from attending and completing school. We analyse changes in financial support for girls' education and caregivers' attitudes towards new roles for girls. While caregivers express high value for girls' education, at ML2, a substantial portion stated that it was acceptable for a child not to attend school if education was too costly or if the child needed to work, although attitudes have improved since baseline. Encouragingly, we find an improvement in attitudes towards investment in girls' education among subgroups who might be expected to have less support—and less financial resources—for their girls to attend school, including for male heads of household and IDPs.

Looking at caregiver attitudes, we find that caregivers have high aspirations for their girl's education and were less likely to say that housework is an acceptable reason for a child to miss school at ML2 than at baseline. However, we find that FE and ABE girls' housework burdens have increased since baseline and were more likely to prevent these girls from attending school at ML2. In contrast, for C1 NFE girls, housework was less likely to prevent girls from enrolling in school at ML2 – a potential sign of the impact of NFE centres. For both groups, housework burdens appear to have increased more for marginalised girls, including IDPs and those with uneducated caregivers, and are very high for girls with disabilities.

Self-efficacy

Original Baseline Cohorts

The data on positive youth development was collected during ML1 and ML2 from the C1 NFE girls using indicators from the Chinese Positive Youth Development Scale (CPYDS) which assesses resilience and confidence. The overall trend of the C1 NFE girls is one of declining self-confidence and belief that they can direct their lives. At ML1, only 28.6 percent of C1 NFE girls agreed with the statement "I cannot do much to change things in my life." However, at ML2, 56.1 percent agreed with the same statement, an increase of 27.5 percent, indicating a decline in the girls' sense of self-efficacy. Other statements that signalled falling self-efficacy in this cohort included, "I have little control of things that happen in my life" (+17.1 percentage point increase), and "I cannot do much to change things in my life" (+13.4 percentage point increase). There were regional differences in responses. In contrast to the aggregate trend, girls in Lower Shabelle and Gedo region saw declines in agreement with statements that indicated a sense of helplessness and fate not being in their hands.

C4 NFE Cohort

The C4 NFE girls also answered questions from the CPYDS during ML1 and ML2, and the overall trends indicated even greater declines in a sense of self-efficacy. Compared to ML1, C4 NFE girls at ML2 tended to agree more with the statements, "I have little control of things that happen in my life" (+23.8 percentage point increase), "I cannot do much to change things in my life" (+23.4 percentage point increase), and "When I face life difficulties, I feel helpless" (+14.7 percentage point increase). Unusually, there were also increases in agreement with statements of high self-efficacy such as "I believe things in my life are mostly determined

by me" (+21.0 percentage points) and "I can finish almost everything that I am determined to do" (+11.8 percentage points). A few important regional differences emerged. Girls in Lower Shabelle were more likely to experience large increases in agreement with statements that indicate low self-efficacy. Conversely, girls in Bay region had less agreement with statements indicating low self-efficacy.

C5 NFE Cohort

The ML2 evaluation point formed the baseline for the C5 NFE girls, and we find that overall they share a similar sense of self-efficacy to C4 NFE girls. They begin with high levels of agreement with the statements "I have little control of things that happen in my life" (72.0 percent), "I cannot do much to change things in my life" (67.8 percent), and "I do not have any solutions for some of the problems I am facing" (67.5 percent). Conversely, we also observe high levels of agreement that indicate a high sense of self-efficacy such as "I believe things in my life are mostly determined by me" (77.3 percent) and "I can finish almost everything that I am determined to do" (87.3 percent). C5 NFE girls in Bay region[4] generally had higher levels of positive youth development while girls from Middle Shabelle, Lower Shabelle, and Banadir regions had substantially lower levels of positive development.

Strengthened Economic Circumstances of Female Youth

Original Baseline Cohorts

Income levels and participation in income-generating activities have been systematically tracked for the C1 NFE cohort since ML1. Over that time, average income reported by a tracked panel sample has increased significantly. This has been coupled with relatively high employment rates among this cohort, nearly doubling from 33.3 percent of girls having a job or occupation of some kind at ML1 to 64.4 percent at ML2.[5] This is supplemented by an increase in the share of girls who report owning their own small business, which has increased by 5.5 percentage points. These gains in employment and self-employment are matched by improvements in average income, which have increased from $17 to $34 per month among this cohort (with greater gains when we exclude outliers). The most likely explanation for increased income is simply the higher rate of C1 NFE girls who are employed, reflecting – in part – their older age, additional time in the labour force since completing their NFE programme, and perhaps a stronger labour market overall.

C4 NFE Cohort

Economic outcomes for the C4 NFE cohort followed those of the C1 NFE cohort, with a significant expansion in employment and self-employment among this cohort since ML1. For instance, the share of C4 NFE girls who report owning their own small business has increased from 8.1 to 18.4 percent. Average monthly incomes also rose, from $15 to $24; this is a smaller increase than observed among the C1 NFE cohort, but is still a substantial increase, which is driven both by an expansion of employment/self-employment and by higher reported incomes among those reporting any income at all. These findings are especially compelling, given that so little time has passed since the end of the C4 NFE learning programmes – gains in employment appear to have materialised quite quickly among this cohort.

Enhanced Social Support for Female Youth

C4 NFE Cohort

The share of girls who have benefitted from additional development opportunities, outside the AGES intervention, has increased substantially since this cohort's baseline, at ML1. For instance, the share of girls

[4] Note from project – the evaluator had mistakenly referred to C4 in this sentence. Corrected by the project.
[5] Employment in this context includes domestic work

participating in a youth group or youth network of some form has increased from 14.7 to 32.2 percent in the last year, and the share who have participated in local political forums or discussions, and discussions related to service provision in their communities have increased similarly over time. Access to humanitarian assistance has also increased, though it is not the main driver of increases in this indicator.

C5 NFE Cohort

Girls in this cohort are starting at a comparatively high level of exposure to other development, governance, and networking opportunities. At the start of their learning programme, C5 NFE girls already have exposure to youth networks and governance discussions at levels achieved by C4 NFE girls after one year of programme exposure. This may have occurred if C5 NFE girls are already benefitting from youth networks and participatory governance opportunities opened up by NFE centres and the cohort that preceded them.

1. Introduction

The Adolescent Girls' Education in Somalia (AGES) Program, funded under FCDO UK's Girls' Education Challenge (GEC) programme and USAID, is being implemented in southern and south-central Somalia. Implementation began in September 2018 with the mobilisation of out-of-school girls, and their further enrolment into a variety of formal and non-formal learning programmes. From its inception, the programme has expanded through the provision of additional donor funding: at baseline, it was being implemented in Banadir, Jubaland, and Southwest State; it has now been expanded into Hirshabelle and its reach in areas where it was already active has increased.

Geographically, the project covers the portions of Somalia that have been most impacted by the three decades of conflict that began in the late 1980s: (i) areas in which pitched battles occurred, (ii) those which underwent multiple rounds of displacement and "clan cleansing", (iii) regions which were flooded by IDPs fleeing violence elsewhere, (iv) and/or areas which have been under al-Shabaab control for significant periods. Finally, the project is being implemented in a complex governance environment.

For a fuller background on the programme and the context in which it operates, we refer readers to the two previous evaluation reports conducted on behalf of the AGES project. The baseline report describes the project's approach, targeting strategy, and activities in detail[6]; meanwhile, the previous midline report (Midline #1, or ML1) describes the political, social, and economic context of the regions where AGES works.[7] In this prelude to the second midline (Midline #2, or ML2) evaluation report, we document the changes in these contextual factors since the ML1 evaluation took place in early-to-mid 2022.

Since 2022, more than 2.6 million Somalis have been internally displaced, primarily due to drought, insecurity and flooding. In 2022, Somalia experienced its longest drought in decades, due to five consecutive failed rainy seasons, which resulted in mass displacement, severe water shortages and a devastating food crisis, exacerbated by the impact of and interplay with a broader global food price crisis. In addition to the drought, the severe flooding caused by heavy Gu rains and the Juba and Shabelle rivers running from the Ethiopian Highlands had catastrophic effects in 2023, particularly in districts such as Beletweyne (Hirshabelle) and Baardheere (Jubaland), where houses and farms suffered extensive inundation, resulting in the displacement of approximately 219,000 people.[8]

This has meaningfully impacted children and education. At the close of 2022, there were an estimated 4.8 million Somali children aged 5 to 17 years who lack access to education, representing a 15 per cent increase over the previous year.[9] Currently, just 48 per cent of Somali children are enrolled in school, a rate that is lower among newly-displaced and persistent IDPs.[10] Severe food shortages and the inability to cover school expenses due to the drought affected the ability of Somali families to send their children to school. For

[6] Machova, Zuzana, Juuso Miettunen, and Brenton D. Peterson. "Adolescent Girls' Education in Somalia: Baseline Evaluation." June 2020. Available at: https://girlseducationchallenge.org/media/gfacfhwb/ages-lngb-baseline-evaluation.pdf.
[7] Madden, Payce, Brenton D. Peterson, Anastasia Vasilyeva, and Nicolas Wicaksono. "Adolescent Girls' Education in Somalia: Midline Evaluation." August 2022. Available at: https://girlseducationchallenge.org/media/khxngspj/ages_ml-1-public-1.pdf.
[8] UNICEF Somalia Humanitarian Situation Report No. 4 - April 2023. *Note from the project: This event occurred after completion of data collection for this evaluation round.*
[9] UNICEF Somalia Humanitarian Situation Report No. 11 – November 2022. Available at: https://www.unicef.org/somalia/media/3296/file/SitRep%20November%202022.pdf
[10] REACH Somalia. 2023 Multi-Sector Needs Assessment: Data Tables. August 2023. Available at: https://repository.impact-initiatives.org/document/reach/fe62d2d9/REACH_SOM_2023_MSNA_Results_Tables.xlsx.

households their main barrier is the costs, both direct and indirect, of education, especially as the drought continues to disrupt family livelihoods.

Humanitarian needs have continued to increase during this period, resulting in an ever-growing number of Somalis in need of humanitarian assistance. At the end of 2022, more than 8.3 million people were in urgent need of humanitarian assistance[11] and, in 2023, an estimated 1.8 million children under five face acute malnutrition[12] while there are 6.5 million people (31.3% of the population) suffering from acute food insecurity. Seven out of ten Somalis currently live on less than USD 1.90 a day.

In addition to climatic and economically driven challenges, the political and security contexts of Somalia remain fraught. The ML1 evaluation took place in the throes of an acute political crisis that threatened to spiral into a clan- and faction-based conflict. On 15 May 2022, the prolonged electoral process in Somalia concluded with the President of the Federal Government of Somalia elected to office. On the one hand, this quelled the growing tensions in the country around President Mohamed Abdullahi *Farmaajo's* term extensions and the transition of power and the inauguration of President Hassan Sheikh Mohamud (HSM) served as a stabilization force – owed in part to the familiarity that comes with a second term in the office. At the same time, HSM's efforts to ramp up the fight against al Shabaab has increased military action broadly in many parts of the country.

In line with the Federal Government's priority of addressing the threat posed by Al-Shabaab, government-led forces conducted a series of offensive operations against the group in Galmudug and Hirshabelle States, together with ¨community defence forces¨ composed of local clan militias. The joint offensives dislodged Al-Shabaab from several areas in Galmudug and Hirshabelle, including around the Xarardheere, Adan Yabaal, Ruun Nirgood and Buulobarde towns in the Mudug, Shabelle Dhexe and Hiraan Regions, respectively. Al-Shabaab has in retaliation launched indiscriminate attacks that have resulted in loss of lives and displacement.

And, while the issue of the office of the President has been resolved, state level elections, the federalism process, and the establishment of broader power sharing agreements remain contentious issues which contribute to ongoing political tension across Somalia (including Somaliland). Conflict arising from political tensions in Laascaanood, Sool region has also caused loss of an unknown number of lives and around 163,000 internally displaced people.[13] In addition, there have been reports of child recruitment in IDP communities and forced taxation of families by non-state armed groups. Furthermore, insecurity along key supply routes has been disrupting the movement of goods and people.[14]

Since 2022 children have continued to be exposed to multiple risks, with hunger, water shortages and outbreaks of diseases such as cholera increasing. Access to children in need has remained a challenge, with most children in need living in inaccessible areas to UN agencies and partners. The role of local partners in delivering services has remained crucial.

[11] UNICEF Somalia Humanitarian Situation Report No. 12 - End-of-Year 2022. Available at: https://www.unicef.org/somalia/media/3416/file/SitRep%20January-December%202022.pdf.
[12] UNICEF Somalia Humanitarian Situation Report No. 4 - April 2023.
[13] UNHCR Somalia Factsheet - March 2023.
[14] OCHA Somalia drought response and famine prevention report.

2. Methodology

2.1 Overall Evaluation Design

The AGES evaluation employs a pre-post, longitudinal research design, tracking girls of varied cohorts – formal school (FE), accelerated basic education (ABE), and non-formal education (NFE) – across evaluation rounds to analyse changes in learning, transition, and intermediate outcomes over time. As was true during the ML1 evaluation round, the present evaluation round actually consists of two or three distinct evaluations, differing in terms of the girls targeted and, to a large degree, the learning tracks targeted. Most important for our discussion in this section, they differ in terms of the research employed and the data they capture. Below, we highlight the key differences between the different components and cohorts in this report.

Baseline FCDO-Supported Cohorts

The first component of the evaluation consists of girls who have been tracked over time since the AGES baseline was implemented in late 2019; this aspect of the programme was funded by FCDO. It consisted of three cohorts of girls: formal education (FE), accelerated basic education (ABE), and non-formal education (NFE) girls and their respective schools and learning centres. These cohorts have been tracked since the baseline, through the ML1 evaluation round in 2022 to the present ML2 round. It is important to note that ABE and NFE girls in this group completed their learning programmes in 2021 and 2020, respectively, but continue to be tracked to understand the trajectory of their lives since that time. This report serves as the second midline evaluation for these cohorts. Throughout this report, we refer to the "baseline cohort"; when referring to the NFE girls from this group, we refer to Cohort 1 (C1) NFE girls.

As noted above, the evaluation utilises a pre-post design without an in-built comparison group. Sampling occurred within centres, meaning that the baseline consisted exclusively of girls who were enrolled by the project at the centre in question. Household surveys were completed with the girls' caregiver and head of household, such that measures of community attitudes are representative of caregivers and heads of household of girls enrolled in AGES centres, rather than the overall adult population in the areas studied.

In the absence of a comparison group, the results of the evaluation are sensitive to both maturation effects (improvements that accrue naturally with age) and exogenous shocks that impact programme outcomes both among the programme sample and in the wider population. To guard against maturation effects, learning outcomes were benchmarked at baseline, with grade-level differences established based on performance of different grade levels in 2019. For instance, girls in Grade 1 at baseline – who should have advanced to Grade 3 by the time of the midline – will have their respective improvements in learning over that period compared to the difference – at baseline – between Grade 1 and Grade 3 girls. The latter difference represents the maturation effect from Grade 1 to Grade 3 that should be expected in the absence of intervention. We make comparisons to the benchmark girls in our analysis of learning outcomes for FE and ABE girls in Section 4 of the report.

Unfortunately, not all outcomes can be benchmarked in this manner, and the evaluation is sensitive to the impact of exogenous events and secular trends in outcomes. Throughout the evaluation, we discuss changes from baseline to midline and our relative confidence in attributing these changes to the impact of the programme itself.

Cohort 4 NFE Girls

The second component of the evaluation focuses on a cohort of girls who were first exposed to the programme in early 2022. These girls – which we refer to as Cohort 4 (C4) NFE girls – are supported by USAID and were first interviewed as part of the ML1 evaluation. They have been tracked to the current round to form a panel sample. The C4 NFE cohort expanded AGES' geographic reach into Middle Shabelle; otherwise, the regional composition of the cohort was similar (Banadir, Bay, and Lower Shabelle), though often covering different districts. Our analysis of this cohort employs a pre-post design without a comparison group; girls were not benchmarked at ML1, so there is no benchmark set of girls for comparison of learning outcomes.

Cohort 5 NFE Girls

The third component of the evaluation is new this year, consisting of Cohort 5 (C5) NFE girls, for whom this evaluation constitutes a baseline study. C5 NFE girls are – broadly speaking – enrolled into the same learning centres as was Cohort 4. They were recruited and enrolled into the learning intervention immediately prior to this round's data collection, and their learning, transition, and other outcomes will be tracked through to the endline, as is true of the other cohorts described above.

As a baseline evaluation of a new USAID-supported cohort of NFE centres and girls, this component mimics most features of the C4 NFE girls' treatment in the previous round. We will employ an identical pre-post design at endline and the outcomes and intermediate outcomes mirror those of the C4 NFE cohort. Analytically, our treatment of C5 NFE girls will be similar to the C4 NFE girls during the ML1 round, assessing baseline outcomes, subgroup differences in learning, gaps in their literacy and numeracy skills, and seeking to understand the demographic composition of the sample and how that might drive patterns of outcomes we observe in the current and future rounds. The only substantial methodological difference between the C4 and C5 NFE cohorts is in the selection of learning centres and girls within them; we discuss these details further in Section 2.3 below.

2.2 Data Collection Tools

The AGES evaluation is a mixed-methods study, employing a range of qualitative and quantitative tools to capture changes over time in learning, attendance, girls' life skills, and other outcomes. The tools target a range of respondent types, including girls, their family members, teachers, head teachers, CEC members, and local religious leaders. Quantitative tools include the following:

- Learning assessments testing numeracy and Somali literacy among cohort girls
- Girl surveys with cohort girls
- Household surveys with the caregivers and heads of household of most cohort girls
- School survey with head teachers of formal (FE) schools
- Attendance headcounts of ABE, NFE, and FE centres
- Classroom observations, consisting of direct observation of teaching practices and teacher behaviour

During the inception period, CARE's technical staff and the evaluation team reviewed the tools from the baseline and ML1 rounds, with the goal of identifying and making improvements, based on the experience of previous rounds. Two changes made during the ML1 round were carried forward in this round. First, the quantitative survey with teachers, which had been included in the baseline round but excluded from the ML1 round, was left out of this evaluation round.

Second, changes made to targeting of the household survey at ML1 were maintained in this round. In short, the household module – completed by a girl's head of household and her caregiver – was administered only

to girls under the age of 18 years. In contrast, girls 18 and over completed a shorter set of questions – capturing some of the same information – themselves, which is consistent with the notion that girls or women of this age are *their own* caregiver, and the best source of information about themselves. As was true during ML1, this decision has implications for our analysis, because the shortened set of questions applied to girls themselves does not capture all the household characteristics normally included in the household survey. As a result, when we conduct subgroup analysis of project outcomes or project impacts, the sample size available for analysis is, at times, smaller.[15] The table below documents the number of girls in each cohort that completed the abbreviated girls-only module of questions (no interview with other household members), the number of girls targeted for the full household module, and the number of girls whose household members completed the full household module. The discrepancies between the last two columns in the table arise from the fact that some heads of household and caregivers did not consent to being interviewed. Consider the example of ABE girls: 232 ABE girls were 18 years or above and the survey did not include a household module for them; 87 ABE girls were under 18 years, and we attempted to complete a survey with their head of household and caregiver; however, consent was refused in 9 cases, resulting in 78 girls with completed household modules.

TABLE 1: SHARE OF ML2 COHORT GIRLS COMPLETING GIRL-ONLY MODULE VERSUS THE FULL HOUSEHOLD SURVEY

Cohort	Completed Abbreviated Module with Girl Only	Targeted for Full Household Survey	Completed Full Household Survey
FE Girls	67 (15.6%)	382 (89.0%)	362 (84.4%)
ABE Girls	232 (74.8%)	87 (28.1%)	78 (25.2%)
Cohort 1 NFE Girls	298 (98.7%)	4 (1.3%)	4 (1.3%)
Cohort 4 NFE Girls	546 (82.6%)	120 (18.2%)	115 (17.4%)
Cohort 5 NFE Girls	372 (72.8%)	159 (31.1%)	139 (27.2%)
Total	1515 (68.5%)	752 (34.0%)	698 (31.5%)

As the project management team noted during the ML1 evaluation, it is surprising that some FE girls – who were reportedly aged 10-12 years in 2019 when they were enrolled in the project – self-report as being 18 or older in the current round of data collection.[16] There are two explanations for this outcome: the first is

[15] Wherever possible, we mitigated this effect by utilising data from the baseline to determine whether a girl fit into a subgroup of interest, though this was primarily possible with relatively fixed characteristics, i.e. those that do not change or change only slowly over time. For instance, if a girl was 18 years or older and, therefore, completed only the shorter household module, we did not capture data at midline on whether her head of household or caregiver had completed any formal education. However, as this characteristic of her parent or caregiver should not change over time, we attributed her status on this outcome from baseline to her midline data. Again, this is only possible in the context of relatively fixed outcomes; for outcomes that can change – such as whether the girl is married or has given birth – we cannot utilise information collected two years prior, and we are forced to exclude girls 18 years or older from subgroup analyses.

[16] Among FE girls, 26.3 per cent have ages at BL and ML2 which are incompatible with one another (i.e. more than 4 years separating the two age reports).

that many Somalis do not know their true birth year or age, owing to years of conflict- and emergency-driven dislocation and displacement, minimal record-keeping, and the lack of any administrative apparatus for establishing ages in a systematic manner. This issue affects both the project's original enrolment of girls, during which girls or their families may have misstated their ages; it also affects the evaluation team's data collection, because it is equally likely that girls could misstate their age during our interviews. The second reason this occurs is due to errors in data entry – again, either by project staff during enrolment in 2019 or by the evaluation team during ML2 data collection. Wherever possible, the evaluation team sought to cross-verify ages based on previous rounds of data, including reporting by heads of household or caregivers in previous rounds. However, this does not resolve all discrepancies, as caregivers also often provide contradictory ages for their children and, given our experience on previous projects, we believe most discrepancies are a function of misreporting by respondents.[17] More importantly, for our purposes here, this data cleaning task occurs after the interview was already completed, so it cannot be used to change which version of the household survey the girl receives.

One important change was made to the tools during this round, concerning the targeting of attendance headcounts and classroom observations. In the ML1 round, the team completed headcounts and classroom observations in ABE centres and C1 NFE centres. This was done although the cohort girls from these centres were no longer engaged in ABE or NFE programmes, in most cases. For instance, while the ABE cohort tracked since 2019 had already completed their learning programmes, the evaluation team completed headcounts in the ABE centres, essentially tracking attendance in *current* ABE cohorts whose girls are otherwise not considered in this evaluation. This data was used only sparingly in the evaluation report and was not combined with headcounts from other types of centres. However, because this data was of minimal value during the ML1 round, the evaluation team and CARE's technical team agreed to restrict headcounts and classroom observations to formal schools and C5 NFE centres, both of which include currently enrolled cohort girls.

The table below documents the achieved sample of "centre-level" tools – head teacher surveys, classroom observations, and attendance headcounts. As the table shows, there are 37 formal schools in our sample, but we completed head teacher surveys and classroom observations in just 35 and 32 schools, respectively. As we note in Annex 1's description of fieldwork-related challenges, this arose due to examinations that were being completed in Dhobley, a lack of cooperation by at least one head teacher (who prevented data collection from occurring *within the school*), and a school that was found to be non-operational. The same issues applied to attendance headcounts in formal schools.

In C5 NFE centres – shown in the bottom panel of the table below – we were able to complete attendance headcounts in every targeted centre (45 centres in total). However, classroom observations were not always possible due to extremely low attendance in a handful of C5 NFE centres. This is reflected in the fact that only 40 C5 NFE centres are represented in the data by classroom observations.

[17] Indeed, this provides further evidence that poor record-keeping is largely to blame for inconsistent age reports across rounds and sources of information. In many previous studies in Somalia, the evaluation team has encountered such contradictory information regarding birth years and grades of children. Moreover, it is widely acknowledged that few Somalis know their true dates of birth, often needing to estimate their age or their children's ages based on events that occurred near the time of birth. This issue is especially prevalent among those living in conflict-affected areas, where even fewer records of birth or age are kept, and those who have been displaced from their original homes.

TABLE 2: SAMPLE TARGETS AND ACHIEVED SAMPLE FOR SCHOOL-BASED DATA COLLECTION TOOLS

Data Collection Tool	Target per Centre	Centres Targeted	Sample Target	Centres Represented in Data	Achieved Sample
Formal Schools					
Head Teacher Survey	1 per formal school	37	37	35	35
Classroom Observations	2 per formal school	37	74	32	72
Attendance Headcounts	N/A	37	N/A	35	175
C5 NFE Centres					
Classroom Observations	2 per centre	45	90	40	80
Attendance Headcounts	N/A	45	N/A	45	129

The final issue of note regarding the quantitative tools concerns measurement of COVD-related outcomes. The evaluation team was asked to incorporate COVID-related questions into the household survey as part of the inception-stage revisions. However, while the evaluation team included the questions in the survey script, a mistake in the scripting or skip/filter logic meant that no girls or households were asked the relevant questions. In other words, while the questions were present in the survey, all respondents were mistakenly filtered out by earlier responses. It was not until fieldwork was complete that the evaluation team realised this data was systematically missing from every interview. Unfortunately, this means the report is unable to assess differential impacts of COVID across schools, or the relationship between COVID exposure and project outcomes.

Turning to the qualitative tools, the table below documents the types of respondents interviewed and their geographic location. In total, the data includes 69 qualitative interviews. These interviews are heavily concentrated in Banadir and Southwest State, reflecting the larger number of cohort girls and learning centres in those areas. For instance, at ML2, Banadir comprises 45.7 percent of the total sample of cohort girls, while Southwest State includes 32.1 percent of the sample. This distribution of cohort girls was intentionally reflected in the distribution of qualitative interviews. Note that, further, qualitative interviews with mothers, teachers, and girls were distributed across the different cohorts – for instance, six FGDs with mothers were completed with mothers of ABE girls, five were completed with mothers of FE girls, and one was completed with mothers of NFE girls.[18] Qualitative sample targets were met for each of the respondent or interview types, including state-level targets established during the inception and fieldwork planning phase.

[18] The distinction between cohorts does not apply to FGDs with CEC members, as CECs are typically attached to formal schools. While they often oversee non-formal education in their communities as well, there is not a separate CEC for NFE centres and a nearby formal school. The distinction also does not apply to KIIs with religious leaders, who are not attached or affiliated with a specific school or centre.

TABLE 3: QUALITATIVE INTERVIEWS COMPLETED, BY RESPONDENT TYPE AND GEOGRAPHIC AREA

Respondent Type	Banadir	Hirshabelle	Jubaland	Southwest State	Total
CEC Members	5	1	2	5	13
Mothers	5	0	3	4	12
Teachers	6	1	2	3	12
Girls – Vignette Exercise	6	1	1	4	12
Girls – Risk Mapping	5	1	2	4	12
Religious Leaders	3	0	0	5	8
Total	30	4	10	25	69

2.3 Cohort 5 NFE Sampling Design

Cohort 5 NFE girls are the only set of respondents being *newly* selected in this round of the evaluation. For other cohorts, it is unnecessary to discuss sampling methodology, because our approach is focused entirely on maintaining the panel samples established when the cohorts were first C5 NFE girls, on the other hand, are newly sampled and several characteristics of the population of C5 NFE centres and the ML2 evaluation merit elaboration of the sampling approach.

First, C5 NFE centres mostly overlap the C4 NFE centres. At the time of sample selection, 80.6 percent of all C5 NFE Centres in the *sample frame* also had girls enrolled in NFE during Cohort 4 sample selection in 2022. Second, the expansion of the AGES evaluation has made it desirable to minimize further, unnecessary geographic expansion. At baseline, the team completed interviews in 108 distinct schools or learning centres; however, the overlap between schools, on one hand, and ABE and NFE learning centres, on the other hand, meant that the teams visited 58 distinct locations. However, the incorporation of the C4 NFE cohort expanded this site-level sample to 151 schools or learning centres and 90 distinct sites. Randomly sampling centres for the C5 NFE sample would needlessly contribute to further expansion in the number of sites visited and, by extension, the cost of the evaluation.

For this reason, the C5 NFE sample was designed to utilise its natural overlap, in terms of sites, with the C4 NFE sample. At the same time, we did not want to limit the sample only to those centres that overlapped, because this would have required visiting comparatively few C5 NFE centres and imposing a higher cluster size per centre. Visiting *only* overlapping centres would have resulted in a sample of just 32 C5 NFE centres and required a cluster size of 16 girls per centre. The drawback of this approach is that fewer, larger clusters decrease statistical precision (an increased margin of error), which we are keen to avoid. In addition, using a uniform cluster size would require post-fieldwork weighting to ensure that smaller centres – those with fewer girls enrolled – are not overrepresented in the sample.

Instead, we opted for a compromise approach: the C5 NFE sample would include all 32 centres that already appeared in the C4 NFE centre, to reduce the added logistical burden of the new cohort. In addition, we assigned 14 additional clusters to the non-overlapping centres, to bring the total number of clusters to 46, with an average cluster size of 11.3 girls. In a strict statistical sense, the two groups of centres – the 32 that overlap the C4 NFE sample and the 14 additional clusters that do not – constitute two different samples (or strata with a sample). The step-by-step approach we took was:

- First, identify the set of C5 NFE centres already included in the C4 NFE centre sample (32)
- Second, calculate the population distribution of C5 NFE girls across states
- Third, determine the number of C5 girls to be sampled from each state to maintain proportionality
- Fourth, because overlapping centres were of different sizes in C4 and C5 (i.e. a centre with many C4 girls does not necessarily have many C5 girls), we scaled the cluster size for each overlapping centre to its share of the C5 girl population. In other words, an overlapping centre in Southwest State with 100 C5 girls enrolled was assigned half as many interviews as a C5 centre in Southwest State with 200 C5 girls enrolled.
- Fifth, to select the remaining C5 centres, we randomly sampled centres within each state, with probability proportionate to their enrolment size and assigned fixed cluster sizes of 11 girls to each.
- Sixth, within each selected C5 centre, selection of girls was random.

This centre-level selection process produced a total sample draw of 519 girls, spread across 46 clusters (average cluster size of 11.3 girls) and 45 unique C5 NFE Centres. Only one NFE centre was assigned two clusters, in line with PPS sampling, which allows for multiple clusters to be assigned to a single location (i.e., samples with replacement).

It is important to be clear about the implication of this sampling approach: it is not strictly representative in the statistical sense, as all girls in the C5 NFE population do not all share a uniform probability of selection. Consider first the set of overlapping centres within a given state. Each girl in an overlapping centre has an equal probability of selection, because the number of interviews assigned to each overlapping centre was proportional to its population. To illustrate: two NFE centres in Benadir with 84 and 141 C5 NFE girls were assigned 6 and 10 interviews, respectively, resulting in selection probabilities for each girl in these centres of 0.071.[19] This outcome is not coincidental – it is a function of scaling cluster size to each centre's share of the state's C5 NFE girl population.

Now consider the remaining centres – not overlapping with the C4 sample – which we selected. In a centre in Banadir with 90 enrolled girls, the centre has a 0.026 probability of being assigned a cluster, while a centre with 252 girls has a 0.074 probability of being assigned a cluster. By using equal cluster sizes (11 interviews) for this portion of the sample, we achieve equal probabilities of selection: the combined probability of selecting the first centre and selecting a given girl within that centre is 0.00324; the equivalent probability for a girl in the larger centre is 0.07418 * 0.0437, or 0.00324.

These examples make clear that girls within the same set of centres – those that do or do not overlap with the C4 NFE sample – have equal probabilities of selection, which is the typical goal of sample selection. However, across these two sets, probabilities are unequal, resulting from the simple fact that every C5 NFE centre that overlapped with a sampled C4 NFE centre was selected automatically (with probability 1), while those that did not overlap with a sampled C4 NFE centre had probabilities of being selected significantly below 1.

This seemingly complex structure is the result of a compromise between logistical and cost considerations, on one hand, sampling rigor, on the other. Our sample maintains proportionality by state and random selection of girls within each centre; it also maintains random selection of centres for 30.4 percent (14 of 46) of the centres. However, most centres were not selected randomly, producing a sample that is not

[19] This is the selection probability conditional on the centre being selected; in the case of overlapping centres, the probability of a centre being selected is 1 (all overlapping centres are selected), such that a girl's total selection probability is equal to her probability of selection within her centre.

representative in the aggregate but has the most important attributes within the constraints imposed. In general, we do not employ weights when analysing the C5 NFE girls and their outcomes.[20] We do, however, test the robustness of our core findings regarding C5 NFE girls (such as state-level differences in learning scores) to the use of sampling weights, to ensure this methodological decision does not drive our results. Where the results differ, we note them explicitly in the text or in footnotes.[21]

The table below documents the zone-level distribution of C5 NFE girls in the full population (the sample frame of C5 NFE centres, overlapping and non-overlapping); in the selected sample, reflecting our sample targets; and in the end sample that was achieved. The gap between the targeted and achieved state-level distribution reflects the fact that we set a sample target of 519 girls – this oversampling was designed to ensure we reached CARE's target of 510 girls. It also accounts for the fact that clusters must be whole numbers of respondents, and sample target of 510 girls was not evenly divisible by the 46 clusters. In contrast to the sample target, our achieved sample consists of 511 girls.

TABLE 4: ACHIEVED SAMPLE AND POPULATION DISTRIBUTION, ACROSS STATES, OF C5 NFE COHORT

Zone	Population Proportion	Sample Target Proportion	Achieved Sample Proportion
Banadir	47.6%	47.8% (n = 248)	47.2% (n = 241)
Hirshabelle	16.5%	16.4% (n = 85)	16.6% (n = 85)
Southwest State	35.9%	35.8% (n = 186)	36.2% (n = 185)

Overall, cluster size targets ranged from 8 to 23 interviews. The largest clusters simply reflect the large number of enrolled girls in some C5 NFE centres – as many as 256 girls, according to the programme's pre-fieldwork enrolment documentation exercise. In practice, the achieved cluster sizes ranged from just 4 girls to a high of 23 girls. In a small number of schools, we were unable to meet the sample target because of a low number of enrolled C5 NFE students. The divergence between the enrolment numbers captured by CARE pre-fieldwork and our experience when visiting the centres reflects the fact that several weeks (or more) had elapsed since CARE's data collection. If girls dropped out or decided against attending the NFE programme, they were included in CARE's enrolment snapshot, but could not be interviewed by the field team. This is an expected outcome of attempting to capture – as is necessary for a valid baseline assessment – enrolment numbers prior to the true start of the intervention; in short, it is not surprising that the enrolment situation had changed, in a few centres, by the time the evaluation field teams arrived to conduct interviews.

[20] In part, this decision reflects the approach to analysis we have adopted with the evaluation cohorts that preceded them, which focuses on maintaining a panel sample, rather than focusing on precise statistical representativeness. In the endline round, when attrition has impacted the C5 NFE sample, the distribution of girls across centres will change significantly, and it will not be entirely clear how to re-weight the sample to adjust for differential cluster size. In the present round, before attrition is relevant, adjusting for uneven cluster size and/or varying probability of selection is possible, but we do not focus on it because it is unlikely to feature heavily in the endline round.

[21] We calculate sampling weights by treating the overlapping and non-overlapping set of centres as separate strata. For the set of centres overlapping with the C4 NFE sample, the within-strata sample is self-weighting, due to the use of scaled cluster sizes. In other words, every girl in the overlapping sample frame had an equal probability of selection without the use of weights. Within the non-overlapping set of centres, the sample is also self-weighting, owing to the use of probability proportionate to size (PPS) sampling. It is *between* strata that girls have differential probabilities of selection. We correct this by applying a weight – fixed within each strata – that upweights girls from the non-overlapping strata. Ultimately, girls in the two strata take weight in the sample equal to the proportion of the sample their strata comprises (22.7 percent of the sample frame are in non-overlapping centres; 77.3 percent are in overlapping centres).

In the next section, we address re-contact procedures and outcomes that are specific C5 NFE's predecessor cohorts – they do not apply to the C5 NFE cohort, as this round constitutes a baseline for this cohort. However, we refer readers to Section 3 of the report, below, which reports the composition and demographic characteristics of the C5 NFE sample and compares them to the C4 NFE cohort originally selected in the ML1 round.

2.4 Re-Contact Procedures and Re-Contact Outcomes

During the evaluation, the same group of girls who were recruited during the AGES baseline (late 2019) or midline #1 (ML1, in early 2022) were contacted again. As noted above, the evaluation team sought to re-contact every girl from either the BL or ML1 rounds, with the goal of maximizing the available sample for analysis. We discuss the logic of our approach to panel attrition and panel maintenance in more detail in Section 2.1, above. This section outlines the re-contact procedures used to maximize re-contact rates and analyzes patterns in successful re-contact. The approach to re-contacting girls involved multiple strategies, including leveraging the local knowledge of enumerators, seeking assistance from teachers and other community members, and utilizing data collected from girls and their families during baseline.

Re-Contact Procedures

Enumerators were instructed to follow specific re-contact procedures, including asking the head teacher and other teachers at the centre whether the girl was present at the time of visiting, calling every phone number on file for the girl and her family three times, with six-hour intervals between each call attempt, and making at least two separate attempts to contact the girl on two different days. Additionally, enumerators were required to visit the girl's household at least twice, using location information obtained from the baseline data, teachers, and other students in the school who knew the girl or her family. They were also instructed to ask the head teacher, teachers, other students, and community leaders for contact information for the girl or her family.

To maximize successful re-contact rates, a decision was made about whether teams would travel to find girls who had moved away from the area, either temporarily or permanently. The approach was designed to balance logistical considerations while maximizing the chances of successfully re-contacting girls. If the girl lived within approximately 30 minutes of the school, an enumerator would travel to her home to interview her. If the girl lived outside that range, the team leader would consider whether the girl could be visited later in the fieldwork when the team was closer to her home. Teams typically moved around the same cities for several days or weeks and could therefore arrange to visit a girl at home later.

In cases where a girl was impossible to contact, refused to participate, lived outside the fieldwork area, or was not reachable safely, the team leaders decided whether to drop her from the sample. For girls who were enrolled in a formal school at baseline, replacement girls were selected randomly by the evaluation team from the original sample frame provided by CARE in 2019, to ensure unbiased sampling. However, ABE, C1 NFE, and C4 NFE girls who could not be located were not replaced as they had typically completed their participation in the educational centre by the time of the midline evaluation. Hence, it was not feasible to select a comparable replacement girl from within the centre.

To ensure a larger sample size for the analysis of transition outcomes, the teams made efforts to record as much information as possible about girls who dropped out of the sample, whether they were replaced or not. Enumerators gathered details such as the girl's current location, enrollment status, and grade level, if possible. As panel attrition is often a significant issue for inferences around transition outcomes, recording such

information was critical as it helped to analyze the transition rates from baseline to midline. Moreover, this information also helped to verify the accuracy and reliability of the primary findings.

Re-Contact Rates

We will centre our discussion on individual-level attrition rather than attrition at a broader level caused by security concerns. To establish a baseline sample, we will begin with the 417 FE girls that we gathered data from in ML1. Additionally, any FE girls that we were unable to find at ML1 but had interviewed at BL were also included in this analysis. This excludes girls from Dinsoor, since they were replaced from the panel at ML1, and it was agreed with CARE's technical team that re-contacting them would be based on security conditions at endline. We are excluding C5 NFE girls because they were not surveyed in previous years, the ML2 survey round is their baseline. As such, our aim at ML2 was to re-establish contact with all FE, ABE, and C1 NFE girls from ML1 as well as the FE, ABE, and C1 NFE girls from BL who we were unable to reach at ML1, and to substitute any who couldn't be located with new girls.

The table below reports re-contact rates from ML1 to ML2 by region in the third column, where Jubaland outperformed relative to Southwest State and even more so compared to Banadir. The fifth column also displays what the region-level re-contact rate was in that region between BL and ML1. In each region the re-contact rate fell meaningfully. But this is to be expected, as in ML2 our survey teams attempted to find girls who we had already tried and failed to find and follow up with in ML1. These girls may have moved between BL and ML1 and so it was unlikely that they would be re-located in ML2. This biased re-contact rates downwards as a result. Column six shows that if we include only those girls who we had successfully followed up with in ML1 then the rate improves meaningfully. It also displays how, although the relative regional rankings (with Jubaland at the top and Banadir at the bottom) hold when looking at re-contact this way, the differences become meaningfully less stark. Jubaland and Banadir were 8.5 percentage points apart before but only 5.5 percentage points apart now. This may be because some regions had a higher proportion of girls who were surveyed in BL, missed in ML1, and then were attempted to be contacted again in ML2.

TABLE 5: RE-CONTACT RATES OF ML1 COHORT GIRLS, IN ML2, BY REGION

Region	Total Possible Sample	Re-contact Rate	Achieved Panel Sample	Re-contact Rate among BL Girls	Re-contact Rate among BL-ML1 Girls
Banadir	685	65.6%	448	78.6%	73.2%
Jubaland	357	74.1%	265	90.2%	78.7%
Southwest State	379	70.6%	266	87.7%	75.8%

We posit that the primary drivers of unsuccessful re-contact are likely to be unenrolling from the school and migration. This is because the school is our survey team's primary avenue for locating the girls. If they have migrated to another area, then they are considered unreachable unless that place is in another community that the survey teams are going to. As such, traits which affect the likelihood that they unenroll or migrate are the primary suspects for failed re-contacts.

Each treatment arm targets girls with specific age profiles while also running for different amounts of time, and these factors may also play into the likelihood of dropping out or migration. To date the FE programme has run for three of its four years and included 10-13 year old girls at baseline, ABE ran for two years before concluding in 2022 and targeted 13-16 year old girls, and the NFE treatment lasted only 11 months for girls

aged 17-20. This final group, the Cohort 1 NFE (C1 NFE) girls, completed their programme in late 2020. There may be varying re-contact rates between programme groups because those on the younger end of the spectrum, FE, may be less likely to get married and therefore move or drop out of school. And whether the programme is still running may affect their willingness to stay in the area as well – the girl and her family may think that because she is gaining valuable skills through the AGES programme there is some benefit to remaining in the area longer.

Our analysis in the table below shows that re-contact rates are higher for those who are still enrolled in or only recently completed the program, FE and C4 NFE girls (73.3% successfully interviewed), relative to the ABE and C1 NFE girls who have already graduated (69.6%). The fact that the 18-20 year old C4 NFE girls were located at a higher rate than their age-group peers (ABE & C1 NFE) but nearly as frequently as the now-13-17 year old FE girls suggests that age is not the driving factor for failed re-contacts. Instead, it indicates that the recency of programme engagement is more likely.

TABLE 6: RE-CONTACT RATES BY COHORT TYPE AND REGION

State	Total Possible Sample	Re-Contact rate	Achieved Panel Sample
Banadir			
FE Girls	217	71.4	155
ABE Girls	217	64.5	140
C1 NFE Girls	251	61.0	153
C4 NFE Girls	440	70.0	308
Jubaland			
FE Girls	142	75.4	107
ABE Girls	108	75.9	82
C1 NFE Girls	107	71.0	76
Southwest State			
FE Girls	106	76.4	81
ABE Girls	142	68.3	97
C1 NFE Girls	131	67.2	88
C4 NFE Girls	334	74.9	250
Hirshabelle			
C4 NFE Girls	139	78.4	109
Aggregate			
FE Girls	465	73.7	343
ABE Girls	467	68.3	319
C1 NFE Girls	489	64.8	317
C4 NFE Girls	913	73.1	667
Total	2,334	70.5	1,646

We also examined whether those who were and were not successfully re-contacted differed on any other observable characteristics beyond programme type. For this we specifically looked at traits that may influence the likelihood that a household migrates and becomes unreachable as a result – see table below. To do so we conducted balance tests between those who were and were not successfully re-contacted among the FE, ABE,

and C1 NFE girls, excluding C4 NFE because they were not asked all the questions as the others. We investigated whether the groups were statistically different on several measures of household poverty, proxy indicators for potential within-community marginalization, disabilities, and land ownership.

These variables drive migration through a myriad of different mechanisms. Household poverty can hinder migration by making travel financially infeasible. IDP families may decide to leave their host communities because of social or economic marginalization. Having children with mental disabilities may cause families to be socially ostracized as well, and physical disabilities directly exacerbate the difficulty of non-vehicular movement. Land ownership is also likely to decrease migration; without a well-established land ownership system, informal and undocumented claims dominate and families who wish to keep their land are effectively rooted to the area.

The table below compares the set of girls who fell out of the sample (first two columns) to those who were successfully re-contacted (columns 3-4). For each characteristic, we compare the two groups of girls to assess whether re-contacted girls are systematically different from those who we failed to re-contact. The comparison is between the 442 girls for whom re-contact efforts were not successful and the 979 who were successfully re-contacted. The right-most column calculates the difference in prevalence of a given characteristic (e.g., the share of households with a poor-quality roof) between the two groups of girls, denoting those differences that are statistically significant using asterisks. To illustrate, the share of girls whose family went without food for many/most nights in the last 12 months is higher among the failure group than the success group – 23 per cent of girls in the former group went without food, compared to 21 per cent of girls in the re-contact success group.

Surprisingly, these balance tests show that those who are successfully re-contacted are overrepresented on each of the marginalization measures (speaking af-Maay, IDP status, and having a mental health disability) relative to those who could not be found. The difference is particularly pronounced for IDP's, with 51% of those re-contacted declaring themselves as being internally displaced but only 24% of those who could not be found. However, because IDP status is a variable that we take from the most recent survey round and then apply backwards, it is possible that this difference is due to IDP's becoming more prevalent since the last round. IDP status does appear to be rising, but its relationship with higher re-contact rates has been constant – in the ML1 survey round 39% of successfully re-contacted girls were IDPs, compared to 23% of those who could not be located. Because IDP status was still associated with higher re-contact rates in ML1, we feel confident that our backwards application of IDP status is not meaningfully unbalancing these results. Lastly, land ownership and self-reported disability status were also overrepresented among those who were re-contacted.

TABLE 7: BALANCE TESTS OF PREDICTORS OF SUCCESSFUL RE-CONTACT AMONG BASELINE COHORTS

Characteristic	Re-Contact Failure Post-BL (n = 442) Pct. in Subgroup	Successful Re-Contact Post-BL (n = 979) Pct. in Subgroup	Difference Between Groups (Pct. points)
HH has poor roof (mud, thatch, cardboard, plastic)	19.0%	18.8%	0.2 points
Went without food for home use most/all days, last 12 months	8.2%	7.7%	0.5 points

Went without food for home use many/most/all days, last 12 months	23.4%	20.6%	2.8 points
Went without water for home use most/all days, last 12 months	5.2%	5.7%	-0.5 points
Went without medicines most/all days, last 12 months	20.5%	16.1%	3.3 points
Went without cash income most/all days, last 12 months	16.3%	15.0%	1.3 points
Family/girl speaks af-Maay at home	21.7%	29.4%	-7.7 points***
HH owns land, either solely or jointly	29.6%	34.9%	-5.3 points**
Family is IDP	24.0%	51.4%	-27.4 points***
Physical Disability	2.7%	6.3%	-3.6 points***
Mental Health Disability	11.8%	18.7%	-6.9 points***

The findings from the previous sections do not indicate any clear connection between age and successful re-contact but cohort and region do show a stronger predictive relationship with successful re-contact. Nevertheless, since many relevant characteristics, like age and household economic status, are interrelated, it is challenging to determine their individual associations with re-contact status using simple models like difference-of-means or cross-tabulations. To gain a more accurate understanding of the factors influencing successful re-contact, we employed regression models that consider a girl's characteristics from their most recent survey round.

Our approach involves predicting the binary outcome of re-contact by considering cohort and region as a series of dummy variables within the model. Additionally, we include measures of household food insecurity, the girl's age at their most recent survey round prior to ML2, IDP and land ownership statuses, as well as both mental and physical health disabilities. These explanatory variables were included based off whether there were statistically significant differences in the balance tests between girls who could not be re-contacted and those who were. We also constructed an indicator variable for whether the girl was surveyed at BL, could not be located at ML1, and contact was re-attempted at ML2. As noted earlier, because our enumeration teams attempt to re-survey all the girls from both BL and ML1, any who had moved between BL and ML1 are much less likely to be in ML2.

Since the schools are the main focal points through which our enumerators find the girls from previous survey rounds, we also aimed to include in our model measures for factors that would drive girls to stop attending school and therefore become more difficult to locate. We constructed two indexes to measure this. The first, the Caregiver's Perceptions Index, measured the degree to which caregivers may or may not be supporting the girl to go to school. The second, the Teacher Quality Index, assessed how many negative teaching habits the girls' teacher exhibited, potentially diminishing her willingness to continue attending. The components

of these indexes are listed below, where one additional point was attributed to the girl for each question that was answered in the affirmative:

Caregiver Perceptions Index

- The girl spends a few hours or less per day doing household chores
- The caregiver aspires for the girl to attain a university education
- Girls' education is a worthwhile investment, even if funds are limited
- Work or household chores are not an acceptable reason for the girl to not attend school
- A high cost of education is not an acceptable reason for girls' failing to attend school

Teacher Quality Index

- Often absent
- Rarely/never encourages participation
- Rarely/never explains how things they're learning are useful in the student's lives
- Lessons move too fast
- Boys & girls treated differently in the classroom
- Students are punished for getting things wrong in a lesson

The figure below reports the regression coefficients (orange dots) and the 95 percent confidence interval (orange horizontal bars) around those coefficient estimates for a series of predictors. This figure is drawn from a single regression, meaning that all of the variables reported were included in a single model. The regression coefficients can be interpreted as the impact of a variable, measured in percentage points, on successful re-contact. For instance, the family being IDP's is associated with a 19-point increase in the likelihood that a girl will be re-contacted successfully, holding all else equal.

The regression results reinforce several of the findings from the less rigorous analysis in the previous sections. First, those who are linguistic minorities or from other areas (af-Maay speakers and IDPs) are 9.5 and 19 percentage points higher, respectively, after accounting for other factors that predict successful re-contact. And households for whom moving is more costly or more difficult, those who own land and those with a daughter with a physical disability, are 5.7 and 18.4 percentage points more likely to be re-contacted. This is all after controlling for whether they were last contacted in the 2019 BL survey round – those girls were 24.6 percentage points less likely to be successfully interviewed in ML2. Lastly, households which reported experiencing hunger for 10 or more of the last 30 days were 21.5 percentage points more likely to be found in ML2. This is in line with the theory described earlier, where poverty reduces a household's financial ability to migrate.

FIGURE 1: IMPACT OF GIRL CHARACTERISTICS ON RE-CONTACT SUCCESS

Another positive finding is that, after adjusting for all other explanatory variables, the programme arm, girl's age, and region were not statistically significant. This rules out that the programmes themselves or their beneficiary-targeting strategies were driving failed re-contacts. And the two indices, measuring household perceptions and teacher quality, were not statistically significant as well. To put it another way, households who underprioritise girls' education and teachers who are not particularly proficient educators are not associated with failed re-contacts. That is reassuring, as we had speculated before constructing the indices that each of these could be driving student dropout and therefore reduce the likelihood that the girls could be successfully re-located.

Based on our analysis, there are no significant indications of any substantial concerns regarding attrition patterns that could introduce bias or significantly change the composition of the sample. The only variable that showed negative impact on re-contact rates was whether the girl was last contacted at BL, and that is almost certainly caused by our survey teams attempting to locate girls who have not been surveyed since 2019.

2.5 Challenges and Limitations

Over the previous two evaluation rounds, the evaluation team has highlighted and discussed several limitations related to the evaluation design. In general, issues related to the overall research design remain a factor in our approach to the analysis and apply to the C5 NFE cohort that was newly recruited in this round. We refer readers to the baseline evaluation report for a discussion of issues related to benchmarking and the lack of a comparison group, to concerns related to the comparability of learning assessments over time, ceiling and floor effects, and panel attrition.

It is important to emphasise the efforts the evaluation team and CARE's technical staff have made to mitigate many of the concerns raised at baseline. Specifically, the comparability of learning assessments over time has been largely maintained through careful review and revision of the assessments in each round, while taking specific care to avoid changing the difficulty of individual test items. Ceiling and floor effects were addressed beginning in the ML1 round; while it would have been preferable to incorporate more difficult numeracy subtasks into the baseline study to allow for over-time comparability with the more difficult assessment, this is a relatively small limitation across four rounds. Panel attrition, to the extent possible, has been reduced through careful re-contact procedures and by proactively re-contacting girls who had fallen out of the sample in previous rounds, to preserve the panel's integrity over time.

Leaving aside pre-existing limitations, the primary new methodological limitation relevant to this round concerns tracking of COVID-19 impacts. During the ML1 round, a miscommunication between the evaluation team and CARE's technical staff resulted in a failure to update the household survey tool with COVID-specific questions. During the current round, the evaluation team made an error in scripting the household survey, which effectively filtered all respondents out of the COVID-19 questions. As a result, we are unable to comment on or document the extent of COVID-19 exposure or impacts within AGES communities and learning centres.

3. Sample Characteristics

3.1 Cohort 5 NFE Sample Demographics

In this evaluation round, a new cohort of NFE girls (cohort 5 or "C5") was recruited and baseline estimates of learning and other key outcomes produced. The C5 girls are predominantly recruited from the same learning centres as the cohort 4 NFE girls ("C4") and are expected to be broadly similar. In subsequent sections, we compare baseline learning and other outcomes of C5 girls produced this round of the reporting to the baseline estimates of C4 girls produced in the previous (ML1) round of reporting. In this section, we discuss and compare the demographic characteristics of the two cohorts to identify any substantive differences which may be expected to result in different learning outcomes.

Cohort 5 is comprised of 511 girls with a mean age of 19.2 years. Almost half (47.2%) of the girls sampled in Cohort 5 are in Banadir, with the remainder distributed across Lower Shabelle, Bay, and Middle Shabelle. The age and geographic distribution of C5 girls is like those of C4 girls at ML1.

TABLE 8: SAMPLE DEMOGRAPHICS OF THE COHORT 4 AND COHORT 5 NFE GIRLS

Note: a complete disaggregation of the sample by disability type is provided in Annex 3.

Sample Characteristics	Cohort 4 NFE Girls (ML1)	Cohort 5 NFE Girls (ML2)
Family Structure and Characteristics		
Girl has only one living parent	18.9	20.1
Girl has no living parents	1.1	0.0
Girl does not live with either parent in her HH	10.7	8.6

Female-headed household	68.6	71.2
HoH has no education of any kind (no Quranic)	16.8	18.7
HoH has no formal education (may have Quranic)	84.6	74.8
Caregiver has no education of any kind (no Quranic)	15.4	20.9
Caregiver has no formal education (may have Quranic)	84.3	67.6
Household Wealth and Socio-Economics		
HH has poor roof (mud/thatch/cardboard/plastic)	21.9	19.8
Went to sleep hungry most/all nights, last 12 months	6.7	11.5
Went without water for home use most/all days, last 12 months	8.1	17.0
HH owns land, either solely or jointly	23.0	21.3
Household owns a phone	28.2	24.9
Household owns a smartphone	3.3	6.7
HoH does not have an occupation or does not earn a wage	14.0	11.2
Household is engaged in pastoralism	0.3	1.0
Disability Status[22]		
Girl has physical disability	3.2	6.1
Girl has physical disability, alternative coding	3.4	7.2
Girl has cognitive, behavioral, or communicative disability	4.4	7.8
Girl has cognitive, behavioral, or communicative disability, alternative coding	5.6	9.6
Girl has mental health disability	15.5	36.2
Girl has mental health disability, alternative coding	17.0	42.9
Displacement and Language		
IDP Household	54.4	37.6
Household speaks af-Maay	33.0	29.5

Family Structure and Characteristics

Generally, the structure and characteristics of families are similar between C4 NFE and C5 NFE girls, as shown in the table above. The most substantive differences between cohorts are the level of education among caregivers and heads of household. At ML1, we found 15.4 per cent of C4 NFE girls' caregivers and 16.8 per

[22] Reporting by difficulty/disability type in this table uses an aggregated approach that combines, for instance, all physical disabilities into a single category. For more fine-grained disaggregation that mirrors the Washington Group definitions, please see Annex 3, Table 107.

cent of heads of household had no education of any kind – including either formal and Quranic. In comparison, these figures are higher for C5 NFE girls. As many as 20.9 per cent C5 girls (more than five points higher than C4 girls) have caregivers with no education of any kind. When we consider all formal education – from *some* primary school to higher education – very few caregivers in either cohort have any exposure. However, there is a slight difference with just 2.3 per cent of C4 NFE caregivers and 5.9 per cent of C5 NFE caregivers having any formal education.[23] There is a similar, though less pronounced difference at the head of household level, with 2.3 per cent of C4 NFE and 4.1 per cent of C5 NFE heads of household exposed to any formal education.[24] Other family structures and characteristics are comparability between cohorts.

Household Wealth and Socio-Economics

Many of the household wealth and socioeconomic characteristics are similar between the two cohorts. Household structure (as indicated by roof type) and household ownership of land are comparable. Household phone ownership is slightly lower among C5 girls (24.9 per cent compared to 28.2 per cent among C4 girls) though smartphone ownership is slightly higher (6.7 per cent compared to 3.3 per cent among C4 girls). Similarly, the share of girls living in households where the HoH does not have an occupation or does not earn a wage is comparable: 11.2 per cent among C5 girls and 14.0 per cent among C4. Given the similarities in assets and employment, it is somewhat surprising, to find meaningful differences between the cohorts in the share of girls who go to sleep hungry (5 points higher among C5 girls than C4 girls) and those who went without water all or most days over the past year (9 points higher for C5 girls than C4 girls).

This may be, at least in part, explained by extraneous factors such as inflation, rising food costs, and/or ongoing drought. In other words, even if employment levels remain consistent, access to food and water may decrease due to economic and environmental conditions. It is also worth noting that other income and socio-economic indicators reveal more substantive differences between the two cohorts. At ML1, 20.1 per cent of C4 NFE girls said that their household went without cash many or most days. At ML2, 34.6 per cent of C5 NFE girls say the same. The hypothesis that this is related to changing conditions, rather than fundamental differences between cohorts, is supported by the changes between ML1 and ML2 amongst C4 NFE girls; there is an 11-point increase – from 20.1 per cent to 31.1 per cent - in the share of C4 girls reporting going without cash many or most between ML1 and ML2. Given the two cohorts look very similar at ML2, we may conclude that they appear to be of the same relative socio-economic status when considered contemporaneously. However, in considering the differences between "Baseline" learning scores (C4 at ML1 versus C5 at ML2), the C5 NFE girls appear worse off in terms of their absolute economic position.

Disability Status

We see consistently higher prevalence of disabilities among C5 NFE girls than C4 NFE girls. Given that the project explicitly sought to enrol more GWDs in this cohort, this trend toward a higher disability prevalence is not surprising. For example, the share of girls with some form of physical disability (vision, hearing, self-care, etc.) has nearly doubled from the C4 cohort to the C5 cohort, from 3.2 to 6.1 per cent, and the share of girls with mental health disabilities is also much higher (36.2 per cent compared to 15.5 per cent) among the C5 NFE cohort.[25]

[23] The difference is statistically significant ($p < 0.001$).

[24] The difference is statistically significant ($p < 0.05$)

[25] The table above documents disability status that has been aggregated to represent difficulties/disabilities that are physical, cognitive/communicative, and mental health-related, respectively. Table 112 in Annex 3 provides a more disaggregated view of the differences between C4 and C5 NFE girls, including all standard types of child functioning, in line with the Washington Group standards. The general trend is still present, with higher rates of every disability – physical, cognitive, behavioural, mental health, etc. – among the C5 NFE cohort, when compared to their peers in the C4 NFE cohort.

At the same time, we again note that our analysis here is focused on both cohorts of girls at their "baseline" round of data collection; for C4 NFE girls, baseline was the prior ML1 round and for C5 NFE girls, this current ML2 iteration serves as the baseline. This is important because we are comparing two different rounds of data collection. In the above discussion, related to economic status, we discussed the potential for overall changes in socio-economic conditions affecting both cohorts of girls similarly. The data appeared to support this. A similar explanation is less intuitive here, as we do not anticipate a far-reaching environmental shift which would result in substantial changes in (both physical and mental) disability prevalence among girls between years. In spite of this, when we look more closely at disability prevalence among the panel of C4 NFE girls contacted at both ML1 and ML2 (n=661), the share of girls reporting any physical disability (excluding mental health) increases from 7.4 per cent to 15.1 per cent.[26] Further, when we compare C5 NFE girls with the ML2 data for C4 NFE girls, the disability prevalence is similar, with 12.1 per cent of C5 and 15.1 per cent of C4 reporting a disability (excluding mental health). Overall, this suggests that a difference in survey administration may explain part or all of the differences in disability prevalence between cohorts 4 and 5, with respondents more likely to self-report moderate/severe difficulties. This is a more likely explanation than systematic differences in characteristics between the cohorts in terms of disability prevalence, though we cannot eliminate this possibility with the available data.

[26] We exclude mental health from this figure as it is somewhat more susceptible to short term variation and, among adolescents, may be expected to increase and/or fluctuate. However, it is also worth noting that the share of C4 NFE girls reporting frequent anxiety / depression increases from 16.9% at ML1 to 42.1% at ML2.

FCDO Cohort Results

- Formal Education (FE) Girls
- Accelerated Basic Education (ABE) Girls
- Cohort 1 Non-Formal Education (C1 NFE) Girls

4. Learning

The AGES programme focuses on two primary outcomes. The first is girls' numeracy and literacy, specifically, girls' command over reading in Somali; the second is "transition": the advancement of girls enrolled in schools from one grade to the next, the transition of girls not enrolled in schools into formal education institutions or alternative learning institutions, and the transition of older girls into gainful employment. In this section, we analyse changes in learning outcomes – Somali literacy and numeracy – among the original cohorts of girls recruited into the programme during the baseline in late 2019.

In the subsections that follow, we first describe aggregate patterns of change across the three cohort types – FE, ABE, and Cohort 1 (C1) NFE girls. Next, we study changes within each of these cohorts relative to benchmarks for year-on-year learning improvements constructed at the baseline. We also analyse subtask-specific or skill-specific changes in learning scores, seeking to identify which components of a broad "reading ability," for instance, the programme was most able to impact. The final two aspects of the analysis consider subgroup-specific programme impacts and test the Theory of Change by assessing the correlation between the programme's intermediate outcomes and changes in learning scores between BL and ML2. Note that, for several of these sections, we analyse the three cohorts – FE, ABE, and C1 NFE girls – entirely separately, as we expect outcomes to be highly cohort-specific at times.

Learning outcomes are measured using a Somali-language Early Grade Reading Assessment (EGRA) and an Early Grade Maths Assessment (EGMA) which is also implemented in Somali. During the baseline, the evaluation team undertook extensive testing of the instrument, including whether individual test items were able to discriminate between low- and high-achieving learners. Since the baseline, only cosmetic changes have been made to the assessments, with the aim of ensuring girls did not remember, e.g., the stories in the literacy assessment. Both the evaluation team and CARE's technical staff have been careful not to make any changes to the assessments that could impact their overall difficulty. We have no reason to believe changes in the tests across rounds can explain any of the findings in this report.

An exception concerns the numeracy assessment, which was supplemented during ML1 by the addition of three more difficult EGMA subtasks. The goal of this addition was to guard against ceiling effects in later evaluation rounds; this change brought the number of numeracy subtasks from 8 to 11. The change was described previously in the ML1 evaluation report. In this section of the report, we note that we continue to use, primarily, the 8-subtask numeracy assessment, as our focus in this report is to make comparisons between baseline and ML2, and doing so requires use of the same assessment across rounds.[27]

Two other methodological points are important to note. First, our analysis uses a panel sample of girls tracked over time. In almost all cases, reported results use the panel of girls tracked between BL and ML2; this can include girls who fell out of the sample temporarily in ML1 but were brought back into the sample in this round. No replacement girls are included in our analysis in this section. The BL-to-ML2 panel sample includes 270 FE girls, 310 ABE girls, and 302 C1 NFE girls.

Second, as articulated in the baseline methodology, our analysis of changes in learning outcomes over time uses benchmarks obtained during baseline from a separate sample of 454 girls, drawn separately from the same schools which were surveyed at baseline. The purpose of the benchmark is to help account for

[27] However, in our analysis of C4 and C5 NFE girls, beginning in Section 8, we exclusively use the 11-subtask "full" version of the numeracy assessment. These cohorts were recruited into the programme and sample during ML1 and ML2, respectively; thus, their baseline assessments during ML1 and ML2 utilised the full 11-subtask numeracy assessment.

maturation effects – the fact that girls on average are likely to develop their skills as they become older, not necessarily due to their education or any specific intervention – and to attempt to separate the programme's effects on girls' learning from learning that would have occurred regardless. This is an imperfect substitute for a comparison group in an experimental or quasi-experimental setting – a point which we note, at times, when explaining occasional ambiguous findings in this section – but benchmarks provide additional robustness above simple, non-benchmarked, pre-post comparisons. Note that benchmarks are only used in the analysis of FE and ABE girls, whereas for C1 NFE girls we utilise a simple pre-post calculation.

4.1 Learning Benchmarks

Before turning to analysis of learning scores, the table below provides the benchmarks established for FE girls, broken down by state. In the top panel, we list the benchmarks for Somali literacy, as a function of a girl's starting grade – FE girls were enrolled into either Grade 1 or Grade 2 at baseline, and their expected progression is conditional on the level at which they started, because learning gains vary across grade levels. As can be seen in the table, benchmark expectations for Grade 1 girls are systematically higher than those for Grade 2 girls, because Grade 2 girls are starting from a higher baseline level of learning. Gains for Grade 2 girls are slightly more difficult to achieve, as a result.

TABLE 9: BENCHMARK EXPECTED GAINS IN LEARNING SCORES FOR THE FE COHORT

	Benchmark – 2 Grade Progression		Benchmark – 3 Grade Progression	
	Grade 1 at BL	Grade 2 at BL	Grade 1 at BL	Grade 2 at BL
Somali Literacy				
Aggregate	38.3	19.0	46.3	29.4
Banadir	48.9	21.2	57.5	31.7
Jubaland	31.9	17.2	43.4	27.6
Southwest	31.7	14.0	35.9	23.8
Numeracy				
Aggregate	25.1	10.1	28.6	15.6
Banadir	31.2	10.7	36.0	17.0
Jubaland	27.7	10.1	29.9	15.0
Southwest	15.9	5.8	19.1	9.0

For each grade level, we have produced two benchmark scores. The first captures expected improvements over the course of three years of consistent advancement, e.g., from Grade 2 to 5. The second captures improvements over two years of advancement. The issue of defining a reasonable expectation for FE girls' progress from late 2019 to present is a difficult one, and it impacts our treatment of transition outcomes for this cohort as well. Since the baseline, just over 3 years have passed, suggesting that girls – if they were enrolled consistently and progressed in school year-on-year – would have advanced three grade levels in school. However, 3 consecutive years of advancement is a relatively high bar in this context, compounded by the school disruptions caused by COVID-19 during this period. Therefore, we also use the less ambitious two-level benchmark (which captures expected gains from, for instance, Grade 2 to Grade 4) at times.

The second panel reports equivalent benchmarks for numeracy. Note that we calculate state-specific benchmarks because – for a wide variety of reasons, including differential school quality – the expected gains from an additional year of schooling seem to vary across states. Of course, we should be cautious about over-interpreting differences in benchmarks across states, because state-specific benchmarks were calculated from a limited sample size of girls. In any case, employing the state-specific benchmark in lieu of applying the average benchmark to all girls simply ensures that we weight our cohort-to-benchmark comparisons by the share of girls in each state, which is an important statistical adjustment in this case.

We also note the broad representation within the benchmark sample. During the baseline, benchmark girls were recruited from nearly every school included in the FE sample, resulting in a benchmark sample with wide geographic coverage that broadly mirrors the composition of the FE sample itself. Indeed, there are just two FE schools in our sample that are not also included in our benchmark sample; there are several schools in the benchmark sample that are not included in the FE sample due to shifts in the FE school-level sample post-baseline. The key point in this discussion, however, is that the benchmarks are not derived from a narrow set of schools, but represent the same communities and schools as the FE sample, on average.

The table below documents benchmarks established for the ABE cohort. These benchmarks are simpler insofar as ABE girls were all enrolled in the same level or stream of ABE programming at BL, so there is no variation in girls' starting points.[28] In addition, the ABE benchmark is calculated as equivalent to the gains expected between Grade 1 and Grade 2 of formal schooling, which we assess using the sample of benchmark girls who were enrolled in formal schooling at the baseline. As with the FE cohort, we calculate state-specific benchmarks to ensure that our calculations of the mean benchmark mirrors our calculation of the mean gain in learning scores among the cohort, in the sense that it is adjusted for the state-by-state distribution of ABE girls in the sample.

TABLE 10: BENCHMARK EXPECTED GAINS IN LEARNING SCORES FOR THE ABE COHORT

	Somali Literacy	Numeracy
Aggregate	22.2	15.8
Banadir	27.9	22.0
Jubaland	27.8	17.8
South West	9.1	5.1

4.2 Aggregate Trends

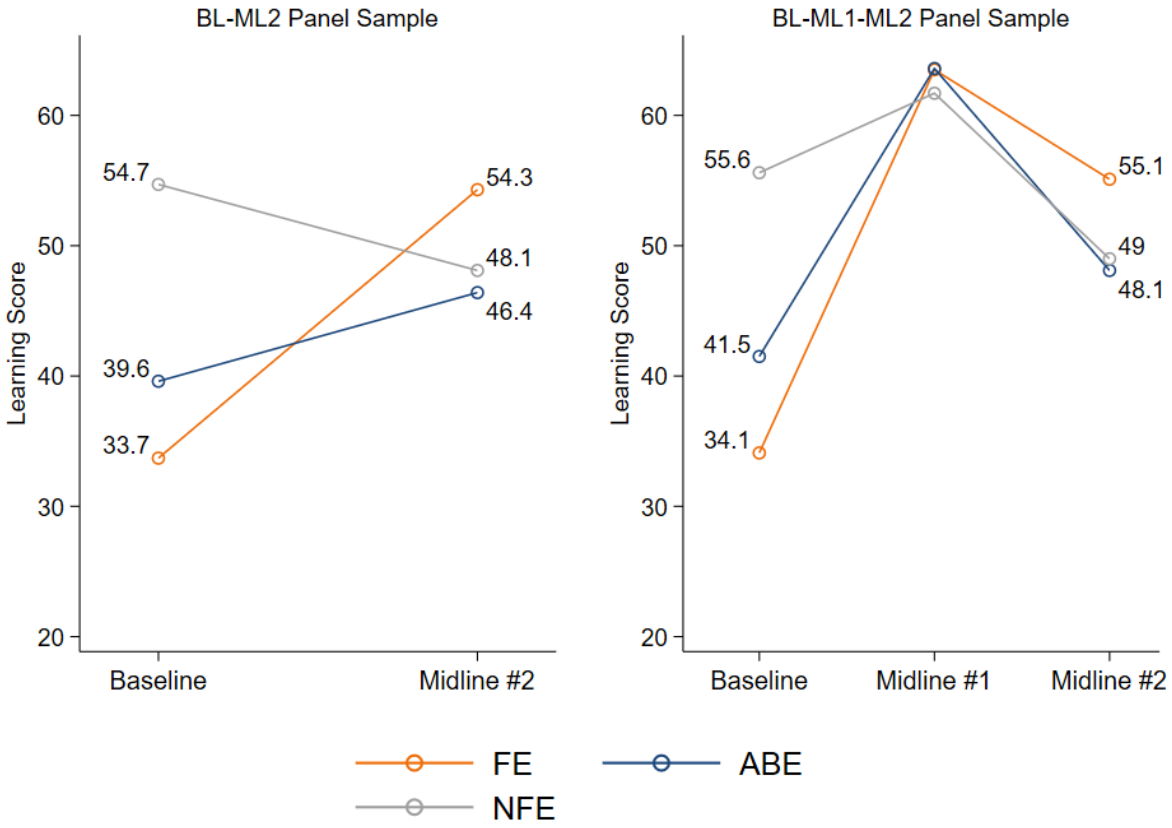
To illustrate the evolution of learning outcomes over time across all three cohorts, the figure below plots average Somali literacy scores across rounds. The left panel of the figure utilises the BL-to-ML2 panel sample, which includes all girls who were successfully interviewed at both BL and ML2. The right panel incorporates data from the ML1 round to provide a slightly more nuanced picture of trends in literacy over time. The right panel uses a slightly different sample – the set of girls who were successfully interviewed across all three rounds, which reduces the per-round sample slightly.[29]

[28] Or, rather, there is no identifiable difference on the basis of grade level assigned.

[29] For instance, the sample of FE girls in the left panel (BL-to-ML2 panel) is 270, while there are 251 girls in the right panel (BL-ML1-ML2 panel sample).

The figure highlights several important trends. First, FE and ABE girls have experienced learning gains since the baseline, though this outcome is far more pronounced among the FE girl cohort. FE girls' outsized gains are not entirely surprising, given that many of them have remained enrolled in school since 2019, whereas the ABE programmes are intentionally designed to be shorter. On the other hand, C1 NFE girls' literacy rates have regressed since baseline, in the aggregate.

FIGURE 2: SOMALI LITERACY SCORES OVER TIME AMONG THE ORIGINAL BASELINE COHORTS (LEFT PANEL = BL-ML2 PANEL; RIGHT PANEL = FULL PANEL ACROSS ROUNDS)

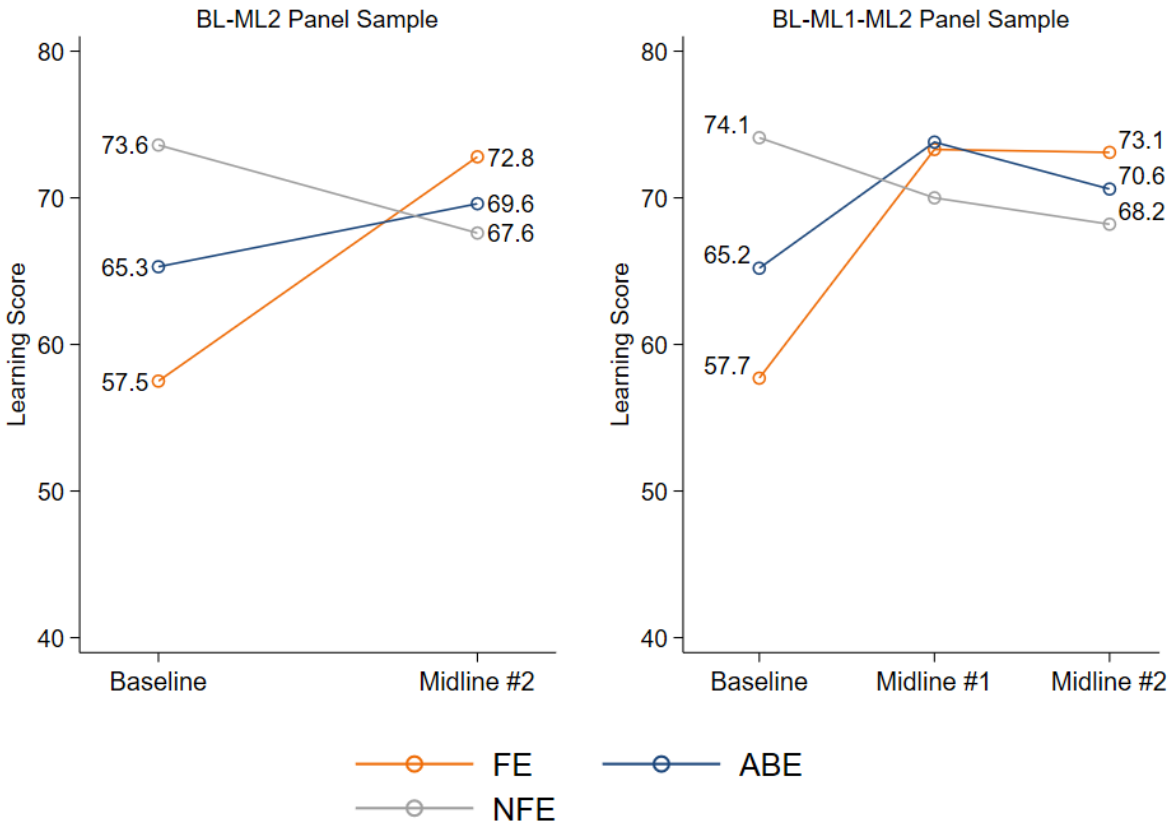


Second, there is a clear pattern, shown in the right panel, of regression toward baseline levels of learning. Put differently, there is literacy backsliding across all three cohorts from ML1 to ML2 – for instance, FE girls' Somali literacy scores – in this specific sample – increased from 34.1 percent to 63.5 percent from BL to ML1 but have since declined to 55.1 percent at ML1. We grapple with this trend further in our specific analysis of FE girls, in which we find an unsurprising gap in the learning gains made by FE girls who remained enrolled and advanced year-on-year and those who did not.

The figure below repeats the analysis for numeracy outcomes, with similar but less stark results. In the left panel, it is clear that there are sizable improvements in numeracy among FE girls and small, but tangible, improvements among the ABE cohort, while the C1 NFE cohort has regressed over time. However, in the right panel, we can see that the gains in the former two cohorts are entirely concentrated in the BL-to-ML1 period and that – since ML1 was conducted in early 2022 – neither ABE or FE girls have experienced further improvements in numeracy. Again, ABE backsliding in numeracy is not entirely unexpected, given that many

ABE girls are no longer in any learning programme; the plateau shown among FE girls is somewhat more concerning. Note that the plateau among FE girls cannot be explained as a function of ceiling effects, because the same stagnation from ML1 to ML2 can be observed when we employ the full 11-subtask numeracy assessment that was implemented beginning to ML1 to avoid such issues.

FIGURE 3: NUMERACY SCORES OVER TIME AMONG THE ORIGINAL BASELINE COHORTS (LEFT PANEL = BL-ML2 PANEL; RIGHT PANEL = FULL PANEL ACROSS ROUNDS)



Formal Education Girls

Building on the overall findings above, we now seek to understand changes in learning outcomes in greater detail within each cohort. We start with formal school (FE) girls, reporting the evolution of literacy and numeracy scores over time. The table below shows the change in scores for the set of FE girls who were successfully re-contacted across all three rounds (top panel, "BL-ML1-ML2 Panel") and for the set of girls who were contacted at both BL and ML2 ("BL-ML2 Panel"). As described previously, we generally prefer the "BL-to-ML2" panel because it provides a larger sample size for analysis, though it may obscure relevant trends by ignoring outcomes in the ML1 round – out of necessity, given that many of the girls in question were not included in the final ML1 data. We provide both literacy and numeracy scores here to summarise the overarching gains made by FE girls, before turning to the benchmarked analysis.

TABLE 11: LITERACY AND NUMERACY SCORES AMONG FE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score
Literacy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	251	34.1	63.5	55.1
Banadir	105	32.7	64	54.9
Jubaland	92	37.6	71.9	61.4
South West	54	30.9	48.2	44.5
Literacy Scores – BL-ML2 Panel				
Aggregate	270	33.7	N/A	54.3
Banadir	117	32.3	N/A	53.6
Jubaland	95	37.4	N/A	61.1
South West	58	30.8	N/A	44.5
Numeracy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	251	57.7	73.3	73.1
Banadir	105	54.6	74.8	72.9
Jubaland	92	61.2	74.4	78.5
South West	54	58	68.3	64.5
Numeracy Scores – BL-ML2 Panel				
Aggregate	270	57.5	N/A	72.8
Banadir	117	54.3	N/A	72.1
Jubaland	95	61	N/A	78.7
South West	58	58.4	N/A	64.7

Of course, gains in learning over time may occur naturally due to maturation effects. For this reason, the AGES evaluation design included benchmarks to account for maturation effects, comparing girls' gains in literacy to the expected gains over the same number of years, established at baseline. In the table below, we compare the 270 panel FE girls – that set of baseline girls that were successfully re-contacted at ML2 – to the two alternative benchmarks described above. First, we note that learning outcomes improved markedly across time and across all three geographic areas in which FE programming is being evaluated. In total, Somali literacy scores have increased 20.5 points since baseline among this cohort.

TABLE 12: LITERACY IMPROVEMENTS AMONG FE GIRLS, RELATIVE TO BENCHMARKS, BL-ML2 PANEL

				Comparison to Benchmark	
	Sample Size	**ML2 Score**	**Gain from BL**	**2 Grade Progression**	**3 Grade Progression**
Aggregate	270	54.3	20.5	-9.2*	-18.2*
Banadir	117	53.6	21.3	-13.4*	-23.0*
Jubaland	95	61.1	23.7	-0.2	-11.0*
South West	58	44.5	13.8	-15.5*	-20.4*

The last two columns match that gain of 20.5 points – or state-specific gains, in the rows below it – against the benchmark expectations for learning gains that arise from either a 2-grade or 3-grade progression (e.g., from Grades 1 or 2 to Grades 3 or 4, in the case of a 2-grade progression, or from Grades 1 or 2 to Grades 4 or 5, in the case of a 3-grade progression). By either standard, FE girls' improvements in learning scores lag behind the benchmarked expectation – in most cases, to a large degree that is statistically distinguishable from a null result.[30]

The table below reports equivalent findings related to numeracy, with FE girls compared to pre-established benchmarks. Gains in numeracy are somewhat less dramatic than those in Somali literacy, but are still substantial – 15.3 points, on average, from baseline to ML2. However, FE girls do not meet or exceed either of the benchmarks calculated, missing the lower (2 grade level change) benchmark by 3.2 points and the higher (3 grade level change) benchmark by 7.5 points.

TABLE 13: NUMERACY IMPROVEMENTS AMONG FE GIRLS, RELATIVE TO BENCHMARKS

				Comparison to Benchmark	
	Sample Size	**ML2 Score**	**Gain from BL**	**2 Grade Progression**	**3 Grade Progression**
Aggregate	270	72.8	15.3	-3.2*	-7.5*
Banadir	117	72.1	17.8	-2.9	-8.4*
Jubaland	95	78.7	17.7	-0.4	-4.0
South West	58	64.7	6.2	-8.3*	-11.5*

As we noted previously, this comparison may be unfairly prejudicial toward the FE girls, however. First, there may be lingering impacts of COVID-related school disruptions – although school closures had ended prior to the ML1 round of data collection, ad hoc school disruptions continued and closures may have long-run impacts on girls' learning by causing them to drop out earlier than they otherwise would have.[31] Likewise, even if girls remained in school and even progressed in grades, the standards used to determine whether promotion is warranted may have declined during COVID, in response to reduced classroom time and other considerations. COVID cannot fully explain a failure to meet learning benchmarks, however, as scores actually declined – markedly, in the case of literacy – since ML1, when the main impacts of COVID should already have been felt.

Another factor that helps explain this pattern of stagnant (numeracy) or declining (literacy) scores from ML1 to ML2 is the worsening drought conditions throughout 2022 (between ML1 and ML2), particularly in several of AGES core programme areas. The long-running drought conditions in Somalia worsened in 2022, with analysis by the Famine Early Warning Systems Network (FEWSNET) predicting famine-level conditions by the end of 2022 in Baidoa and among IDPs residing in Mogadishu, both target beneficiaries of AGES.[32] These conditions are reflected in the frequency of coping mechanisms used by AGES households: for instance, between ML1 and ML2, the share of households that reduced their daily meals in the previous month

[30] We denote results that are statistically significant at the 5 percent level or below with a single asterisk.

[31] Note that the sample of girls studied for learning outcomes include those who dropped out of their school/learning programme, such that any shock driving higher dropout rates also has impacts on learning scores.

[32] Famine Early Warning Systems Network (FEWS NET) and Food Security and Nutrition Analysis Unit (FSNAU). "Nearly 8.3 million people across Somalia face Crisis (IPC Phase 3) or worse acute food insecurity outcomes." December 2022. Available at: https://reliefweb.int/attachments/fc2d405c-ca29-4526-ad96-6618c2756192/Multi-Partner-Technical-Release-on-Updated-IPC-Analysis-for-Somalia-fo-October-2022-to-June-2023-Final-%28English%29-13-Dec-2022.pdf.

increased from 61.4 per cent – already elevated due to drought conditions in early 2022 – to 71.8 per cent. Relative to either BL or ML1, the share of girls reporting that they went to bed hungry frequently in the previous year has also risen substantially. Declining climatic and economic conditions have a range of direct impacts on girls' learning, either through reduced nutrition (and, more specifically, protein consumption) or through the economic pressures placed on girls and their families.

More fundamentally, the benchmark girls constitute a relatively poor counterfactual for the FE girls. The first reason is that the benchmark girls in higher grades are *uniformly* in higher grades; in other words, the benchmark Grade 5 girls all reached Grade 5. This is not true of all FE cohort girls. In an important sense, FE girls' learning outcomes are conditional on continued enrolment or successful transition year-on-year; by comparison, all benchmark Grade 5 girls have, by definition, had relatively continuous enrolment. Therefore, the finding that FE girls do not meet the benchmark is, in part, a reflection of their comparatively lower exposure to schooling, as some of them have not reached the level of schooling implied by our benchmark grade levels.

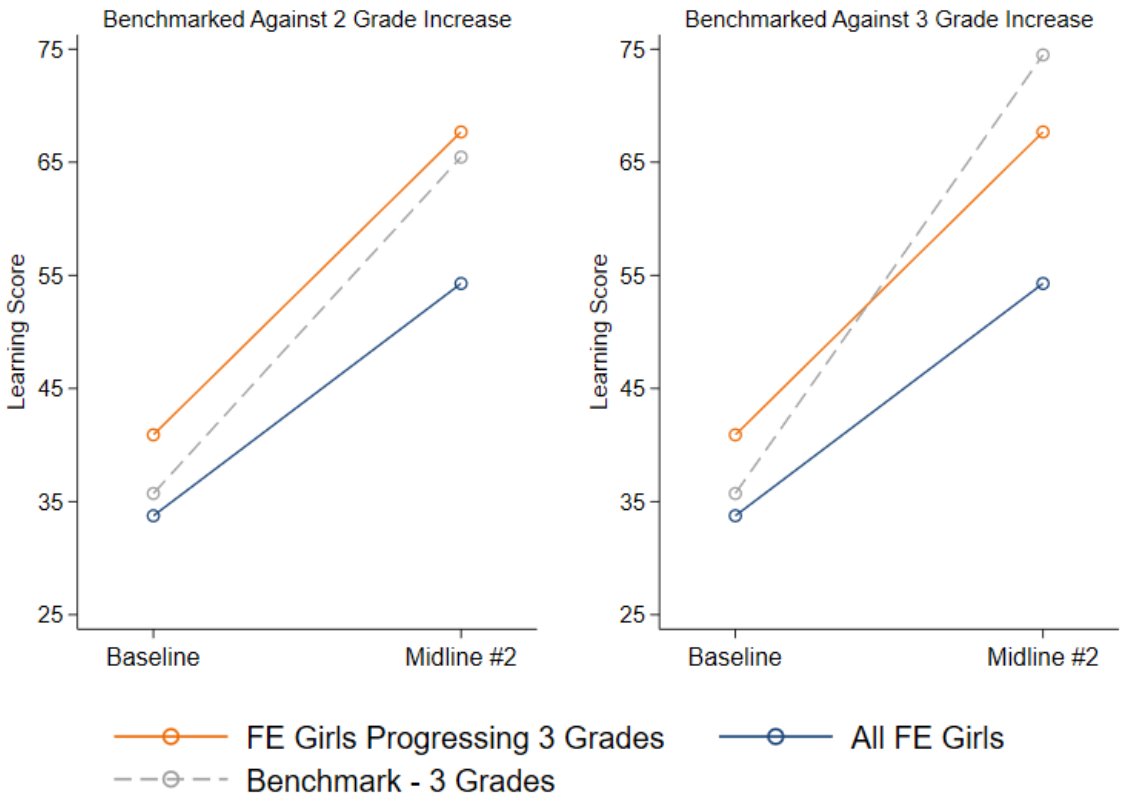
The second reason is even more problematic, because it concerns the underlying *type* of girl that reaches, for instance, Grade 5. Even if we limit our FE girl sample to those girls who remained in school and progressed the expected three grade levels – as we do in further analysis below – benchmark Grade 5 and FE cohort Grade 5 girls are not necessarily the same *type* of girl. To illustrate: benchmark Grade 5 girls remained in school up to Grade 5 in the absence of the AGES intervention; they did so without an explicit programme. To the extent that FE cohort girls reach Grade 5, we should attribute at least a portion of their longevity in school to the AGES intervention. It would be reasonable to assume, then, that benchmark Grade 5 girls might come from wealthier families, have parents who value education more highly, or have greater interest in obtaining an education themselves. To the extent that this is true, the higher-grade benchmark girls do not constitute a precise like-for-like comparison for the FE girls in terms of learning outcomes.

In the figure below, we separate the Somali literacy growth rates of FE girls who progressed 3 grades between BL and ML2 (e.g., if they started in Grade 1, they are now in Grade 4) from FE girls as a whole. The figure shows Somali literacy scores over time for FE girls who achieved the expected grade promotion, for all FE girls, and for the 2-grade (left panel) and 3-grade (right panel) benchmarks.

The trends over time highlight two findings. First, FE girls who were promoted three times since 2019 do experience a slightly steeper increase in literacy than the overall set of FE girls. In fact, the difference in the trend lines – comparison orange to blue lines in the figure – understates this point, because the blue trend line includes all FE girls, including those who were promoted three times. When we treat these groups separately, FE girls in the promoted group gained 26.8 points in the literacy assessment from BL to ML2, while FE girls who were not promoted 3 times gained just 10.9 points. In short, FE girls who maintained enrolment and were consistently promoted across years achieved *closer* to the expected benchmarks than those who dropped out or who maintained enrolment but were not consistently promoted from grade to grade.

Second, comparing the left and right panels of the figure illustrates the degree to which a "3 grade level progression" standard for benchmarking is much stricter. Comparing grade-promoted girls to the 2-grade benchmark in the left panel, the grade-promoted FE girls do not achieve the benchmark learning gains, but the difference in slopes between their learning gains (orange line) and the benchmark (gray dotted line) is minimal. When compared to a 3-grade benchmark in the right panel, however, they lag behind dramatically.

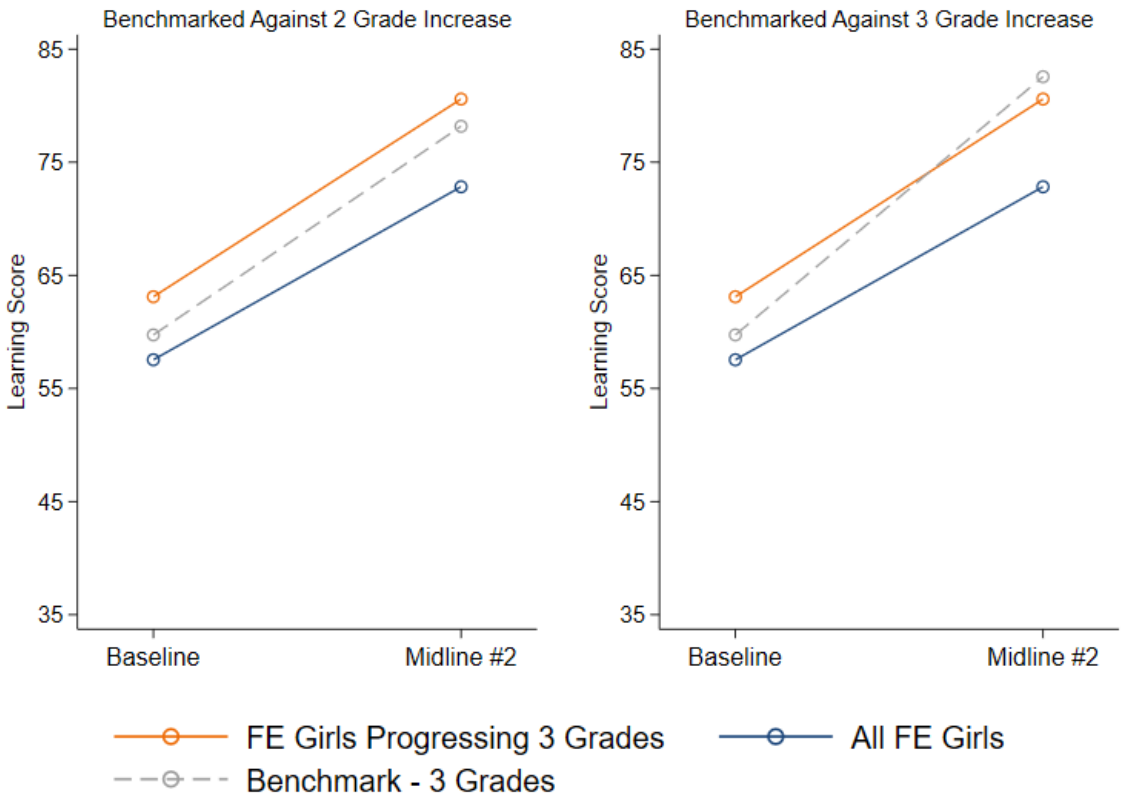
FIGURE 4: BENCHMARKED IMPROVEMENTS IN LITERACY AMONG FE GIRLS FROM BL TO ML2



The figure below reports the same analysis for numeracy outcomes. In numeracy, the gap between the promoted FE girls – those who achieved consistent promotion since BL – and the 2-grade benchmark is similarly minimal. However, they lag far behind the 3-grade benchmark in the right panel. It is also worth noting that the gap between promoted FE girls and those who have not been consistently promoted is smaller in the context of numeracy. Gains in literacy scores from BL to ML2 among the grade-promoted FE girls were 2.5 times higher (26.8 points of improvement versus 10.9) than those among other FE girls in the sample. In the case of numeracy, grade-promoted FE girls achieve gains 1.5 times higher (17.5 points of improvement versus 11.9) than the other FE girls in the sample.

Overall, FE girls recruited at baseline showed significant improvements in both Somali literacy and numeracy from baseline to ML2. However, this trend is tempered by the fact that FE girls – even those who remained in school and were promoted year-on-year in line with expectations – failed to meet the expected learning improvements established as benchmarks. FE girls with more exposure to schooling experience greater gains, but still do not meet the benchmarks. In addition, learning among FE girls has plateaued or even declined since ML1. Our view is that the benchmarks established may have constituted unrealistic expectations, strained further by the COVID-19 pandemic. However, the plateau in learning outcomes between ML1 and ML2 is cause for deeper concern about the learning trajectory of this cohort.

FIGURE 5: BENCHMARKED IMPROVEMENTS IN NUMERACY AMONG FE GIRLS FROM BL TO ML2



Accelerated Basic Education (ABE) Girls

Turning to the second of the original baseline cohorts, we now consider gains in learning among the ABE girls. To start, the table below reports mean literacy and numeracy scores, by state and in the aggregate, for ABE girls across each round of the evaluation. It is important to note that the table covers both of the panel samples of ABE girls previously mentioned: the BL-ML2 panel captures the set of girls who were recruited at BL and successfully re-contacted at ML2, regardless of their status at ML1, providing the largest panel sample for studying gains between those two rounds. Meanwhile, the BL-ML1-ML2 ("full") panel includes only girls who appeared in all three rounds of data collection, yielding a smaller sample size but one which provides richness in the form of results at ML1.

TABLE 14: LITERACY AND NUMERACY SCORES AMONG ABE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score
Literacy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	262	41.5	63.6	48.1
Banadir	117	39.8	65.8	45.2
Jubaland	65	43.3	57	45.4
South West	80	42.7	65.7	54.6
Literacy Scores – BL-ML2 Panel				

Aggregate	310	39.6	N/A	46.4
Banadir	134	37.7	N/A	42.7
Jubaland	82	40.6	N/A	46.2
South West	94	41.6	N/A	52.1
Numeracy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	262	65.2	73.8	70.6
Banadir	117	65.1	77.1	66.6
Jubaland	65	67.1	62.2	75.7
South West	80	63.9	78.3	72.1
Numeracy Scores – BL-ML2 Panel				
Aggregate	310	65.3	N/A	69.6
Banadir	134	65.4	N/A	65.5
Jubaland	82	67.3	N/A	74.6
South West	94	63.4	N/A	71.1

ABE girls were recruited in late 2019 and enrolled in accelerated learning programmes designed to facilitate re-entry into formal school for girls who had dropped out and fallen behind. In contrast to formal school and the FE girls described in the prior section, the ABE intervention was time-limited – in other words, it had an expected ending in 2021, at which point beneficiaries would either transition into the formal education system, continue to a different education pathway, or move into the labour market, among other possibilities. This distinction is important to emphasise, because it suggests that our expectations of learning over 3 or more years should be tempered in the case of ABE (and C1 NFE) girls, as their exposure to specific learning interventions is less than FE girls.

This distinction is reflected, somewhat in the benchmarks established for ABE girls, which are derived from the expected change in learning between Grade 1 and Grade 2 girls – a less ambitious benchmark compared to that applied to the FE girls in the previous section. The table below reports gains in Somali literacy among ABE girls between BL and ML2, broken down by state, and compared to this benchmark. As the table shows, ABE girls experienced relatively marginal gains in literacy – as high as 10.5 percentage points in Southwest State, but just 6.8 points in the aggregate. They also systematically failed to achieve the benchmarked literacy gains, except in Southwest State, where literacy improvements were higher and the benchmark was less ambitious (9.1 points expected, compared to 27.8 or 27.9 points in Banadir or Jubaland, respectively).

TABLE 15: LITERACY IMPROVEMENTS AMONG ABE GIRLS, RELATIVE TO BENCHMARKS

	Sample Size	ML2 Score	Gain from BL	Comparison to Benchmark
Aggregate	310	46.4	6.8	-15.4*
Banadir	134	42.7	5.0	-22.9*
Jubaland	82	46.2	5.6	-22.2*
Southwest	94	52.1	10.5	1.4

These findings are generally replicated in the case of numeracy outcomes, as the table below demonstrates. Improvements in numeracy were concentrated exclusively in Jubaland and Southwest State and, aggregating across the entire sample, ABE girls fell 11.5 points below the expected gains in numeracy.

Our analysis of ABE girls and their failure to meet the benchmark is complicated somewhat by the fact that ABE girls regressed – especially in terms of literacy – from ML1 to ML2. This reflects the structure of the ABE programme, as the learning-focused intervention was generally no longer present between ML1 and ML2; as a result, learning loss is not entirely surprising. However, even this regression, documented earlier in this section, does not explain the failure to meet the benchmark, as ABE girls did not experience the expected learning improvements over the BL-to-ML1 period either.

TABLE 16: NUMERACY IMPROVEMENTS AMONG ABE GIRLS, RELATIVE TO BENCHMARKS

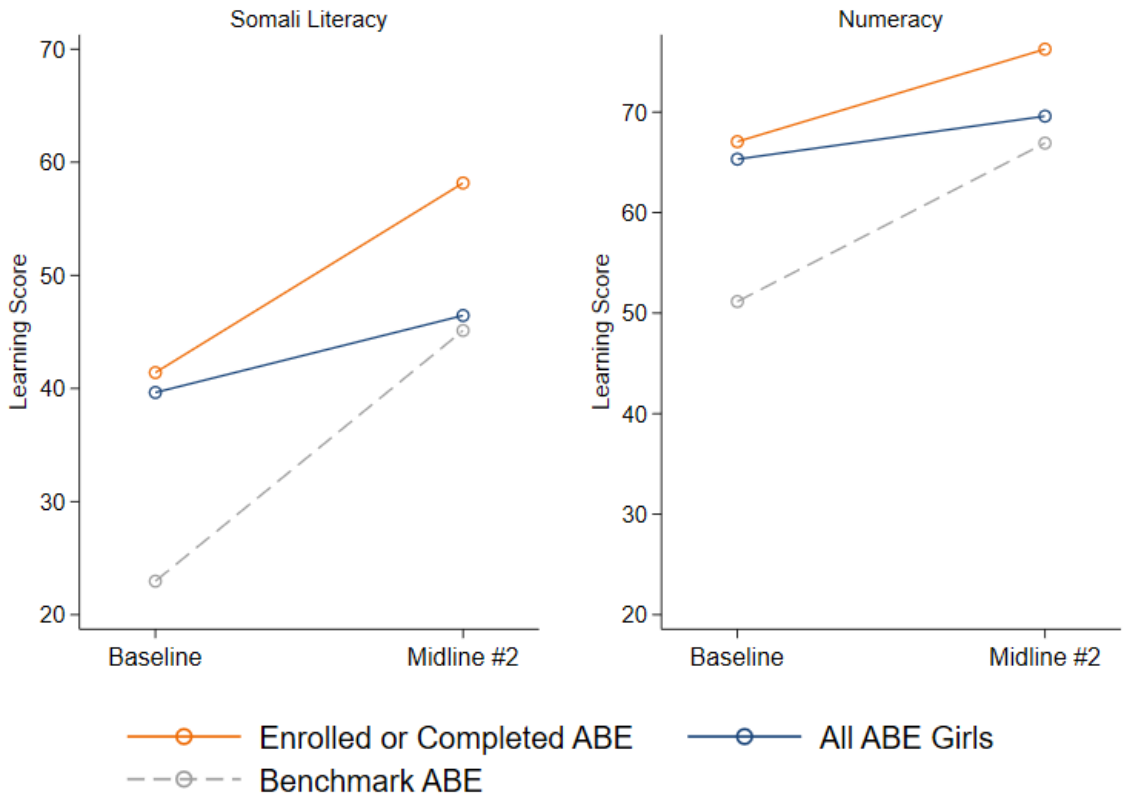
	Sample Size	ML2 Score	Gain from BL	Comparison to Benchmark
Aggregate	310	69.6	4.3	-11.5*
Banadir	134	65.5	0.0	-22.0*
Jubaland	82	74.6	7.3	-10.4*
Southwest	94	71.1	7.7	2.6

The failure to achieve the benchmarks laid out are largely a function of sharply diverging outcomes within the ABE cohort. In the figure below, we plot the Somali literacy and numeracy scores at BL and ML2 of two groups of ABE girls. The first are those who self-report completing an ABE programme or self-report current, ongoing enrolment in either ABE, NFE, or formal school. The second is the full set of ABE girls. This latter group includes 135 girls who did not complete the ABE programme, according to self-reports during the current round of data collection. In other words, we can consider two sets of ABE girls in our panel sample: 175 who report ongoing enrolment or completion of the ABE programme and the 135 who report non-completion of their programme.

When we disaggregate learning changes in this way, girls who are currently engaged or completed the ABE programme achieve greater learning gains – gains that fall slightly short of, but are of a similar magnitude to, the benchmarked expectations. On the other hands, girls who do not report having completed ABE and who are no longer enrolled in any learning programme, have significantly worse learning outcomes over time.

As was true of our presentation of FE girls' learning trajectories, this figure understates the distinction between the two groups of girls. From BL to ML2, ABE-completers – those who completed ABE or remain engaged in a learning programme – gained 16.8 points in the Somali literacy assessment and 9.2 points in numeracy. ABE non-completers, on the other hand, regressed by 6.1 points in terms of literacy and by 2.2 points in terms of numeracy over the same period. The former group did not meet the benchmarks established at baseline but, as the figure below shows, their trajectory (denoted in orange) is roughly similar to the benchmarks (dotted gray line). When we include *all* ABE girls, including non-completers, learning gains (denoted in blue) are significantly worse than the benchmarks.

FIGURE 6: BENCHMARKED IMPROVEMENTS IN LITERACY AND NUMERACY AMONG ABE GIRLS



As in our discussion of a similar dichotomy in the context of FE girls, this point suggests that girls who are properly exposed to the full intervention *do* register learning gains. This is encouraging; however, it means that the programme's impact is conditional on girls remaining engaged in the programme – this distinction is similar to the methodological differentiation between an "average treatment effect" and the "intent-to-treat" effect, in which an intervention might have an impact among those who are effectively exposed to it, but – once non-compliance is factored in – the intervention's impact is reduced or nullified entirely. This is, potentially, an important area of further research for the programme, especially because there may be critical differences in outcomes between the following four pathways through an ABE intervention:

- Dropped out of ABE, remained out of any education programme
- Completed ABE
- Completed ABE, remained in higher-level ABE streams
- Completed ABE, transitioned into the formal education system

Our sample can only offer suggestive evidence regarding the learning outcomes associated with these different pathways, due to the small size of the ABE panel sample. This issue is worsened by the relatively poor documentation, on the part of the evaluation team, of girls' post-ABE educational exposure – while the data indicates whether a girl is enrolled in school, it does not indicate when she enrolled or the extent of any gap between ABE and enrolment.

Considering those limitations, our analysis provides suggestive evidence of differentiation across pathways. Among ABE girls who report, at ML2, that they are enrolled in formal school, literacy scores at 25.2 points

higher than at BL and numeracy scores are 8.5 points higher than at BL. ABE girls who report they continue, at ML2, to attend an ABE programme show similar gains over time. In contrast, girls who report being employed at ML2 show no tangible improvement in literacy or numeracy between BL and ML2, and girls who are reportedly idle – not enrolled in a learning programme and not engaged in employment – have lost both literacy and numeracy skills since baseline. These results suggest that ABE programming can have substantial learning benefits in cases where girls follow the targeted pathway of using ABE to transition into formal school or remain in higher-level ABE (or equivalent) courses. Unfortunately, the rate of transition from ABE into formal schooling is relatively low; focusing on maintaining enrolment in ABE and facilitating transition into formal schooling, above all else, may be one option for increasing the net impact of ABE on learning outcomes.

Cohort 1 Non-Formal Education (C1 NFE) Girls

The final group of girls tracked as part of the original baseline cohorts consists of girls participating in non-formal education (C1 NFE). Learning and numeracy scores over time for this cohort are provided in the table below, broken down by geographic zone. For each of the two outcomes, we analyse two different panel samples, as described in the context of FE and ABE girls previously.

TABLE 17: LITERACY AND NUMERACY SCORES AMONG C1 NFE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score
Literacy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	263	55.6	61.7	49
Banadir	130	61.2	69.8	52.5
Jubaland	64	50.5	49.9	44.5
South West	69	49.5	57.4	46.6
Literacy Scores – BL-ML2 Panel				
Aggregate	302	54.7	N/A	48.1
Banadir	143	60.9	N/A	52.3
Jubaland	76	48.1	N/A	44.2
South West	83	50.1	N/A	44.6
Numeracy Scores – BL-ML1-ML2 ("Full") Panel				
Aggregate	263	74.1	70	68.2
Banadir	130	76.1	76.4	69.7
Jubaland	64	70.5	50.3	72.1
South West	69	73.7	76.5	61.8
Numeracy Scores – BL-ML2 Panel				
Aggregate	302	73.6	N/A	67.6
Banadir	143	75.1	N/A	69.2
Jubaland	76	69.6	N/A	70.6
South West	83	74.6	N/A	61.9

Unlike the FE and ABE girls discussed previously, C1 NFE girls were not explicitly benchmarked for learning improvements at baseline. Our focus in this section will be in documenting changes in learning outcomes since baseline, without reference to a particular benchmark.

In practice, the lack of a benchmark is of marginal importance, because learning outcomes among this cohort have generally gotten worse over time. Among our panel sample of 302 C1 NFE girls, Somali literacy scores have declined by 6.5 points since BL, as shown in the table below. Importantly, this decline was broad-based geographically – we observe roughly similar declines in literacy across all three geographic zones.

TABLE 18: LITERACY IMPROVEMENTS AMONG C1 NFE GIRLS SINCE ML1

	Sample Size	Baseline Score	ML2 Score	Gain from BL
Aggregate	302	54.7	48.1	-6.5
Banadir	143	60.9	52.3	-8.6
Jubaland	76	48.1	44.2	-3.9
Southwest	83	50.1	44.6	-5.5

Numeracy scores have evolved similarly, as shown in the table below. Across all girls, numeracy scores have declined by 6 points, with a particularly large decline in Southwest State.

It is important to note that this finding is not new to the ML2 round. From baseline to ML1, C1 NFE girls experienced a significant decline in learning skills, a fact that is documented in the ML1 evaluation report as well as the figures – presented earlier in this section – that showed trends from BL, through ML1, to ML2. Among the three cohorts tracked, C1 NFE girls were the only group to experience decline over both the BL-to-ML1 and the ML1-to-ML2 periods.

TABLE 19: NUMERACY IMPROVEMENTS AMONG C1 NFE GIRLS SINCE ML1

	Sample Size	Baseline Score	ML2 Score	Gain from BL
Aggregate	302	73.6	67.6	-6.0
Banadir	143	75.1	69.2	-5.9
Jubaland	76	69.6	70.6	1.0
Southwest	83	74.6	61.9	-12.7

The same trend documented in the case of ABE girls can also be observed, though less starkly, among the C1 NFE girls. Specifically, C1 NFE girls who transitioned into formal school show significant gains in literacy and more modest gains in numeracy since baseline. It is important to note that NFE girls' enrolling in formal school was not an explicit goal of the programme, given that NFE girls were either older adolescents or young adults when initially enrolled in the programme. Therefore, the set of girls in the sample who have moved into formal school is very small (n = 18, out of 302 girls in the panel sample). Girls who self-report continued enrolment in an NFE programme also demonstrate marginal improvements in learning outcomes since BL.

At the same time, the majority of NFE girls have either shifted into employment (n = 157) or are out-of-school and not employed (n = 78). Among these girls, there has been a sharp decline in literacy and numeracy over time. For instance, among girls who are currently employed or self-employed, literacy scores have declined by 9.1 points since the baseline, and numeracy has declined by 6.3 points. Girls who are "idle" –

neither employed nor in any form of learning programme – have experienced the steepest decline in literacy and numeracy skills. One interpretation of this outcome is that some form of continuing education, even if it is through informal means, is essential to maintain rudimentary literacy and numeracy skills while transitioning into adulthood. Utilisation of these skills through employment may be too sporadic to have much impact on long-run skills maintenance.

4.3 Subtask-Specific Changes in Learning

The analysis in the previous section focused on aggregate changes in literacy and numeracy outcomes. In this section, we focus on individual subtasks, with the goal of understanding the nature of learning improvements made via the programme's educational interventions. For instance, while ABE girls demonstrated moderate improvements in numeracy between BL and ML2, it is possible that these improvements were concentrated among one or two key numeracy skills, such as addition involving two-digit numbers.

Throughout this section, we will refer to subtasks by their number, and note the specific skill being tested when discussing the results. In the context of Somali literacy, the subtasks assess the following broad skills:

- Subtask 1: Identification of numbers
- Subtask 2: Quantitative Discrimination – identifying the largest number in a set
- Subtask 3: Missing number identification
- Subtask 4: Addition with 1 digit
- Subtask 5: Addition with 2 digits
- Subtask 6: Subtraction with 1 digit
- Subtask 7: Subtraction with 2 digits
- Subtask 8: Word Problem

For this analysis, we do not include the additional 3 subtasks added during the ML1 evaluation round, as our focus is on comparing change from BL to ML2.

The literacy subtasks assess the following skills:

- Letter Sound Identification
- Words Commonly Used
- Reading fluency
- Reading Comprehension - Level 1
- Reading Comprehension - Level 2
- Reading Comprehension - Level 3

We begin by focusing on FE girls and their progression since baseline. During the baseline evaluation, we found that there was a significant drop-off in performance among FE girls between Somali literacy subtasks 1 and 2, as the assessment shifted from letter identification to identification and reading of common words. This was the sharpest breakpoint in performance on EGRA for FE girls; following this shift, girls' performance declined but only very gradually, and there was a small subgroup of girls who performed relatively well across all six subtasks.

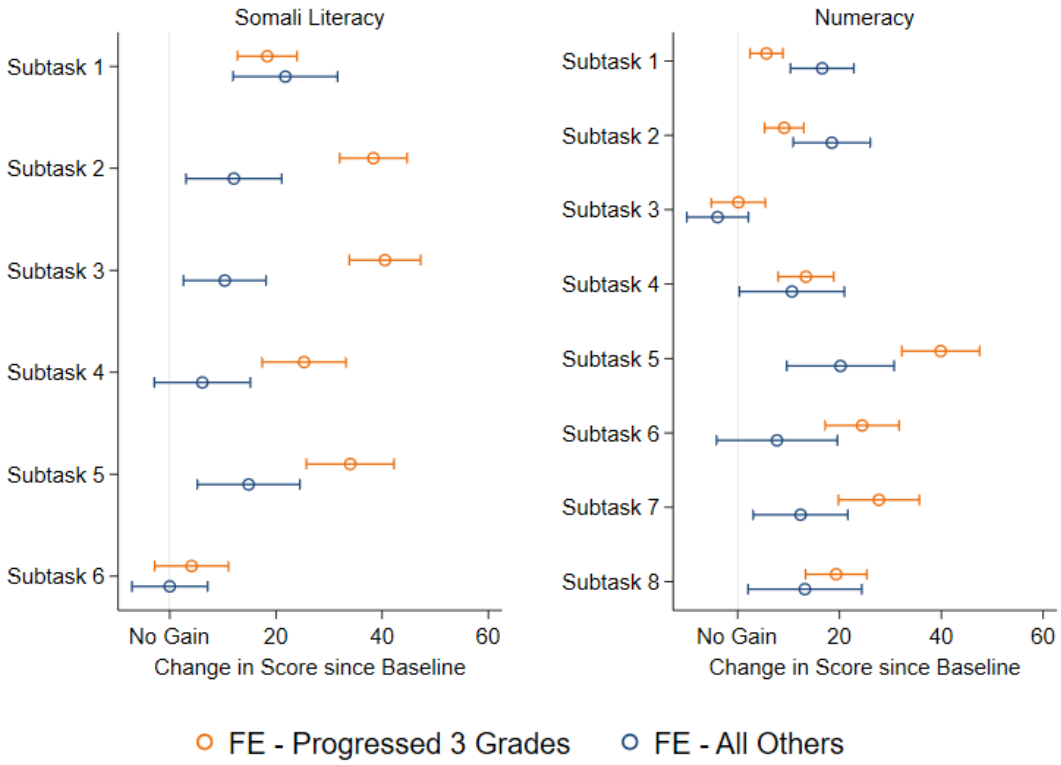
In the figure below, we report the changes in subtask-specific scores for FE girls in literacy in the left panel. We report results for two groups of girls – those who remained enrolled and were promoted across 3 grade

levels between 2019 and 2023, and those who either dropped out or were not promoted in line with expected year-on-year progress. For each subtask, we report the mean *change* in scores from BL to ML2, disaggregated by these two groups. The vertical line labelled "no gain" indicates that there has been no change since baseline in terms of proficiency on a given subtask. The change in mean scores is denoted by a hollow circle; the horizontal bars around each circle denote the 95 percent confidence interval around the change score; where the confidence interval does not intersect the vertical line at 0 ("no gain"), the change from BL to ML2 is statistically significant.

To ensure the interpretation is clear, consider an example using subtask 2 of the literacy assessment. For this subtask, the change in scores among FE girls who progressed 3 grade levels since BL is 38.4 points. In contrast, among the FE girls who did not progress 3 grade levels, the increase in scores on subtask 2 is just 12.1 points. In both cases, the change since baseline is statistically significant.

Interestingly, the figure shows that both types of girls in the sample improved in terms of letter identification (subtask 1). Indeed, on both the easiest and hardest subtasks, the two groups of FE girls are virtually indistinguishable from one another. On the hardest subtask, focused on more difficult aspects of reading comprehension, neither group of girls improved substantially since the baseline. Notably, this suggests that, even among FE girls who have been enrolled and promoted year-on-year consistently, there is a plateau in learning between subtasks 5 and 6, at the level of more difficult reading comprehension.

FIGURE 7: SUBTASK-SPECIFIC LEARNING IMPROVEMENTS, BL TO ML2, AMONG FE GIRLS



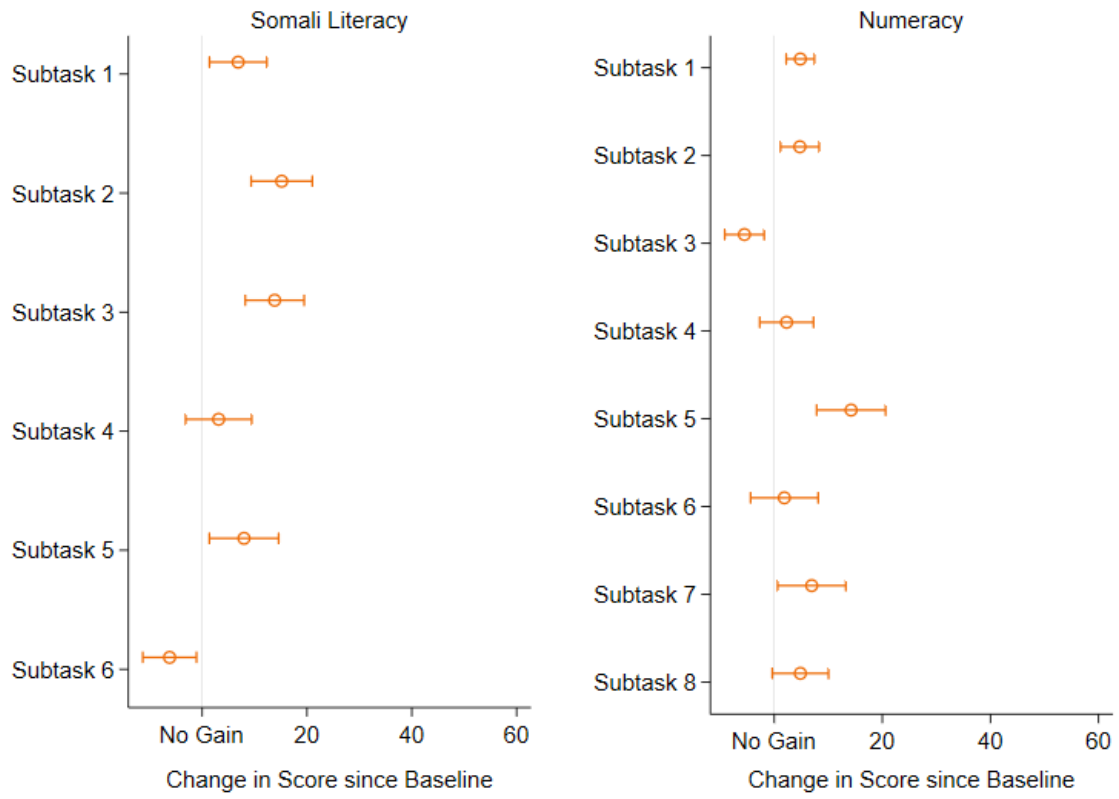
It is within the middle four subtasks – subtask 2 through subtask 5 – that we observe an important divergence between girls who progressed in school and those who did not. Across these subtasks, there are large increases

in proficiency among promoted FE girls. The other FE girls also progressed on these subtasks, but at a much lower rate.

To illustrate the magnitude of the gaps reported within literacy subtasks, consider subtask 3, which focuses on reading fluency. The scoring of this subtask consists of counting the number of words a girl successfully reads within a short story, over the course of one minute of reading time. Focusing on the group of promoted FE girls, they were able to read 26.2% of the words in the story at baseline; by ML2, this had increased to 66.8%. This can be interpreted as approximately equivalent to a girl's reading speed, in words per minute, for the story. Among the non-promoted girls, their reading speed increased from 14.9% to 25.3% over the same period.

The patterns shown in the right panel of the figure are less clear-cut and less encouraging, in some ways. The most positive finding concerns the general increases in proficiency in mid-range subtasks that focus on 2-digit addition and subtraction. At baseline, the shift from 1-digit addition or subtraction to 2-digit problems was a major breakpoint, with 66 percent of girls showing proficiency in 1-digit addition but just 23 percent showing proficiency in 2-digit addition. From BL to ML2, girls improved across all addition and subtraction subtasks, but these gains were most pronounced in the 2-digit versions of these problems.

FIGURE 8: SUBTASK-SPECIFIC LEARNING IMPROVEMENTS, BL TO ML2, AMONG ABE GIRLS



Shifting to ABE girls, we now consider the changes in subtask-specific scores across the entire cohort of ABE girls, without disaggregating as a function of progression through the programme or transition into other educational opportunities. As was the case with the FE girl's cohort, ABE girls demonstrate no improvement in the most difficult literacy subtask (#6); indeed, performance – among a panel sample of the same girls

across rounds – declined from BL to ML2. ABE girls show improvement in word identification and reading fluency, suggesting that the programme was effective in helping them identify words; however, they show less improvement in reading comprehension (subtasks 4-6), even when reading comprehension questions target a story they read for fluency (as in subtask 4's questions about the content of the story read for speed/fluency in subtask 3).

The right panel of the figure documents subtask-specific improvements in numeracy scores among ABE girls. The results of this analysis are extremely mixed. The most positive outcome is on subtask 5, which tests one's ability to complete simple (1-digit) subtraction. Girls improved markedly on this subtask but made no aggregate gains in addition and little improvement in slightly more difficult addition or subtraction problems involving two digits.

Of course, it is natural that there are fewer large gains in subtask-specific literacy or numeracy skills among ABE girls, given that they showed less significant aggregate improvement on EGRA or EGMA in our earlier analysis. It is notable that ABE girls tended to retain or improve on the most basic skills in both literacy and numeracy (e.g., letter and word identification, number identification, and number discrimination). On slightly higher-order skills, such as 1-digit addition, ABE girls already tended to have proficiency, reducing the scope of improvement from BL to ML2; ABE girls scored 82.4 percent on 1-digit addition (subtask 4) at baseline. To the extent that accelerated basic education is intended to impart practical numeracy skills, ABE girls tend to achieve that. However, if the goal of the intervention was to ensure girls could complete simple multiplication or reliably subtract 2-digit numbers, it is less reasonable to say the project achieved this objective.

4.4 Testing the Theory of Change

Youth Leadership Index

The Youth Leadership Index (YLI) is scored from 0 to 100 and provides an estimate of girls' self-assessed leadership skills. The girls answer 21 questions about how they think about the consequences of their own actions, their belief in their ability to express their thoughts clearly and organizing others to take part in an activity to achieve a common goal. To test the Theory of Change, in this section we determine whether YLI is a predictor of increases in learning outcomes and investigate the relationship between the YLI score of FE, ABE, and C1 NFE girls and the change in their learning outcomes from BL to ML2.

TABLE 20: EFFECT OF YLI SCORES ON CHANGES IN LEARNING OUTCOMES, BY COHORT

Cohort	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
FE girls	0.17	0.001	0.25	0.001
ABE girls	0.10	0.032	0.17	0.006
C1 NFE girls	-0.01	0.856	0.00	0.971

We use a linear regression framework with girls' age and region as control variables framework to understand the effect of YLI scores on the change of learning outcomes since BL. As the table above shows, an additional point in YLI is associated with a modest but significant increase in numeracy and literacy scores among both FE and ABE girls but had a statistically insignificant effect for C1 NFE girls.

GEF Participation

In connection with the life skills discussed in the previous section are girls' participation in the Girls' Empowerment forum. This section discusses Girls' Empowerment Forum (GEF) participation among FE, ABE, and C1 NFE girls and its relationship to learning outcomes. GEFs are the primary mode through which AGES seeks to increase self-esteem, leadership capacity, and life skills among girls. The after-school programmes offered by GEF provide girls with a peer support network, positive female role models, and opportunities for tutoring and direct mentoring.

Research from GEFs in the SOMGEP-T programme as well as AGES show that girls who participate tend to perform better on learning assessments. Our study on surfaces similar findings for the ABE and C1 NFE cohort who participated in the programme in BL and who continue to remain in contact with GEF members. Surprisingly, we found that the positive effect on learning outcomes of ABE and C1 NFE cohort's participation can still be observed approximately two years after their participation in the ABE or C1 NFE has ended. Furthermore, the positive effect on learning outcomes is even stronger among those who have continued contact with other GEF members.

We should note here that GEF participants are not randomly assigned, and so some of the effect on the learning outcomes may be due to an unmeasured variable such as motivation which may be higher in girls who participated in GEF.

TABLE 21: EFFECT OF GEF PARTICIPATION ON CHANGES IN LEARNING OUTCOMES, BY COHORT

Participation - Cohort type	Cohort	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
GEF Participation	FE girls	0.72	0.833	4.67	0.314
	ABE girls	7.85	0.006	14.56	0.003
	C1 NFE girls	4.81	0.049	11.73	0.009
GEF Participation Continued	FE girls	1.50	0.698	7.50	0.199
	ABE girls	8.00	0.012	11.58	0.031
	C1 NFE girls	8.64	0.012	14.81	0.008

FE girls who participated in GEF did not improve their learning outcomes significantly more than other FE girls who did not participate in GEF. Given that FE girls are of school-age and the FE programme is intended to help them stay in school, the FE girls have received support from formal schooling to improve both their

learning outcomes as well as their leadership skills. This level of support at formal school regardless of their participation in GEF may be the reason the effect of their GEF participation is muted.

Teaching Quality

The AGES Theory of Change suggests that an improved pedagogy will lead girls to improve their learning outcomes. If teachers are better trained to teach numeracy and literacy skills to girls, are consistently present, make students feel safe and welcomed in the classroom, and encourage their schooling, then students will learn more in class and will be more likely to attend school attend school. The midline assessment asked girls a series of questions presented below about their teacher's pedagogical practices. In this section, we analyse the relationship between student-reported measures of teaching quality and gains in learning scores from BL to ML2.

For each of the teaching quality measures shown below, we use a linear regression to predict change in literacy and numeracy scores with the teaching quality measure as the predictor variable and age, state, and cohort as control variables.[33] We do not present the findings by cohort because there were no significant differences between the cohorts with respect to the relationship between teaching quality and learning outcomes.

TABLE 22: EFFECT OF TEACHING PRACTICES ON CHANGES IN LEARNING OUTCOMES, BY COHORT

Teacher quality	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
My teacher does not make me feel welcome in classroom	-0.29	0.871	-1.503	0.637
My teachers are often absent	-2.26	0.242	-3.292	0.325
My teacher rarely/never encourages participation	1.29	0.493	-0.060	0.986
My teacher explains how learning things is useful in our lives	6.88	0.044	2.513	0.693
My teacher's lessons move too fast for me	3.03	0.052	3.905	0.092
My teacher punishes students who get things wrong in a lesson	2.28	0.082	2.325	0.271

[33] When a girl is no longer enrolled in school or a learning programme during data collection, the girl was asked to reflect on their teacher's practices from when they were in school or in the programme which typically occurred in late 2019 and 2020 for ABE and C1 NFE girls.

My teacher used corporal punishment in last week	0.33	0.850	1.62	0.553

Our findings presented in the table above generally do not indicate a clear relationship between the teaching quality measures and learning outcomes. Girls who indicated that their teacher explains how learning things is useful in their lives improved their numeracy score 6.88 points more compared to those who did not. Other measures of teaching quality did not have a clear statistical relationship to changes in learning scores. This finding is somewhat less surprising among ABE and NFE girls who may have had a few years to compensate for the poor pedagogical practices they experienced at BL, but even among FE girls who are in school, there is not a clear relationship between teacher quality and learning scores.

Community Attitudes

The Theory of Change posits that changes in community attitudes around girls education will lead girls to both attend and learn more at school. When caregivers believe that girls' education is worthwhile vis a vis other demands to the household, then caregivers are more likely to send their daughter to school and support their education pursuits instead of helping with household chores, allowing her to marry, or to work. When caregivers place value on education, then they are to provide commensurate support, financially or otherwise, to their daughter.

In this section, we assess whether community attitudes, specifically those of their caregivers, towards girls' education are predictors of improvements in learning outcomes from BL to ML2. As with other intermediate outcomes, we take a linear regression approach to understanding the relationship between community attitudes and changes in learning outcomes and control for age and region.

TABLE 23: EFFECT OF COMMUNITY ATTITUDES ON CHANGES IN LEARNING OUTCOMES, BY COHORT

Caregiver attitudes	Cohort	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
Aspires to university education for girl	FE girls	5.37	0.244	6.71	0.299
	ABE girls	-3.51	0.333	-2.27	0.589
	C1 NFE girls	4.39	0.164	14.09	0.011
Girls' education is worthwhile investment even if funds limited	FE girls	3.18	0.403	-1.02	0.816
	ABE girls	-2.55	0.508	-3.44	0.597
	C1 NFE girls	-1.56	0.662	-5.86	0.182
Work/chores acceptable reason for non-enrolment	FE girls	-0.79	0.807	-3.60	0.399
	ABE girls	1.27	0.628	2.24	0.608
	C1 NFE girls	-4.81	0.079	-8.40	0.044

Expense acceptable reason for non-enrolment	FE girls	4.09	0.131	2.50	0.477
	ABE girls	1.81	0.520	0.76	0.797
	C1 NFE girls	0.76	0.696	3.11	0.322

We find that most caregiver attitudes are not significantly correlated with either numeracy or literacy scores for the FE, ABE, and C1 NFE cohorts. However, for C1 NFE girls, we find that the attitude of a caregiver in which the caregiver aspires for a university education for their girls is positively correlated with a 14.09-point increase in the literacy scores. The caregivers who believe work or chores are an acceptable reason for non-enrolment at school are associated with an 8.4-point *decline* in literacy scores. Both are significant effects on changes in scores and may speak to the sensitivity of this cohort to community attitudes given that they are out of school already and rely on their caregivers to support new educational opportunities for them.

5. Transition

The second outcome of focus for the AGES project concerns retention and life pathways, broadly construed. "Transition" is a core outcome for all GEC, GEC-T, and LNGB projects, and it seeks to capture how the project affects girls' pathways through life. In this section, we assess transition outcomes among the original baseline cohorts recruited into the project in late 2019. As noted in the analysis of learning outcomes above, there are critical differences between FE, ABE, and C1 NFE girls, all of whom are included in this section.

Operationalising successful transition is always complex, because there are many possible pathways that can be considered a success, including retention in school, shifting into employment, or re-enrolment in formal school for those who were not enrolled previously. Transition is more than a measure of retention; it incorporates alternative education, vocational training, and even the possibility that an AGES beneficiary will start a small business. It is further complicated by the fact that successful transition is defined by a girl's starting point: for a girl enrolled in formal school, continued enrolment in the same grade level is insufficient for a successful outcome; in contrast, for a girl who was out-of-school or enrolled in an NFE programme, transitioning into formal education is a success, regardless of her grade level. In these ways, transition accounts for the complex environment in which AGES is implemented.

The analysis in this section is broken into three parts, each with a distinct goal. In the first section, we define transition – as it is used by AGES – in more detail and report the overall transition rates for girls who were initially recruited into AGES programming in 2019. In the second section, we analyse subgroup-specific transition outcomes, assessing whether transition rates differ as a function of a girl's household characteristics, marginalisation, or other factors. Finally, we analyse the relationship between the programme's intermediate outcomes – such as teaching quality, attendance, and self-esteem – and transition rates, with an eye to providing evidence for or against the programme's Theory of Change.

5.1 Aggregate Transition Outcomes

As we alluded to in the previous section, defining a successful transition outcome is complex, because transition is an inherently complicated, contextual, and multidimensional outcome. In much of our analysis, we will define transition in a binary fashion – success or failure – even though this obscures significant variation in how girls' lives evolve in response to the programme. At times, we will describe and present

results for more specific pathways, to provide additional depth. We also consider one important alternative conception of transition, in the context of ABE and NFE girls, which recognises that ABE and NFE girls who remain as participants in those types of learning programmes may or may not be considered a "success" – a point we unpack in greater detail below.

The table below defines transition outcomes for the three cohorts of girls, classifying them as either successful or unsuccessful – in binary terms – as a function of their cohort.[34] Recall that FE girls were recruited between ages 10 and 13 (primarily 10-12 years), while ABE and C1 NFE girls averaged 14.1 and 17.9 years of age at baseline. As the table makes clear, the starting point of the girl is central to defining successful transition. A formal schoolgirl who drops out of school to pursue employment is not considered a success because she is comparatively young and employment after completing just 2-3 years of primary schooling is not sufficient. In contrast, an NFE girl who has transitioned into employment or self-employment is considered successful because she is older and employment is a desirable outcome for older girls, and because NFE courses are shorter, such that her learning programme, begun at baseline, has certainly ended by the time of the midline evaluation. Similarly, enrolment in formal school is sufficient – regardless of grade level – to be considered a successful transition outcome for ABE and NFE girls, but, for FE girls, upward progression in grades is required.

TABLE 24: TRANSITION PATHWAYS, ACCORDING TO STARTING POINT OR COHORT

Starting Point	Successful Transition	Unsuccessful Transition
FE Girl Enrolled in grades 1-2 at baseline	• Retention in formal school, with progression through the grades (e.g., a girl in grade 1 has reached grade 3 two years later) • Drops out but is enrolled in a technical or vocational education programme	• Drop out • Retention in formal school without appropriate grade progression • Transition into employment or self-employment
ABE Girl Enrolled in ABE at baseline	• Enrolment in formal school, at any grade level • Transition into a technical or vocational education programme • Transition into age-appropriate, non-exploitative employment • Transition into self-employment	• Drop out • Idleness after programme completion • Transition into NFE • Retention in ABE
NFE Girl Enrolled in NFE at baseline	• Enrolment in formal school, at any grade level • Transition into a technical or vocational education programme • Transition into age-appropriate, non-exploitative employment • Transition into self-employment	• Drop out • Idleness after programme completion • Transition into ABE • Retention in NFE

[34] The text in red in the table indicates transition outcomes that are slightly ambiguous, and which we code according to two different standards, discussed in greater detail below.

An idiosyncratic aspect of transition in this context is the time that has elapsed since baseline, which was between 3 and 3.5 years prior to data collection for this round. This has two implications for defining transition outcomes. First, FE girls who have remained in school consistently during that time should, theoretically, have advanced three grade levels – a girl enrolled in Grade 1 in late 2019 should be nearing the end of their time in Grade 4 in early 2023. However, this is a high bar for them to clear, as it requires consistent, year-on-year progression for three straight years. In similar programmes in the region, it is common for girls to progress grades in some years but not others, and defining transition based on progression in *every year* since baseline appears overly restrictive. It also requires progression during COVID-related school closures.

In practice, a surprising number of girls meet the 3-year progression standard. Despite this, our approach to this issue is to define a more liberal version of successful transition for FE girls based on advancing two grade levels since baseline. This means that girls could have schooling disrupted for a year by COVID and still be eligible to be considered a successful transition case. Likewise, if girls were enrolled throughout the period but advanced grades in two out of three years, they are considered a successful transition case. Our core measure of transition employs the more stringer standard, which demands three years of grade progression since baseline; however, we also employ the 2-year standard at times, where noted.

Second, girls recruited into ABE and NFE programmes in late 2019 should have completed their learning programmes by 2021 and mid- or late-2020, respectively. At ML1, we found that many girls self-reported continued enrolment in these programmes, often in higher levels or streams; there is a similar pattern this year – for instance, out of 310 ABE girls successfully re-contacted from the baseline cohort, 85 (27.4 percent of the sample) indicated that they are currently enrolled in an ABE programme. The question this invite is whether continued enrolment in ABE or NFE over three years later should be considered a successful outcome, especially as the ideal pathway for ABE and NFE girls, respectively, was to transition into formal school or shift into gainful employment. It is also possible that respondents erroneously report continued enrol because it was the last learning project, they were part of or because they understand that the programme is tied to our evaluation.

In the analysis below, we consider both definitions of successful transition for ABE and C1 NFE girls.[35] We define two versions of a binary transition outcome: the first considers continued enrolment in ABE (for ABE or NFE girls) or NFE (for NFE girls) a successful outcome; we refer to this as a "liberal" definition of transition.[36] The second does not consider continued enrolment – unless it involves transition into formal school – a successful outcome; in this definition, girls who self-report enrolment in ABE or NFE can still be classified into the successful transition category if they are employed or self-employed, in line with the table above. However, they are not categorised into the successful transition category based on their continued enrolment alone – it requires appropriate employment. This is our standard

Our approach to analysis considers transition *since* baseline; for this evaluation round, we set aside transition outcomes that are specific *between* ML1 and ML2. The reason is hinted at by our discussion of FE girls' grade

[35] This logic does not apply to C4 NFE girls, considered in a later section of the report. C4 NFE girls have been enrolled in NFE programming since early 2022 and it would not be surprising for their programme to be ongoing *or* for them to be continuing their education in a structured manner while looking for employment. This is an issue that can be revisited during the endline evaluation, when C4 NFE girls will have been in the programme for over two years and should, theoretically, have moved on from the NFE programme.

[36] We never consider enrolment in NFE to be a successful outcome for ABE girls. Because ABE is generally a higher level of education than NFE, an ABE girl shifting into an NFE programme is an unambiguously negative outcome.

progression above: girls' transition pathways can be varied, with girls moving in and out of school, or shifting between pathways. Our primary interest is in whether girls are presently engaged in education, training, or gainful employment; if they experienced setbacks or deviations along the way to the present outcome, it is not directly relevant to our main analysis.

Our sample for analysing transition outcomes is limited to the set of girls who have been successfully re-contacted from the baseline round. This approach reflects our interest in aggregate transition outcomes since baseline; it also simplifies much of the discussion that follows, because we do not incorporate FE girls who were selected as cohort replacements during the ML1 round. To ensure our results are not driven by this decision – which excludes 43 FE replacement girls – we also check our main findings related to FE with the expanded sample and its more complicated structure.[37]

The panel of girls tracked from baseline to ML2 includes 270 FE girls, 310 ABE girls, and 302 C1 NFE girls. In total, our conservative or strict approach to defining transition results in an aggregate transition rate of 61.3 percent, while the liberal approach outlined above – which treats ABE and NFE girls who remain enrolled in ABE or NFE programmes as successful cases, as well as considering a 2-year progression standard for FE girls – results in a rate of 72.7 percent. The gap in results between these two coding schemes are broadly shared – it results in an increase of between 10.1 and 12.7 percentage points in each of the three cohorts.

As the table above made clear, there are many transition pathways relevant to the three cohorts studied here. Beyond dichotomous "success" versus "non-success" results, the tables below document the rate of different transition outcomes at ML2. The first table reports on FE girls. The two sets of results are similar – the distinction is in the number of years of grade progression required for an FE girl who remains enrolled in school to be considered a case of successful transition.

The first two columns report results using our more liberal definition, in which two years of grade progression are considered sufficient progression since baseline. In this version, the most common outcome among FE girls is continuing enrolment, with grade progression since BL in line with expectations. Among FE girls, 69.3 percent remain enrolled in school and have progressed in grade level; dropping out, in this version, is more common than being held back, with just 6.3 percent of girls simultaneously remaining enrolled but failing to be promoted consistently.

TABLE 25: TRANSITION OUTCOMES AMONG FE GIRLS, USING ALTERNATIVE TRANSITION DEFINITIONS

Transition Outcome	Transition Defined using Liberal Standard		Transition Defined using Conservative Standard	
	Number of Girls	Share of Sample	Number of Girls	Share of Sample
Positive grade progression	187	69.3%	164	60.7%
Now OOS	55	20.4%	55	20.4%
Enrolled, Held Back	17	6.3%	40	14.8%

[37] This structure is slightly more complicated by the fact that replacement girls were recruited in ML1 and we lack information about them at baseline. However, we do know that they were enrolled in formal school in either grade 1 or 2 at that time, so it is possible to reconstruct their transition pathway between BL and ML2 to a broadly accurate degree.

Now employed	5	1.9%	5	1.9%
underage employ	3	1.1%	3	1.1%
Self-employed	2	0.7%	2	0.7%
Now in NFE	1	0.4%	1	0.4%

The right-most two columns switch to our stricter standard of defining successful transition. In this version, fewer girls successfully meet the grade progression standard. It is still, by far, the most common outcome for FE girls, while dropping out (and not having gainful employment) is the second most common outcome, at 20.4 percent of the sample.

Few FE girls are engaged in employment of any kind. This is not entirely surprising, because FE girls remain young at ML2. The more important fact, though, is that girls remain enrolled at very high rates – ignoring whether a girl was promoted across grades, 75.5 percent of girls who were enrolled into FE in 2019 remain enrolled. While it is difficult to judge precisely how impressive this outcome is, in the absence of a comparison group or national statistics on enrolment consistency across years, it appears impressive, given the cohort of girls in question and the context of the COVID-19 pandemic. It is important to remember that FE girls were enrolled *as part* of the project and were not already engaged in schooling. The fact that the programme brought FE girls into primary school and so many remain enrolled 3+ years later, despite COVID-related disruptions, is suggestive evidence of the programme's impact on enrolment outcomes.

The table below provides the equivalent results for ABE (top panel) and C1 NFE (bottom panel) girls. The first two columns report on the "liberal" definition of transition, which girls who remain enrolled in ABE/NFE learning programmes are classified as successful cases of transition. The most common outcome among ABE girls, at 28.7 percent of the sample at ML2, is to remain enrolled in ABE. However, the next two most common outcomes – being out-of-school or idle, and being employed – follow behind closely, with 27.1 and 24.5 percent of the sample, respectively. As we reported at ML1, a significant share of ABE girls has also transitioned into formal education, though this outcome is not as common as one might expect, given that facilitating re-entry into formal education is an explicit goal of the ABE intervention.

The impact of our coding decisions can be seen in the right-most columns, which ignores continued enrolment in ABE as a pathway, classifying girls only on whether they transitioned into formal school, are employed/self-employed, or are out-of-school without employment. In this version, "idleness" is the dominant outcome, with 46.5 percent of ABE girls now unemployed, without being enrolled in any type of established learning programme.[38] At the same time, a significant number of ABE girls are engaged in gainful, age-appropriate (18 years or older) employment.

TABLE 26: TRANSITION OUTCOMES AMONG ABE AND NFE GIRLS

Transition Outcome	Transition Defined using Liberal Standard		Transition Defined using Conservative Standard	
	Number of Girls	Share of Sample	Number of Girls	Share of Sample
ABE Girls (n = 310)				

[38] We use the term "idle" although it is far from accurate, as out-of-school girls tend to perform significant household labour and may be engaged in other activities. We use this term simply to indicate those girls who are not in school and are simultaneously not engaged in paid labour of some kind.

Continued enrolment in ABE	89	28.7%	N/A	N/A
Out-of-School, Idle	84	27.1%	144	46.5%
Employed	76	24.5%	103	33.2%
Enrolled in formal school	57	18.4%	57	18.4%
Self-employed	2	0.6%	4	1.3%
Employed, but underage (under 18 years)	2	0.6%	2	0.6%
NFE Girls (n = 302)				
Employed	156	51.7%	180	59.6%
Out-of-School, Idle	78	25.8%	95	31.5%
Continued enrolment in ABE or NFE	43	14.2%	N/A	N/A
Enrolled in formal school	18	6.0%	18	6.0%
Self-employed	6	2.0%	8	2.6%
Employed, but underage (under 18 years)	1	0.3%	1	0.3%

Employment rates are even higher among the NFE cohort, as shown in the bottom panel of the table. Of all C1 NFE girls, nearly 60 percent are employed, and a further 2.6 percent are self-employed (operating their own business), while 6.0 percent have shifted into formal schooling. Relatively few NFE girls – 31.5 percent – are idle.

The relatively high employment rates among ABE and – especially – NFE programme participants suggests that the programme has been effective in helping girls find jobs or providing them the skills necessary to engage in productive labour. It is important to note that ML2 transition outcomes represent a single snapshot in time, however, girls enter and exit employment, as they do with schooling. For instance, out of 78 ABE or NFE girls who were employed at the time of the ML1 evaluation in early 2022, 49 of them (62.8 percent) remain employed at ML2. At the same time, positive labour market outcomes may take time to materialise; this is reflected in the fact that, out of 156 women now employed at ML1, the majority (53.9 percent) were unemployed and out-of-school (i.e., "idle") at ML1. This outcome suggests the need for patience when analysing programme outcomes like employment; a woman may have gained valuable skills but may be unemployed because she has not found an appropriate position or because she is temporarily between positions. This also suggests that the endline evaluation should carefully document girls' employment status

across the years since the baseline, to ensure it captures a complete picture of women's employment outcomes, including consistency of employment and rates of "churn."[39]

Although the lack of a comparison group tempers the strength of our conclusions, the project produced the expected outcomes among FE and NFE girls. The outstanding question is whether these outcomes – consistent and continued enrolment and grade promotion for most FE girls 3 years on and gainful employment for a majority of NFE girls – can be attributed to the programme itself or would have obtained in the absence of the programme. In either case, the pathways we expected to observe for ABE girls were less commonly achieved, with many ABE girls either remaining – nominally or otherwise – enrolled in ABE learning programmes or remaining idle, without employment and without pursuing further education. Transitioning from ABE into formal schooling was arguably the most ambitious transition pathway possible, given the barriers to formal school attendance, and this is confirmed by the relatively small share of girls who followed this pathway.

5.2 Subgroup Transition Rates

In this section, we analyse subgroup-specific transition outcomes. Given the different barriers girls face across geographic space, as a function of their household's economic position, and their own characteristics, we would expect variation in transition outcomes as a function of these characteristics and challenges. Due to the small sample size of each individual cohort, we report subgroup results that aggregate across the FE, ABE, and C1 NFE cohorts. However, we also disaggregate the subgroups into cohort-specific effects, to study whether any subgroup-specific findings are shared across subgroups or concentrated among only one cohort.

We generally define subgroups as a function of characteristics at baseline. For many outcomes this is non-controversial. In other cases, this is not true, however. Consider disability status, which can and does change over time. To define a girl's disability status or her household's economic circumstances, we use the information collected at baseline. This is necessary to keep the subgroup of a fixed size and composition over time, to ensure that our results are not a function of changing subgroup composition over time. Our interest is in how the programme impacted different subgroups; in some cases, the programme may have contributed to shifting a girl from one group to another (e.g., by supporting her household economically, reducing their economic vulnerability), and our interest is in capturing this mechanism of impact as part of the subgroup-specific effect of the programme.

In the table below, we report subgroup-specific transition outcomes, with panels of the table organised according to topical area. For each subgroup, we report the aggregate sample size of the subgroup, its overall transition rate, and the gap or difference between transition rates in the subgroup and transition rates in the remainder of the sample. This latter column labelled "subgroup difference," can be interpreted as a regression coefficient capturing the relationship between membership in the subgroup and transition rates. Subgroup differences marked with an asterisk are statistically significant at the 0.05 level, after accounting for school-level clustering.

[39] This is similar to a deliberate decision during the endline evaluation of SOMGEP-T. Recognising that there were some girls who would be interviewed at endline who had *not* been interviewed in both the interim midline rounds, the team was concerned about our ability to systematically track transition outcomes across time. To supplement transition information collected during each round, the team retrospectively asked respondents about their enrolment status and grade level for each of the previous five years. This allowed analysis of the consistency of enrolment and year-on-year grade progression, providing depth to discussion of "varied pathways" over time.

The latter three columns disaggregate transition rates, within each subgroup, by cohort. Note that we do not report p-values or statistical significance for these results, as the sample sizes are generally too small to be reliable.

The first panel captures geographic differences in transition rates and presents some of the starkest subgroup-specific results. Girls in Banadir have systematically worse transition outcomes, while those in Jubaland outperform the sample average by 15.4 points. This outcome is somewhat surprising, given that we would expect the labour market to be most active in Banadir. However, Banadir's underperformance is reflected across all three cohorts – though it is particularly concentrated among ABE and NFE girls – suggesting that it is not simply labour market considerations that are driving lower transition rates.

Beyond geography, girls with disabilities do not have dramatically different transition outcomes from the broader sample. In fact, across all types of impairment, these girls marginally outperform the sample average, though none of the differences reported are statistically significant. For instance, girls that report experiencing anxiety or depression have transition rates 4.7 points higher than the sample average, and girls with a physical disability also transition at higher rates than the sample average.

TABLE 27: SUBGROUP-SPECIFIC TRANSITION RATES, IN AGGREGATE AND AMONG COHORTS

				Cohort-Specific Transition Rates		
Subgroup	Sample Size	Aggregate Transition Rates	Subgroup Difference	FE Cohort	ABE Cohort	NFE Cohort
Overall	882	61.3	N/A	63.3	52.9	68.2
Geography						
Banadir	394	54.3	-12.7*	59.8	43.3	60.1
Jubaland	253	72.3	15.4*	77.9	59.8	78.9
Southwest State	235	61.3	-0.1	46.6	60.6	72.3
Disability Status						
Any physical disability	66	68.2	7.4	73.7	50.0	85.7
Any physical disability, alt. coding	72	65.3	4.3	66.7	48.1	85.7
Any cognitive disability	81	67.9	7.2	66.7	65.5	71.0
Any cognitive disability, alt. coding	96	60.4	-1.0	54.8	55.9	71.0
Any mental health disability	317	64.4	4.7	62.3	56.3	72.3
Any mental health disability, alt. coding	396	60.1	-2.2	61.5	47.5	71.7
Any non-mental health disability	132	67.4	7.2	69.4	54.0	80.4
Any non-mental health disability, alt. coding	148	62.8	1.8	59.2	50.9	80.4
Any disability	383	63.4	3.7	61.9	53.1	73.7
Any disability, alt. coding	456	60.1	-2.6	60.6	46.9	73.2

Parental Educational Attainment						
HoH has no education of any kind (no Quranic)	191	61.3	-0.1	55.7	57.8	71.9
HoH has no formal education	724	62.2	4.6	62.5	55.8	68.4
Caregiver has no education of any kind (no Quranic)	249	61.8	0.7	61.8	54.1	70.7
Caregiver has no formal education	749	61.9	4.1	62.8	55.3	68.1
Household Economic Characteristics						
HoH does not earn a regular wage	326	60.4	-1.4	64.7	48.5	66.7
HH has a poor-quality roof	174	60.3	-1.2	56.1	48.6	77.0
HH went to sleep hungry most nights, last 12 months	52	46.2	-16.1	47.1	50.0	40.0
HH went without clean water most days, last 12 months	47	51.1	-10.9	37.5	47.4	75.0
HH went without medicine most days, last 12 months	142	50.7	-12.7*	48.7	37.0	63.2
HH went without cash income most days, last 12 months	129	51.2	-11.9	59.5	42.4	57.6
HH owns lands	298	65.8	6.7	65.5	58.0	75.0
Caregiver has savings of some form	16	31.3	-30.6*	33.3	0.0	66.7
HH owns a phone	756	61.4	0.3	63.3	53.4	67.8
HH owns a smartphone	161	62.7	1.7	64.3	62.2	61.7
Parental Circumstances						
Girl has only one living parent	110	70.9	10.9*	78.6	55.2	75.5
Girl has no living parents	9	33.3	-28.3	0.0	25.0	50.0
Girl does not live with either parent in her HH	44	63.6	2.4	57.1	46.7	77.3
Female-headed household	330	66.7	8.5*	73.0	54.8	71.5
Linguistic Status						
Household speaks af-Maay	256	63.3	2.7	44.4	66.7	75.3

School Facilities and Characteristics						
Girl will not use drinking facilities at school	214	67.8	8.5*	75.6	55.1	74.7
Girl will not use toilet facilities at school	157	68.2	8.3	77.8	50.0	73.3
Girl cannot use books/learning materials at school	247	65.2	5.3	62.8	56.2	75.0
Girls are able to take textbooks/materials home at night	223	62.8	-8.0	62.8	N/A	N/A
Not enough seats for every student in class	135	68.9	8.9*	73.9	N/A	N/A
School has reliable electricity	203	60.6	-14.8*	60.6	N/A	N/A
School has water access within 1 km	128	60.2	-7.5	60.2	N/A	N/A
School provides at least one meal for students	123	71.5	14.1	71.5	N/A	N/A
School has only cement floors (no dirt)	207	64.3	1.1	64.3	N/A	N/A
School has separate toilets for girls	251	64.5	10.7	64.5	N/A	N/A
No female teachers, either FT or PT	7	57.1	-7.1	57.1	N/A	N/A
Short instructional time per day	32	53.1	-12.4	53.1	N/A	N/A
Agree: my teachers are often absent	241	59.8	-2.2	57.3	N/A	N/A
Textbooks are shared between students	176	67.0	9.1	67.0	N/A	N/A
School charges school fees	129	55.0	-17.6*	55.0	N/A	N/A

The educational attainment of the adult members of a girl's household – her head of household and caregiver – do not have large effects on her propensity to remain in school or obtain employment. This may seem surprising, but it likely reflects the mixed nature of the transition outcome. Even if the educational attainment of one's parents predicts more consistent enrolment or greater educational attainment, our transition outcome is only partially driven by enrolment rates. Employment is at least as large a factor in determining overall transition rates, and it is less clear that greater educational attainment of one's parents would directly and significantly impact employment prospects for a young woman.

Outside of geographic disparities, perhaps the most compelling subgroup-specific finding concerns household economic circumstances. Part of the programme's built-in Theory of Change was the idea that economically marginalised girls would be less likely to remain enrolled in learning programmes. This appears to be borne out by the data, with girls from households that have experienced frequent hunger or other forms of short- and medium-term deprivation (e.g., going without medicine, without cash income, and without access to clean water for household use) all predicting substantially lower transition rates. It is possible that economic circumstances are so closely correlated across geographic space that these results are driven entirely by regional disparities in transition rates. However, we investigated this possibility further, finding that household economic deprivation predicts lower transition rates even when controlling for region or district of residence.

5.3 Testing the Theory of Change

Moving beyond the disaggregation of transition outcomes across subgroups in the previous section, we now subject the AGES Theory of Change, as it pertains to transition, to a tentative test. According to the theory of change, the outputs of the programme are hypothesized to have a positive impact on transition outcomes via a series of intermediate outcomes. Based on this, we can hypothesise that transition outcomes should be correlated with self-esteem and leadership skills of the girls, the relative support for education expressed by their caregivers, and their participation in Girls' Empowerment Forums (GEFs). The latter is not an intermediate outcome, but is the programme output with, arguably, the strongest theoretical links to transition rates, due to the GEFs' expected effects on girls' confidence and peer networks.

We execute these tests in a regression framework, focusing on transition outcomes in the current evaluation round, ML2, while setting aside the transition status girls had achieved at ML1. As was true in our analysis of aggregate transition outcomes, success is defined partially by the learning track in which the girl was enrolled at baseline, with different standards of transition applied to FE and NFE girls, for example. We estimate a series of linear regression models "predicting," or explaining variation in, our binary measures of transition. In each case, we control for region, age of the girl, and learning track or girl type. We cluster standard errors by school code.

The core results of our regressions are summarised in the table below. Each row represents a single multivariate regression explaining variation in transition rates as a function of one intermediate outcome or output (listed in the left column) and a variety of control variables. We report only the regression coefficient for the intermediate outcome of interest, and do not include results for the ancillary control variables. Results that are statistically significant at the .05 and .01 level are denoted by one or two asterisks, respectively.

Starting with caregiver attitudes, we see no impact of positive caregiver attitudes on transition outcomes. The first metric of caregiver attitudes consists of a set of five questions caregivers were asked at baseline; caregivers indicated under which circumstances they felt it was acceptable for a girl to not attend school. These circumstances included the need for the girl's help with housework, school being too costly, and the girl having reached an age where she is too old for school. We construct a simple index aggregating these questions into a 0-5 scale, with higher scores representing more positive attitudes – a caregiver with a score of 5 did not feel it was acceptable for a girl to be kept out of school under any of the five circumstances.

More positive attitudes are not associated with higher transition rates. Indeed, in at least one model, they are associated with a statistically significant decline in successful transition outcomes. This outcome is difficult to explain, though it is worth noting that the older age of NFE girls, especially, may reduce the importance of caregiver attitudes on schooling and employment decisions. Caregiver attitudes may have more important

impacts among younger girls, though even this is not supported by subsequent analysis, not reported in the table. We also do not find any systematic relationship between successful transition outcomes and a girl's household chore burden during the baseline round of data collection. In other words, girls who are responsible for fewer chores at home – a proxy, in one sense, for the importance adults in the household place on the girls' education – do not have better transition outcomes at ML2.

TABLE 28: RELATIONSHIP BETWEEN KEY INTERMEDIATE OUTCOMES AND TRANSITION RATES AT ML2

Independent Variable of Interest	Liberal Coding of Transition Outcome	Strict Coding of Transition Outcome
Positive caregiver attitudes toward girls' education at BL	-12.0 points*	-3.5 points
Household chore burden at BL	-0.7 points	+0.7 points
YLI score at BL	+0.03 points	+0.06 points
GEF Participation (binary indicator)	+10.6 points**	+9.5 points**
GEF Participation (ordinal 0-3 score)	+6.2 points** (per 1-unit increase)	+5.0 points* (per 1-unit increase)

One possible explanation for the lack of correlation between more positive caregiver attitudes and positive transition outcomes relates to the importance of exogenous factors in shaping transition at ML2. Consider employment as an outcome: finding a job is not strictly, or even mostly, a function of a girl's effort level. Whereas enrolment in school is almost entirely under the control of a girl and her family, who must make decisions about her time use and whether to spend money on school fees, employment requires matching to an available job, and an employer finding her to be the best candidate for the job. These factors, external to a girl or her family and almost entirely outside her control, introduce "noise" into the outcome, making a successful outcome a function, in part, of luck or randomness.

The second intermediate outcome we study is a girls' self-esteem and leadership skills, as measured by the Youth Leadership Index (YLI). In some ways, we might expect self-efficacy – discussed in Section 7.6 and measured using the Chinese Positive Youth Development Scale (CPYDS) – to be a better predictor of transition outcomes, as it focuses more directly on self-esteem, and less on leadership skills. However, the CPYDS was only introduced to the evaluation at ML1, after most members of the original baseline cohorts had extensive exposure to the programme's interventions (i.e., our measure would be post-treatment).

Our regressions analysing the importance of YLI scores show ambiguous links to transition outcomes. A 10-point increase in YLI scores at baseline is associated with either a 0.3 or 0.6 percentage point increase in the likelihood of successful transition at ML2. Neither of these results approach statistical significance and are of extremely small magnitudes. In additional regressions not reported here, we also analysed the impact of a positive *change* in YLI between BL and ML2 on transition outcomes, theorizing that strong gains in self-esteem could drive positive transition outcomes. However, there is no evidence for such a link using our approach.

Finally, we turn to participation in Girls' Empowerment Forums or girls' clubs. We define a binary variable indicating whether a girl ever reported participating in a GEF; we also define an ordinal variable that captures

increasing levels of participation on a 0-3 scale.[40] We report regressions using these two respective independent variables in the bottom rows of the table above. Using the binary measure, 37.4 percent of the sample (or 330 out of 882 girls) had participated in a GEF in the past. This is an important distinguishing factor in shaping transition outcomes, as shown in our results – girls who participate in a GEF in some forms are either 9.5 or 10.6 percentage points more likely to have a successful transition outcome, depending on how the definition of transition we employ in the model. This positive impact is scaled by the *level* or *depth* of GEF participation – girls scoring 2 or 3 on our proxy measure for GEF participation are somewhat more likely to have a successful transition than girls with a score of 1; both groups are significantly more likely to transition successfully than girls who do not report *any* GEF participation.

The finding regarding GEF participation is extremely encouraging, especially given the magnitude of the effect size. GEF participation is associated with a major increase in transition rates, equivalent to large differences across states reported in the previous subsection. However, the findings should be interpreted with caution – girls self-select into GEF participation, and GEF participants may have greater underlying motivation or other latent characteristics that would produce better transition outcomes, even in the absence of GEF participation. Disentangling underlying motivation from the causal impact of GEFs would require a tailored research design that controls selection into and participation in GEFs; however, it is clear from this analysis, the relationship between GEF participation and learning outcomes reported above and in the ML1 evaluation report, and similar analysis conducted in the context of the SOMGEP-T programme, that GEF participation is associated with outsized, positive impacts on girls.

6. Sustainability

This section presents a narrative analysis of AGES programme sustainability divided into subsections, one for the each of the seven key sustainability indicators identified in the AGES logframe.

The analysis was carried out based on all data sources available, including quantitative data (from household surveys and school survey) and qualitative data (from FGD responses with CECs, teachers and mothers). The resulting triangulated analysis was used to generate one qualitative sustainability score (ranging from 0 to 4) for each indicator.

Sustainability scores for each indicator can range from 0 to 4, as follows:

- 0 – Negligible
- 1 – Latent
- 2 – Emerging
- 3 – Becoming established
- 4 – Established

While the rationale for each indicator's score is provided in the detailed narrative sections that follow, it is important to note that the scores are admittedly subjective. We assigned scores based on the totality of

[40] A girl receives one point on this scale if she reports having ever participated in a GEF in either the ML1 or ML2 round. She receives an additional point if she reports that she continues to be in contact with the GEF at ML1, and an additional point for the same outcome at ML2. Thus, a girl with a score of 3 on this metric self-reported participation in a GEF and reported being in contact with the GEF in both the ML1 and ML2 rounds. This is a proxy indicator for the depth of engagement a girl has with a GEF.

information available – in some cases, this was limited to a few qualitative interviews, while in other cases it included quantitative data from multiple respondent groups and a range of qualitative data. As a broad rule of thumb, where noticeable but not dramatic changes have been observed since the baseline, we tended to score indicators as one grade above the baseline.

Self-replication rate of village savings and loans (VSL) groups

The AGES programme focused, among other interventions, on the strengthening of the economic situation of local communities to increase family support for girls' enrolment, attendance, and retention. As part of these economic interventions, AGES established and supported Village Savings and Loans Associations (VSLAs) for girls' caregivers and girls enrolled in NFE. VSLAs are community self-managed groups that, especially in rural areas and amongst poor households, often provide the only opportunity for families to save money and obtain loans. The borrowed money can be used to finance new businesses or large household expenses, including expenses related to children's education such as school fees or school materials.

Informal village-level savings groups are relatively widespread in Somalia, however even in this favourable cultural context, there are concerns regarding the sustainability of the established VSLAs after the conclusion of the programme. This concern is particularly relevant considering that in March 2023, only 7.9 percent of caregivers interviewed reported being part of a VSLA. Using a comparable sample of respondents, we can observe an increase in 5 percentage points in caregivers' participation in VSLA from BL to ML2, pointing to a positive albeit small effect of the interventions.

The low rate of participation to VSLA amongst caregivers is concerning, considering that only 4.2 percent households that are not part of a VSLA have separate savings. VSLA not only provide with the opportunity to save and borrow money, but also support household resilience to adverse economic shocks. In case of future reduction in income, loss of livelihood or large expenses, these families will likely not be able to support an increased financial burden, possibly negatively affecting the investment in girls' education and affecting the sustainability of girls' educational outcomes.

The creation and support to VSLA was also part of the programme for girls that are enrolled in the first cohort of NFE. Relative to the caregivers of ABE and FE girls, the C1 NFE girls' participation to VSLA is significantly higher, with 26.9 percent of them reporting being part of a VSLA. While this data was not collected at BL, girls newly recruited in the programme are unlikely to be enrolled in other savings group at an early age, so that the current participation rate calls attention to the positive effect of the programme. Also, for C1 NFE girls, the VSLA is one of the few mechanisms to save money, as only 5% of them report having a bank account.

These trends are relatively stable, with minor differences due in part to a small sample size and to resampling of some respondents, across ML1 and ML2, suggesting that, while the programme undoubtedly had an initial positive impact on VSLA participation, it hasn't been able to capitalize on it to increase further the participation rate. At the same time this stable trend might also indicate that the current groups are sufficiently established to survive after the end of the programme.

When asked about whether their VSLA was still active, a different trend emerged between caregivers and C1 NFE girls. Almost three caregivers out of four (83 percent) who are part of a VSLA reported it still being active – however this number should be only considered as indicative of a general trend as the overall participation is particularly low. On the other hand, only one in three C1 NFE girls (34.6 percent) report their group being still active. Both figures remain relatively consistent with respect to those of ML1 (respectively 83.1 percent and 36.4 percent) and confirm the indication of stability of the intervention evidenced in the overall rate of participation to VSLA.

The willingness to be part of a VSLA group and its longer-term sustainability depends not only on it having an active membership, but also on effective leadership and mechanism for saving and loan disbursement. While there is no information on loan rotation, savings amount, membership count and other such indicators are available to gain better understanding on the efficacy of the VSLA for programme participants, the stability of participation rate and active membership between ML1 and ML2 might support the claim that the few currently active group will continue their operations beyond the end of the programme.

While we cannot link VSLA participation to educational outcomes causally, there is suggestive evidence that VSLAs are beneficial to girls' education beneficial to girls' education. Specifically, 69 percent of the caregivers that are part of a VSLA used the savings from the VSLA on their daughters', their sons' education, or both. On the other hand, a FGD respondent reported that a girl part of the AGES programme, after finding a job, started saving her income to in turn support her brother education[41].

Sustainability Indicator Score: 1.5 (Latent/Emerging)

Proportion of parents able to support costs of girls' education

In addition to the economic interventions mentioned in the previous section on VSLA, the AGES project also focused on changing the attitudes and practices of parents, caregivers, and community at large towards supporting girls' education. As part of this effort, a significant indicator of sustainability of the AGES programme is the proportion of parents and caregivers whose financial situation allows them to support the costs of girls' education. The participation to VSLA contributes to this indicator as well as the disbursement of cash support, bursaries, households' livelihood, and income generating activities, and general economic context.

The positive change in caregivers' attitudes towards girls' education that was observed from BL to ML1 persists in the findings at ML2: costs of education and families' financial difficulties are progressively not considered sufficient reasons to withdraw a girl from school or to avoid providing her with basic education. Looking at a comparable sample between BL and ML2, the proportion of caregivers that believe it's acceptable for a girl not to attend school due to "education being too costly" has fallen from 62.9 percent at BL to 44.6 percent at ML2, confirming the result of ML1. The result of household economic scarcity might be observed indirectly through girls' lack of school attendance because of having to work or support the family to allow other family members to be engaged in income-generating activities. Also in this case, the attitudes of principal caregivers improved from BL to ML2: respondents that believe it's acceptable to withdraw a girl from school due to "having to work" decreased from 25.2 percent at BL to 16.3 percent at ML2. The attitude of girls' themselves suggest that, while likely marginally less concerned with household finances than their caretakers, AGES programme participants hold strong beliefs regarding the importance of education. The almost totality (97.2 percent) of girls interviewed agrees or strongly agrees that "even when funds are limited it is worth investing in your education".

This positive attitude translated also in positive practices towards supporting girls' enrolment and attendance so that, while dropout from schools still occurred, the proportion of dropout motivated by financial concerns is limited. Considering girls (ABE, FE, C1 NFE) that had to drop out of school, or that moved from FE to ABE or NFE, after their participation in the AGES programme, in only 6.7 percent of the cases one of the motivations was the need to work or earn money. For 5 percent of the girls, one of the reasons was indirectly

[41] See FDG with mothers, Banadir, Int. 203.

related to reducing the financial burden on their family (having to help at home, in the farm or with the livestock). There has been no significant change in between ML1 and ML2.

On the other hand, prior to enrolment in the AGES programme, a much higher proportion of caretakers reported economic challenges as a factor driving the previous lack of enrolment of girls in school (the same is true for girl having to work as reason). At BL 56.6 percent of caretakers reported that their girl was previously not attending school because the family didn't have enough money to cover the cost of education. On the other hand, a lower but still significant proportion of caretakers (22.6 percent) reported the girl having to work, earn money or help at home as a reason for her lack of attendance.

While this data is not directly comparable between BL and ML2, it however suggests that the combination of interventions in the AGES programme, including those related to economic support and change in social norms and awareness raising, have achieved a beneficial effect in promoting girls' enrolment and maintaining it throughout the programme duration even when households are facing adverse economic circumstances.

The project paid for the salary of one teacher per school in exchange for free seats in formal education for participants. This indirect financial support – which essentially subsidised enrolment for participants – ended in early 2022, but had a clear benefit. Specifically, it appears to have increased the likelihood a girl would enrol in school, reported by 94.1 percent of the caregivers, as well as the likelihood they would attend school consistently, reported by 93.2 percent of all caregivers. At the same time, the overall financial condition of households that are part of the programme, while still very poor, has improved since the BL, as denoted by the increase in households that never went to bed hungry in the past 12 months (17.4 percent at ML2, compared with 21 at BL), although there has been a slight increase of 1.5 percentage point from the situation at ML1.

To the extent that AGES support for teacher salaries, in exchange for free enrolment for programme participants, promoted enrolment and attendance, these gains may be partially lost when AGES support ends. The overall improved financial conditions relative to the pre-programme period, and the change in attitudes confirmed from BL to ML1 and ML2, might point to a sustained improvement in willingness and ability of families to keep the girls in school.

Sustainability Indicator Score: 2.5 (Emerging/Becoming Established)

Parental support for girls' and boys' participation in GEFs and BEFs

The community-level uptake of AGES interventions and ideals, and their maintenance beyond the end of the programme is also strengthened through community-level networks. Among the networks that play an important role in ensuring the sustainability of achieved results are the GEFs, platforms for social change that bring girls together and empower them to take active roles in their community. GEFs carry out awareness raising activities focused on topics and causes important to women, including financial empowerment, education, and health.

To assess their long-term viability, we asked AGES girls whether they were GEF members and, if so, whether they were still connected with fellow members and what types of activities they have participated in under the GEF umbrella. While only one in four girls reported being part of a GEF (25.3 percent), in most cases these groups are still active, with 53.6 percent of girls part of a GEF reporting that they are still in touch with other members. While maintaining contact with other members presents a lower bar than full engagement

(with meetings and official status), even this form of network maintenance is critical for girls' social support.[42] In general, compared to the situation at ML1, there has been a slight improvement in GEF participation rate (6.5 percentage points more), however we observed a slight reduction in the continued connection with fellow members (4.5 percentage points less)[43]. This increased participation vis-à-vis a decrease in level of engagement might be indicative of a natural evolution of GEFs during the course of the programme implementation, as new GEF groups are being established and promoted, while there is a "natural" decadence rate of older groups. Additionally, as the ABE and C1 NFE programme has concluded in previous years and GEFs were established as a component of these programmes, connections to other GEF members and participation in the network itself might also contextually decrease. This is especially likely to be the case for GEFs being established in urban areas, where the catchment area for the school – and therefore the GEFs – might include girls from relatively far neighbourhoods who, after the end of the programme, have limited incentives to get back to the previous location. This might therefore point to a limited sustainability of educational attainments past the end of the AGES programme.

As a core number of GEFs however remain active (and possibly increase their number of participants), there might be signs of hope that the GEFs with higher member engagement will prolong their existence in the longer term. At the same time, GEFs are only one possible platform for girls. We asked C1 NFE girls if they were part of other youth networks and 28 percent responded affirmatively; this included many girls (68.2 percent) who are not currently part of a GEF. Thus, it appears girls have other groups or opportunities for networking and generating peer support; while it is possible this could undermine the long-term sustainability of GEFs, if these groups serve broadly similar purposes, this should not be considered a negative outcome. Even if negative, we can only speculate regarding the relationship between and impact of external groups on GEFs, other than to note that GEFs are not alone in providing opportunities for Somali girls.

TABLE 29: PARTICIPATION RATES AND ENGAGEMENT IN GEFS, BY COHORT

Indicator	Total	FE girls	ABE girls	C1 NFE girls
Participation Rate	25.6%	21%	31%	25.5%
Percent Still in Contact with GEF Members	53.6%	58.9%	56.3%	44.2%

The various activities implemented by GEFs cover areas related to business and finances, changing community perceptions regarding topics of interest and promoting girls' access to education. While the proportion of GEFs with active engagement has only decreased slightly with respect to ML1, in the past year the number of activities carried out by GEF with active engagement has decreased significantly, with each GEF carrying out a smaller range of activities (see details in the section below on "Proportion of GEFs and BEFs implementing community actions to support attendance and retention").

When we interviewed CECs, teachers and mothers regarding the support provided to girls' education by various networks and organisations, none of the respondents mentioned GEFs, which might indicate that,

[42] Note, also, our data does not allow us to assess the depth of engagement in GEFs beyond asking girls whether they participate and whether they remain in contact with girls from their GEF.

[43] The percentage points are calculated on a reduced sample that considers only respondents that were interviewed both at ML1 and ML2

although there are some active GEFs, their impact and therefore the support for their continuation after the end of the programme among caregivers and other adults is limited.

Sustainability Indicator Score: 1.5 (Latent/Emerging)

Proportion of GWDs who remain in school

The AGES programme supports the availability of quality learning opportunity tailored to the needs of vulnerable and marginalised girls by targeting its interventions to the needs of girls with disabilities (GWD), who face additional challenges in enrolment and participation in schools. Additionally, to the barriers to education experienced by other girls, GWD education is further adversely affected by lack of assistive equipment for girls, adapted teaching materials, inadequate physical infrastructure in school and to reach schools. This is often also compounded by generalised stigma regarding disabilities, which however has been the object of continued efforts by teachers and seem to have achieved positive results in schools targeted by the programme.

Overall, in the fourth year of programme implementation the enrolment of GWD remains slightly lower than that of girls without disabilities (43.2 percent compared to 57.4 percent). While the proportion of GWD currently enrolled appears to be slightly lower than last year, we need to interpret this result in view of the significantly higher absolute number of GWD enrolled (146 at ML2 compared to only 91 at ML1) which points to a positive impact of the programme interventions on promoting schooling for GWD.

For GWD attending education programs, the distribution across types follows the same pattern of girls without disabilities, with the majority participating in formal schooling (65.6 percent), followed by ABE (19.6 percent) and lastly NFE (12.2 percent)[44]. However, while the enrolment of GWD in FE is lower than that of girls without disabilities (78.3 percent), the proportion of GWD enrolled in NFE is three times higher than that of girls without disabilities (only 4.6 percent of all girls without disabilities enrolled in school). While the number is too small to be statistically significant, it might point to the importance for GWD and their families to provide these girls with skills that can be directly applied to find a job or generate income, as disabilities pose an additional financial burden on these households. In line with this possibility, we find that the retention rate of GWD in NFE, although low at 15.6 percent, is slightly higher than for girls without disabilities, while the opposite is true for retention rates of girls enrolled in FE or ABE, pointing again to the relative importance of non-formal education for GWD. These findings however need to also be seen considering the significant mobility of girls targeted by the programs, including GWD, across education types, as overall one in four GWD was enrolled in a different type of education programme at the start of AGES.

TABLE 30: ENROLMENT RATES OF GWDS, BY DISABILITY TYPE[45]

Disability Type	% Enrolled
Difficulty seeing, even if wearing glasses	60%
Difficulty hearing, even if using a hearing aid	54.5%
Difficulty walking or climbing steps	47.8%

[44] Enrolment in other forms of schooling has not been included in the list.

[45] The following disability types follow the WG definition, note that the same GWD might have multiple disabilities and is therefore counted for all reported disabilities.

Difficulty with self-care such as washing all over or dressing	50%
Difficulty communicating (e.g. understanding or being understood)	52.6%
Difficulty learning or remembering things and concentrating on activities that the child enjoys doing	25%

The enrolment of girls with physical disabilities is hampered by the general economic situation of most household in programme areas that limits the possibility for these girls to access needed assistive devices and pay for transport to and from school. Additionally, the limited funding available for CECs and schools also prevents improvement of school facilities, so that most teachers interviewed as part of the FGDs reported inadequate school facilities and lack of money as some key barriers preventing the inclusion of GWD in schools[46]. While collaboration with and referral to other NGOs might help in lowering the financial and logistic barriers to GWD attendance, their impact is still limited. This is evidenced by the fact that we observe the highest enrolment for girls with visual and hearing disabilities, respectively 60% and 54.5%.

Reports by teachers and CECs of schools covered by the AGES programme often report supporting attendance and quality learning of GWD by adapting the seating arrangement in classes according to disability types, for girls with visual and mobility disabilities[47].

At the same time, while less clearly identified, the challenges faced by girls with mental health disabilities are also significant, as evidenced by the fact that only 39.7 percent of girls with mental health disabilities are enrolled in school compared to 57.8% of girls with other disabilities. The same is true looking specifically at girls with cognitive disability (see table above), for whom the enrolment rate is the lowest amongst other disability types at 25%. While the majority of GWD that are part of the AGES programme have mental disabilities (355 in the ML2 sample) against a much smaller sample of 83 without mental health disabilities, in Africa in general, children with mental health or cognitive disabilities are still disproportionately stigmatised. Until recently, support towards children with disabilities has focused on more "visible" physical disabilities, while interventions that focused on girls with mental health and cognitive disabilities were relatively limited, also because supporting education of those girls might require more specialised training of teachers and provision of adapted learning materials. The persistence of stigma for mental health disabilities, even among teachers that are part of the AGES programme is clearly highlighted by a CEC member that referred to them as "sick girls who are possessed by demons when they are in school"[48].

Awareness raising activities carried out across-the-board and teacher trainings on inclusive teaching methodologies and classroom management (see section below "Proportion of teachers implementing inclusive education strategies in class") have also positively impacted the overall classroom environment. Attitudes by peers is also an important factor for GWD to feel welcome and accepted, therefore reported improvements in social attitudes of other children towards children with disability in schools and prompt address of GWD concerns by teachers also indicate positive developments for the enrolment and attendance of GWD[49].

Sustainability Indicator Score: 2.5 (Emerging/Becoming Established)

[46] See FDG with CEC member, Bay, Int. 106, See FDG with CEC member, Bay, Int. 109.
[47] See FDG with CEC member, Bay, Int. 106, FDG with CEC member, Lower Juba, Int. 111.
[48] See FDG with CEC member, Banadir, Int. 104.
[49] See FDG with CEC member, Banadir, Int. 102, FDG with CEC member, Bay, Int. 106, FGD with CEC member, Bay, Int. 108.

Proportion of GEFs and BEFs implementing community actions to support attendance and retention

The AGES programme focuses on empowerment of girls to be change-makers in their respective communities and GEFs are a key platform for girls to achieve this both through direct interventions aimed at improving enrolment, attendance, and retention of girls, and through indirect interventions related to economic support and change in community perceptions.

GEF activities at community-level include awareness raising on the importance of girls' education, prevention of early marriages, community discussions and trainings for community members on various topics, including health, girls' rights etc. Business and financial support activities contribute to the other AGES programme activities aimed at strengthening the economic situation of girls' households, reducing the need to withdraw girls from school due to monetary reasons. GEFs also directly provides support for girls' education by bringing girls' instances to institutional platforms through participation in CECs, enrolling out-of-school girls and easing the learning burden on new students as well as pre-existing ones through tutoring and study groups. Outreach and mobilisation of out-of-school girls by GEFs constitutes the most widespread activity of this type (19.8 percent of all girls enrolled in a GEF reported carrying it out). As the operation of the GEF extends beyond the programme participants to their communities of origin, this points to a positive direct spill-over effect of the interventions and, if maintained, of an increased sustainability of community-level improvement in girls' school attendance.

TABLE 31: PROPORTION OF GEFS ENGAGED IN ACTIVITIES, BY COHORT

Indicator	Total	FE girls	ABE girls	C1 NFE girls
Girls Education Support				
Enrolling out-of-school girls	19.8%	21.1%	19.8%	18.2%
Participation in CECs	8%	5.6%	12.5%	5.2%
Teaching other girls/ study groups	8%	2.2%	12.5%	9.1%
Business & Finances				
Savings group	7.2%	4.4%	9.4%	7.8%
Joint business	2.3%	-	4.2%	2.6%
Changing Community Attitudes				
Preventing early marriage	11%	10%	11.5%	11.7%
Awareness raising	41.4%	46.7%	42.7%	7.8%
Trainings (health, girls' rights, gender, etc)	6.1%	3.3%	10.4%	3.9%
Community discussions	4.2%	1.1%	9.4%	1.3%

Overall, the GEFs are more active in the area of awareness raising for girls of all groups (FE, ABE, C1 NFE), as it might be perceived by participants are the core objective of the network and, at the same time, requires

less organisational effort than activities such as participation in CECs (8 percent overall) or a more limited investment of time compared to training others, respectively only 8 percent and 6.1 percent of girls participating in GEFs reported teaching other girls or study group or carrying out community trainings on various topics. At the same time the activities related to business and finance are relatively less implemented by GEFs as well, with only 7.2 percent of GEFs implementing savings groups and 2.3 percent establishing joint business. While there might be some overlap with VSLAs, discussed previously, which could account for a low rate of savings groups, it is significant that only very few C1 NFE groups are engaged in GEF joint businesses.

As mentioned previously, together with a decrease in number of GEFs with active engagement, the range of activities carried out has also decreased. We can observe this trend for all activities covered in the survey, including the most popular activity, awareness raising, which went from 90 percent at ML1 to 41.4 percent at ML2. Of course, a decrease in the variety of activities carried out might simply indicate an increased focalisation of efforts and resources on fewer activities, with higher effectiveness. It is much more likely, though, that GEFs have reduced their activity levels, since support to many GEFs has ended. For example, support for GEFs embedded in C1 NFE centres ended in 2021; as such, we would expect a natural decline in their activity levels by 2023. In this case, the sustainability of GEFs community-level impact will be severely at risk as programme support for other GEFs ends along with the end of the broader AGES programme.

Sustainability Indicator Score: 1 (Latent)

Proportion of umbrella schools adopting new methodologies

To help improve learning outcomes for students, as part of AGES activities, teachers in targeted schools have been trained on new teaching methodologies. If these new approaches are adopted and proved to be beneficial for both students and teachers, the sustainability of improved learning outcomes for students will be greatly improved. Such interventions do not only have an impact on teachers trained but can also extend through informal sharing and observation to teachers that were not trained and new teachers, who after seeing the success of these new methodologies might be incentivised to adopt them in their own teaching. Considering a certain natural rate of teachers' mobility, this spill-over effect of new teaching methodologies will also likely extend to other schools in the area that were not originally targeted by the AGES programme.

Furthermore, CARE's continued coordination and involvement with the Federal/ States Ministries of Education is expected to promote this ministry efforts to train teachers or disseminate information on new teaching practices, also contributing to programme sustainability.

The table below reports the number of teachers (male and female) trained on different skills in the schools targeted by the programme in the previous year. The training of female teachers is especially important as it can further empower girls to stay in schools by providing positive role models and being able to better understand difficulties faced by female students in the classroom and at home.

TABLE 32: NUMBER OF TEACHERS TRAINED IN THE PAST YEAR

Skill	Total Teachers			Average (per school)	
	Female	Male	TOTAL	Female	Male
Gender sensitive teaching methods	50	103	153	1.4	2.9

Teaching methods for maths	71	121	192	2.0	3.5
Teaching methods for reading and writing	89	170	259	2.5	4.9
Inclusive education	101	184	285	2.9	5.3
Child protection	75	140	215	2.1	4.0

While the absolute number of teachers trained is still lower for female teachers in all skills, the proportion of female teachers trained in the past year out of all full-time female teachers in the schools is higher for all skills than for their male colleagues. When we interviewed head teachers in the programme schools, they reported the rate of female teachers that attended trainings on teaching methods for maths and for reading and writing respectively as 40.6 percent and 50.9 percent. The same figures are as low as 22.5 percent and 31.7 percent for male teachers. On the other hand, the average number of teachers trained (female or male) per school is still quite low, ranging from 1.4 to 5.3. The total number of teachers trained in gender sensitive teaching methods is also the lowest among other skills, which is concerning considering the focus of AGES programme on supporting girls' education.

TABLE 33: PERCENTAGE OF TEACHERS TRAINED (AT ML1 AND ML2[50],[51])

Skill	Trained (ML1)	Trained (ML2)
Gender sensitive teaching methods	21.5%	44.6%
Teaching methods for maths	27.0%	31.7%
Teaching methods for reading and writing	36.4%	42.0%
Inclusive education	40.0%	52.0%
Child protection	30.2%	48.8%
Other	18.8%	20.1%

The proportion of teachers being trained on new teaching methodologies has further increased with respect to ML1 for all skills, with the most significant improvements being observed in skills on which a relatively lower proportion of teachers reported been trained at ML1: gender-sensitive teaching methods, inclusive education, and child protection (all with increases above 10 percentage points).

As detailed further in the Teaching Quality section of the report, the quantitative analysis points to some success in the actual adoption of new teaching practices by trained teachers, including formative assessments and engagement of students who are not participating in class. At the same time, the full transformation of teaching approaches remains slow, with many teachers still relying on traditional teaching practices.

[50] The percentages are calculated only on the sample that is comparable between ML1 and ML2

[51] The schools employ both full-time and part-time teachers. As the training targets in particular full-time teachers, the proportion of teachers trained has been calculated on the total of full-time teachers. In case some part-time teachers also participated to the training, the figure might be slightly overestimated.

Qualitative evidence obtained from teachers and CEC members interviewed as part of FGD on the other hand suggests the successful adoption of new practices taught, however interviewees selected are likely to be among the most active teachers (possibly suffering from selection bias). Except for new teachers, which we expect to be able to observe these methodologies from already-trained teachers, most older teachers interviewed mentioned an improvement in their attitude towards students, including children with disabilities and girls, and in their lessons planning, resulting in increased motivation for the teachers[52] and increased teaching effectiveness. As reported by one of the teachers "one of the many things I've changed is the number of lessons I used to teach at first, and sometimes I couldn't find the time to explain them. But in the past two years, I've found the time to both teach and explain lessons"[53]. This improvement in time management has been mentioned by many teachers as a success of the training, allowing them to have a longer-term lesson plan rather than preparing the materials on a lesson-by-lesson basis[54].

The increased number of teachers trained in child protection and gender-sensitive teaching methods has also been reflected in a reported change in both attitudes and practices of teachers, with a decrease in corporal punishments and increase understanding of each child's situation. Teachers often report having more patience, explaining again topics that were not understood by the child[55], and generally promoting a positive classroom environment by increasing interactions and times for discussion. This finding is also confirmed by the more detailed analysis present in the *Teaching Quality* section that, while recognising the persistence of corporal punishment, points to their overall decline in favour of other forms of disciplinary practices.

Sustainability Indicator Score: 3 (Becoming Established)

Proportion of teachers implementing inclusive education strategies in class

As part of its focus on supporting education of vulnerable girls, one of the key modules that teachers covered by the AGES programme are trained on is inclusive teaching methodologies. This training covers approaches tailored to make education inclusive across gender, displacement status, disabilities etc. The sustainability of programme outcomes related to the inclusion of diverse students is likely to be increased by the acquisition of additional skills that, as for the general discussion on the importance of teacher trainings, once proven successful, can be replicated by teachers, and spread to new teachers as well. Advocacy work carried out by CARE with the Ministry of Education will also further support the programme sustainability.

At present, among teachers trained in the previous 12 months, training on inclusive teaching methodologies represents the highest rate of completion, with 52 percent of teachers trained. Absolute number of teachers trained as well as average teachers trained per school is also the highest among all skills covered by this activity. A total of 285 teachers (male and female) were trained in the past year, with an average of 5.3 male teachers and 2.9 female teachers trained per formal school.

Training on teaching methodologies specifically over the inclusion of GWD has been especially successful, as evidenced by many teachers and CEC members mentioning making adjustment to the classroom set-up to support the learning of GWDs. Beyond classroom logistics, teachers also specifically mentioned having an equally supportive attitude towards girls with and without disabilities, and motivating GWD by showing them

[52] See FGD with teachers, Gedo, Int. 509.
[53] See FGD with teachers, Banadir, Int. 501.
[54] See FGD with teachers, Banadir, Int. 503.
[55] See FGD with teachers, Banadir, Int. 501, FGD with teachers, Banadir, Int. 504, FGD with teachers, Gedo, Int. 509.

that "that they can compete with other students and achieve first place"[56]. Awareness raising by teachers and motivating students is a general result of the training in inclusive teaching methodologies students coming from various marginalised groups are all benefiting from this change in teachers' attitudes[57]. Teachers increasingly encouraging discussions in class and interactions between the students is also likely to result in less discrimination and a more favourable classroom environment for girls coming from vulnerable communities.

When looking at the reported reasons for dropping out of formal or non-formal education, teachers' poor attitudes towards the students or lack of skills are rarely, if ever, indicated by girls as a motivating factor in their decision to leave school. Less than 1 percent of girls indicated mistreatment or discrimination by the teachers as a reason for dropping out and 95% of FE girls state that their teachers make them feel welcome in the classroom (see *Teaching Quality* section). The same finding is true when we asked girls if their withdrawal from school or other non-formal education programmes was due to the teacher being unable to teach them[58].

The more detailed quantitative analysis of pedagogical practices, including the adoption of inclusive teaching methodologies present in the *Teaching Quality* section of the report finds some positive evidence that teachers treat male and female students more equally, as evidenced by the increase in the proportion of classrooms where teachers directed questions at both genders. At the same time, perceptions of girls themselves still indicate that further support is needed to strengthening teaching equality.

Given the limited amount of information present in the data beyond gender-lenses, especially in the household survey and head teacher survey, regarding inclusivity along non-gender-related lines, although these results are positive, they are not conclusive.

Sustainability Indicator Score: 3 (Becoming Established)

7. Intermediate Outcomes

7.1 Attendance

Attendance is the first intermediate outcome tracked for the original baseline AGES cohorts, supported by FCDO. The project's focus on attendance is premised on the link between consistent attendance and the project's core outcomes, learning and transition. The link between attendance and learning outcomes is the strongest, theoretically, with better attendance resulting in greater exposure to instructional time and more consistent learning, without major interruptions that can produce learning loss. The connection between attendance and transition is most direct in terms of its impact on grade-level progression; however, it is also related to continuing enrolment, insofar as better attendance maintains a girls' link to the school, her teacher(s), and peers. Perhaps more importantly, girls with inconsistent attendance tend to fall behind the learning curve in their classes, a strong predictor of eventual dropout.

To measure attendance rates, we rely on attendance headcounts conducted during site visits by the external evaluator's field teams. This approach is likely to be the most reliable in the context of AGES implementation

[56] FDG with teachers, Banadir, Int. 501.
[57] FDG with teachers, Banadir, Int. 504.
[58] Girls were asked to agree or not to the following statement "The teacher does not know how to teach someone like me".

due to poor record-keeping at schools and learning centres. Headcount surveys are conducted for each grade or stream, and in cases where there are multiple classes in the same grade or stream, team leaders' complete headcounts in every class or stream/level in the centre. This maximizes the sample of headcounts available for analysis.

Throughout this section, we limit our analysis of changes from baseline (BL) to the first midline (ML1) and now the second midline (ML2) to a comparable sample of schools or centres. This excludes schools or centres visited at BL but excluded at ML1 or ML2, which would introduce bias in attendance rate calculations. Instead, we construct a comparable sample of schools that appeared in all three samples and calculate attendance rates from the classes surveyed in those schools only. This may make some results (e.g., attendance rates) appear slightly different than they were presented in reports from previous rounds, because the sample of schools from any given survey round may not be the same.

While attendance headcounts are generally reliable, there are two shortcomings that need to be considered. First, they provide a snapshot of attendance on a single day, and attendance rates can fluctuate significantly from day to day, increasing the variability in results and decreasing confidence slightly. Second, the measurement of attendance is based on the total enrollment numbers, which serves as the denominator in the calculation, meaning that the validity of the attendance rate calculation depends on the reliability of enrollment records.

Aggregate Analysis of Attendance

As noted above, our analysis focuses primarily on headcounts conducted during school visits; secondarily, we analyse attendance counts conducted by teachers on the day of our visit and the day prior to our visit. While girls in the FE cohort were enrolled into grades 1 or 2 at baseline, they are now concentrated in grades 4 through 7. However, our analysis focuses on attendance in classrooms for grades 1-5, as these are the grades on which AGES interventions were concentrated.

Before turning to aggregate attendance results, it is important to note a shortcoming of the ML2 data collection. In previous rounds, we collected data from every classroom in grades 1-5, recognising that schools often have multiple classes of students within a single grade, especially in lower primary grades, where enrolment numbers are highest. However, a mistake during training and fieldwork implementation during the ML2 round meant that field teams were generally instructed to complete just one headcount per grade level; if a school had multiple grade 2 classrooms, for instance, the team leader would randomly select one classroom for the headcount. This differs from primary rounds, during which *all* classrooms in the target grade ranges were assessed for attendance. The result is that our sample size of classrooms is slightly smaller than expected (n = 175, compared to 260 at ML1) and it includes slightly fewer lower-primary classrooms than we would expect, as these are often the levels where there are multiple classrooms. The impact on our analysis is minimal, however, in part because the smaller sample size within clusters and therefore has a small effect on precision or statistical power. In addition, there was already variation across rounds in the grade-level composition of the sample, making it necessary to control for grade level in regression models regardless of the mistake in the current round.

In the table below, we present attendance rates for formal schools across the three rounds, using a naïve average (i.e., without correcting for differences in grade-level composition across rounds). For girls (top panel) and boys (bottom panel), we report attendance rates derived from our physical headcount – conducted by a member of our field teams – and teacher-recorded attendance rates from the day of our visit and the day prior to our visit, across rounds.

The first pattern that can be readily observed in the table is the fact that teacher-recorded attendance has increased for both boys and girls over time. For instance, girls' attendance, as recorded by the teacher on the day of our team's visit, has increased from 72.7 percent on average at BL to 83.0 percent at ML2, while limiting the sample to the same set of schools in all rounds. This contrasts with a small decline in attendance derived from physical headcounts over the same period and in the same schools. Because this same pattern obtains for boys and for attendance from the day *prior* to our visit, this suggests that what is changing is teacher behavior. If actual attendance was increasing, this would also be reflected in the physical headcounts; moreover, the divergent trend between measures cannot be explained by shifting enrolment levels, because enrolment serves as the denominator in calculations of the attendance rate using either physical headcounts or teacher records.[59]

The second pattern that stands out is the decline in attendance – based on the physical headcounts – among both boys and girls, but with an outsized decline among girls. As was true in previous rounds of the evaluation, our preference is to rely on physical headcounts as our primary measure of attendance, given the sometimes poor record-keeping of teachers and the high rates of missingness when using teacher counts.[60] Overall, girls' attendance has fallen by 4.5 percentage points since BL, while boys' attendance has fallen by about half that rate (2.2 percentage points, from 86.7 percent to 84.5 percent). For both girls and boys, this is a continuation of the trend observed at ML1, though most of the decline occurred between BL and ML1.

TABLE 34: CHANGES IN FORMAL SCHOOL ATTENDANCE RATES OVER TIME, BASED ON HEADCOUNTS

Attendance Outcome	BL	ML1	ML2
Girls			
Attendance – Physical Headcount	86.7%	82.7%	82.2%
Attendance Today – Teacher-Reported	72.7%	82.3%	83.0%
Attendance Yesterday – Teacher-Reported	66.4%	72.5%	85.7%
Boys			
Attendance – Physical Headcount	86.7%	84.8%	84.5%
Attendance Today – Teacher-Reported	71.2%	85.0%	85.0%
Attendance Yesterday – Teacher-Reported	66.4%	70.7%	85.9%

[59] The data seem to suggest that higher attendance rates, from teacher reports, over time stem from fewer teachers entering "0" instead of taking attendance. At BL, teachers documented zero female attendance in 10 classes; at ML2, no classes had less than 30 percent female attendance. A similar pattern applies to boys' attendance, so this cannot be explained by improved attendance among girls as a result of the programme. Our view is that teachers, in lieu of recording attendance accurately, may have occasionally written "0", and this information made it way into our own data. This is a further reason for us to prefer physical headcounts over teacher reports.

[60] This issue was more pronounced in previous rounds. For instance, at ML1, our sample of headcounts for analysis is 202, after accounting for erroneous values that had to be excluded from the data. In contrast, our sample of teacher-reported headcounts from the day of our visit includes 165 observations, owing to the number of teachers who do not consistently record attendance.

The results reported in the table do not control grade level or the grade level composition of the sample. To account for these differences across rounds, we estimated linear regression models in which the outcome was the attendance rate based on physical headcounts. Our models account for geographic zone and grade level, allowing us to understand how attendance rates have evolved over time while controlling for these other predictors of attendance.

Focusing first on girls' attendance, the regression approach generally mirrors the results in the table. First, attendance rates have fallen by 3.8 percentage points from BL to ML2, though ML2 attendance rates are slightly higher – in the regression framework – in ML2 than in ML1. In other words, the biggest decline appears to have occurred between BL and ML1, with a very slight (0.7 percentage point) recovery to ML2. While the drop from BL to ML1 was statistically significant at the 10 percent level ($p = 0.06$), the drop from BL to ML2 fell slightly outside the bounds of statistical significance ($p = 0.12$). We observe a similar pattern in boys' attendance, with a 2.2-point drop from BL to ML1, and an increase of 0.6 points from ML1 to ML2. Overall, boys' attendance in a classroom of equivalent grade level is 1.6 percentage points lower at ML2 than at BL. Substantively, the regression results mirror those of the naïve analysis reported in the bottom panel of the table above.

Geographically, there are moderate differences across zones in terms of attendance trends over time. Declining girls' attendance is driven by declines – from BL to ML2 – in Banadir and Southwest State, while attendance among Jubaland FE girls has remained steady over the same period. As with the differences in aggregate trends across genders, boys' attendance by zone has largely tracked with the trends among girls – decreasing attendance primarily in Banadir and Southwest State.[61]

Digging more deeply into grade-level differences, we estimated a further model that incorporated interaction terms between the round of data collection and the grade level of the classroom assessed in the headcount. This analysis allows us to understand *which* grades experienced the most significant decreases or increases in attendance over time. In the table below, we report the change in headcount-based attendance rates for a specific grade, between BL and (alternately), ML1 or ML2. For instance, the results show that girls' attendance in Grade 4 classes declined by 8.1 points from BL to ML1. Importantly, all results are *relative to the BL* round; thus, girls' attendance in Grade 4 classes declined by 4.1 points from BL to ML2, but this implies that girls' attendance in Grade 4 classes increased between ML1 and ML2, as the decline at ML2 is smaller than that seen at ML1.[62]

TABLE 35: CHANGES IN GRADE-SPECIFIC ATTENDANCE RATES, BY ROUND

Grade Level of Classroom	Girls' Attendance Relative to BL		Boys' Attendance Relative to BL	
	ML1	**ML2**	**ML1**	**ML2**
Grade 1	-4.3	-1.9	0.3	2.0

[61] One difference is that boys' attendance has *increased* substantially in Jubaland from BL to ML2, while attendance remained flat over this period among girls. Still, in terms of general trends, this is consistent with girls' attendance, insofar as Jubaland's trends are more positive – or less negative – than those in the other two zones.

[62] Note that there are few consistent patterns regarding attendance by class level over time. At baseline, Grade 4 classes had the highest average girls' attendance rate; at ML1, this distinction belonged to Grade 2 classes; at ML2, it belonged to Grade 1 classes. There are good theoretical reasons to expect both higher attendance among higher grades (motivated or dedicated girls have self-selected into those grades through continued enrolment) and lower attendance among higher grades (older girls face more demands on their time from household chores, for instance). These countervailing forces help explain the ambiguous relationship between grade level and attendance rates, which we also observe for boys' attendance based on headcounts.

Grade 2	-0.2	-2.6	1.5	-2.3
Grade 3	-3.5	-4.7	-1.3	0.4
Grade 4	-8.1**	-4.1	-4.8	-1.9
Grade 5	-5.0	-4.0	-3.0	-2.5

** = statistically significant at the 5 percent level.

Moving beyond attendance itself, a review of classroom records revealed a few encouraging trends. First, the dropout rate among both boys and girls has declined since baseline. During classroom headcounts, field teams report the number of students registered in the class, the number who have dropped out, and – from these figures – derive the total number of students who *should* be in attendance. From BL to ML2, the dropout rate among girls has declined from 8.0 to 5.7 percent, and a slightly smaller decline can be seen among enrolled boys. These results are positive, but only in a suggestive way, because the definition of a dropout is ambiguous, and teachers may be recording this information using different standards over time.[63]

7.2 Teaching Quality

We now discuss four dimensions of teaching quality and practices: professionalism, gender equity, the use of physical punishment, and pedagogical practices. AGES is expected to strengthen teaching quality through a combination of teacher training addressing gaps in literacy and numeracy teaching, improved capacity for quality assurance, and increased knowledge of inclusive education. Improved teaching quality is then expected to strengthen learning, as girls are expected to learn more when teachers use high-quality teaching practices. The use of positive teaching practices may also improve transition as girls may be more incentivized to attend and stay in school; in contrast, if teachers are often absent or do not engage students, girls may be more inclined to drop out.

We utilise two data sources for our analysis: classroom observations conducted to understand teachers' practices during lessons and the survey with girls. Data from classroom observations provides direct insight into teaching practices, but we note that our analysis is limited to FE schools and C5 NFE schools (discussed in a subsequent section).

Teacher Professionalism

We consider two dynamics of teacher professionalism: the extent to which teachers make students feel welcome and teacher absenteeism, as reported by girls in the girl survey. Results are reported for the full group of students assessed at BL, ML1, and ML2 in the table below.

Across all rounds, we find that most girls—over 95%—state that their teachers make them feel welcome in the classroom. This suggests that teachers have consistently been effective at creating a welcoming learning environment. In contrast, girls report more challenges with teacher absenteeism: At ML2, only around 86% of girls agreed that their teacher was not "often" absent. This figure, however, represents a significant improvement from baseline, when only 69% of girls stated that their teacher was not often absent. Indeed, teacher absenteeism has steadily improved across evaluation rounds, increasing by almost 11 percentage

[63] As we noted above, declining attendance rates may be a function of girls who – in the past – may have dropped out of school but, in the present, remain enrolled but with poor attendance rates. This would simultaneously explain lower dropout *and* attendance rates. As a result, we urge caution when interpreting changes in dropout rates over time.

points from BL to ML1 and an additional 6 percentage points from ML1 to ML2. This improvement was echoed by a teacher in an FGD, who stated that, "My attendance and the way I treat students has changed".[64]

TABLE 36: CHANGE IN TEACHER PROFESSIONALISM, FE SCHOOLS

Outcome	BL	ML1	ML2	Difference (BL to ML2)
Feels welcome	95.2%	98.4%	96.0%	0.8
Not often absent	69.0%	79.5%	85.7%	16.7*

To better understand changes in teacher absenteeism, in the table below, we disaggregate results by zone, focusing on changes from BL to ML2. While no change was significant, we find the largest change in South West State, of 19.0 percentage points, and the smallest in Banadir, of 14.8 percentage points. The similarity of these results, however, belies substantial differences by district. Most notably, in Gedo district (Jubaland), the rate at which teachers were reported as not often absent increased from only 25% at BL to 92% at ML2,[65] while in Lower Juba district (Jubaland), rates decreased from 98% at BL to 82% at ML2. The reason for these trends is not entirely clear; we do not, for example, find substantially different reported rates of insecurity due to conflict or environmental disasters (such as flooding) in either of these districts, nor were any major barriers to teacher attendance mentioned in FGDs.

TABLE 37: CHANGE IN TEACHER ABSENTEEISM IN FE SCHOOLS, BY ZONE

Region	BL	ML2	Difference
Banadir	68.1%	82.9%	14.8
Jubaland	69.9%	86.3%	16.4
South West State	70.7%	89.7%	19.0

Classroom Gender Equity

To understand gender equity in classrooms, we analyse teachers' differential treatment of girl and boy students and their attitudes towards gender equity. We do not assume that teachers hold positive views about gender equity or equality, given that AGES is implemented in highly marginalised areas of Somalia where girls face major cultural and other barriers to education. Given this dynamic, teachers' attitudes towards gender roles and girls' education may have an important impact on the success of interventions.

Our analysis addresses both gender equality and gender equity. Equality, in this case, refers to a situation where girls and boys are given the same opportunities and treatment. Equity, in contrast, refers to differential treatment of girls to give them additional opportunities to help overcome the effects of systemic discrimination or barriers. AGES seeks to not just ensure gender equality, but to be gender transformative through an emphasis on equity.

We first analyse metrics of gender equality, starting with girls' responses to the statement, "My teachers treat boys and girls differently in the classroom". Agreement with this statement—i.e., belief that classrooms are

[64] FGD with teachers, Gedo, Int. 509
[65] n = 38.

not gender equal—has increased since baseline, from around 37% of girls agreeing that teachers treated girls and boys differently at BL to 53% at ML2. This decline in perceptions of gender equality is significant.

To better understand this indicator of potential inequity in classrooms, we now analyse data from classroom observations. The below table shows mixed results. We find a large and significant increase in the percent of classrooms where teachers directed questions at both genders. However, we find decreases—though not significant—in whether teachers gave boy and girl students equal amounts of time to answer questions and whether boys and girls were asked equally difficult questions.

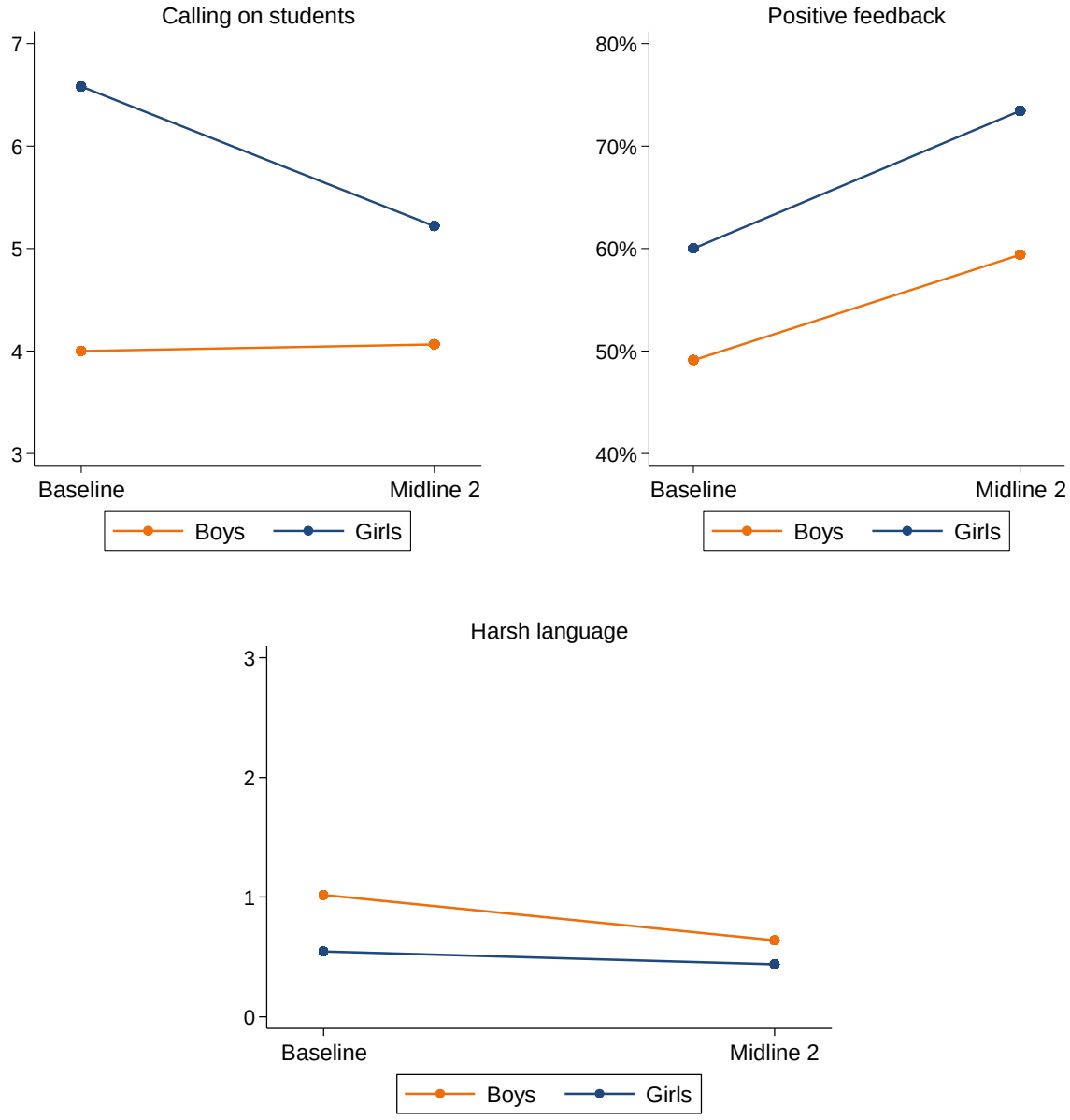
TABLE 38: CHANGE IN GENDER EQUALITY IN CLASSROOMS, FE SCHOOLS

Outcome	BL	ML2	Difference	P-value
Equal access to learning materials	94.4%	92.7%	-1.7	0.75
Questions directed at both genders	65.5%	93.6%	28.1	0.003*
Same amount of time to answer questions	51.4%	40.0%	-11.4	0.39
Questions equally difficult	36.1%	31.5%	-4.6	0.70

Rather than providing any broad conclusions about teachers' perceptions of gender equality, these results may suggest that strengthening teaching equality is a gradual process. For example, training and sensitisation may have successfully improved teachers' knowledge of the importance of gender equality. However, teachers may still have ingrained habits and beliefs—such as, for example, the belief that boys are better students than girls, and can thus answer more difficult questions—which they find difficult to change. As such, the improvement in one metric of teaching equality may suggest that teachers are indeed trying to improve their gendered behaviours, but that continued reinforcement to overcome ingrained biases is needed.

With these results in mind, we now move on to measures of greater relevance for AGES' goal of gender equity. Our main measures of equity include the number of times teachers called on boy and girl students, whether positive feedback was provided to girls and boys, and how often harsh language was used with girls and boys. The figure below shows the change in these measures from BL to ML2. We find that in general, teachers tend to use practices favouring girls in class. Across both BL and ML2, on average, teachers called on girl students more frequently than boys, were more likely to use positive feedback with girls than boys and used harsh language less frequently on girls than on boys. Indeed, for this last practice, we find that on average, teachers used harsh language with girls less than once per class.

FIGURE 9: CHANGE IN USE OF GENDER EQUITABLE PRACTICES IN CLASSROOM, FE SCHOOLS



However, it is also important to note that there appears to have been little *relative* change in teachers' use of equitable practices for girls over time. The use of harsh language directed at both boys and girls has declined, and the use of positive feedback has increased – nearly in parallel – for both boys and girls. The one exception is the rate at which students are called on to answer questions (top-left panel): we find a decline in the number of times teachers called on girl students from BL to ML2, compared to a slight increase in the frequency with which teachers called on boy students. Given that girls were called upon by teachers much more frequently

at baseline, this shift represents greater equality of participation overall and we do not believe it reflects negative on gender equity in AGES classrooms.[66]

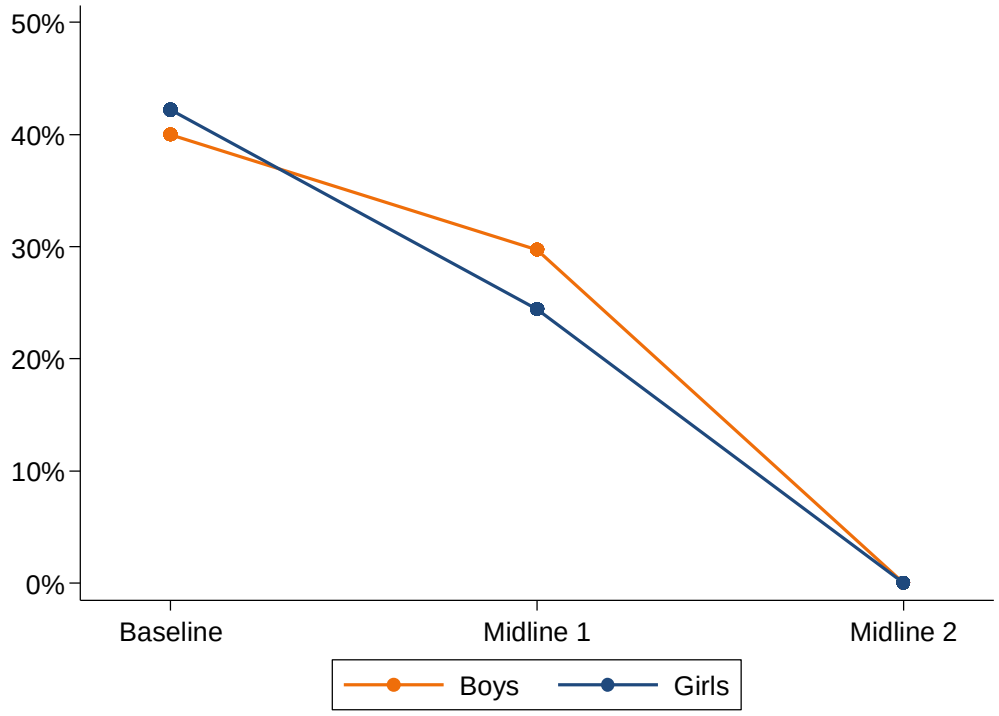
Disciplinary Practices

We now analyse negative disciplinary practices used by teachers, including punishing students for wrong answers and use of corporal punishment. These practices may deter learning by making students afraid to take part in class or to attend school. We first analyse negative practices observed during classroom observations, noting that the use of negative practices is likely to be under-observed due to social desirability bias (i.e., teachers are less likely to use a negative practice if someone is watching them). We thus validate our findings with data from the girls' survey.

The below figure shows encouraging change over time in use of physical punishment against boys and girls. We find a significant and consistent decline in the use of physical punishment since baseline, from being used against boys in 40% of classrooms and against girls in 42% of classrooms at baseline to being used by zero teachers at ML2. This is a very promising finding, suggesting that teachers in FE schools have gained an understanding of the harmful effects of physical punishment.

[66] The top-left panel also shows a small overall decline in the frequency with which teachers call on students, aggregating across male and female students. This finding would be consistent with the increased enrolment documented elsewhere in this report and corroborated by records from other education evaluations (the Education Sector Improvement Grant, ESPIG, study) and directly from the Federal Education Management Information System (EMIS). In more populous classrooms, teachers may be reverting to less participatory methods. On the other hand, they may be using participatory methods that are more appropriate to larger class sizes, such as group work or games, rather than calling on students, as direct engagement of this kind can be construed as less valuable in very large classrooms.

FIGURE 10: CHANGE IN OBSERVED USE OF PHYSICAL DISCIPLINE BY TEACHERS, FE SCHOOLS



As noted above, however, we would expect teachers to use physical punishment infrequently when observed; as such, the below table provides validating information from the girls' survey on teachers' use of physical punishment as well as punishment for wrong answers. These results reported by girls still show progress, but less so than in the figure above. We find a significant decline in the percent of girls reporting that their teachers have used physical punishment, from 70% to 31%, and a smaller (and insignificant) decline in reported use of punishment for wrong answers. However, teachers' use of both negative practices remains far more prevalent than ideal, particularly given the potentially severe consequences of these practices on learning and transition. The use of punishment for wrong answers is particularly common, with around three-quarters of teachers using this practice.

Furthermore, discouragingly, use of both physical punishment and punishment for wrong answers has increased significantly since ML1, by around 15 percentage points for the former practice and almost 20 percentage points for the latter. In combination with the results discussed above, this suggests that teachers may now know that physical punishment is perceived as a negative practice—hence not using this practice while under observation—but may not have fully internalised this concept, instead still relying on negative discipline within their day-to-day teaching.

TABLE 39: CHANGE IN TEACHERS' USE OF NEGATIVE DISCIPLINARY PRACTICES, REPORTED BY FE GIRLS

Outcome	BL	ML1	ML2	Difference (BL to ML2)
Punishment for wrong answer	82.1%	56.2%	75.7%	-6.4
Use of physical punishment	70.1%	16.0%	31.4%	-38.7*

Along these lines, in FGDs, responses on the use of physical discipline were mixed. For example, in an FGD with teachers, one teacher stated that, "Formerly, we would hit them with sticks, but now the punishment is to water the plants".[67] Another teacher in Banadir stated that, "In the past, I would punish the students if they couldn't answer the questions, but I no longer do so. The student will dislike education if he is punished".[68] Similarly, teachers in Gedo stated the following:

> *Teacher 1: Most of the time we give them advice and we don't punish them for their mistakes. If the student becomes too stubborn, we call the parents. Yes, we have made changes and one of the ways we deal with them now is no physical harm or beating.*
>
> *Teacher 2: The student is given lectures and admonitions; if they do not take the admonitions, they will face punishment, a punishment that will not cause them any problems. A lot has changed in the last two years, and there are new methods. Previously, punishment was the top priority; now, we do not rush to punish students.*
>
> - FGD with teachers, Gedo, Int. 509

CEC members also described initiatives to reduce the use of physical punishment. A CEC member in Bay, for example, stated the following:

> *We ensure that the child is not abused, and everyone gets their rights, is not discriminated against, or punished. We raise awareness, we tell teachers not to hit students or children.*
>
> - FGD with CEC members, Bay, Int. 108

Along these lines, a CEC member in Banadir described the positive effects of efforts to reduce physical punishment, stating that "students now leave school pleased rather than unhappy as they did in the past when they were punished".[69]

These positive sentiments, however, were not universal. Indeed, several girls mentioned fearing school because they could be punished for wrong answers.[70] Some teachers, parents, and CEC members also mentioned continued use of physical punishment despite increasing knowledge about its negative effects. In an FGD in Lower Shabelle, for example, teachers stated the following:

> *Teacher 1: [Students] are disciplined to hit them when they make a mistake, such as being hit on the hand with a small stick.*
>
> *Teacher 2: They are given advice for the first time, and if they make another mistake, they will be physically punished.*
>
> - FGD with teachers, Lower Shabelle, Int. 511

[67] FGD with teachers, Lower Juba, Int. 510.
[68] FGD with teachers, Banadir, Int. 501.
[69] FGD with CEC members, Banadir, Int. 102.
[70] E.g., FGD with girls, Banadir, Int. 404; FGD with girls, Bay, Int. 406; FGD with girls, Bay, Int. 407.

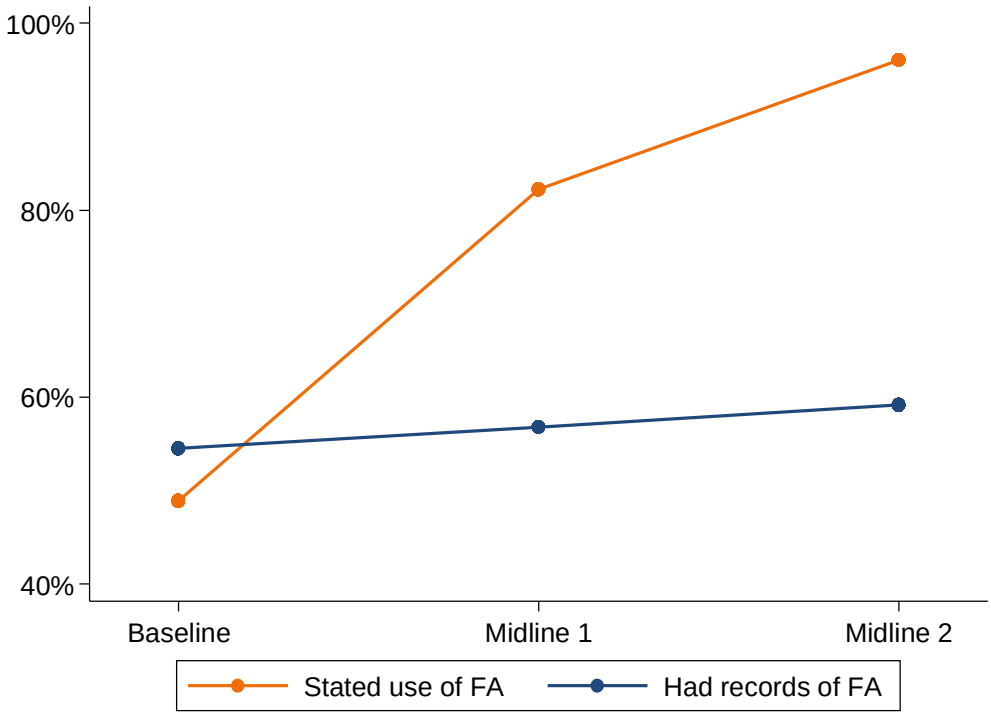
Overall, these qualitative findings validate those reported by girls in the table above, suggesting that while use of negative disciplinary practices has declined it has by no means been eliminated. There remains a clear need for further interventions to continue reducing the use of negative practices in classrooms.

Pedagogical Practices

Lastly, we examine the pedagogical practices used by teachers, including the use of formative assessments and positive teaching practices such as student-centred activities or games, allowing students to instruct each other, asking open-ended questions, soliciting students' opinions, involving students who are not participating, and using group work. As in above sections, we cross-reference data from classroom observations and the survey with girls to validate findings.

The figure below shows the change in the rate at which teachers stated they use formative assessments (FA) across evaluation rounds and the rate at which they were able to show records of the formative assessments. We find a large and significant increase in the rate at which teachers state they use formative assessments, increasing by more than 23 percentage points from 49% at baseline to 96% at ML2. However, we find a much more modest increase in the rate at which teachers were able to show records of formative assessments, from 55% at baseline to only 59% at ML2. This pattern may occur for two reasons: First, if teachers are overstating their use of formative assessments (potentially due to social desirability bias), and second, if teachers are not keeping proper records. In either case, the results point to a clear need for improvement.

FIGURE 11: CHANGE IN STATED AND RECORDED USE OF FORMATIVE ASSESSMENTS, FE SCHOOLS



Looking now at the use of participatory and student-centred practices in classrooms, the below table shows observed use of positive teaching practices during classroom observations in FE schools, focusing on the

change from BL to ML2. This table reports the percentage of teachers who were observed to have used a positive practice in at least two out of three observation blocks.

We find mixed results overall. Use of student-centred games and peer-to-peer instruction declined, though not significantly, from baseline to ML2, while use of other positive practices increased. Only one of these increases, however, was significant; teachers were significantly more likely to attempt to involve students who were not participating at ML2 than at baseline. Indeed, at ML2 this was the most frequently observed positive teaching practice.

TABLE 40: CHANGE IN USE OF POSITIVE TEACHING PRACTICES, FE SCHOOLS

Outcome	BL	ML2	Difference	P-value
Use of student-centred activities or games	27.3%	17.2%	-10.1	0.28
Students instruct each other	54.6%	46.9%	-7.7	0.49
Use of open-ended questions	45.5%	67.2%	21.7	0.10
Solicitation of student opinions	45.5%	64.1%	18.6	0.12
Involvement of students who are not participating	47.3%	81.3%	34.0	0.01*
Use of group work	27.3%	29.7%	2.4	0.80

Overall, like results at ML1, these results show a slight but not definitive positive trend. We also note that the use of most positive practices increased from ML1 to ML2, except for peer-to-peer instruction and group work (the latter of which declined very slightly). However, overall, the use of positive teaching practices remains somewhat limited. Indeed, on average at ML2, teachers in formal schools used only 8 to 9 positive teaching practices total per class—i.e., only around 1 to 2 of the six practices listed above per observation block. This represents only a marginal increase from baseline, when teachers used almost 8 positive practices total per class on average.

Furthermore, Table 41 shows that use of traditional, ineffective teaching practices remains prevalent at ML2. These practices include having students copy from the board or repeat after the teacher for much of the lesson; as in Table 40, results below show the percent of teachers using these practices in at least two out of three observation blocks. We find that around two-thirds of teachers rely heavily on ineffective teaching practices. In contrast, only two positive practices were observed more frequently than these ineffective practices, the use of open-ended questioning and involvement of students who were not participating. While use of ineffective practices has declined somewhat since baseline, the decrease is not significant.

TABLE 41: CHANGE IN USE OF INEFFECTIVE TEACHING PRACTICES, FE SCHOOLS

Outcome	BL	ML2	Difference	P-value
Students mostly copy from board	70.9%	65.6%	-4.4	0.67
Students mostly repeat teacher	80.0%	67.2%	-12.8	0.22

These findings suggest that there is a clear need to continue strengthening the use of positive teaching practices to improve teaching efficacy, learning outcomes, and student attendance. However, while classroom observations provide useful data on teaching practices, they may not show the full picture as teachers are likely to "overperform" relative to their norm when observed. As such, we lastly analyse data from the girls' survey on positive practices used by teachers. Results are presented in the table below.

Contrary to our expectation that positive teaching practices may be over-reported above, this table shows a more decisively positive picture of teaching practices than results from classroom observations. At ML2, FE girls were significantly more likely to report that lessons moved at the right speed (rather than too fast or too slow) and that their teacher often encouraged participation. This latter finding also validates results from

classroom showing greater encouragement for students to participate in class. Additionally, girls were also more likely at ML2 than at baseline, though not significantly so, to report that teachers explained the use of subjects for life outside of school, gave ideas for students to learn outside of class, and used different language to help students understand lessons.

TABLE 42: CHANGE IN USE OF POSITIVE TEACHING PRACTICES REPORTED BY FE GIRLS

Outcome	BL	ML2	Difference	P-value
Teacher often explains use of subjects	71.1%	80.0%	8.9	0.12
Teachers often gives ideas to learn outside of class	50.4%	60.1%	9.7	0.08
Lessons move at the right speed	40.4%	69.6%	29.2	<0.001*
Teacher often uses different ways of explaining	51.5%	58.9%	7.4	0.18
Teacher often encourages participation	54.1%	77.8%	23.7	<0.001*

Additionally, we find an increase, though not significant, in the percentage of girls reporting that lessons move at right speed, teachers use different ways of explaining, and teachers encourage participation from ML1 to ML2. However, we find a slight (and insignificant) decline in teachers explaining the relevance of subjects and giving students ideas to learn outside of class from ML1 to ML2. While results are still positive overall, this emphasizes the need for continuous reinforcement of the importance of positive teaching practices so that teachers do not backslide into old, ineffective or harmful habits.

Qualitative data further validates that teaching practices have improved since baseline. Many teachers described changing their practices to incorporate new, positive methods; for example, a teacher from Banadir stated the following:

> *I used to keep on teaching regardless of whether all the students understood, but now I focus on explaining to the kids who don't understand the lessons, and I've also modified how I explain or present the lessons.*
>
> - FGD with teachers, Banadir, Int. 501

Teachers in a Lower Shabelle FE school described using a variety of positive practices, including formative assessments, asking questions on the previous day's lesson at the start of class, and using practical demonstrations.[71] In Bay and Gedo, teachers similarly described the importance of asking questions about the previous day's lesson to engage students and ensure understanding; for example:[72]

> *When I arrive at class, I usually do a recap first thing in the morning and check to see if the students understood the previous day's lesson well. I check each student individually.*
>
> - FGD with teachers, Bay, Int. 507

However, as with results from the classroom observations, analysis of the qualitative data still points to room for improvement. A teacher from Lower Shabelle, for example, stated that he writes lessons on the board for students to copy, claiming that this was an improvement over his previous method of writing directly into

[71] FGD with teachers, Lower Shabelle, Int. 511
[72] See also FGD with teachers, Gedo, Int. 509

students' books.[73] Continued emphasis on positive teaching practices and sensitization on the inefficacy of traditional teaching practices may help further bolster learning and transition rates by endline.

7.3 Leadership and Life Skills

This section aims to analyse the progress made in leadership skills, self-confidence, and life skills throughout the programme. These aspects are assessed through self-perception, primarily using the Youth Leadership Index (YLI) as the main measurement tool. In addition to reporting overall scores, the analysis will disaggregate the changes from baseline to ML2 based on various factors. This will help determine the main influences on self-confidence and leadership while providing a detailed understanding of the program's successes and areas for further improvement. A baseline analysis will also be conducted for a newly recruited group of NFE girls in 2023. The Life Skills Index (LSI) will also be included in the analysis for this section.

The theory of change underlying the programme posits that improved quality learning opportunities tailored to the needs of ultra-marginalised girls, along with social norm change towards broader life opportunities, will influence the acquisition of life skills. Enhanced life skills, in turn, are expected to improve literacy, numeracy, financial literacy, and facilitate the transition to further education, employment, or self-employment.

The importance of girls' self-leadership skills for enhancing learning and transitional outcomes is grounded in broader findings from evaluations and lessons learned from projects such as SOMGEP-T and Phase 1 of the GEC initiative. The GEC thematic review on self-esteem indicates that interventions focused on building girls' self-esteem led to positive changes in their attitudes, including increased motivation to attend school, new aspirations, and a sense of school belonging. These changes, in turn, have a positive impact on attendance, class participation, and overall learning outcomes.

The Youth Leadership Index is a composite indicator based on a set of 21 questions, developed by CARE International and successfully piloted and used across several countries. The indicator measures a respondent's self-confidence, decision-making, voice, vision and organizational skills (including the ability to motivate others and work with them to address common issues). The table below contains the list of the questions used for the construction of the YLI.

TABLE 43: LIST OF YOUTH LEADERSHIP INDEX (YLI) QUESTIONS

Questions
I like to try new activities that I may not know how to do.
My friends ask me for advice.
I recognise when people have different skills to contribute to a task.
I am comfortable when my teacher calls on me to answer a question.
I contribute ideas to discussions at home even if they are different from others' ideas.
I ask questions at school when I don't understand something.
I can describe my thoughts to others.
The things I do set a good example for my peers.
I consider the possible outcomes of my decisions before making them.

[73] FGD with teachers, Lower Shabelle, Int. 511

I accept responsibility for the outcomes of my decisions.
I recognise when the choices I make today can affect my life in the future.
I can show what is important to me with my actions.
If someone does not understand me, I try to find a different way of saying what is on my mind.
I encourage others to join to help my community.
I cooperate with others to get things done at home.
If someone treats me unfairly at school, I am comfortable telling an adult.
I am willing to work hard to achieve my dreams.
I am better able to finish a task when I plan.
When I have the opportunity, I can organise my peers to do an activity.
I am interested in being a leader at my school.
I try to understand the cause of a problem before trying to solve it.

The YLI is calculated based on 21 self-reported questions on a 4-point Likert scale. All the cohort groups of girls were asked to indicate how often (rarely, sometimes, most of the time and almost always) they acted in a certain way, depending on the question asked. Lower values indicate more negative outcomes and higher values indicate more frequent instantiations of the behaviour and, by extension, more positive outcomes. The score ranges between 21 and 84 points and for the purposes of the analysis the score was standardised on the scale of 0 to 100. When a girl scored the lowest possible number of points (21) by responding 'rarely' to all questions, the standardised YLI score will take the value of 0%.

This section aims to analyse the changes in YLI scores from baseline to ML2 for the cohorts of girls in the original baseline. In addition to presenting overall figures, the analysis will disaggregate the data by key demographic variables and characteristics, including the main cohort groups (Formal Education, Non-Formal Education, and Accelerated Basic Education), regions, and districts within regions as necessary. The primary focus will be on the YLI scores themselves, rather than solely examining the proportion of girls who achieve a YLI score of 70 percent. However, increasing this proportion remains an important objective and a key target of the program, so it will also be reported.[74]

It is important to note that throughout this section, all findings should be interpreted with caution due to the absence of a comparison group. This limitation is particularly significant for this analysis because the main variable of interest, YLI scores, is expected to naturally increase with age. To demonstrate this, a regression analysis was conducted using the baseline data, regressing YLI scores on age while controlling for cohort. The results indicated a statistically significant positive effect, suggesting that an increase in scores would be challenging to attribute solely to the program's impact, as higher scores could be attributed to age-related factors.[75]

[74] The programme target is a 70% score on the YLI scale. Using the standardised 0-100 scale calculated by the evaluation team, this is equivalent to a score of 70 points; however, if employing the non-standardised YLI scale from 21-84 points, a 70 percent score is equivalent to 65.1 points.

[75] This regression was replicated using the cross-sectional data, i.e. the new group of NFE girls that were recruited in 2022. Again, a statistically significant positive effect was found, meaning that in the same cohort YLI scores and therefore leadership skills tended to be higher for older girls.

As seen in the table below, the total mean score for the girls who were in the baseline survey round, FE, ABE and C1 NFE, was 49. This rose to 52.9 at ML1 but between then and ML2 scores grew tremendously, to 66.5. This is a statistically significant jump of around 36%, a very meaningful success for the programme in increasing leadership skills from a relatively low base.

TABLE 44: CHANGE IN YLI SCORES BY ROUND AND SCHOOL TYPE

Outcome	BL	ML1	ML2	Difference (Earliest Round - Latest Round)
FE girls	46.0	49.2	66.3	20.3**
ABE girls	48.1	53.2	66.0	17.9**
C1 NFE girls	52.8	56.2	67.0	14.2**
Total (FE, ABE, C1 NFE)	49.0	52.9	66.5	17.4**

*** significant at 99% level, ** significant at 95% level, * significant at 90% level

The proportion of girls meeting the target of 70% on the YLI has also made meaningful strides since ML2. Using the panel sample of girls, at baseline it began from a low base of 11.5% and far from the endline target of 80% of girls reaching this target. At ML1 the figure stood at 16.1%, an increase of 4.6 percentage points or nearly 40%. But progress has truly been made since then, rising 26 percentage points to 42.1%. The gains are particularly notable among FE girls, whose baseline scores were merely 5.6%. This points to some success in the programme of increasing youth leadership skills, is a positive finding for the intervention. While the new score at ML2 still falls far short of 80%, at the current ML1-ML2 growth rate it is possible that this ambitious target may be met.

TABLE 45: CHANGE IN PROPORTION OF GIRLS WITH A YLI ABOVE 70 BY ROUND AND SCHOOL TYPE

Outcome	BL	ML1	ML2	Difference (Earliest Round - Latest Round)
FE Centres	5.6%	14.3%	43.8%	38.2%***
ABE Centres	11.5%	16.0%	40.1%	28.6%***
NFE Centres (Cohort 1)	17.1%	17.9%	42.6%	25.5%***
Total (FE, ABE, C1 NFE)	11.5%	16.1%	42.1%	30.7%***

For both the aggregate YLI scores as well as the proportion of each main cohort group the youngest girls (FE) exhibited the lowest scores at baseline. Subsequently, they also had the greatest gains since the programme's inception. To analyse whether these gains are partially due to an age effect, where maturing naturally increases girls' tendency to exhibit confidence and youth leadership skills, we also attempted to separate the

improvement from the programme and the improvement from age. The observed mean increases are compared against the expected increase over two years from the age effect as measured from the baseline. For FE centres the regression coefficient was 1.7, meaning we expect to see a 5.1-point increase in the YLI index over three years. For girls in ABE centres, we would expect an 8.3-point increase, while for NFE centres the coefficient was slightly negative and not statistically significant and as such will not be used in the analysis. Overall, then, the increases are still magnitudes larger than what we would expect based on the difference in YLI scores by age at each group's baseline.

The quantitative data clearly demonstrates a significant improvement in girls' leadership skills and confidence. This positive trend is reinforced by the qualitative data gathered from focus group discussions. During these discussions, girls shared their personal experiences in school and how it impacted them. When asked about the changes they experienced after attending school, a girl from Lower Juba expressed how she had "never written anything in my life, now I read and write everything and people appreciate it, I speak my mind, I express my thoughts. People believe that women can't do anything, but I do everything"[76], while another commented "once I learnt the math and understood it, I opened a small business, and it was great for me"[77]. These responses overwhelmingly reflect the positive outcomes witnessed in the positive scores mentioned above, offering personal perspectives on the matter.

The impact of the programme on the development of leadership skills is underscored by the relationship between participation in GEFs and YLI scores. Among the three baseline cohorts, there is a small but meaningful relationship between the two: self-reported participation in a GEF is associated with a 2.8-point higher YLI score (on a 0-100 scale) at ML2, an effect that is statistically and substantively significant. This relationship holds when we consider *changes* in YLI scores over time as well – a better indicator, because it controls for pre-existing differences in YLI scores between girls who self-select into participating in GEFs and those who do not.[78] In this model, GEF participation is associated with greater *gains* in YLI scores between BL and ML2 among girls in our panel sample. Girls who did not participate in a GEF experience a substantial gain – 24.8 points – in YLI scores from BL to ML2; however, GEF participants experience a larger gain of 26.2 points over the same period.[79] The impact of GEF participation is most concentrated among the FE cohort, which may stem from the more consistent, long-run exposure to GEFs experienced through a long-term formal school, as compared to a shorter-term NFE programme; it is also possible that GEFs have an outsized impact among FE girls because they tend to be younger and benefit most from the socio-emotional skills development and support they receive.

The scores broken down by zone – Banadir, Jubaland, South West State, and Hirshabelle – reveal distinct impacts of the programme across each region. The most notable change occurred in Banadir, where there was a substantial increase from 49.7 to 60.3 between BL and ML1, followed by another significant jump to 68.1 by ML2. This represents a statistically significant increase of 18.3, corresponding to a percentage gain of 37%. Jubaland, on the other hand, witnessed a modest rise to ML1, but has now grown to the same level

[76] Vignettes FGD with ABE Girls, Lower Juba, Int. 609
[77] Vignettes FGD with ABE Girls, Lower Juba, Int. 608
[78] In other words, our initial analysis looked at the correlation between GEF participation and YLI scores in a cross-sectional model, but it is possible girls who self-select into participating in GEFs are naturally more confident or otherwise would have higher YLI scores regardless of their participation in a girls' club. By assessing the change in YLI scores among the same girls between BL or ML1 and ML2, we are able to mostly account for this underlying difference between participants and non-participants, yielding more robust conclusions.
[79] This finding is robust to regression-based control variables for age, region, and other demographic characteristics. This finding is also consistent with the results of Section 4.4, which showed that both YLI scores and GEF participation are predictive of learning scores. The fact that YLI scores and GEF participation are correlated helps to explain this pattern of joint impact on learning outcomes.

as Banadir with a score of 68. In South West State there was initially a noticeable decline in scores between BL and ML1. But it has since rebounded, experiencing an impressive increase of nearly 15 points, a statistically significant net gain of 9.3 since baseline.

TABLE 46: CHANGE IN YLI SCORES BY ROUND AND AREA

Outcome	BL	ML1	ML2	Difference (Earliest Round - Latest Round)
Banadir	49.7	60.3	68.1	18.3**
Jubaland	44.6	46.5	68.0	23.4**
South West State	52.7	47.1	62.0	9.3***

*** significant at 99% level, ** significant at 95% level, * significant at 90% level

The results are even stronger when looking at changes in the proportion reaching the target of 70 or more. Among the original cohort of girls, Banadir showed a strong increase in proportion from 14.2% at BL to 46.6% at ML2, and Jubaland from 2.3% to 44.3%. This increase in Jubaland is especially promising given their low baseline. And while the proportion reaching the target fell sharply in South West State between BL and ML1, from a comparatively high 16.7% to 2.5%, it has since rebounded to 32%.

To summarise, there were increases in scores for the Youth Leadership Index over the period from BL to ML2 and in the proportion meeting the target of at least 70. The target set at the baseline – 80% of girls reaching the target – will still be difficult to reach but, given the enormous strides made since ML1, may be met over the course of the programme. This rise occurred across each of the original cohorts – FE, ABE, and NFE schools. The rise in scores was not evenly shared across regions, largely driven by a staggering 42.1 percentage point increase in Jubaland, but all regions saw large and statistically significant gains since baseline.

7.4 School Management and Governance

We now discuss the quality of school management and governance. Strong school management has many benefits for student learning, enrolment, retention, and motivation. Well-managed schools strengthen learning by employing qualified teachers, providing teachers with training opportunities, monitoring teaching practices and teacher attendance, and providing teachers with sufficient and timely pay. Effective management may also engage with communities to encourage parents to enrol their children in school, monitor student attendance, promote support for education among the community, and address barriers to enrolment and retention for marginalised students. Lastly, well-managed schools may use financial resources to improve infrastructure, design and implement school development plans, protect children, ensure inclusivity, and manage response to crises.

To improve school management and governance, AGES targets the activities of Community Education Committees (CECs) and Ministry of Education representatives working with programme schools. CECs are provided with capacity-building support to improve their ability to reach and assist marginalised girls, especially girls with disabilities, to enrol and remain in school. Specifically, AGES works with CECs to improve their community mobilisation capacities and adopt inclusive practices. AGES also supports government officials to increase their knowledge of the needs and rights of marginalised girls and to design

and implement policies addressing inclusive and special needs education. Lastly, AGES supports government officials to improve quality assurance and school monitoring procedures.

In this evaluation, we first focus on dynamics of management captured in the survey with head teachers. We then analyse indicators of interest to AGES' work with CECs. To evaluate the efficacy of CECs, we first assess the work undertaken by the CEC to address barriers to girls' education and the extent to which it corresponds to the actual barriers faced by girls in accessing education. We also analyse data from qualitative interviews on school management and governance to better understand the efficacy of CEC initiatives.[80]

School Management

We first briefly discuss management modalities of schools. At ML2, we find that 100% of surveyed head teachers reported that their school had a management plan. This is an increase since both baseline, when 91% of schools had a management plan, and ML1, when 79% of schools had a management plan.

To further understand management, the table below shows the group or individual reported by head teachers to have authority for a six school-related tasks at ML2. This table first shows that—as expected—school directors have a large influence over school management, with respondents most frequently reporting that directors have authority over the school budget, purchase of supplies and equipment, school calendar, student discipline policies, and personnel and hiring. CECs, however, also have a relatively high level of involvement, particularly regarding personnel and hiring and the purchase of supplies and equipment. Lastly, it is worth noting the substantial involvement of the national (and, to a lesser extent, FMS) government over the school curriculum and school personnel; this dynamic emphasizes the need for continued engagement with government.

TABLE 47: RESPONSIBILITY FOR MANAGEMENT OF VARIOUS TASKS, FE SCHOOLS

Responsible	Budget	Purchase of supplies and equipment	School calendar	Student discipline policies	Curriculum	Personnel and hiring
CEC	14.7%	20.6%	0.0%	11.8%	0.0%	23.5%
School director	47.1%	44.1%	67.7%	55.9%	17.7%	38.2%
Teachers	2.9%	0.0%	2.9%	23.5%	5.9%	0.0%
NGO or charity	11.8%	14.7%	2.9%	0.0%	0.0%	2.9%
FMS government	2.9%	2.9%	5.9%	0.0%	17.7%	8.8%
National government	14.7%	11.8%	8.8%	2.9%	52.9%	20.6%
Other	5.9%	5.9%	11.8%	5.9%	5.8%	6.0%

One key responsibility not included in the above table is supporting the provision of training to teachers. The below table shows the change in the average number of female and male teachers receiving trainings on a variety of topics within FE schools. We find a consistent increase in the number of teachers attending trainings on almost all topics at ML2 compared to baseline. While only one of these increases is significant, considering the limited sample size for analysis and the consistency of positive patterns, these results suggest that school

[80] We do not address AGES' work with the MoE in this section as no interviews were conducted with MoE officials at ML2.

management has effectively improved the extent to which they train teachers or support teachers to attend trainings.

Looking at the types of trainings received by teachers, at ML2, the greatest average number of female and male teachers attended trainings on inclusive education, followed by literacy teaching materials. Head teachers were not asked who provided the trainings received by teachers; as such, it is difficult to determine the reason why these trainings were prioritised.

Lastly, we note that fewer female teachers received training than male teachers. However, schools generally have far fewer female teachers than male teachers; at ML2, only around one-quarter of teachers in FE schools were female. As such, it appears that female teachers actually receive training at a higher rate than male teachers.

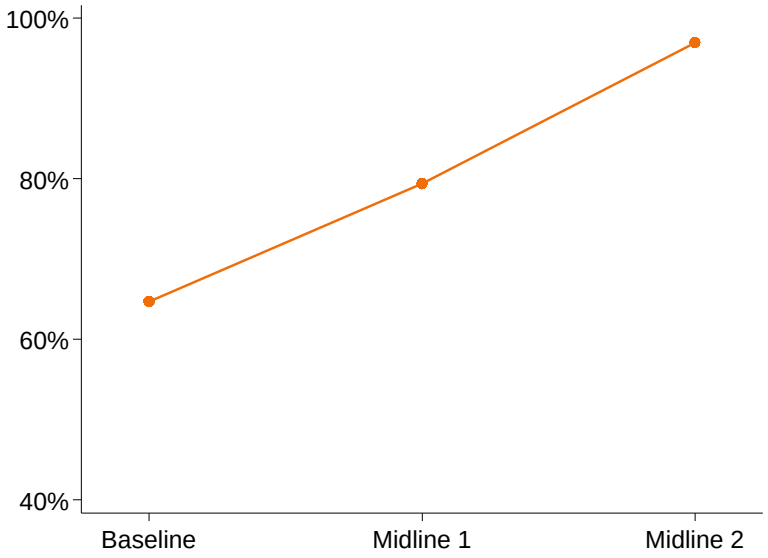
TABLE 48: AVERAGE NUMBER OF TEACHERS RECEIVING TRAININGS, FE SCHOOLS

	Female teachers			Male teachers		
	BL	ML2	Diff.	BL	ML2	Diff.
Gender sensitive teaching methods	1.3	1.5	0.2	2.5	3.1	0.6
Maths teaching methods	1.4	2.1	0.7	3.2	3.4	0.2
Literacy teaching methods	2.5	2.6	0.1	4.0	4.9	0.9
Inclusive education	2.8	3.2	0.4	5.5	5.3	-0.2
Child protection	1.6	2.3	0.7	3.2	4.0	0.8
Other	0.7	1.6	0.9	0.7	2.7	2.0*

CEC Initiatives to Address Barriers to Education

We now discuss involvement of CECs in school management as a key indicator of interest for AGES. At ML2, head teachers in 97% of FE schools (all but one school in which a head teacher survey was administered) reported that the school had a CEC. This represents a slight decline from ML1, when 100% of schools had a CEC, but still points to widespread existence of CECs. Furthermore, the below figure shows that levels of CEC activity have steadily increased since baseline. At ML2, 97% of CECs had come to the school during the past year to monitor facilities, a 32 percentage point increase since baseline.

FIGURE 12: CHANGE IN PERCENT OF CECS MONITORING FE SCHOOL FACILITIES



The below table shows the activities undertaken by CECs, as reported by head teachers and by FE girls. For data from head teachers, the table shows change from baseline to ML2. For data from FE girls, questions on CEC activities were not asked at baseline; as such, the table shows the change from ML1 to ML2.

For both head teachers and FE girls, the most commonly reported activities of CECs were to promote the enrolment of out-of-school children, including by supporting dropouts to return to school. Compared to baseline, head teachers reported an increase in these activities at ML2. However, notably, FE girls were significantly less likely to report that CECs supported dropouts and out-of-school girls at ML2 than at ML1. While the reason for these contradictory results is unclear, they may point to over-reporting of CEC activity by head teachers (potentially due to social desirability bias) or to greater knowledge of CEC activities among head teachers than girls. In either case, results suggest a potential need to continue supporting CECs to increase their activity levels.

Head teachers also reported frequent involvement in CECs in tracking student and teacher attendance and in undertaking child protection activities. For tracking of student attendance and child protection, CEC involvement had substantially and significantly increased since baseline. However, it is again worth noting that girls were substantially less likely to report that CECs monitored student and teacher absenteeism than head teachers. Furthermore, girls reported a significant decline in CEC monitoring of student absenteeism at ML2 compared to ML1.

TABLE 49: INITIATIVES UNDERTAKEN BY CECS IN FE SCHOOLS

Initiative	BL	ML1	ML2	Difference
Reported by head teachers:				**BL-ML2**
Promoting enrolment of out-of-school children	60.6%	73.5%	82.4%	21.8*
Tracking student attendance	45.5%	61.8%	76.5%	31.0*
Tracking teacher attendance	51.5%	55.9%	67.7%	16.2
Following up on cases of dropout	45.5%	67.7%	67.7%	22.2

Child protection activities	18.2%	26.5%	61.8%	43.6*
Raising funds for the school	24.2%	47.1%	35.3%	11.1
Hiring teachers	12.1%	35.3%	11.8%	-0.3
Reported by FE girls:				**ML1-ML2**
Support dropouts to return to school	-	70.1%	42.8%	-27.3*
Enrolment of out-of-school girls	-	81.6%	41.1%	-40.5*
Awareness raising on girls' education	-	21.8%	40.0%	18.2*
Check when a girl is absent	-	50.3%	33.9%	-16.5*
Address cases of harassment, abuse, and violence	-	18.4%	17.2%	-1.2
Provide hygiene materials for girls	-	9.5%	15.6%	6.0
Monitor teaching quality	-	16.3%	12.8%	-3.5
Monitor teacher absenteeism	-	5.4%	7.8%	2.3
Hire female teachers	-	2.7%	7.8%	5.1

In qualitative interviews, CEC members also described activities conducted by their groups. Many members mentioned monitoring student attendance and teaching quality (including teacher attendance) through regular school visits:

> *We are constantly present because we go to school and learn [what is happening] there. We keep an eye on the absent students, then go to their homes and bring them to school.*
>
> - FGD with CEC members, Banadir, Int. 102

> *[The CEC's] job is to make a regular inspection of the school, in terms of cleanliness, teacher attendance, and they also monitor students who are absent. And if the teacher takes any student outside, his parents are called so that he does not miss his class. This way parents are aware of what is happening and how his child is doing in school.*
>
> - FGD with CEC members, Bay, Int. 107

While not reported as a primary activity in the table above, child protection activities were also frequently mentioned in FGDs. In Banadir, for example, a CEC member stated that, "We make sure that the teachers do not hit the children and we stop the punishment of the children".[81] In Lower Shabelle, another member stated the following:

> *Many teachers like to punish the students who are late. The CEC committee will prevent the teacher from punishing the student who will be affected or hurt or may cause injuries to the student, which may cause the student to drop out of school.*
>
> - FGD with CEC members, Lower Shabelle, Int. 112

[81] FGD with CEC members, Banadir, Int. 102

In addition to these activities, around one-third of head teachers stated that CECs were involved in raising funds for schools at ML2. Qualitative data provides validating information to this finding, with some CECs stated that they were involved in fundraising. In Lower Juba and Lower Shabelle, for example, CEC members described the following fundraising activities:

> *As a committee, we do indeed raise money. We attempt to find solutions for the sake of the children and to ensure that the teachers continue to teach when they are not paid or given modest allowances. Yes, we raise funds from the local community and parents to pay the teachers' wages.*
>
> - FGD with CEC members, Lower Juba, Int. 110

> *We have raised donations to maintain the school. Once we collected about 350 to 400 dollars. We did repairs, replaced windows, painted walls, and things like that.*
>
> - FGD with CEC members, Lower Shabelle, Int. 112

Around half of interviewed CECs, however, stated that they were not involved in fundraising.

The below table provides additional information on CECs' monetary contributions to schools, showing the percent of CECs reported to contribute to teachers' salaries. At ML2, around half of head teachers reported that CECs contribute to teacher salaries, a substantial increase since baseline. The average contribution amount had also increased substantially since baseline, to around US$80 per month at ML2 (although sample size is highly limited, so results should only be taken as indicative). As a result, at ML2 among schools where CECs contributed to teacher salaries, these contributions made up around 30% of teacher salaries.

TABLE 50: CEC CONTRIBUTIONS TO TEACHER SALARIES, FE SCHOOLS

Outcome	BL	ML2	Difference
Percent contributing to teachers' salaries	20.7%	48.5%	27.8
Of whom, average contribution (USD per month)	22.3	81.4	59.1
Of whom, average percent of female teacher's salaries contributed by CEC	18.3%	37.9%	19.6
Of whom, average percent of male teacher's salaries contributed by CEC	29.8%	35.9%	6.1

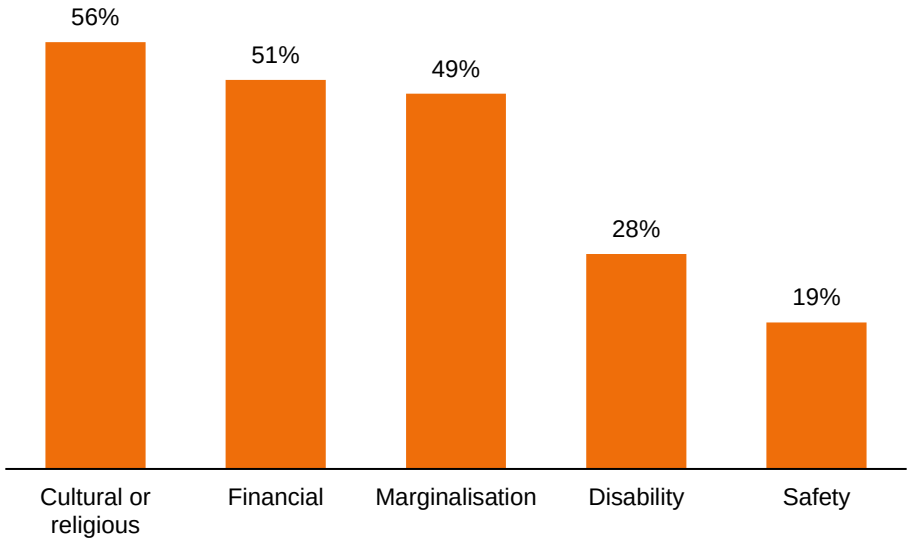
Overall, these results, though mixed, suggest an active role for CECs in school management. However, CEC activity is insufficient to ensure positive learning and transition outcomes; rather, CECs must also be addressing issues of relevance to the challenges faced by marginalised girls. To better understand the relevancy of CEC activities, in Figure 13 below, we show the frequency with which caregivers stated it is acceptable for a child not to attend school in various scenarios.[82] We categorise these scenarios into five groups: (1)

[82] In past reports, we have reported the reasons why girls were not attending school for girls who were out of school. At ML2, there were no FE girls who were out of school, and thus no respondents were asked these questions. As such, we instead report results for acceptability. While this does not give as direct of a picture into reasons why girls may not attend school, it is still indicative of potential barriers to girls' education.

cultural or religious, if caregivers believe school teachings are culturally inappropriate; (2) financial, if education is too costly or the child needs to work; (3) safety, if the child may be harmed or harassed at or on the way to school; (4) disability, if the school cannot meet the child's physical or learning needs; and (5) other dimensions of marginalisation, including if the child is needed to help at home, is married, is too old, or is a mother.

At ML2, we find that caregivers were most likely to state that cultural and religious, financial, and other marginalisation-related scenarios were acceptable reasons to withhold a child from school. For marginalisation factors, caregivers were most likely to state that it was acceptable for a child not to attend school if she was a mother (29%) or if she was too old (23%). In contrast, caregivers were less likely to express acceptance for students to miss school due to disability- or safety-related factors.

FIGURE 13: SITUATIONS WHEN IT IS ACCEPTABLE FOR A CHILD TO NOT ATTEND SCHOOL



Regarding financial barriers, it is important to note that at ML2, 47% of schools reported that no fees were charged for enrolment, a decrease since baseline when 58% of schools charged no fees. Among schools that did charge a fee, the average fee was around US$7 per month, a potentially substantial amount given the prevalence of extreme poverty in Somalia. Furthermore, 24% of schools at ML2 reported that families had to pay other costs for their child to attend school, an increase from 12% at baseline. These results suggest that a large portion of students may face financial barriers to enrolment and continued attendance in school. Qualitative data supports this finding; within FGDs, many CECs members mentioned that financial challenges prevent children from attending school, either due to inability to pay school fees or to afford other materials needed for school. Financial barriers also interact with other dimensions of marginalisation, as evidenced in an FGD with CEC members in Bay:

> *The lack of finance is a challenge for the family. If they don't have money, the boys are educated, but most of the girls are not educated.*
>
> *One of the biggest problems [displaced students] face is financial conditions. They can't buy notebooks, they can't buy textbooks.*
>
> - FGD with CEC members, Bay, Int. 106

Unfortunately, many CECs member expressed limited ability to help students with financial challenges. A CEC member in Banadir, for example, stated that, "The CEC can't provide them with financial support; all we can do is raise awareness".[83] Another member stated that, "The things [the CEC] doesn't do are related to finances but we advocate as much as we can. What hasn't worked is because of lack of money".[84] Indeed, CEC members frequently stated that a lack of financial resources limited their activities to support children and schools more broadly. In Lower Shabelle, for example, a CEC member stated that, "The CEC can't do anything about the school challenges because they don't have the resources".[85]

Referring back to the table above – on initiatives undertaken by CECs, as reported by cohort girls – we note that only 40% of girls reported that CECs were involved in awareness raising on girls' education. The prevalence of cultural and religious barriers to education shown in the above table suggests that it may be useful to increase CEC activity in this area; however, FGDs provide a contrasting picture of the involvement of CECs in awareness-raising, with many CEC members stating that this was a primary activity of the groups.[86] As a result of these awareness-raising activities (and those of other organisations), many CEC members stated that support for girls' education had massively increased:

> *A lot has changed, really. Many once disapproved of girls going to school and leading better lives since it was believed that girls should stay at home. People nowadays are different than they were in the past because they understand that it is preferable to teach females than boys since an educated girl reflects an educated society.*
>
> - FGD with CEC members, Banadir, Int. 102

CEC members also reported awareness raising on the need to educate marginalised girls, including girls with disabilities, pastoralists, IDPs, and married girls, and pregnant girls. However, many respondents described substantial barriers remaining to these students' education. The following quotes provide indicative examples of these challenges for girls facing various dimensions of marginalisation:

> *Because of the school building, the students with disabilities cannot access to attend the class… Also, a student with a disability needs special equipment to learn, as well as a special teacher who can make the student understand. Also, there are problems they may encounter from the community or in the class where they're studying.*
>
> - FGD with CEC members, Lower Juba, Int. 110

> *It is challenging for the children because most of them are refugees who have relocated from outlying regions and live in rented homes. If the rent goes up, they move to other low-rent properties. Because of these movements and the rent, itself, it makes it difficult for the child to focus on what they are learning.*
>
> - FGD with CEC members, Lower Juba, Int. 110

[83] FGD with CEC members, Banadir, Int. 103
[84] FGD with CEC members, Banadir, Int. 105
[85] FGD with CEC members, Lower Shabelle, Int. 112
[86] E.g., FGD with CEC members, Banadir, Int. 103; FGD with CEC members, Bay, Int. 106; FGD with CEC members, Lower Shabelle, Int. 112; FGD with CEC members, Middle Shabelle, Int. 113

> *If the girls marry, they drop out of school; if they find work, they drop out of school; and if they become mothers, they drop out of school.*
>
> - FGD with CEC members, Banadir, Int. 103

A variety of activities were described to address these challenges for marginalised students. Several CECs in Banadir described initiatives to help students with disabilities, for example, with one CEC stating that they supported children (with help from an NGO) to have eye surgery and purchase a wheelchair.[87] A few members also described supporting children with eyesight challenges to receive appropriate accommodations by encouraging teachers to sit those children at the front of the class.[88]

Several CECs also described outreach activities specifically focused on pastoralist children. In Bay, for example, the CEC took a very active role in supporting these children:

> *One of the committee's duties is to go through the neighbourhoods until they reach the [pastoralists] to tell them that school is free… We support [pastoralists] through awareness, we also support pastoralist children that go to school to be a parent for them. In some cases, there are 20 children under the committee's [care]. We have given full support to those 20 children… If the teachers has issues with any of the children, the committee acts as a parent figure because pastoral parents aren't anywhere to be reached.*
>
> - FGD with CEC members, Bay, Int. 107

Overall, these findings suggest an active role for CECs despite resource challenges. They also suggest that CECs are addressing issues relevant to ultra-marginalised children, and thus to AGES programming. However, more work is likely needed to support CECs and ensure that they are actively addressing issues of high relevance to girl students.

7.5 Community Attitudes

In this section, we assess changes in community attitudes around girls' education that can play an instrumental role in helping marginalised girls overcome the barriers preventing them from attending and completing school. The perceived benefit of education may influence caregivers' decisions to send their daughter to school instead of keeping her at home to help with housework, allowing her to marry, or engaging her in income generating activities. It may also influence the extent to which communities provide monetary or other support to girls to stay in school. Because the environment where AGES is implemented is characterised by high levels of poverty and constrained resources, investment in schooling may come at the cost of basic necessities for households. Caregivers' decisions to keep girls in school are thus influenced both by their views on the importance of education but also on their belief that the economic and social benefits of education outweigh its costs.

AGES expects that positive shifts in community attitudes and practices towards girls' education will emerge from the participation of different community actors in platforms for social change on gender and support for girls with disabilities. To reach this goal, the programme focuses on social norm change, including training religious leaders on Islamic principles to support gender equity and education for girls and inclusive education,

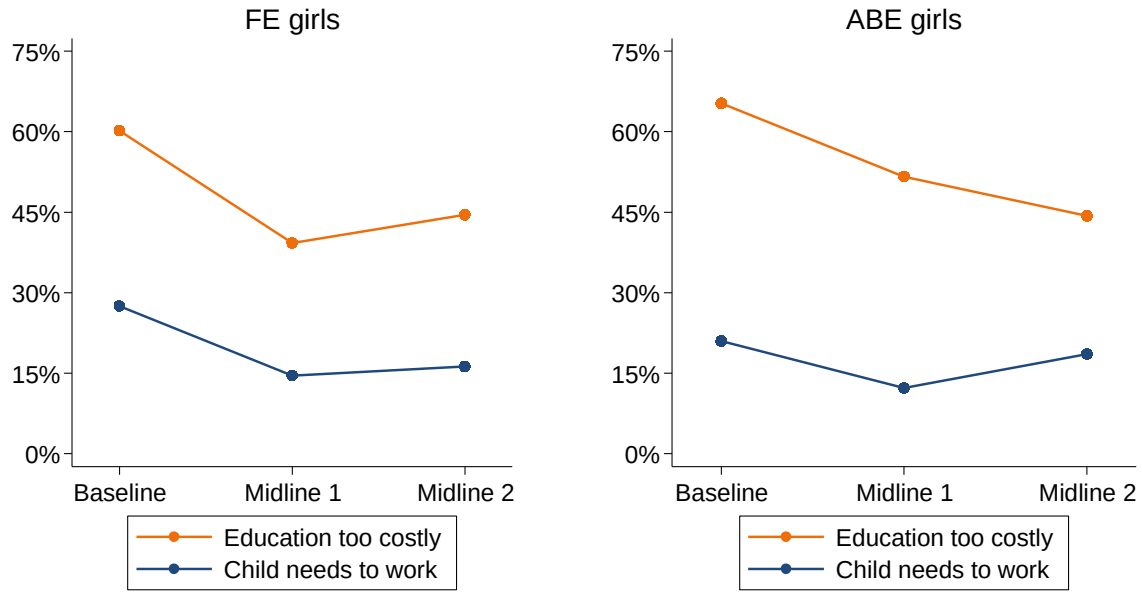
[87] FGD with CEC members, Banadir, Int. 101
[88] FGD with CEC members, Banadir, Int. 104

mobilising female caregivers through village savings and loans associations and enrolment drives, and messaging about education and rights of girls, especially girls with disabilities. To track progress in community attitudes, AGES assesses changes in (1) the percent of parents investing profits from income-generating activities in expenses related to girls' education and (2) changes in mothers' attitudes and practices towards new roles for girls. Along these lines, in this section, we first analyse changes in caregivers' willingness to financially support girls' education. We then discuss caregivers' aspirations for the level of education their daughter should achieve, and the relative importance placed on education.

Investment in Girls' Education

To better understand the value caregivers place on girls' education despite limited financial resources, we first analyse responses to the statement, "Even when funds are limited it is worth investing in [my daughter's] education".[89] We find extremely high rates of agreement with this statement across all evaluation rounds, at 97% at baseline, 98% at ML1, and 99.5% at ML2 for FE girls and 98% at baseline, 99% at ML1, and 97% at ML2 for ABE girls. As we will discuss further below, this result likely reflects social desirability bias, as caregivers are likely aware that opposition to girls' education is seen as undesirable.

FIGURE 14: ACCEPTABILITY FOR CHILDREN TO NOT ATTEND SCHOOL GIVEN FINANCIAL CHALLENGES



Indeed, the figure above shows a much different picture of financial support for girls' education. This figure shows responses to two questions asking about situations when it is acceptable for a child not to attend school: (1) if education is too costly and (2) if the child needs to work. As discussed in the above section on School Management and Governance, we find a relatively high level of acceptability for children to not attend school due to financial challenges: At ML2, 45% of FE caregivers and 44% of ABE caregivers stated it was acceptable

[89] For this and subsequent questions in this section, we examine results for FE and ABE girls, as an insufficient sample size of C1 NFE girls were asked questions about financial support for girls' education at ML2.

for a child not to attend if education was too costly and 16% of FE caregivers and 19% of ABE caregivers stated it was acceptable if the child needed to work.

In a positive sign, however, these attitudes appear to have improved since baseline, when the relative levels of acceptability were 60% and 27% respectively for FE girls and 65% and 21% respectively for ABE girls. For attitudes towards the cost of education, these declines were significant. This may reflect improved attitudes towards girls' education; however, it may also reflect a strengthening of economic circumstances which places less pressure on households and communities and thus leads to improving attitudes around financial constraints.

To better understand these dynamics, which may be highly contextual, in the below table we analyse the percent of FE caregivers stating that it is acceptable for a child not to attend school if education is too costly by a range of demographic characteristics.[90] We first find that the reported level of acceptability has declined from baseline to ML2 for every demographic group analysed; in most cases, the decline was substantial.

We find significant declines from baseline to ML2 for households headed by men, IDP households, caregivers who have some primary education or above, and for both households who have faced food insecurity over the past year and those who have not. Encouragingly, this pattern suggests an improvement in attitudes towards investment in girls' education among subgroups who might be expected to have less support—and less financial resources—for their girls to attend school. For example, we might expect male HoHs to have less support for girls' education than female HoHs, and indeed this pattern was seen at baseline. However, the significant and large decrease for male HoHs means that at ML2, their level of purported financial support for girls' education is now on par with that of female HoHs. Similarly, due to the higher levels of marginalisation and (potentially) economic distress faced by IDP households, we might expect those households to value investment in girls' education less; however, at ML2, levels of support are now similar for IDP and non-IDP households.

TABLE 51: ACCEPTABILITY FOR FE CHILDREN TO NOT ATTEND SCHOOL IF TOO COSTLY, BY DEMOGRAPHIC CHARACTERISTICS

Subgroup	BL	ML2	Difference	P-value
Zone				
Banadir	53.0%	43.1%	-9.9	0.19
Jubaland	64.2%	49.3%	-14.9	0.10
South West State	70.7%	42.2%	-28.5	0.06
Caregiver/household characteristics				
Female HoH	47.2%	45.2%	-2.0	0.73
Male HoH	67.4%	44.7%	-22.7	0.002*
IDP	64.3%	46.4%	-17.9	0.01*
Non-IDP	56.9%	44.0%	-12.9	0.08
Caregiver has no education or Madrassa only	58.3%	47.5%	-10.8	0.11

[90] We do not analyse results by pastoralism or girls' disability status due to low sample size for pastoralist households and girls with disabilities.

Caregiver has some primary education or above	65.5%	44.2%	-21.3	0.01*
Household has been hungry many or most days in past year	65.0%	40.3%	-24.7	0.004*
Household has never or rarely been hungry in past year	59.8%	46.7%	-13.1	0.05*

Interestingly, at baseline, we found that educated caregivers were *more* likely to state that it was acceptable for a child not to attend school if it was too costly. The higher support from non-educated caregivers may have come from the desire to provide their children with opportunities that they did not have. However, due to the significant decline in results for educated caregivers, at ML2 levels of support were similar across both of these groups.

Looking at results for ABE girls in Table 52 below, we find similar patterns to those for FE girls but with an even steeper decline for almost all subgroups. As with FE girls, we find significant declines in perceptions of the acceptability of withholding children from school if too costly for male HoHs, IDPs, and both food insecure and food secure households. Additionally, we find large and significant declines within Banadir and for female HoHs. These results are particularly noteworthy because of FE girls, these were two of the subgroups that saw the least change from baseline to ML2.

TABLE 52: ACCEPTABILITY FOR ABE CHILDREN TO NOT ATTEND SCHOOL IF TOO COSTLY, BY DEMOGRAPHIC CHARACTERISTICS

Subgroup	BL	ML2	Difference	P-value
Zone				
Banadir	73.1%	39.5%	-33.6	0.02*
Jubaland	59.8%	44.4%	-15.4	0.25
South West State	56.4%	50.0%	-6.4	0.52
Caregiver/household characteristics				
Female HoH	62.5%	35.7%	-26.8	0.02*
Male HoH	65.5%	48.0%	-17.5	0.04*
IDP	68.3%	37.5%	-30.8	0.002*
Non-IDP	59.2%	53.3%	-5.9	0.58
Caregiver has no education or Madrassa only	61.2%	47.8%	-13.4	0.21
Caregiver has some primary education or above	66.7%	41.7%	-25.0	0.25
Household has been hungry many or most days in past year	74.4%	35.0%	-39.4	0.02*
Household has never or rarely been hungry in past year	61.0%	46.6%	-14.4	0.04*

Despite these positive results, it is again worth noting that for both FE and ABE girls, 35% to 55% of caregivers in all subgroups stated that it was acceptable to withhold a child from school if education was too costly. This remains a very high level of agreement with this statement which may have negative implications for girls' attendance and enrolment in school. However, this result may also reflect the very real financial limitations faced by marginalised households in Somalia; indeed, in an FGD, one mother stated that, "There are no threats that prevent girls from going to school now, other than financial issues".[91] As such, while efforts to further improve community attitudes towards girls' education are warranted, there is also a need to enhance the affordability of schooling (including school materials) and support broader initiatives for income generation and livelihoods of households.

Lastly, caregivers were asked if they participated in a savings group, and if so, what the savings were used for. We report results for ML2 only due to limited sample size. We find that only 5% of the 429 caregivers who reported participating in a savings group at ML2 said that they spent savings on education. Among these 20 caregivers, 70% stated that they invested in both their sons' and daughters' education, 15% on only their daughters' education, 10% on only their sons' education, and 5% on other education expenses. These results suggest that spending on education was relatively equitable, but that education was not a highly prioritised use of savings for caregivers. Of course, this could be explained by the circumstances faced by households, especially over the previous year, with drought intensifying in programme areas. Food purchases were a slightly more common use of savings, which is at least somewhat consistent with a drought-induced drawdown of household savings. In addition, it is worth noting that the fungibility of money means that allocation of savings to food or other purchases can still impact education spending by freeing up resources to pay educational expenses later, or by reducing the need to remove children from school to earn money for the family.

Caregiver Attitudes and Practices

We now discuss changes in caregivers' expectations for their girls' education and attitudes around gender norms that may influence decisions about girls' schooling. First, we note that at ML2, 100% of FE and ABE caregivers agreed that girls were just as likely to use their education as boys. Furthermore, in the table below, we show changes in caregivers' aspirations for the level of education they want their daughter to achieve. At both baseline and ML2, we find that most caregivers aspire for their girls to attend college or university—around 90% at ML2 for FE and ABE girls. For FE girls, caregivers' desire for girls to reach college or university increased slightly from baseline to ML2, but not significantly. There was almost no change in caregivers' aspirations for their girl's education for ABE girls.

TABLE 53: CAREGIVERS' ASPIRATIONS FOR GIRLS' EDUCATION

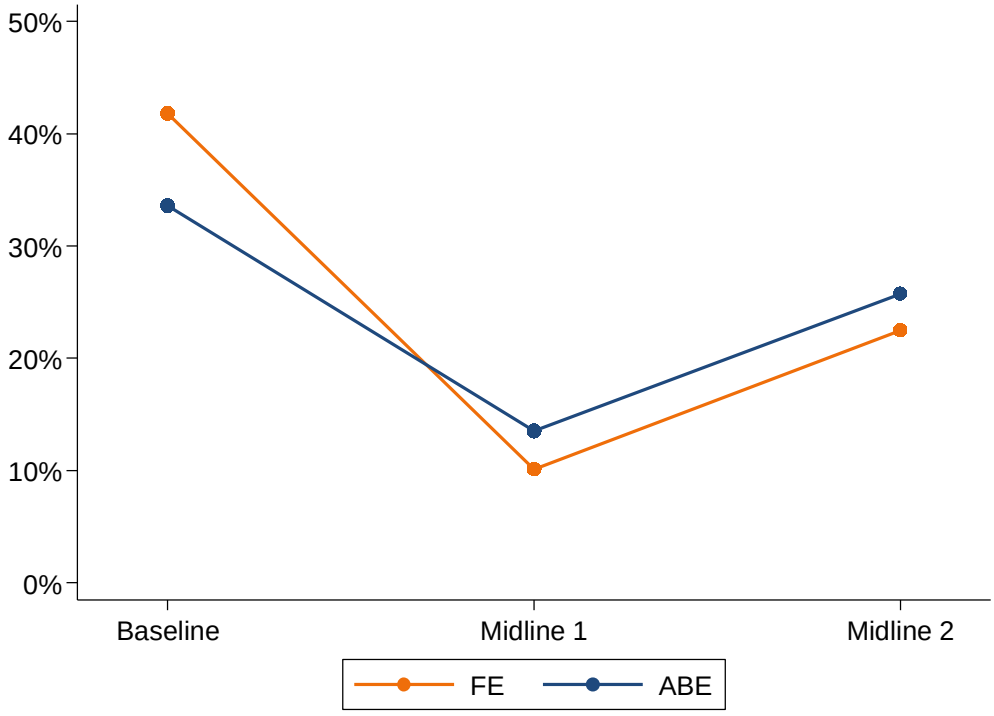
	FE			ABE		
Education level	BL	ML2	Difference	BL	ML2	Difference
Primary or less	0.4%	0.0%	-0.4	0.3%	0.0%	-0.3
Secondary	11.2%	7.6%	-3.6	10.0%	10.4%	0.4
College or university	88.4%	92.4%	4.0	89.7%	89.6%	-0.1

[91] FGD with mothers, Bay, Int. 208

While these results suggest a high level of support for girls' education, it is important to note that this support may be susceptible to social desirability bias and is also phrased very abstractly. In other words, caregivers may wish that their girl is highly educated in the absence of any constraints to do so; however, when faced with the reality of financial, cultural, security, and other limitations, actual levels of support for girls' schooling may substantially decline. This is also evidenced in the above subsection where we find a high level of abstract support to finance girls' education, but much lower levels of support when caregivers are asked about more tangible situations where decisions must be made about constrained financial resources.

To further understand community attitudes towards girls' education, we thus examine caregivers' perceptions of the acceptability of withholding a child from school because they need to do housework. Figure 15 shows that FE caregivers' perceptions of acceptability have significantly declined since baseline, when 42% of caregivers agreed that this was an acceptable reason for a child to miss school. However, rates have increased from the ML1 value of 10%, to reach 22% at ML2. Furthermore, rates did not decline significantly for ABE girls' caregivers from BL to ML2, and similarly increased from ML1 to ML2.

FIGURE 15: ACCEPTABILITY FOR CHILDREN TO MISS SCHOOL DUE TO HOUSEWORK RESPONSIBILITIES



To better understand why this may be the case, we analyse changes in the housework responsibilities of girls and in the frequency with which girls have been unable to attend school due to housework. The results in Table 54 suggest that girls' housework burdens may have substantially increased since baseline. For FE girls, we find that significantly more caregivers reported that their girls spend a whole day doing housework, somewhat more caregivers reported that their girls spend half a day, and somewhat fewer caregivers reported that their girls spend a few hours a day or less. Furthermore, caregivers were significantly more likely to say that housework sometimes stops their girl from attending school and significantly less likely to say that housework does not stop her from attending school. Similarly, for ABE girls, we find that significantly more caregivers reported that their girls spend the whole day on housework, significantly fewer reported that their

girls spend little time on housework, and significantly fewer reported that housework does not stop their girl from attending school.

TABLE 54: FE AND ABE GIRLS' HOUSEWORK RESPONSIBILITIES AND IMPACT ON SCHOOL ATTENDANCE

	FE			ABE		
	BL	ML2	Difference	BL	ML2	Difference
Average time spent by girl doing housework						
Whole day	0.0%	8.1%	8.1*	1.4%	11.7%	10.3*
Half day	28.8%	34.4%	5.6	36.8%	53.3%	16.5
Quarter day/a few hours	37.8%	33.9%	-3.9	30.2%	22.1%	-8.1
Little time/an hour	33.3%	23.5%	-9.8	31.6%	13.0%	-18.6*
Housework prevents girl from attending school						
Yes, not enrolled because of this	4.4%	1.8%	-2.6	2.4%	3.9%	1.5
Yes, stops her often	1.5%	4.0%	2.5	2.0%	5.2%	3.2
Yes, stops her sometimes	5.8%	16.0%	10.2*	10.6%	20.8%	10.2
No, does not stop her	88.4%	77.3%	-11.1*	85.1%	70.1%	-15.0*

In the below table, we further report results for C1 NFE girls who self-assessed their housework responsibilities at ML1 and ML2. First, we note that NFE girls reported substantially higher housework burdens than FE and ABE girls, with over three-quarters of NFE girls reporting that they spend half a day or a full day on housework at ML2. As with FE and ABE girls, we also find a higher reported housework burden at ML2 than in prior rounds, with the percent of girls reporting they spend a whole day on housework increasing significantly and the precent spending a quarter day on chores decreasing significantly.

However, we find more promising results for the impact of housework on NFE girls' schooling. At ML2, significantly fewer C1 NFE girls stated that they were not enrolled in school due to housework. Given that NFE centres specifically target out-of-school marginalised girls, this may be directly suggestive of programme impact.

TABLE 55: C1 NFE GIRLS' HOUSEWORK RESPONSIBILITIES AND IMPACT ON SCHOOL ATTENDANCE

	ML1	ML2	Difference	P-value
Average time spent by girl doing housework				
Whole day	24.7%	34.4%	9.7	0.02*
Half day	47.1%	44.4%	-2.7	0.42
Quarter day/a few hours	23.1%	12.7%	-10.4	0.01*
Little time/an hour	4.7%	8.5%	3.8	0.10
Housework prevents girl from attending school				
Yes, not enrolled because of this	12.2%	1.9%	-10.3	<0.001*
Yes, stops her often	9.8%	13.5%	3.7	0.31
Yes, stops her sometimes	30.2%	32.8%	2.6	0.48

No, does not stop her	47.5%	51.7%	4.2	0.36

Below, we further disaggregate results for whether housework prevents FE and ABE girls from attending school. For both FE and ABE girls, we see that across almost all subgroups, caregivers were more likely to report that housework prevented their girl from attending school at least occasionally. The one exception to this pattern is for FE girls in Jubaland, where caregivers were slightly (but not significantly) less likely to report an impact of housework on attendance. In contrast, caregivers of FE girls reported a significantly greater impact of housework on attendance in Banadir; ABE girls also had the highest burden of housework responsibilities in Banadir. The reason for this is unclear; indeed, these patterns contradict expectations of Banadir as a more "progressive" region due to its proximity to Mogadishu.

Looking at caregiver characteristics, for FE and ABE girls, Table 56 shows a significant increase in the impact of housework on attendance for male-headed households and IDP households. For the former case, this data may discouragingly suggest that households headed by men place lower value on girls' education and instead prioritise girls' involvement in household duties. For the latter, the findings may reflect the fact that IDP families have less access to services and facilities, and thus tend to have a greater burden of domestic responsibilities. Additionally, for ABE girls, we find a significant increase for girls whose caregiver has no education or who attended Madrassa only. This may reflect the more challenging household and economic circumstances faced by these households, a lack of prioritisation for education, or a combination of both factors.

Last, while sample size is insufficient to analyse the change in housework responsibilities of girls with disabilities since baseline, it is important to note that at ML2, caregivers of these girls in FE schools were substantially more likely to report that housework prevented school attendance than in other subgroups. As with the findings for male HoHs, this may discouragingly suggest that caregivers tend to underprioritise education for girls with disabilities.

TABLE 56: IMPACT OF HOUSEWORK ON FE AND ABE GIRLS' SCHOOL ATTENDANCE, BY SUBGROUP

	FE			ABE		
Subgroup	**BL**	**ML2**	**Difference**	**BL**	**ML2**	**Difference**
Zone						
Banadir	7.1%	26.2%	19.1*	25.5%	40.5%	15.0
Jubaland	16.7%	14.1%	-2.6	6.3%	16.7%	10.4
South West State	11.4%	24.4%	13.0	9.5%	22.7%	13.2
Caregiver/household characteristics						
Female HoH	12.7%	16.4%	3.7	14.3%	25.0%	10.7
Male HoH	11.1%	24.7%	13.6*	15.2%	32.7%	17.5*
IDP	10.6%	24.1%	13.5*	14.4%	34.0%	19.6*
Non-IDP	12.9%	20.7%	7.8	15.7%	23.3%	7.6
Caregiver has no education or Madrassa only	11.8%	21.3%	9.5	15.9%	34.8%	18.9*

Caregiver has some primary education or above	12.5%	22.1%	9.6	6.7%	16.7%	10.0
Girl characteristics						
Disability	-	31.3%	-	-	32.1%	-
No disability	11.7%	18.2%	6.5	14.3%	28.6%	14.3*

Table 57 additionally shows results for C1 NFE girls. In line with results in Table 55, we find that the impact of housework on school attendance declined—though not significantly—for most subgroups, except for Banadir. Despite these declines, however, at ML2, almost all C1 NFE subgroups were more likely to report that housework was detrimental to their school attendance than FE and ABE subgroups. The impact of housework was particularly acute for IDP and girls with disabilities, similar to results in Table 56, showing how the interaction of various dimensions of marginalisation may limit girls' ability to attend school and successfully learn.

TABLE 57: IMPACT OF HOUSEWORK ON C1 NFE GIRLS' SCHOOL ATTENDANCE, BY SUBGROUP

Subgroup	ML1	ML2	Difference	P-value
Zone				
Banadir	51.6%	56.3%	4.7	0.48
Jubaland	44.3%	35.5%	-8.8	0.41
South West State	61.5%	44.9%	-16.6	0.11
Girl characteristics				
IDP	57.0%	53.3%	-3.7	0.62
Non-IDP	47.6%	44.8%	-2.8	0.62
Disability	60.3%	50.0%	-10.3	0.13
No disability	49.5%	46.4%	-3.1	0.58

Overall, in contrast to results found regarding financial support for girls' education, these results suggest a potentially concerning decline in caregiver attitudes towards the value of girls' education and a high housework burden which impacts school attendance for FE, ABE, and NFE girls. It is worth noting, however, that in qualitative interviews, most respondents placed high value of girls' education. Mothers in Bay, for example, stated the following:

> *I find it wonderful that my daughters are learning. Girls used to be told to stay at home, but nowadays they go to school instead. The advantages for their kids will last throughout their lives. They claim that an educated girl represents an educated society. Education for girls is important for the future.*
>
> - FGD with mothers, Bay, Int. 207

> *In the past, girls weren't taught anything, and I thought that all they should do was housekeeping, but I've since understood that girls and boys should educate equally, and they should both work hard in the classroom.*
>
> - FGD with mothers, Bay, Int. 208

A religious leader in Lower Shabelle—an important stakeholder with influence over broader community attitudes regarding girls' education—similarly expressed support for girls' education. However, he also stated that parents tended to prioritise boys' education over girls':

> *I have positive feelings about the education of Somali women. God granted girls the same rights to education as he did to boys when it comes to education, [though] parents prioritise boys more. In the past six years, I have opened a school and worked as a teacher at various schools. I strongly believe that girls learn better than boys, and I have taught many girls.*
>
> - KII with religious leader, Lower Shabelle, Int. 308

Once again, however, the impact of social desirability bias on these answers must be considered. As such, the overall picture from this section is of mixed community attitudes towards the value of girls' education. A key takeaway is that while many individuals may express support for girls' education in the abstract, many barriers to this education remain which reduce *actual* levels of support. These barriers seem to differentially affect girls facing various dimensions of marginalisation—such as IDP girls or girls with disabilities—thus serving as a major challenge for the success of AGES programming.

7.6 Increased Self-Efficacy

This section examines questions and attitudes related to positive youth development as well as access to protection services. The reader should note that the following section focuses exclusively on girls from the NFE Cohort 1 that responded to the questions related to self-efficacy in both the ML1 and ML2 evaluations.

Positive Youth Development

During the Midline 1 and Midline 2 evaluation rounds, data on positive youth development was collected from the C1 NFE girls using indicators from the Chinese Positive Youth Development Scale (CPYDS). This globally recognised scale assesses various aspects of youth development such as resilience and confidence. Seven indicators related to self-efficacy were chosen from the CPYDS. The table below lists the seven indicators and tracks the percentage of girls who either agreed or strongly agreed with the statements from ML1 to ML2.

TABLE: CHANGE IN AGREEMENT WITH CPYDS QUESTIONS OVER TIME

CPYDS Question	ML1	ML2	Difference
Lack of control of life	40.5%	57.6%	+17.1
Lack of solutions to problems	40.8%	54.2%	+13.4
Inability to change life	28.6%	56.1%	+27.5
Helplessness	41.2%	49.6%	+8.4
Fate not in hands	54.6%	63.4%	+8.8

Determine own life	67.2%	72.5%	+5.3
Ability to complete tasks	83.6%	84.7%	+1.1

Control of own life

The first indicator measures the girls' perception of control over their lives. When prompted with the statement, "I have little control of things that happen in my life," 40.5% agreed or strongly agreed during the ML1 round, while 57.6% agreed or strongly agreed in the ML2 round. Across the first five indicators, girls demonstrate a weaker perception of self-efficacy. For instance, the percentage of girls who feel they have little control over their lives increased from a minority (40.5%) to a majority (57.6%) across the rounds.

The increase in the perception of a lack of control occurred in four out of the five regions for the C1 NFE girls including in Banadir, Lower Shabelle, Gedo, and Lower Juba. Most notably, in Lower Shabelle the percentage agreeing with the statement increased from 14.3% in ML1 to 85.7% in ML2 (while this is a large percentage increase, the sub-sample of C1 NFE girls in this region who responded to the CPYDS questions consisted of only 21 respondents). The exception is Bay region, where the share of girls who feel they lack control over their life *decreased* from 39.6% to 29.2%.

Disaggregating the results by other key variables produces similar trends. For girls without disabilities, 42% agreed in ML1, while just over 50 percent agreed in ML2. For girls with disabilities, 36.2% initially agreed, while 63.7% agreed during the ML2 round. Households with female heads of household transitioned from 34.7% to 54.1% while those without a female head of household changed from 45.1% to 60.6%. For non-Maay speakers, 43.1% initially agreed, while 61.2% agreed in the most recent round. Maay speakers retained a minority in agreement with the statement, switching from 33.5% to 48.6%.

Solutions to problems

The second indicator measures the girls' perception of how they can solve problems. When given the statement, "I do not have any solutions for some of the problems I am facing", the percentage of girls in agreement increased from 40.8% to 54.2% between rounds ML1 and ML2.

The regions of Bay and Lower Juba saw increases mirroring the overall change in response (from 29.1% to 39.6% and from 38.3% to 61.7%, respectively). The percentage of girls believing they have a lack of solutions in Banadir remained consistent. Once again, Lower Shabelle saw a large increase in percent agreeing from 28.6% in ML1 to 71.4% in ML2. For this indicator, Gedo also demonstrated a significant increase from 17.6% to 64.7%.

Girls without and with disabilities began with similar percentages in agreement during ML1 (40.9% and 40.6%, respectively). However, during the ML2 round, only 49.6% of girls without disabilities agreed with the statement, while 58.5% of girls with disabilities now believed that they did not have any solutions for some of the problems they were facing. Households with female heads of household reported 34.7% of girls in agreement with the statement in ML1, and this increased to 54.2% in ML2, a similar result to households without a female head of household which saw an increase from 44.3% to 56.3% in agreement. Both Maay and non-Maay speakers transitioned from a minority in agreement to a majority in agreement with the statement between the two rounds.

Ability to change

The third indicator measures the extent to which girls think they are able to change the course of their life. When prompted with the statement, "I cannot do much to change things in my life", the initial percent in

agreement was only 28.6%. During the ML2 evaluation, this percentage had increased to 56.1% agreeing they lacked the ability to change, which is the largest increase across all indicators for the C1 NFE girls.

The transition from a minority of the sample agreeing that they cannot do much to change their own life to a majority in agreement with the statement transcended all sub-groups. All regions saw increases in the percentages. In Gedo, of the seventeen C1 NFE girls who responded to the CPYDS questions, initially none of them agreed with this statement, but during the ML2 evaluation, nine out of the seventeen switched to agreeing that they did not have the ability to change their own life.

Girls with and without disabilities, Maay and non-Maay speakers, and households with both female and non-female heads of households all saw percentage increases that resulted in most of the girls in each sub-group agreeing that they lack the ability to enact changes.

Helplessness

The fourth indicator measures the girls' perceptions of helplessness. When prompted with the statement, "When I face life difficulties, I feel helpless", 41.2% initially agreed during the ML1 round, while 49.6% agreed during ML2. This change diverges from the first three indicators, as it is the first instance in which a majority *disagreed* with the statement in both rounds.

In terms of region, four out of the five regions still saw increases in percentage agreeing from ML1 to ML2: Banadir (51.2% to 51.9%), Lower Shabelle (19.0% to 57.1%), Bay (29.2% to 43.8%) and Lower Juba (34.0% to 55.3%). However, Gedo saw a substantial decrease (47.1% to 23.5% agreeing).

Girls without a disability transitioned from 38.3% agreeing that they feel helpless in the face of difficulties to 43.3% agreeing with the statement. Girls with a disability changed from 49.3% to 55.6% in agreement, indicating that more than half of girls with disabilities felt helpless in the face of life challenges during the ML2 round. Disaggregating the results by Maay versus non-Maay and female head of household versus non-female head of household results in similar trends to the overall change in the sample.

Fate not in hands

The fifth indicator evaluates how the girls perceive their fate. Overall, 54.6% of the individuals initially agreed with the statement, "I feel my life is determined by others and fate". During the ML2 round, 63.4% of girls agreed with the statement. This indicator differs from the preceding four because there was already an initial majority in agreement with the statement in the ML1 round.

While the regions of Banadir and Lower Juba saw increases in the percentage of girls agreeing with the fact that their fate is not in their own hands (from 61.2% to 76.7% and from 31.9% to 63.8%, respectively), the other three regions of Lower Shabelle, Gedo, and Bay saw *decreases* in the percentages agreeing with this statement. Bay started off with only 39.6% in agreement with the idea that their fate was out of their hands, but during the ML2 round this had decreases to only 27.1% in agreement, indicating that an almost three-fourths majority felt their fate was self-determined in this region.

Non-Maay speakers saw an increase from 58.5% to 71.8% in agreement, while Maay speakers increased from 31.0% to 41.9%, indicating a majority still *disagreed* with the statement during ML2. Girls with disabilities and without disabilities, as well as girls belonging to female and non-female led households also saw increases in the proportion of girls perceiving their life as out of their control.

Determine own life

The sixth indicator measures the girls' perception of their ability to determine their own life, which differs from the preceding indicators because it is a positive measure of self-efficacy. When faced with the question,

"I believe things in my life are mostly determined by me", 67.2% initially agreed, while 72.5% agreed in ML2. This result is important because it demonstrates how positive self-efficacy traits are also increasing; these increases are not limited to negative self-efficacy traits such as the belief that one's fate is outside of their own hands.

All regions saw increases in the percent agreeing that the girls determine their own lives, except for Gedo, which saw a decrease from 88.2% in agreement to 29.4% in agreement (similar to Lower Shabelle, Gedo has a very small sub-group size of only seventeen girls responding to the questions in this cohort, so a few girls changing their answers between ML1 to ML2 affected the proportions greatly).

The sub-groups such as girls with a disability, girls without a disability, Maay speakers, non-Maay speakers, girls in female-led households, and girls in non-female led households all saw increases in the percentages agreeing that they determine their own life between rounds ML1 and ML2.

Task completion

The seventh and final indicator measures the girls' belief in their ability to finish tasks. The statement provided was, "I can finish almost everything that I am determined to do", which is once again an indicator of positive self-efficacy. There was a small increase in the overall percentage of girls agreeing with this statement from 83.6% in the ML1 round to 84.7% in the ML2 round.

The breakdown by region shows that the percentage agreeing that they can complete tasks increased in two regions: Banadir (from 85.3% to 89.9%) and Lower Juba (from 66.0% to 78.7%). While the percent who disagreed increased in Lower Shabelle, Gedo, and Bay.

Like the sixth indicator, the sub-groups such as girls with a disability, girls without a disability, Maay speakers, non-Maay speakers, girls in female-led households, and girls in non-female led households saw small increases or stability in the percentages agreeing that they can finish tasks.

Synthesis

All seven of the statements related to the self-efficacy indicators experienced increases in the percentages of girls agreeing with them from ML1 to ML2. These increases were evident with statements that were indicative of a *weaker* perception of self-efficacy and self-confidence (e.g., more girls agreed with the statement "When I face life difficulties, I feel helpless"), but there were also higher percentages of girls agreeing with statements that were indicative of a *stronger* perception of self-efficacy (e.g., "I believe things in my life are mostly determined by me"). Breaking down these overall trends into sub-groups such as region, language, disability presence, and head of household characteristics does reveal certain trends as well as anomalies. For example, Lower Shabelle and Gedo experienced decreases in agreement percentages for indicators 4 and 5 while the other regions experienced increases.

While there are some trends and anomalies at the sub-group level, these groups all appear to follow the overall trend of a declining perception of self-confidence for the C1 NFE girls. This rise in agreements with statements related to helplessness and an inability to change one's life transcends region, and all the other variable groups discussed in this section. Additionally, the two indicators reflecting positive aspects of self-perception (indicator six on determining one's own life and indicator seven on completing tasks) saw very small increases in the percentage of girls agreeing with the statements compared to the indicators reflecting negative aspects of self-perception. Due to this transcendence, it is possible that the girls are becoming less optimistic about their capacity to reach certain goals and decide their own fate as they become older and experience additional life challenges. The initial average age of the C1 NFE girls that were interviewed in the ML1 round was 20.2 years old, with a minimum of 10 and a maximum of 27 years old. It is possible that as

these girls become older and encounter obstacles related to life and work changes, their perception of their ability to affect change in their lives may worsen.

It is also important to note that the statements related to the CPYDS indicators focus on the lives of the girls. When asked about helplessness, the statement refers to "life difficulties" in general. It is possible that splitting up the difficulties into different sectors such as social, school, family, and work difficulties may also result in different levels of self-confidence and self-efficacy depending on the sector of the girls' lives under examination. Similarly, it can be difficult to parse the extent to which changes in sentiments reflect external fluctuations in the environment versus internal feelings. This is true broadly – as short-term events or issues are likely to create recency biases among respondents. It is especially relevant in Somalia, a particularly dynamic environment where short term changes can include substantive changes in personal safety and/or food security – which are likely to affect ideas such as control, helplessness, and fate.

Access to Protection Services

This section evaluates the C1 NFE girls' access to protection services across ML1 and ML2 in both the school and the community, and the channels through which they can report abuse, harassment, or exploitation.

School

During the ML1 round, 95% of this cohort answered "Yes," when asked if they have an avenue to report any form of harassment, abuse, or exploitation at the school. During the ML2 round, this percentage decreased, with 81.7% of girls responding affirmatively that they have a channel to report abuse in the school, which is a sizeable decrease.

When asked who their top choices are for reporting said abuse, the top choice remained the head teacher or other teachers across ML1 and ML2. The other popular choice was the "Other" category, which included specified responses such as parents, husbands, and sisters.

Community

During the ML1 round, 88.9% of the girls responded "Yes," when asked if they have an avenue to report any form of harassment, abuse, or exploitation in the community. During the ML2 round, this percentage decreased to 77.5%. This decrease in the girls' perceived ability to report community abuse mimics the trend seen above in the school scenario.

Also like the school scenario, the top choices for reporting abuse in the community included teacher and head teacher, but the third most popular option in this case was the police. During a risk mapping exercise with a group of C1 NFE girls, one individual in Bay reported being afraid of certain people under the influence of alcohol in the community shops. When asked how this could be solved, she replied that soldiers or police should be notified and be responsible for maintaining order and protecting the peace of the community. Another girl expressed her fear regarding her walk home through an alley in which boys attempt to harass the schoolgirls. Her proposed solution involved widening the alley, and another individual suggested that parents should be responsible for the protection of their children in such circumstances.[92]

Synthesis

Despite the drop in percentage of girls agreeing that they have avenues to report abuse in both the schools and the community, the risk mapping exercises demonstrate how the interviewed girls do have solutions regarding potential issues, and these mostly correspond to the top choices for reporting the abuse (e.g.,

[92] Risk Mapping FGD with Girls, Bay, Int. 407.

reporting the issues to the police, head teacher, or their parents to deal with the root cause). Similar to the ML1 choices for who to report abuse to, the CEC and GEF options remain underutilised resources in the ML2 findings. It appears that the teachers, police, and family members remain the preferred option for reporting abuse.

7.7 Strengthened Economic Circumstances

The final intermediate outcome for the original (FCDO-supported) baseline girls focuses on their economic and employment outcomes, specifically among girls who completed the NFE programme, whose purpose was to provide skills that would allow girls to enter the labour market or start small enterprises. Our focus in this section is on assessing the extent of employment among C1 NFE girls and the per-month income they report earning.

It is important to note limitations in the measurement of employment. both employment and, to a greater extent, income. As we point out in our discussion of the results below, many NFE girls are currently engaged in domestic work, but it is not always clear whether they are compensated for this work and what form (in cash or in-kind) that compensation took.[93] It is also not clear whether respondents understood domestic work (described as "non-agricultural domestic chores inside the home, such as child-raising and cooking") to include domestic work in *their own homes* or exclusively domestic work performed in the homes of other families. This was a limitation of the survey we employed, which captured information on occupation from a panel of 253 NFE girls across the two rounds.

The table below reports the share of NFE girls who are employed in different sectors or types of job, by round. At the broadest level, it is clear – from the top row – that the share of C1 NFE girls who do not have a job has declined significantly since the last evaluation round. Domestic work remains the most common form of work for girls, and the share of girls engaged in domestic work nearly doubled; this is the most important source of new employment, accounting for 45.1 percent of all new employment in the sample.

TABLE 58: EMPLOYMENT, BY JOB TYPE, AMONG C1 NFE GIRLS IN THE ML1 AND ML2 ROUNDS

Job Category	Share of Girls. ML1	Share of Girls, ML2
No occupation or job	62.1	29.6
Domestic work	18.6	33.2
Unskilled sales/service worker (e.g., hawker, shoe cleaner, domestic helper, cleaner)	9.5	13
Student	4.7	5.9
Sales/Service worker (e.g., waiter, retailer or clerk at shop, etc.)	2	7.5
Other	0.4	5.1
Trades, craft workers, extractive industries	1.6	3.2

[93] While we are able to assess how many girls engaged in domestic work *also* report income in the previous month, this income may come from other sources, so it is not a clear indicator that domestic work was remunerative.

Farming, fishing, pastoralism	1.2	2
Professional or managerial positions	0	0.4

Outside domestic work, gains in employment numbers were smaller but still important – there are increasing numbers of girls engaged in service-sector work, in catch-all other forms of work, and in a broad category of physical trades, crafts, etc. Combining the two service-sector categories – which are ambiguously distinguished by their degree of formality – 20.5 percent of girls are engaged in this form of work. This aligns with much of the qualitative data from the ML2 round, in which mothers often indicated that their daughters are employed in shops or vendor stalls, often assisting their mother or taking over portions of the business.[94] Several women described having their daughters staff the store when they are not personally present, and one reported that her daughter handles the accounts, presumably doing routine bookkeeping for her mother's shop.[95] We provide these examples, in part, to solidify the idea that service-sector work is of particular importance in the overall composition of jobs.

The trends in the table do raise a question regarding the nature of domestic work, as noted above. Average income – over the previous month – among girls employed in domestic work is lower than that of girls engaged in other forms of employment, but simultaneously higher than that of girls who are unemployed. This suggests that domestic work *might* generally be compensated; on the other hand, the 68.3 percent of girls employed in domestic work reported no income over the past month. It is likely that a significant share of girls engaged in domestic work are performing this work in their own households or otherwise not receiving monetary compensation for the work, though the exact share cannot be determined from our data.

The last trend in income-generating activities, not captured in the table above, concerns the share of NFE girls who have their own small business. At ML1, 9.9 percent of NFE girls had their own small business, rising to 15.4 percent at ML2. These businesses continue to centre on the sale of prepared food and other forms of retail vending.

Turning to overall income, we note an important caveat in our measure of income which consists of self-reported income over the past month. Issues related to accurate recall, volatility in income (especially as we captured information about just a single month of income), and other issues are well-known impediments to accurate data collection regarding income. A bigger problem is that a large proportion of girls did not know their income over the past month – out of 1474 girls asked about their income in this evaluation round (all NFE girls, of all cohorts), 42.2 percent did not know their income. Among the C1 NFE cohort analysed here, just 191 girls (of 302) reported their income, including girls who reported earning no income at all. This issue is exacerbated by the fact that we seek to compare changes in income over time, meaning that we need to compare girls' income from ML1 to ML2; because additional girls did not know their income at ML1, our panel sample of girls with information in both rounds is limited to just 111 girls.

Respondents reported income in a variety of currencies (US Dollars, Somali Shillings and, in Gedo, Kenyan Shillings), which we converted to US dollars. We consider two sets of girls – both panels tracked from ML1 to ML2. The first consists of all girls who reported their income (even if they reported having no income) in both rounds. The second restricts this set of girls by excluding outliers – girls whose income changed dramatically across rounds. Outliers may represent inaccurate data. In the table below, we report results for

[94] FGD with Mothers, Lower Juba, Int. 212; FGD with Mothers, Banadir, Int. 204; FGD with Mothers, Banadir, Int. 205.
[95] FGD with Mothers, Banadir, Int. 204.

the first set of girls – the full panel of C1 NFE girls for whom we captured income information in both rounds – in the first three columns. In the right-most three columns, we report the same analysis excluding outliers, where we define outliers as girls whose monthly income changed by more than $500 from ML1 to ML2.[96]

TABLE 59: MEAN MONTHLY INCOME AMONG C1 NFE GIRLS, BY ROUND

	Full "Income Panel"			Income Panel, Excluding Outliers		
Zone	**Obs.**	**ML1**	**ML2**	**Obs.**	**ML1**	**ML2**
Overall	111	$51.95	$155.17**	99	$16.82	$33.68
Banadir	56	$6.63	$86.57*	53	$6.46	$25.62**
Jubaland	29	$169.56	$400.93	20	$42.21	$61.35
South West State	26	$18.41	$28.82	26	$18.41	$28.82

As shown in the table, income increased in all three geographic zones from ML1 to ML2, though to somewhat varying degrees. Measured as a share of their starting point, gains were largest in Banadir, where monthly earnings were particularly low in ML1. The gains in Banadir are statistically significant regardless of which version of the results we use, but this is due, in part, to the larger sample size available for Banadir. A conservative estimate of increases in mean income is that they increased by $16.86 from ML1 to ML2; larger gains can be seen in the analysis that includes outliers, but we urge caution when relying on those results, given their inclusion of some particularly large outliers.[97]

When we dig deeper into patterns of income across girls, the strongest predictor of increased income from ML1 to ML2 is being engaged in *either* education or work. Girls who are idle – out-of-school and not employed – have monthly incomes $219 lower than other girls of the same age and in the same geographic area. Older girls also experienced greater gains in income, though this effect is not statistically significant.

While the gains in income reported above are notable, and employment rates have increased – even if we discount the increased share of NFE girls engaged in domestic work – key barriers to income generation remain. The most important, according to qualitative interviewees, is a lack of startup capital for opening new businesses.[98] Mothers we interviewed described new skills gained by their daughters, including the ability to count, keep accounts, and so forth. However, without savings or credit, it is not possible to start a small business.[99] Other women cited security concerns that might affect one's ability to start a business, and one mother reported that her daughter – who helps run a shop/stall for the family – needs to be accompanied

[96] For instance, a girl whose income increased from $100 to $800 across rounds would be excluded. This logic applies to declines in income as well – we treat a decline from $800 to $100 as an outlier and exclude it from analysis.

[97] While findings related to average income and employment are encouraging, other indicators are less so. The share of girls who report a non-zero income has not increased over time. This binary measure provides an alternative to mean income, which avoids the issue of uncertainty regarding total earnings, as we assume respondents are better able to remember whether they earned any money than the precise amount they earned.

[98] Many of the qualitative interviews cited were conducted with mothers of cohort girls. These girls were not exclusively engaged in NFE programmes. However, the insights provided by mothers of ABE or FE girls are often still useful, because they are drawn from the same communities and – in the case of mothers of FE girls, who are often still in school – are describing what they see as barriers to starting a business in hypothetical terms.

[99] FGD with Mothers, Banadir, Int. 201; FGD with Mothers, Banadir, Int. 202; FGD with Mothers, Banadir, Int. 205; FGD with Mothers, Bay, Int. 207.

home by one of her brothers if she is at the shop too late into the evening.[100] One group of mothers also noted that community attitudes affect their daughters when they try to work or start businesses – as they described, people in the community say "they are not like males" and "ask them why they are working."[101] Notably, these mothers described a shift when a girl becomes successful: "as soon as she makes progress, everything is asked of her."[102]

[100] FGD with Mothers, Banadir, Int. 201; FGD with Mothers, Banadir, Int. 204.
[101] FGD with Mothers, Banadir, Int. 204.
[102] FGD with Mothers, Banadir, Int. 204.

USAID Cohort Results

- Cohort 4 Non-Formal Education (C4 NFE) Girls
- Cohort 5 Non-Formal Education (C4 NFE) Girls

8. Learning - C4 NFE Cohort

8.1 Aggregate Learning Outcomes

One of the primary outcomes of the AGES programme for the C4 NFE cohort of girls is improving the girls' numeracy and literacy skills, specifically their ability to read in Somali. The purpose of the 11-month NFE programme was to provide C4 NFE girls with basic literacy and numeracy skills that could prove useful in the labour market. In this section, we present findings on the aggregate changes in learning outcomes observed among C4 NFE girls from ML1 to ML2 followed by changes in the specific subtasks in numeracy and literacy assessments. Next, we present the findings on the trends among girls with disabilities and other demographic sub-groups. Finally, we analyse the extent to which the intermediate outcomes are predictors of changes in numeracy and literacy.

Before we proceed with the findings on learning outcomes, we describe the composition of the C4 NFE cohort of girls in terms of region and age. The table below shows the distribution of 661 C4 NFE girls across regions and age groups. The 661 C4 NFE girls participated in the ML1 and present ML2 evaluations. As shown in the table, the C4 NFE cohort is slightly older than school-age.

TABLE 60: DISTRIBUTION OF C4 NFE GIRLS BY REGION AND AGE

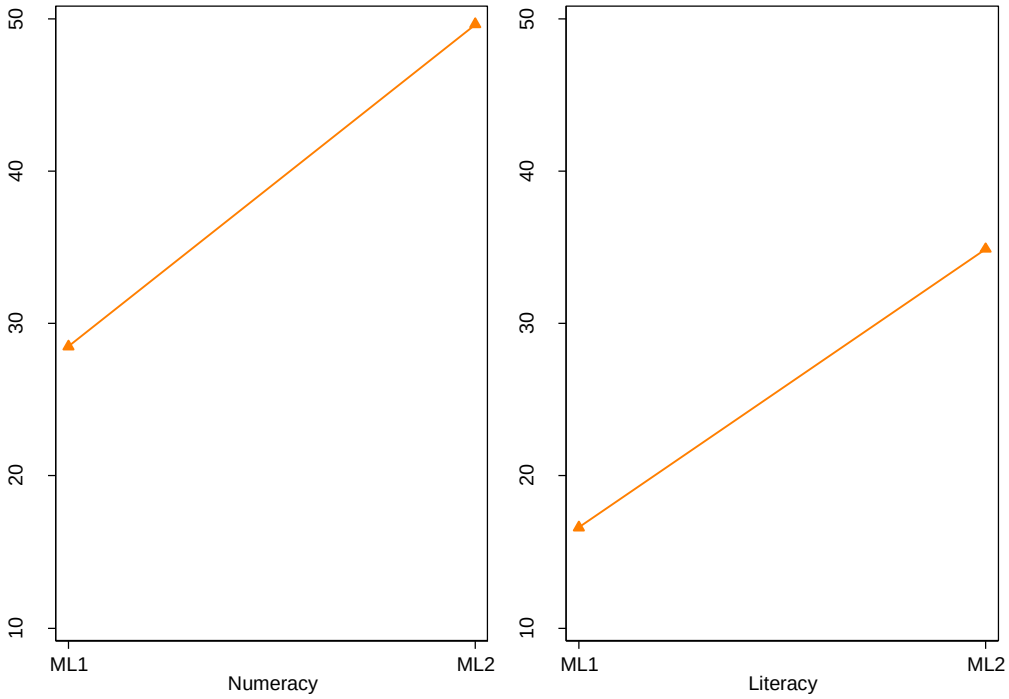
Characteristic	Number of C4 NFE girls	Percent
State – Region		
Banadir -- Banadir	304	46.0
South West State -- Bay	165	25.0
South West State -- Lower Shabelle	84	12.7
Hirshabelle -- Middle Shabelle	108	16.3
Age groups		
Age 15 years and under	53	8.0
16-17 years	67	10.1
18-19 years	232	35.1
20-22 years	123	18.6
23+ years	186	28.1
Total	**661**	**100.0**

The graph below presents a high-level cursory review of how girls scored on numeracy and literacy tests across ML1 and ML2 rounds. While we lack a benchmark for C4 NFE girls, we can observe an overall improvement in both numeracy and literacy scores across rounds. On average, the C4 NFE girls scored initially scored lower on literacy tests than numeracy tests, but they improved their score substantially between ML1 and ML2.

To interpret this improvement in learning outcomes appropriately, we should note that for C4 NFE girls there are no benchmarks or counterfactual cases to estimate the impact of the 11-month NFE programme against a comparison group, but since the C4 NFE girls were not enrolled in school, their improvement in numeracy and literacy on average would likely have increased only slightly or stagnated across rounds. With the assumption that numeracy and literacy skills of C4 NFE girls would have stayed the same without participation in the 11-month NFE programme, we can interpret the changes in aggregate numeracy and literacy because of participation in said programme.

It should also be noted that the 11-month NFE programme provided to C4 NFE girls ended by the time of the ML2 evaluation, and so the improvements observed in learning outcomes are outcomes observed after the completion of the programme. However, as will be discussed in the section on transition, some C4 NFE girls completed the programme and continue to remain in some form of non-formal or formal education, which may also be contributing to the rising learning scores documented here.

FIGURE 16: CHANGES IN AGGREGATE NUMERACY AND LITERACY SCORES FOR C4 NFE GIRLS



The table below provides greater detail on the extent to which C4 NFE cohort achieved improvements on numeracy and literacy tests from ML1 to ML2. The observed improvement among C4 NFE girl in numeracy and literacy were both substantially large and statistically significant.

TABLE 61: DIFFERENCE IN LEARNING OUTCOMES

Type	ML1 Mean	ML1 SD	ML2 Mean	ML2 SD	Diff.	P-Value
Numeracy scores	28.5	25.9	49.6	25.3	21.1	0.000

Literacy scores	16.6	26.8	34.9	32.9	18.3	0.000

Geography

In this section, we investigate the trends in learning outcome changes of C4 NFE girls by region and observe geographic patterns in the changes. As the table below highlights, C4 NFE girls in Banadir, Lower Shabelle, and Middle Shabelle improved their numeracy scores from 24.6 to 28.2 points on average, however in Bay, C4 NFE girls only improved by 2.3 points on average, which was not a statistically significant difference from their ML1 numeracy score. This smaller increase in the numeracy assessment is in part owed to the higher level in numeracy skills that C4 NFE girls in Bay region possessed at ML1 at which point their average score in numeracy was nearly double that of any other region. By ML2, the gap in numeracy scores was reduced as C4 NFE girls in other regions have largely caught up, with C4 NFE girls from Middle Shabelle lagging slightly.

TABLE 62: DIFFERENCE IN NUMERACY OUTCOMES BY REGION

Type	ML1 Mean	ML1 SD	ML2 Mean	ML2 SD	Difference	P-Value
Banadir	24.4	22.4	52.6	22.7	28.2	0.000
Lower Shabelle	20.2	21.2	48.2	26.5	28.0	0.003
Bay	47.0	28.4	49.3	28.0	2.3	0.665
Middle Shabelle	18.3	19.9	42.9	25.8	24.6	0.000
Aggregate	28.5	25.9	49.6	25.3	21.1	0.000

We observe a similar geographic trend in literacy outcomes among C4 NFE girls, but the extent of the improvement by C4 NFE girls from other regions than Bay was even greater. The C4 NFE girls in Bay outscored their counterparts in other regions at ML1. However, at ML2, the C4 NFE girls from other regions improved their literacy scores by a significant margin and surpassed the scores of C4 NFE girls in Bay on average. Meanwhile, the literacy skills of C4 NFE girls from Bay appear to have stagnated with only a slight improvement in literacy scores which was not statistically significant.

One plausible explanation for this pattern is the increasing impact of drought over the course of 2022 and early 2023. As noted in the introduction to this report, Bay and IDP communities in Banadir faced the most severe food shortages as a result of drought during mid-to-late 2022, food stress which was expected to rise to famine level by the end of 2022. Our data indicates that C4 NFE girls in Banadir faced the most significant deterioration in food security between ML1 and ML2 – much more so than any other group. For example, the share of households reducing the number of daily meals eaten over the last 30 days – a common household coping strategy – increased from 46.4 to 83.6 percent in Banadir between ML1 and ML2. The impact of the drought can help explain underperformance in Bay, where the drought was particularly severe; however, it is important to note that our data only partially supports this notion, because we do not find a significant

decrease in protein consumption in Bay between 2022 and 2023,[103] and only a moderate uptick in the use of household coping strategies. Nonetheless, it is quite clear from extensive third-party reporting that drought and food insecurity was particularly severe in Bay, and we expect this to have significant effects on learning outcomes, especially over a short-term NFE-style learning programme.

TABLE 63: DIFFERENCE IN LITERACY OUTCOMES BY REGION

Type	ML1 Mean	ML1 SD	ML2 Mean	ML2 SD	Difference	P-Value
Banadir	12.3	24.0	35.5	32.9	23.2	0.000
Lower Shabelle	15.2	23.6	40.1	32.8	24.9	0.008
Bay	30.1	31.9	31.8	32.1	1.8	0.572
Middle Shabelle	9.3	20.4	33.8	34.2	24.5	0.001
Aggregate	16.6	26.8	34.9	32.9	18.3	0.000

The aggregate findings above paint a picture of substantive improvements in the literacy and numeracy among C4 NFE girls after completion of the 10-month NFE program, suggesting that the NFE programme made important progress towards its goal of providing them with basic literacy and numeracy skills that could transfer to the workplace. While C4 NFE girls in Banadir, Lower Shabelle, and Middle Shabelle appear to have experienced improvements in literacy and numeracy skills, their peers in Bay, who had better learning outcomes at ML1, appear to have not improved significantly since ML1. The reason for this difference is unclear, but the finding may indicate that the lessons for C4 NFE girls in Bay were not sufficiently challenging for them to expand their literacy or numeracy skills.

8.2 Foundational Skill Gaps

In this section, we move from aggregate scores to numeracy and literacy scores disaggregated by their constituent subtasks, with the goal of understanding the patterns of skill achievement among the C4 NFE girls. Our analysis here focuses identifying performance gaps on specific learning subtasks at ML2; in Section 8.3, we assess changes in girls' scores on each subtask, essentially asking where girls have or have not improved over the previous year.

In the table below, we report the share of girls who meet different achievement thresholds for each of the 11 numeracy subtasks with which they were presented. Subtasks on EGMA and EGRA both tend to become more difficult over the course of the assessments, but there are additional nuances with regard to performance, as subtasks require discrete skills that can be learnt in semi-isolation.

[103] Note from the project: 41% of the Bay respondents reported not consuming protein-rich foods of animal origin on the day prior to the survey, compared to an average total of 24%. While the proportion of respondents not consuming protein-rich foods did not increase over time, the very low consumption may have affected cognitive processing. Prior data from Somalia (Miettunen, Peterson, and Robert, 2020) shows that the low consumption of protein-rich foods is associated to significantly lower learning outcomes for girls.

We define four bands of achievement: non-learners received a score of 0 on a given subtask, which means that they were unable to answer any of the individual test items correctly. Emergent learners scored between 1 and 40 percent on the subtask, established learners achieved scores between 40 and 80 percent, and proficient learners scored above 80 percent on a given subtask.

TABLE 64: FOUNDATIONAL SKILL GAPS IN NUMERACY, C4 NFE COHORT AT ML2

Subtask #	1	2	3	4	5	6	7	8	9	10	11
Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	4.8	6.2	28.3	21.6	37.4	33.6	45.7	16.6	69.9	63.2	45.1
Emergent Learner	1.8	3.6	66.3	3.5	11.0	4.8	19.5	2.7	8.0	11.2	22.8
Established Learner	3.6	16.9	3.8	12.0	21.6	11.0	17.4	22.8	11.5	11.5	26.5
Proficient Learner	89.7	73.2	1.7	62.9	30.0	50.5	17.4	57.8	10.6	14.1	5.6

As the results in the table show, performance in numeracy does not drop off very quickly across subtasks. NFE girls perform best on the number identification and number discrimination (the latter involves selecting the largest of two numbers) tasks, but also perform reasonably well on 1-digit addition, 1-digit subtraction, and equivalent word problems (subtask 8, which involves relatively simple addition and subtraction presented as word problems). Girls experience a substantial decrease in performance when addition or subtraction are complicated by employing two-digit numbers (subtasks 5 and 7), especially when compared to their 1-digit counterpart subtasks. With that said, declining performance between 1-digit and 2-digit tasks is not "all-or-nothing". For example, while proficiency in addition drops from 62.9 to 30.0 when moving from 1-digit to 2-digit tasks, much of the movement is from proficiency to the emergent/established learner categories. This suggests that a significant subset of girls, proficient in 1-digit addition, are able to add 2-digit numbers with a degree of accuracy, but may not be able to apply the procedure (e.g., carrying numbers) to more difficult items within subtask 5.

In the table below, we report proficiency levels for Somali literacy, across the EGRA's six subtasks, using the same classification scheme. Here the picture appears to be less positive, with a significant number of girls unable to consistently recognise letters and very few girls able to recognise a significant share of common words. Unsurprisingly, girls who are unable to identify a majority of common words also fare poorly in reading fluency: in general, word recognition is where a critical gap emerges between girls. Among girls who recognise 50 percent or fewer words in subtask 2 score a 4.9 percent on subtask 3, equivalent to reading just a few words of a short story. In contrast, girls who are able to recognise more than 50 percent of the words in subtask 2 apply that skill well in subtask 3, averaging scores of 66.8 percent.

TABLE 65: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, C4 NFE COHORT AT ML2

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	19.4	38.1	47.5	55.1	54.6	62.9
Emergent Learner	16.5	19.4	16.5	5.0	4.2	12.1
Established Learner	17.2	23.1	23.1	26.2	21.2	22.5
Proficient Learner	46.9	19.4	12.9	13.8	20.0	2.4

8.3 Subtask-Specific Gains in Learning

Building on the analysis of subtask-specific learning outcomes, we now evaluate *changes* or *gains* in performance on specific subtasks between rounds. Our analysis focuses on changes in girls scores in each subtask from ML1 to ML2 and discerns trends in those changes.

Numeracy subtasks

In analysing the numeracy subtasks, we review the changes in scores for the numeracy assessment administered at both ML1 and ML2 which are each composed of 11 numeracy subtasks. The subtasks include questions to assess girls' abilities to identify, add, subtract, multiply and divide numbers. In addition to these abstract mathematical calculations, the girls are also asked to solve word problems to assess their ability to apply their numeracy skills to practical problems that they might encounter in everyday life. The table below presents each subtask with a description and the corresponding mean score that the C4 NFE girls achieved at ML1 and ML2, the difference between the rounds of evaluation, and the P-value of that difference.

TABLE 66: CHANGES IN NUMERACY SUBTASK SCORES FOR C4 NFE GIRLS

Subtask Number	Description	ML1 Mean	ML2 Mean	Difference	P-value
1	# Identification	69.2	91.9	22.7	0.000
2	Quant. Discrimination	62.6	85.4	22.8	0.000
3	Missing #	12.6	12.7	0.1	0.956
4	Addition (1 digit)	39.0	71.3	32.3	0.000
5	Addition (2 digit)	16.5	46.2	29.7	0.000
6	Subtraction (1 digit)	29.2	58.4	29.2	0.000
7	Subtraction (2 digit)	14.2	33.1	18.9	0.000
8	Word Problem (Add. & Sub.)	40.9	73.9	33.0	0.000

9	Multiplication	4.9	21.2	16.3	0.000
10	Division	10.0	25.2	15.3	0.000
11	Word Problem (Mult. & Div.)	14.2	26.7	12.4	0.000

The above table highlights several important findings. First, the C4 NFE cohort of girls improved their mean score for each of the 11 subtasks. The increase in scores for each of the subtasks were statistically significant except for the "Missing #" subtask. There were substantial improvements in subtasks 1,2, 4, 5, 6, and 8 and afterwards the improvements as well as scores begin to taper downward. Improvements in subtask 8, the addition and subtraction word problem were the largest of any subtask, suggesting that the NFE programme helped to improve their ability to apply mathematical principles to real world problems.

Second, the lack of a statistically significant improvement in the "Missing #" subtask follows a trend also observed from BL to ML1 in which NFE girls only experienced minimal improvement on the subtask. This lack of improvement on the "Missing #" subtask among NFE girls in ML1 and among C4 NFE girls in ML2 may be indicative of the emphasis of the NFE programmes on imparting numeracy and literacy skills that could be useful in the labour market and less focused on developing a more abstract sense of numbers and patterns.

Literacy subtasks

The C4 NFE girls completed the six subtasks that comprise the literacy assessment at both ML1 and ML2. The literacy subtasks included fundamental questions on the sounds of letters, reading individual words, reading fluency, and reading passages in Somali and answering comprehension questions. Like numeracy subtasks, the literacy scores of C4 NFE girls improved significantly. The changes in literacy scores were meaningful, and the C4 NFE girls improved most with the subtasks assessing fundamentals such as "Letter Sound Identification" in which there was 25-point improvement and reading "Words Commonly Used" which improved by 21.3 points. The C4 NFE girls also made substantial but smaller improvements with more challenging subtasks. The improvements in reading comprehension indicate that the improvement in fundamental literacy skills translated to better understanding of reading materials.

TABLE 67: CHANGES IN LITERACY SUBTASK SCORES FOR C4 NFE GIRLS

Subtask Number	Description	ML1 Mean	ML2 Mean	Difference	P-value
1	Letter Sound Identification	33.8	58.7	25.0	0.000
2	Words Commonly Used	15.7	37.0	21.3	0.000
3	Reading fluency	11.6	28.9	17.4	0.000
4	Reading Comprehension 1	14.3	31.7	17.4	0.000
5	Reading Comprehension 2	14.8	34.7	19.9	0.000

6	Reading Comprehension 3	9.6	18.3	8.7	0.000

In ML1, gaps in foundational literacy were observed among the C4 NFE girls in which their performance along all subtasks was relatively poor. Fewer than 20 percent of girls achieved scores above 40 percent on any of the subtasks. These gaps indicate a lack of fundamental literacy skills, and so it is unsurprising that much of the improvements from ML1 and ML2 among C4 NFE girls were in those subtasks which assessed fundamental literacy skills and in which there remained a lot of room for growth.

Note from the project: *The distribution of the C4 NFE girls by proficiency levels is provided in Annex 4, Learning Tables.*

8.4 Subgroup Programme Impact

In this subsection, we present key findings on differences the changes observed in learning outcomes from ML1 to ML2, analysing differences among subgroups in terms of aggregate differences of numeracy and literacy scores. Given the many subgroups of the C4 NFE panel cohort, we have chosen subgroups which have demonstrated impacts on learning trajectories. These subgroups of concern to the programme include girls with disabilities, girls from socio-economically disadvantaged households, and girls from schools lacking in resources and infrastructure. Our analysis approach will present the differences in numeracy and literacy scores from ML1 to ML2 alongside the significance of that difference.

TABLE 68: CHANGES IN LEARNING SCORES FOR C4 NFE GIRLS BY SUBGROUP

Subtask Number	Literacy Difference	Literacy P-value	Numeracy Difference	Numeracy P-value	Sample Size (n)
Overall	18.3	0.000	21.1	0.000	661
Language					
Maay speaker	6.3	0.1	7.8	0.090	234
Disability Status					
Mental health disability	19.1	0.0	21.8	0.000	112
Non-mental health disability	13.9	0.0	17.0	0.001	49
Any disability	19.9	0.0	21.7	0.000	143
Household and Demographic Characteristics					
No parents	22.4	0.0	30.9	0.001	20
Female-headed household	21.4	0.0	23.8	0.000	152
HoH no education	21.6	0.0	22.9	0.000	37

HoH no formal education	21.0	0.0	21.5	0.000	178
Caregiver no education	21.3	0.0	22.6	0.000	33
Caregiver no formal education	21.1	0.0	21.7	0.000	176
Neither HoH or caregiver has education	23.6	0.0	24.9	0.000	31
Neither HoH or caregiver has formal education	21.2	0.0	21.7	0.000	171
HoH does not earn wage	23.0	0.0	22.7	0.000	99
HH has poor roof	10.4	0.0	15.9	0.000	151
Went to sleep hungry most/all nights	16.9	0.0	22.0	0.000	51
Went without water at home most/all days	16.2	0.0	15.2	0.000	59
Went without meds most/all days	17.7	0.0	21.6	0.000	95
Went without cash income most/all days	15.2	0.0	22.8	0.000	139
HH owns land	12.4	0.0	12.0	0.013	156
Girls' Characteristics					
Girl owns phone	21.6	0.0	22.6	0.000	195
Girl owns smartphone	15.3	0.1	16.3	0.076	21
Girl lives with her husband (proxy for marriage)	7.9	0.0	14.1	0.0	214
Girl spends a few hour or more on chores	17.0	0.0	20.2	0.000	607
School Resources and Infrastructure					

Girl won't use drinking water at school	9.3	0.0	19.0	0.000	118
Girl won't use toilet at school	13.8	0.0	20.6	0.000	140
No computers at school	17.4	0.0	20.5	0.000	587
Girl cannot use books/learning materials at school	16.4	0.0	19.9	0.000	42
Not enough seats in class	15.2	0.0	24.1	0.000	26

Girls with disabilities have some of the most serious barriers preventing them from entering and participating in school. The 11-month NFE programme appears to have helped the C4 NFE girls who had difficulty accessing education previously significantly improve their numeracy and literacy. Girls with non-mental health disabilities who may have visual, hearing, mobility, use of arms, self-care, communication, cognitive, or behavioural disabilities experienced substantive and significant improvements on their numeracy and literacy scores. C4 NFE girls with mental health disabilities which included girls with anxiety or depression similarly had gains in their learning outcomes.

Among the C4 NFE girls' characteristics, girls who belonged to households in which the caregivers or heads of households who did not have education, did not earn wage, or were female-headed saw particularly large improvements in literacy and numeracy scores. Similarly, for C4 NFE girls who faced potential barriers such as long hours of household chores or marital obligations in their personal lives as well as those who attended schools which lacked resources we also observed increases in literacy and numeracy skills following the NFE programme.

8.5 Testing the Theory of Change

The Theory of Change posits that programme outputs lead to changes in a series of intermediate outcomes which in turn have a positive impact on learning outcomes. With numeracy and literacy C4 NFE girls, the intermediate outcomes are acquisition of life skills and improved teaching quality. In this section we assess the extent to which the intermediate outcomes can be predictive of girls' learning trajectories by examining the relationship between each intermediate outcome and the sample's changes in learning outcomes.

Youth Leadership Index

The Youth Leadership Index (YLI) is a score scaled from 0 to 100 that measures girls' self-assessed leadership skills in the context of school. The index is based on answers to 21 questions asking girls questions about thinking about the consequences of their own actions, belief in ability to express their thoughts clearly, and organizing others to organize their peers to participate in an activity with a shared goal. To determine whether

YLI is a predictor of increases in learning outcome data, we analyse the relationship between the YLI score C4 NFE girls and the change in their learning outcomes.

TABLE 69: EFFECT OF 1 POINT INCREASE IN ML1 YLI SCORE AND CHANGES IN LEARNING OUTCOMES

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
BL score	0.30	0.000	0.35	0.000

Using a linear regression model with girls' age, cohort type, and region as control variables, we find that an additional point in YLI leads to a 0.3 point increase in numeracy and 0.35 point increase in literacy, and both coefficients are statistically significant.

GEF Participation

Related to the life skills discussed in the previous section is participation in the Girls' Empowerment Forum. This section discusses C4 NFE Girls' Empowerment Forum participation and its relationship with improving learning outcomes over time. The Girls' Empowerment Forum is held after-school and is a programme in which girls develop a peer support network, work with a positive female role model, and are provided with opportunities for tutoring as well as mentoring, among other benefits. Previous research on GEFs as part of the SOMGEP-T programme as well as AGES has shown that girls who participate in perform better on learning assessments.

Our study for C4 NFE girls similarly surfaces findings that show that participation in GEF leads to higher literacy and numeracy scores. We use a regression framework that controls for region and age and find that GEF participation as well as continued contact with GEF members is associated with significantly higher numeracy and literacy scores.

TABLE 70: EFFECT OF GEF PARTICIPATION AND CHANGES IN LEARNING OUTCOMES AT ML2

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
GEF participation	16.1	0.000	15.4	0.000
GEF participation continued	16.5	0.000	15.2	0.000

It should be noted that GEF participation and continued participation are not randomly assigned and C4 NFE girls self-select into it. Girls who are motivated to participate in GEF may be more motivated to do well in the NFE programme and as such improve their scores even more.

The findings from YLI and GEF participation are consistent with research from SOMGEP-T and AGES we find that GEF participation may lead to improved YLI scores which in turn have a positive effect on learning outcomes. A regression model of YLI and GEF participation indicates that GEF participation is associated with a 12.1 increase in YLI score.

Teaching Quality

The AGES Theory of Change draws a causal line from improved pedagogy to higher learning outcomes. Teachers are trained on inclusive pedagogy that allows girls to feel comfortable and engaged in the classroom with the belief that it will foster a better learning environment for girls. In this section, we examine the relationship between teaching quality and the learning changes experienced by C4 NFE girls.

For each of the teaching quality measures within the table below, we use a linear regression framework with age and region as control variables to understand the effect of the teaching quality indicator on learning outcomes. As shown in the table below, since the aggregate improvements in numeracy and literacy led to relatively large increases in learning scores, even the negative teaching quality measures shown below are associated with large increases in learning outcomes. However, the negative effects of the teaching practices described below are captured in the reduced size of the improvement in the learning outcome as observed for the C4 NFE girls who said that their teacher rarely/never encourages participation in which the size of the effect on learning score change is relatively lower and are not statistically significant improvements.

TABLE 71: EFFECT OF TEACHING PRACTICES ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
My teacher does not make me feel welcome in classroom	10.1	0.000	11.2	0.018
My teachers are often absent	11.5	0.001	13.3	0.002
My teacher rarely/never encourages participation	3.6	0.448	1.8	0.766
My teacher explains how learning things is useful in our lives	12.3	0.038	11.8	0.167
My teacher's lessons move too fast for me	17.4	0.000	15.6	0.000
My teacher punishes students who get things wrong in a lesson	16.5	0.000	16.5	0.000
My teacher used corporal punishment in last week	8.9	0.046	6.6	0.245

Furthermore, improvements in literacy scores were also relatively smaller among C4 NFE girls who said that their teacher used corporal punishment in the last week as well as those who said that their teacher rarely/never explains how learning things is useful in their lives. The large improvement in learning outcomes among C4 NFE girls presents challenges in disentangling the effects of potential barriers to learning, but the

lack of improvement in the presence of negative teaching practices provides evidence that teaching quality remains an important predictor of learning outcomes.

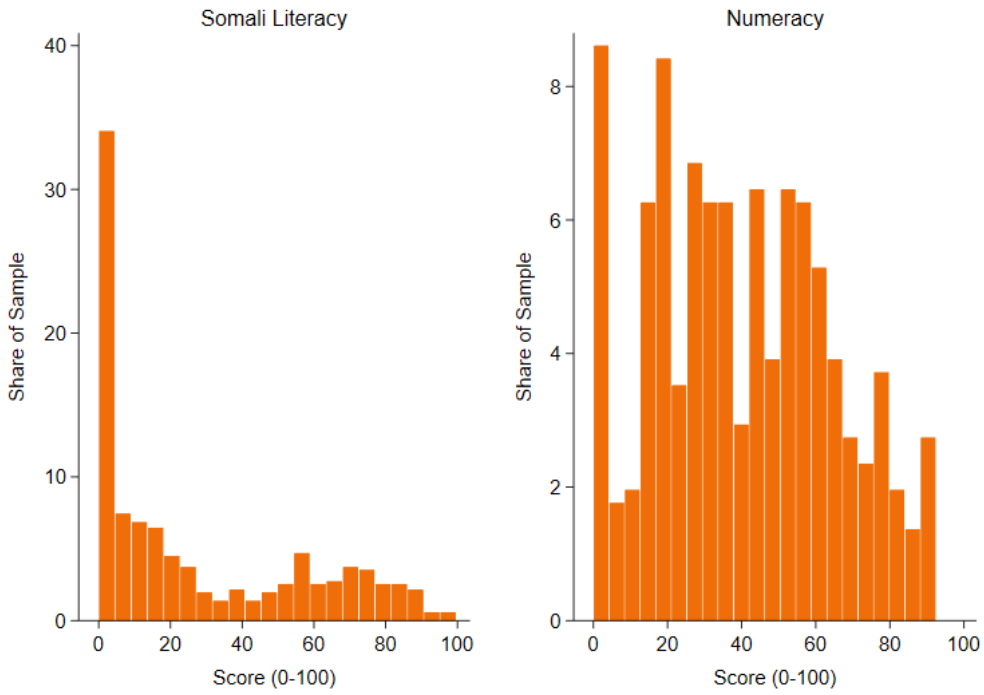
9. Learning – C5 NFE Cohort

In this evaluation round, a new cohort of NFE girls (cohort 5 or "C5") was recruited. This cohort of girls will be tracked from ML2 through the endline round, to understand the evolution of their learning scores and transition outcomes, among other programme outcomes. The C5 NFE cohort is the latest group of NFE girls to be brought into NFE learning centres, following on the C4 NFE girls who were recruited and enrolled in early 2022. The C5 NFE group was enrolled into the learning centres immediately prior to this round of data collection. In this section, we document their learning scores at ML2 – which constitutes a baseline for them – including subgroup differences in literacy and numeracy, and gaps in their literacy and numeracy skills.

9.1 Aggregate Learning Outcomes

Cohort 5 NFE girls have a mean Somali literacy score of 27.4. There is a negative and statistically significant relationship between age and literacy scores, with lower scores among older girls. There are also substantial differences in scores among C5 girls across locations. Scores are highest in Middle Shabelle (36.2) followed by Banadir (30.5), Lower Shabelle (21.3) and Bay (14.5). Elsewhere in the survey tool, we asked NFE girls if they can read, to which 64.8% said "yes." Confirmation of literacy in the question is strongly correlated with Somali literacy scores; girls who said they can read have a mean literacy score of 38.6 while girls who said they cannot have a mean score of 6.9. Overall, there is a large share (22.3%) of C5 girls who score 0 on the Somali literacy assessment. Cohort 5 NFE girls have a mean numeracy score of 40.2. There is also a negative and statistically significant relationship between age and numeracy scores, with lower scores among older girls.

FIGURE 17: COHORT 5 NFE GIRLS' LITERACY AND NUMERACY SCORES



9.2 Foundational Skill Gaps

Beyond the general aggregate scores of NFE girls, we are interested in the different components of the learning assessments and the variation in performance both within and across them. In this section, we continue to focus on the new cohort ("C5") of NFE girls recruited by the programme in 2023. To do so, we analyse performance on individual subtasks, categorising girls according to their level of proficiency on each subtask. The categories are as follows:

- Non-Learner – was unable to complete any of the subtask (scored 0%).
- Emergent Learner – was able to answer a portion of the subtask correctly, but generally unable to perform the tasks demanded (scored 1-40% on the subtask).
- Established Learner – was able to answer around half or more of the individual test items correctly, showing facility with the subtask but not mastery (scored 41-80% on the subtask).
- Proficient Learner – was able to answer most of the individual test items correctly, showing a degree of mastery of the subtask (scored 81-100% on the subtask).

In the table below, we report the share of girls who fall into each of the four categories (listed down the first column) for each subtask (listed across the header of the table) in the numeracy assessment. Included along the header is a brief description of the skills each subtask tests.

TABLE 72: NFE COHORT 5 NUMERACY SUBTASKS

Subtask #	1	2	3	4	5	6	7	8	9	10	11

Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	6.7	9.4	34.4	31.3	52.6	48.5	62.4	30.3	85.1	79.8	62.2
Emergent Learner	5.7	4.7	64.0	5.7	7.2	3.9	17.6	2.2	4.3	6.8	17.4
Established Learner	2.3	16.8	1.6	11.0	16.4	8.6	9.8	20.0	6.7	4.5	17.0
Proficient Learner	85.3	69.1	0.0	52.1	23.7	38.9	10.2	47.6	3.9	8.8	3.3

Subtasks 1 and 2 – number identification and discrimination – have the highest performance among C5 NFE girls. The first task requires naming a series of single- and double-digit figures, while the second asks girls to identify the larger figure in sets of pairs. Similarly, most (52.1%) C5 girls meet the 'proficiency' threshold for 1 digit addition (subtask 4) and substantial share similarly perform well on 1 digit subtraction (subtask 6) and word problems involving addition and subtraction (subtask 8). However, it is also worth noting the polarization of girls in these subtasks as either proficient learners or non-learners, with few in between. For instance, while more than half of C5 girls are proficient leaners at 1 digit addition, almost a third (31.3%) are categorised as non-learners on the same task. Similarly, scores on subtask 6 (1 digit subtraction) are concentrated in the highest proficiency category (38.9%) and lowest proficiency / non-learner category (48.5%). This suggests there is a substantial gap between learners more broadly, however this pattern does not appear in the broader numeracy trends. In other words, we do not see clustering of high and low numeracy scores, with a low prevalence of "middle" numeracy scores. This is because the 'proficient learners' are not a consistent group across numeracy subtasks. Many girls (27%) are proficient in one or two subtasks and few (11.2%) are proficient in more than six of the eleven subtasks.

Unsurprisingly, the subtasks requiring multiplication and division (subtasks 9 – 11) have the lowest performance, with most girls categorised as non-learners, or unable to complete any of the subtasks. As the most advanced arithmetic within the subtasks, this is expected. Somewhat more surprisingly, we see fewer non-learners in the word problems than we do in other questions; while 85.1% and 79.8% of students were categorised as non-learners on the multiplication and division subtasks, respectively, just 62.2% are categorised as the same on the word problems involving multiplication and division.

This pattern was also observed at the ML1 among C4 girls, whereby girls' performance on word problems exceeded comparable arithmetic-only problems. In the last report, we consider potential reasons for this trend, the most likely being that NFE girls have likely done much of their learning in practical and applied settings, rather than in school. As such, they may be more comfortable thinking through arithmetic grounded within a practical context, rather than as an equation with symbols on a page.

We take the same approach in breaking down subtask performance of C5 NFE girls in the literacy assessment. The share of girls in each learning category, by subtask, is presented in the table below.

TABLE 73: NFE COHORT 5 LITERACY SUBTASKS

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	**22.5**	43.8	59.5	65.2	66.3	74.2
Emergent Learner	**20.2**	19.8	15.3	5.1	4.9	10.4
Established Learner	**18.0**	22.5	15.3	20.5	16.8	13.9
Proficient Learner	**39.3**	13.9	10.0	9.2	11.9	1.6

To some extent, the trends are clearer and more intuitive within the literacy subtasks. Items are relatively more difficult from task to task, and we see the share of non-learners in each increase with each new subtask. Most girls (57.3%) are proficient or established learners on the letter recognition subtask, which requires girls name both capital and lower-case letters on a page. More than a quarter of the cohort (26.2%) scored 98% or higher on the subtask. Far fewer girls reached the proficiency threshold on any of the following subtasks and a substantial share (43.8%) of the cohort were unable to read a single word in task 2.

We may also consider again the question posed elsewhere in the survey tool which asked girls if they can read / are literate. Of the girls who confirmed, "yes" they are literate, 9.4% are categorised as non-learners at subtask one: letter recognition. As many as 26.9% of the girls who said they are literate are categorised as non-learners in subtask two and almost half (42%) are non-learners at subtask three: reading fluency. While we generally anticipate respondents are likely to over-report literacy in such questions, the difference between self-reported and demonstrated literacy levels is somewhat surprising. The over-estimation of one's literacy levels may be related to a genuine misunderstanding of what is meant by literacy and/or a desirability bias incentivizing claims to literacy. However, the nature of the assessment may also underestimate actual literacy levels, which under different circumstances (ie receiving an SMS, using social media, taking in information while shopping) may be higher. This hypothesis is similar to one presented in the previous numeracy section, which anticipates NFE girls are more comfortable applying their knowledge in practical settings and/or familiar contexts than in 'academic' assessments.

9.3 Subgroup Learning Scores

In this section, we explore the girl and household characteristics which appear to relate to difference in learning scores. We refer to girls with characteristics of interest as belonging to a subgroup and are interested in whether learning scores of girls within a subgroup differ significantly from others. The subgroups we examine revolve around girl's disability status, household and girl's social characteristics, household education, and household economic status. Our quantitative analysis proceeded in two steps. First, we regressed numeracy and literacy scores for C2 NFE girls on the sub-group variable of interest through a linear regression model without controls. For those sub-group characteristics that were significant at at least the 90% confidence interval, we re-ran the regression model with the addition of control variables: the girl's age and her state of residence, both of which we have found to have an effect on C5 NFE girls' learning outcomes.

Due to the lengthy list of possible sub-groups to measure, this sub-section presents only the characteristics whose effects on assessment scores remained significant even after the introduction of the two control variables, though the numbers reported are from the simpler, no controls, model.

TABLE 74: NFE COHORT 5 SUBGROUP LEARNING

		Literacy			Numeracy		
Subgroup	Sample Size	Learning Assessment Score	Coefficient	P-Value	Learning Assessment Score	Coefficient	P-Value
All C5 NFE girls	511	27			40		
HH poor roof quality	101	14.27	-16.41	0.00	34.22	-7.46	0.01
No food most days	59	15.32	-13.71	0.00	32.96	-8.20	0.02
No water most days	87	19.20	-9.94	0.00	32.17	-9.69	0.00
Disability, excl. mental health	31				26.73	-14.35	0.03
Married	4	4.39	-23.23	0.06	20.53	-19.84	0.07
IDP	192	21.04	-10.25	0.00	37.08	-5.02	0.02
Maay speaker	151	19.37	-11.46	0.00	33.91	-8.95	0.00

Most household characteristics have no significant relationship to learning scores. However, it is worth noting that many of the household characteristics – such as being an orphan, having a female head of households, and/or level of education among household heads and caregivers - which serve as predictors for girl learning outcomes are not relevant or applicable to NFE girls who are older and no longer under the care of another adult. However, some of the socioeconomic indicators of the household, such as the quality of the roof and frequency of experiencing low access to food and water, remain relevant. As may be expected, learning scores are lower among girls who are worse off, based on these household socioeconomic indicators. This tracks with broader trends which link income and financial well-being to learning outcomes, for a range of potential reasons from biological (students who are hungry are less able to focus) to social (households with greater financial strain may compel girls to contribute more to domestic and/or income-generating activities which detract from school attendance and studying). We may also interpret this information while considering the particularities of the NFE group, who are generally older and have not progressed through the formal education system. It is possible that these socio-economic household factors – which may or not be answered in relation to the household of the girls' parents and/or caregivers – also correlate with their level of exposure to the formal education system. In other words, girls who are worse off economically now, may have also been worse off during primary school years, and had less quality time in school overall, therefore starting 'further behind' other girls in their cohort. Future rounds of evaluation may consider incorporating a survey question which captures number of years attending formal school, to better understand girls' path towards this baseline.

Though the sample size is small, girls with physical disabilities appear to perform worse on the numeracy assessment. Though the range of disabilities captured in this variable is wide – including hearing, visual, mobility, and communication challenges, among others – the linkages are relatively intuitive; many interact

with both quality learning – ie accessing and maximizing participation in school – as well as performance on the assessment itself – ie hearing questions and interacting with enumerators. However, it is interesting to note that in both ML1 and ML2, disability status was significantly correlated with numeracy but not literacy scores. The reason for this is unclear and warrant further exploration of how the learning assessment and/or teaching practices may differ in ways which affect numeracy but not literacy.

Among the C5 NFE cohort, IDP status is a statistically significant predictor of both numeracy and literacy scores. Among those who are IDP's, most (53%) have been displaced for more than two years. Although we do not have data on the precise timing and effects of displacement on the sampled girls, it is likely that girls either spent school-aged years living in a camp – where access to quality schools may be limited – and/or were relocated during school-aged years, disrupting their enrolment and attendance.

10. Transition - C4 NFE Cohort

Alongside improving learning outcomes, improving successful transition rates for girls is a primary outcome sought for within the AGES programme. Transition seeks to capture how the NFE programme affects the C4 NFE cohort's pathways in life whether it be retention in the NFE, transition into employment, or enrolment in a formal school.

In the first sub-section, we define the successful and unsuccessful pathways for girls recruited into the C4 NFE programming and estimate the transition rates for girls who were recruited into the NFE programme in 2022 before ML1. In the second section, we analyse sub-group specific transition outcomes. We end with an analysis of the programme's Theory of Change and provide an analysis of whether intermediate outcomes are predictors of transition rates.

10.1 Aggregate Transition Outcomes

While transition is a multidimensional outcome in the lives of girls who participate in the programme, within this study, we will define transition as either a success or failure. The table below provides a description of pathways that we define as successful and in line with programmatic goals and other as unsuccessful. As shown below, continued education or employment are regarded as successful transitions, and dropping out or staying at a similar educational level is considered an unsuccessful transition.

TABLE 75: TRANSITION PATHWAYS, ACCORDING TO STARTING POINT OR COHORT

Starting Point	Successful Transition	Unsuccessful Transition
C4 NFE Girl Enrolled in NFE at ML1	• Enrolment in formal school, at any grade level • Retention in NFE, with progression through levels (where available) • Transition into a technical or vocational education programme • Transition into age-appropriate, non-exploitative employment • Transition into self-employment	• Drop out • Transition into ABE • Retention in NFE without progression in levels • Non-gainful employment, or employment if girl is under 18 years of age

The C4 NFE programme began in February 2022 lasted approximately 11 months and has concluded. However, C4 NFE girls can stay in the programmes longer than the 11 months because some learning centres reportedly have multiple "NFE levels" or offer other learning programmes that girls may construe to be a continuation of NFE programming. Since the C4 NFE programme has only been run for a year, we regard continuing in such an "NFE programme" to be a successful transition. However, if at endline C4 NFE girls

are still in NFE and have not progressed to higher levels, then this would be considered an unsuccessful transition.[104]

The transition outcomes of the C4 NFE girls are shown in the table below. A plurality of the C4 NFE girls are retained in the NFE programme, which – as mentioned previously – atML2 is considered a successful transition. Over a quarter of C4 NFE girls are now employed. Employment for any girl under the age of 18 years is considered a negative transition outcome; in total, just 3 (0.2 percent) of girls fall into this category. Gainful, age-appropriate employment was more common, with 27.5 percent of girls employed in this way. Very few C4 NFE girls reported being self-employed.

TABLE 76: TRANSITION OUTCOMES FOR C4 NFE GIRLS

Transition	N	Percent
Retention in NFE	246	37.2%
Age-appropriate, non-exploitative employment	182	27.5%
Now OOS	169	25.6%
Enrolment in formal school, at any grade level	50	7.6%
Self-employment	11	1.7%
Underage employment	3	0.2%
Total	661	100.0%

Approximately a quarter of C4 NFE girls said that they were now out of school but not employed. This outcome for C4 NFE girls is regarded as an unsuccessful transition as they are neither employed, self-employed, or in a learning programme. While a substantial number of girls are not employed, the C4 NFE cohort at ML2 experienced better transitions than NFE girls at ML1 of whom 68.7 percent dropped out or completed the programme without an alternative outcome and only 16.5 percent dropped out into employment or self- employment. This improvement suggests that the NFE programme is finding greater success at ML2 in guiding NFE girls to more successful transitions.

[104] We emphasise that the AGES-provided NFE programme consists of a single level of NFE, which takes place over 11 months. The discrepancy between this reality and the self-reported enrolment of girls in other "levels" of NFE (or continued enrolment after the end of the 11-month intervention) likely arises from girls' misunderstanding or mischaracterization of the learning programme they are participating in. For example, a girl who has completed NFE may transition into a vocational programme but may not distinguish sharply between these two types of classes, indicating to an enumerator that she continues to attend a non-formal education programme. This conjecture is supported by discussions with several field team leaders and enumerators, some of whom expressed confusion themselves about the precise difference between non-formal education, accelerated basic education, and other forms of learning/training that occurs outside of a formal school. In other cases, field team members described girls who self-reported enrolment in NFE, but described a programme that differed from the AGES intervention in intensity, likely because it is simply a continuing education or adult education programme hosted in a community, which girls may shift into at the end of the AGES NFE programme.

10.2 Subgroup Transition Rates

The previous section explored the specific transition pathways that the C4 NFE girls followed as a cohort. Here, we examine differences in transition rates among groups within the C4 NFE girls cohort in order to identify the subgroups that may be more or less impacted by the programme. Specifically, we explore how geography, family characteristics, household socioeconomic status, disability, marital status, and indicators of household marginalization impact transition rates within the C4 NFE girls cohort.

For this analysis, successful transition is binary. A girl within the cohort is defined as having made a successful transition if she has transitioned into a formal education setting, has acquired gainful employment, *or* has remained enrolled in the NFE programme. Any other outcome is defined as an unsuccessful transition. The transition rate is thereby defined as the proportion of C4 NFE girls who have successfully transitioned. By comparing the rate at which various subgroups transition, relative to the transition rate of all other girls in the cohort who do not belong to that subgroup, we can identify the subgroups for whom the program, as currently constituted, may be benefiting more or less. The table below compares the transition rates of cohort subgroups of interest relative to overall transition rate.

Regional differences had the strongest impact on the transition rates of all the subgroups included in this analysis. Girls from Banadir were 11.5 percentage points less likely to transition than the rest of the cohort (p=0.06), while girls from Middle Shabelle are more than 20% more likely to transition than the rest of the cohort. Transition rates for girls from Lower Shabelle and Bay did not differ significantly – in statistical terms – from the rest of the cohort. Given that regional location is a variable that integrates a wide range of other explanatory factors, we first analysed some of the more fine-grained subgroups to see if they may be driving the regional differences observed.

Sub-groups related to family characteristics, disability status, and indicators of household marginalisation did not predict differences in transition rates for the C4 NFE girls. There was a large amount of variation in the number of girls belonging to specific subgroups and as a result it is possible that a number of these subgroups have real variation in their likelihood of transitioning that is not detectable. As a result, we are cautious about drawing even weak inferences about the impact of specific sub-group categories on transition rates. For example, none of the sub-groups within the marginalisation category are statistically significant, while the observed changes in transition rates are typically in the positive direction. Likewise, disability sub-groups also tend to show statistically insignificant, but positive changes. We highlight these examples to caution the reader from drawing general conclusions about subgroups, which have statistically insignificant results, that vary in a way that you might expect. Additionally, the subgroups within the marital status and pastoralist categories have very small sample sizes ($n \leq 4$) and as result do not provide any new information on the impacts of marriage or pastoralist livelihood strategies on transition rates.

TABLE 77: SUBGROUP TRANSITION OUTCOMES AMONG THE C4 NFE COHORT

Sample Characteristics	Observations	Transition Rate	Difference from Aggregate
Aggregate Transition Rate	661	74.0	
Geography			
Banadir	304	67.8	-11.5[105]

[105] Significant at the 10 percent level only (p = 0.06).

Lower Shabelle	84	78.6	5.3
Bay	165	71.5	-3.3
Middle Shabelle	108	91.7	21.1*
Family Characteristics			
Girl has only one living parent	**53**	**64.2**	**2.4**
Girl has no living parents	4	50	-12.6
Girl does not live with either parent in her HH	24	62.5	0.2
Female-headed household	169	61.5	-3.7
HoH has no education of any kind (no Quranic)	56	57.1	-7.0
HoH has no formal education (may have Quranic)	199	61.3	-13.7
Caregiver has no education of any kind (no Quranic)	53	58.5	-5.1
Caregiver has no formal education (may have Quranic)	198	62.1	-2.6
Household Wealth and Socio-Economics			
HH has poor roof (mud/thatch/cardboard/plastic)			
Went to sleep hungry many nights (10+), last 12 months	116	69.8	-5.0
Went to sleep hungry most/all nights, last 12 months	51	84.3	11.2*
Went without water for home use most/all days, last 12 months	59	84.7	11.8*
HH owns land, either solely or jointly	156	70.5	-4.5
Household owns a phone	195	62.6	-16.2*
Household owns a smartphone	21	71.4	-2.6
HoH does not have an occupation or does not earn a wage	99	63.6	-12.2*
Household is engaged in pastoralism	3	100	26.1*
Disability Status			
Girl has physical disability	**49**	**77.6**	**3.9**
Girl has physical disability, alternative coding	51	74.5	0.6

Girl has cognitive, behavioral, or communicative disability	67	73.1	-0.9
Girl has cognitive, behavioral, or communicative disability, alternative coding	81	69.1	-5.5
Girl has mental health disability	278	74.8	1.5
Girl has mental health disability, alternative coding	312	74.4	0.7
Girl's Marital Status			
Ever married	4	50	-24.2
Currently married	2	50	-24.2
Is a mother	1	0	-74.2*
Displacement and Language			
IDP Household	302	77.5	6.5
Household speaks af-Maay	234	74.8	1.3

* Denotes that the observed difference is significant to the $p<0.05$ level

While there were several sub-groups within the household wealth and socioeconomics category that differed significantly in their likelihood of transitioning, the direction of the change was often counter-intuitive. For example, girls who reported going to sleep hungry most nights in the last 12 months, or who reported that their household went without water most days were significantly *more likely* to successfully transition than girls who were not part of that subgroup. Similarly surprising, girls from households with a cell phone were less likely to transition than girls from households who did not. In fact, the only statistically significant subgroup difference that comports with a common-sense expectation is that girls from households where the head of household is unemployed and does not earn a wage were 12.2 points less likely to transition compared with other girls in the cohort.

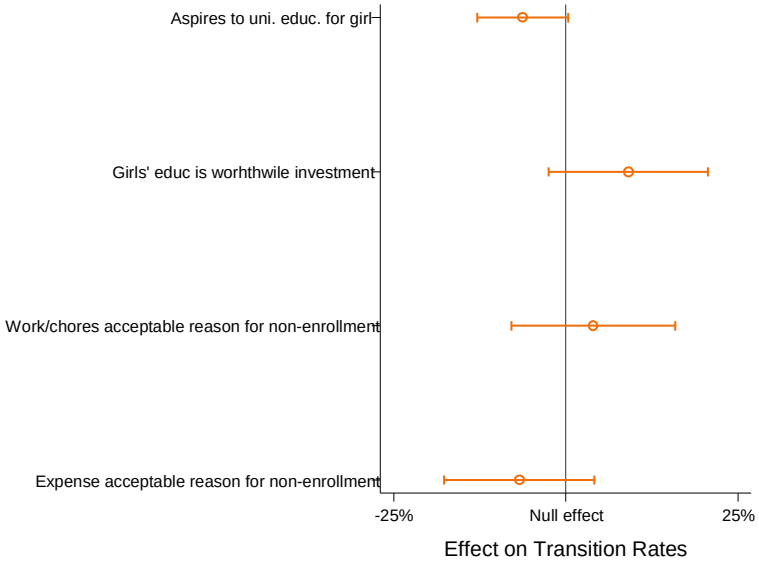
While these results seem surprising at first glance, disaggregating by region helps to make more sense of these results. Recall that girls from Banadir were significantly less likely to transition while girls from Middle Shabelle were much more likely to transition. When disaggregating the significant sub-group variables by region, we found that the subgroups for which we found a surprising, negative impact on transition, the sample of girls within that subgroup were dominated by girls from Banadir. For example, 70 of the 109 girls who live in a household with a mobile phone are from Banadir. Conversely, for subgroups where the transition rates were surprisingly higher, such as being from a household that did not have enough food or water most of the time, the sample was dominated by girls from Middle Shabelle. One interpretation of these results is that there are one or more factors, that covary with region, that have a significant impact on the likelihood of a girl successfully transitioning and those factors are not captured within the subgroups utilised in this study. Another possibility is that one or more of the factors examined are important for determining the likelihood of transition, but that there are important regional differences in those factors, and our inability to conduct the analysis while controlling for region is limiting our ability to detect the effect.

10.3 Testing the Theory of Change

The goal in this section is to assess whether the programme's intermediate outcomes are predictors of transition rates as hypothesized in the Theory of Change. The Theory of Change suggests that caregiver attitudes toward girls' education should be positively correlated with higher transition rates, all else equal.

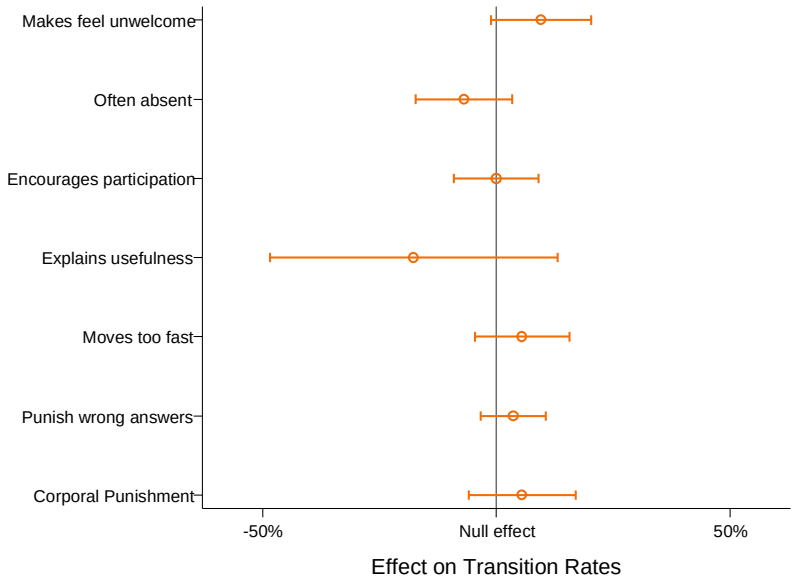
As such we take a linear regression approach with intermediate outcomes as predictor variables and age and region as control as control variables to predict binary transition success. The figure below reports these regressions. The dot in the figure indicates the relationship between the caregiver attitude and transition rates and the horizontal bars extending from the dot in both directions indicate the 95 percent confidence interval for that relationship. When the horizontal bars overlap with the vertical null effect line, as it does for all the caregiver attitudes indicators shown below, then the effect of the attitude does not have a discernible effect on transition rates either positively or negatively. While none of the caregiver attitudes indicators are statistically significant, they generally estimate relationships in the direction that we would expect: caregivers who believe girls' education is a worthwhile investment have more of positive association with successful transition rates whereas caregivers who believe expense is a an acceptable reason for non-enrolment are more associated with unsuccessful transitions.

FIGURE 18: CAREGIVER ATTITUDES AS PREDICTORS OF SUCCESSFUL TRANSITION



Turning to indicators of teaching quality, as assessed by C4 NFE girls who answered questions about their teacher's practices in the classroom, we find that the relationship between teacher characteristics and transition outcomes is generally weak. As shown in the figure below, we find that none of the indicators of teaching quality are significant predictors of successful or unsuccessful transitions.

FIGURE 19: ASSESSMENT OF TEACHING QUALITY AS PREDICTORS OF SUCCESSFUL TRANSITION



Beyond teaching quality, we also assessed the relationship between leadership skills (as measured by YLI scores) and GEF participation – expected to, among other things, promote the development of leadership skills in girls – and transition outcomes. Overall, YLI scores are a consistent, but weak, predictor of transition outcomes in the C4 NFE cohort. To illustrate, a 10-point (on a 0-100 scale) difference in YLI scores is associated with a 2.1 percentage point increase in the likelihood that a C4 NFE girl will transition successfully at ML2. Gains in YLI scores from ML1 to ML2 are similarly associated with increased transition rates, though only the former effect is significant at any conventional level of statistical significance.

GEF participation has a similar impact on transition outcomes, likely in part because GEF participation and YLI scores tend to move together. Both participation – at any point, whether at ML1 or ML2 – and continued GEF participation at ML2 was linked with higher transition outcomes. If a girl participated in a GEF in *either* ML1 or ML2, they are 4.1 percentage points more likely to transition successfully following the end of their NFE programme, a finding that is far from statistical significance ($p = 0.29$), but substantively important.[106] It seems likely that GEFs promote leadership skills, which can produce more positive transition outcomes; it is also possible, however, that the higher underlying or preexisting motivation of girls who engaged with and continued to engage with GEFs at ML2 is driving the link to successful transitions.

In the above analysis, we find few strong predictors of transition outcomes at midline. Indicators of community attitudes, teaching quality, and YLI had null relationships with transitions. Continued GEF participation was positively linked to successful transitions but may be a proxy for other substantive personal characteristics driving transitions such as motivation.

11. Intermediate Outcomes

[106] Among girls who specifically continued to engage with the GEF at ML2, transition rates are even higher – this form of participation is associated with an 8.1 point increase in transition, a difference that is marginally significant, at the 10 percent level.

11.1 Attendance

Cohort 5 NFE Girls

Attendance as an intermediate outcome is only tracked for Cohort 5 NFE girls, who began their programme immediately prior to the start of data collection for this evaluation round. Cohort 4 NFE girls are not included in our consideration of attendance rates, because C4 NFE programmes ended prior to the start of data collection. The results in this section are intended to inform the programme regarding attendance rates for C5 NFE girls, and can be used to understand the impact of the NFE intervention on learning and transition outcomes in the next evaluation round.

We refer readers to Section 7.1 for a more detailed discussion of our measures of attendance, which consist of physical headcounts performed in NFE classes during field visits by the evaluation team. The headcount is compared to the class's enrolment records to derive the attendance rate. Field teams also recorded any attendance information captured by the teacher the day of the visit and the day prior to the visit, though we focus on our own headcounts for most analysis.

The table below reports overall attendance rates and rates disaggregated by geographic zone. Our primary measure is based on physical headcounts, reported in the first row. Our secondary measures rely on teacher-reported attendance on the day of our visit and the day prior. In general, these three measures tell similar stories regarding geographic variation in attendance rates: attendance is lowest in Banadir, followed by Hirshabelle, with the highest attendance rates – across all measures – in South West State. Overall, attendance on the day of our visit, based on physical headcounts, was 64.9 percent.

TABLE 78: ATTENDANCE RATES AMONG C5 NFE GIRLS, BY ZONE

Attendance Measure	Banadir	Hirshabelle	South West State	Overall
PHYSICAL HEADCOUNT	52.6%	61.9%	79.3%	64.9%
ATTENDANCE TODAY – TEACHER-REPORTED	53.4%	64.7%	81.8%	66.8%
ATTENDANCE YESTERDAY – TEACHER-REPORTED	56.3%	71.6%	82.8%	70.3%

These results are based on a moderately-sized sample of classrooms, covering 129 C5 NFE classrooms across 45 centres. It is important to note that attendance rates are a function, to a significant extent, of enrolment levels and record-keeping practices related to enrolment. As is true of formal schools, if enrolment records are kept with extreme accuracy – documenting students who have dropped out – this can increase attendance rates, because it excludes girls for whom there is no expectation of attendance. But centres or schools in which enrolment records are not updated will include students who are not actually enrolled, biasing their attendance rates downward.

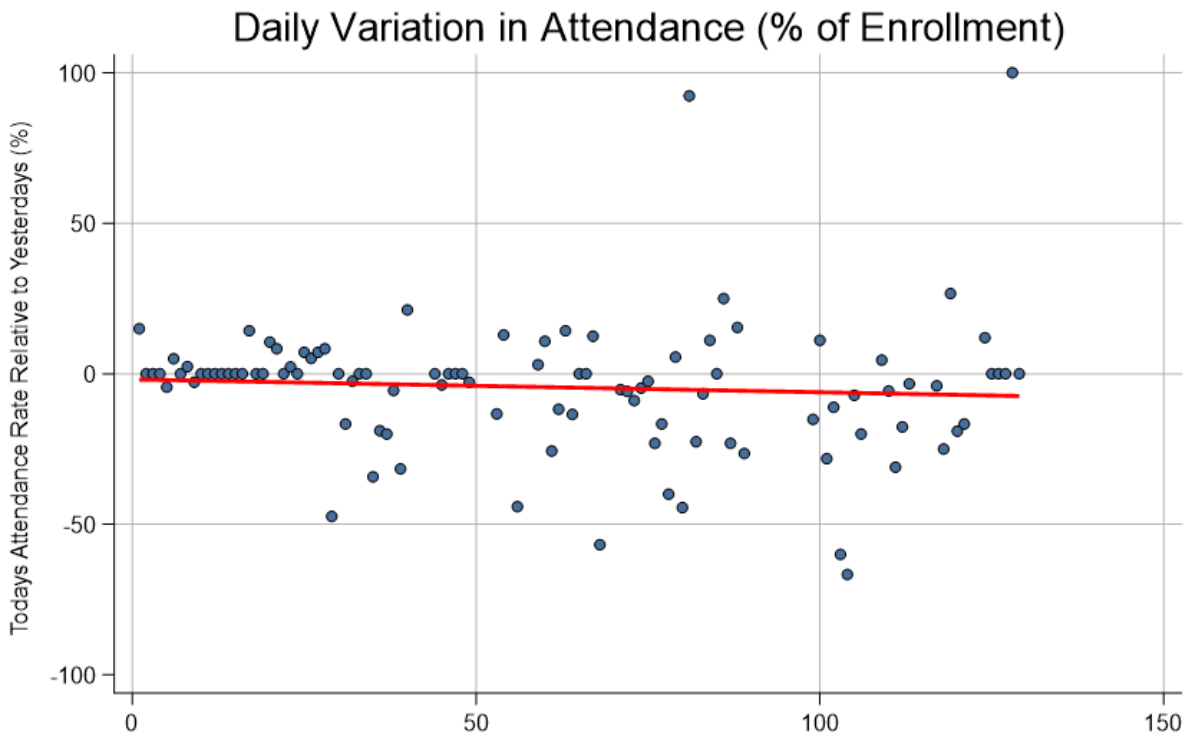
The data available does not allow us to determine whether enrolment numbers are accurate. Average enrolment levels across zones were remarkably similar, ranging from 44.7 to 45.1 students per classroom in Hirshabelle and Banadir, respectively. However, this does not shed light on the quality of enrolment records. Nor does comparison of enrolment records to the enrolment numbers reported by CARE as part of the C5 NFE cohort sample frame; large differences between a centre's enrolment records and CARE's records – if

they exist – may reflect high-quality record-keeping by the school (fastidiously documenting girls that dropped out after CARE's enrolment and verification exercise ended) or low-quality record-keeping by the school (if girls who are actually enrolled and verified by CARE are simply not documented in the school's records).

The main indicator the evaluation team has regarding the quality of record-keeping comes from documentation of the last 5 days of attendance records and a subjective assessment of the completeness and neatness or organisation of the attendance records. On these metrics, C5 NFE centres generally excel, with 71.3 percent of classrooms having complete attendance records for each of the past 5 days, and only 1 centre out of 129 (0.9 percent) having no attendance records available over the previous 5 days. Completeness and neatness were generally rated highly, with centres in South West State generally performing best on these metrics.

It is also possible that our enumerators came during days with lower than average attendance. Figure 3 shows the variation in attendance according to teacher's headcounts on the day our enumerators came to the school relative to the day before. Positive values indicate that the percentage of enrolled girls in attendance on the day our enumerators came were higher than the percentage attending the day before, and negative values indicate that the day before had greater reported attendance rates. The line of best fit suggests that "todays" attendance rates are, on average, lower than yesterdays. Fully 30% of schools reported that yesterday's enrollment was at least 10% greater than todays, and for 13% it was at least 25% higher.

FIGURE 20: VARIATION IN ATTENDANCE BASED ON TEACHER HEADCOUNTS, YESTERDAY VS TODAY



11.2 Teaching Quality

Cohort 5 NFE

We now discuss three dimensions of teaching quality and practices—professionalism, the use of physical punishment, and pedagogical practices—for C5 NFE girls. We do not discuss gender equity, as all C5 NFE classrooms assessed had only girl students; however, we analyse teachers' treatment of girl students under pedagogical practices. This analysis provides a baseline for this cohort of girls to allow for further analysis of changes in teaching quality in NFE programmes at endline. As with the prior teaching quality section, we utilise two data sources for our analysis, classroom observations and the survey with girls.

Teacher Professionalism

The table below shows results for two aspects of teacher professionalism, the extent to which teachers make students feel welcome and teacher absenteeism, disaggregated by zone. We find that over 90% of girls state that their teachers make them feel welcome in the classroom; teachers were most effective at making students feel welcome in South West State, where almost 96% of girls reported positively on this dynamic.

TABLE 79: TEACHER PROFESSIONALISM, C5 NFE CENTRES

Outcome	Overall	Banadir	Hirshabelle	South West State
Feels welcome	92.3%	90.4%	90.1%	95.7%
Not often absent	76.6%	72.3%	78.3%	81.4%

Results were more mixed for teacher absenteeism, with only around three-quarters of girls reporting that their teacher is not "often" absent; in other words, nearly 25% of C5 NFE girls reported that their teacher *is* often absent, a high rate especially considering the potentially severe impact of teacher absenteeism on learning and student attendance. Teachers were reported to be absent most frequently in Banadir and least frequently in Hirshabelle, although the difference was not significant; similarly, within these zones, there were no significant differences by district, suggesting that teacher absenteeism may be a widespread issue rather than solely influenced by security or environmental factors.

Disciplinary Practices

We now analyse negative disciplinary practices used by teachers, including punishing students for wrong answers and use of corporal punishment. First, during classroom observations, enumerators observed the use of physical punishment in classrooms. We find that 0% of C5 NFE teachers used physical punishment during observed classes, a promising finding. However, we expect teachers to use physical punishment infrequently when observed, especially if they are aware that it is an undesirable practice; as such, this finding represents a lower bound on the use of physical punishment.

We thus use data from the girls' survey to validate these findings. The below table shows the prevalence of punishment for wrong answers and physical punishment by teachers, as reported by girls. We find substantially higher rates of both these negative disciplinary practices than suggested with classroom observations, with 15% of girls reporting the use of physical punishment in classrooms and over half of girls reporting that they are punished for wrong answers. Use of negative disciplinary practices—and particularly physical punishment—appears most acute in Hirshabelle.

TABLE 80: USE OF NEGATIVE DISCIPLINARY PRACTICES, C5 NFE TEACHERS

Outcome	Overall	Banadir	Hirshabelle	South West State
Punishment for wrong answer	51.3%	49.0%	54.1%	53.0%
Use of physical punishment	15.0%	16.9%	23.1%	9.2%

While these reported rates of negative disciplinary practices are high, we note that rates are significantly lower than for FE girls. A potential explanation for this was stated by a teacher in an FGD: "This school enrols 17-year-olds and older people who are also women, so there can be no punishment or harsh discipline".[107] Indeed, on average, C5 NFE girls were around five years older than FE girls at ML2. Regardless, for both FE and NFE girls, there remains a need for interventions to reduce the use of negative practices in classrooms—particularly given that these practices can have a significant negative effect on learning and student attendance.

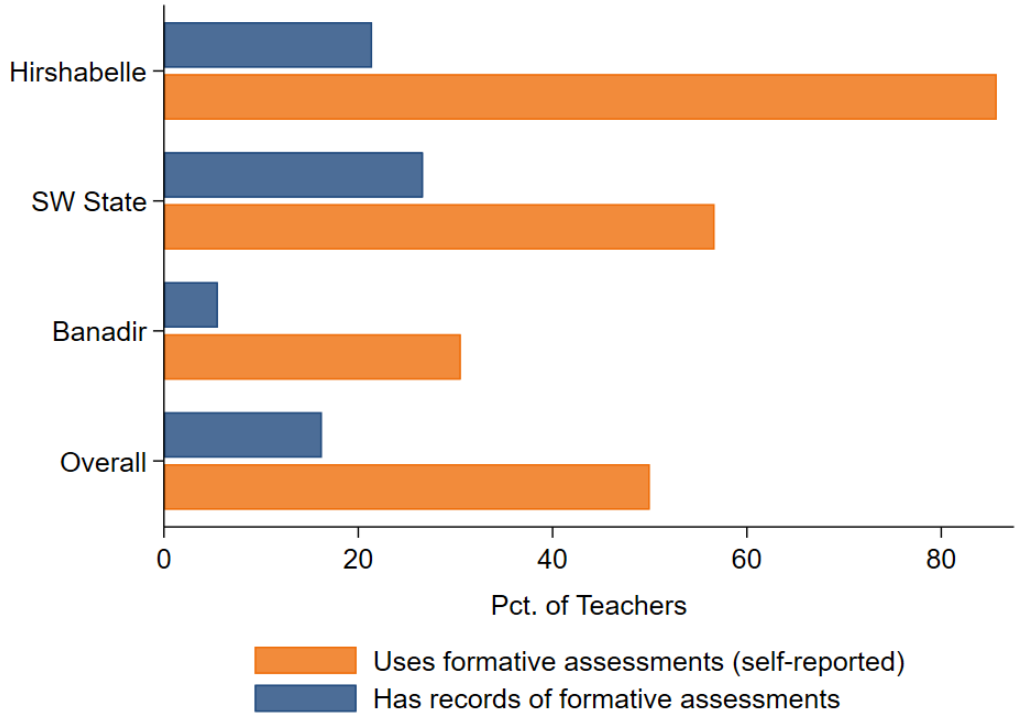
Pedagogical Practices

Lastly, we examine the pedagogical practices used by teachers, including the use of formative assessments and positive teaching practices such as student-centred activities or games, allowing students to instruct each other, asking open-ended questions, soliciting students' opinions, involving students who are not participating, and using group work. We also analyse the frequency with which teachers call on girl students, provide positive feedback, and use harsh language in class. As in above sections, we cross-reference data from classroom observations and the survey with girls to validate findings.

Figure 21 shows the rate at which teachers stated that they use formative assessments and the rate at which they were able to provide records of formative assessments by zone. Across all zones, 50% of teachers stated that they used formative assessments, ranging from a high of 86% in Hirshabelle to a low of 31% in Banadir. In all zones, however, a substantial portion of teachers were not able to produce records of formative assessments. The gap was particularly large in Hirshabelle, where only 25% of teachers who stated that they used formative assessments had records; in contrast, in South West State, around half of teachers were able to produce records of their formative assessments. Overall, however, these results point to a need to both increase the use of formative assessments and improve record-keeping.

[107] FGD with teachers, Middle Shabelle, Int. 512

FIGURE 21: USE OF FORMATIVE ASSESSMENTS IN C5 NFE CENTRES



We examine the use of participatory and student-centred teaching practices in the below table, which reports the percentage of teachers who were observed to have used a positive practice in at least two out of three observation blocks. We find that no teachers used activities or games in at least two observation blocks; group work was also used very infrequently. Regarding group work, however, we note that 36% of teachers used group work in at least one observation block; this may suggest that some teachers view group work as valuable but time-consuming, and thus as a teaching practice best used selectively in class.

In contrast, in more than half of classrooms, teachers were observed asking open-ended questions and soliciting student opinions in at least two blocks. Seventy percent of observed teachers also attempted to involve students who were not participating. While there is thus still substantial room for improvement in the use of positive teaching practices, this suggests that many teachers are indeed attempting to engage their students.

TABLE 81: USE OF POSITIVE TEACHING PRACTICES, C5 NFE CENTRES

Outcome	Overall	Banadir	Hirshabelle	South West State
Use of student-centred activities or games	0.0%	0.0%	0.0%	0.0%
Students instruct each other	36.3%	25.0%	50.0%	43.3%
Use of open-ended questions	52.5%	41.7%	42.9%	70.0%
Solicitation of student opinions	53.8%	50.0%	50.0%	60.0%

Involvement of students who are not participating	70.0%	50.0%	85.7%	86.7%
Use of group work	3.8%	5.6%	7.1%	0.0%

Disaggregating by zone, we find the lowest use of participatory teaching practices in Banadir; indeed, in this region, teachers used an average of only 6 positive teaching practices per class, compared to 7 in Hirshabelle and 8 in South West State. As such, it may be useful to focus on strengthening teaching practices in NFE centres in Banadir.

Table 82 shows the frequency with which NFE teachers interacted with girl students positively, by calling on them or providing positive feedback, and negatively, by using harsh language. We find relatively high rates of teachers interacting positively with students and low usage of harsh language in class. Teachers called on students most frequently in South West State but also used harsh language most frequently in this state.

TABLE 82: INTERACTIONS WITH STUDENTS, C5 NFE CENTRES

Outcome	Overall	Banadir	Hirshabelle	South West State
Average number of times teacher called on girl students	10.2	7.2	7.1	15.3
Percent of teachers providing positive feedback to girl students	71.3%	55.6%	92.9%	80.0%
Average number of times teacher used harsh language with girl students	0.4	0.1	0.0	0.9

Lastly, we validate these results from classroom observations with data from the girls' survey. Table 83 shows that girls generally had positive views of the practices used by their teachers, with over half of girls reporting the use of each of the five listed practices. Girls were least likely to state that their teachers often gave them ideas for how to study outside of class; in contrast, girls frequently stated that teachers explain the relevance of subjects to their lives and encourage participation. This last finding corresponds to results from the classroom observations, when 70% of teachers were similarly observed to encourage students to participate (see Table 81).

TABLE 83: USE OF POSITIVE TEACHING PRACTICES REPORTED BY C5 NFE GIRLS

Outcome	Overall	Banadir	Hirshabelle	South West State
Teacher often explains use of subjects	76.9%	75.6%	77.4%	78.3%
Teachers often gives ideas to learn outside of class	56.4%	51.9%	61.5%	59.8%
Lessons move at the right speed	65.6%	69.3%	65.9%	60.5%

Teacher often uses different ways of explaining	64.5%	66.4%	57.7%	65.2%
Teacher often encourages participation	72.7%	76.8%	79.8%	64.1%

There was relatively little variance in reported use of positive teaching practices, with no consistent under- or overperformance, by zone. Overall, these results suggest that teachers in C5 NFE centres across all zones are attempting to engage students by using participatory, positive teaching practices, and that many girls have positive perceptions of the quality of teaching in their NFE centres. However, there remains substantial room for improvement across all assessed pedagogical practices.

11.3 Leadership and Life Skills

The Youth Leadership Index is a composite indicator based on 21 questions, measured on a 4-point likert scale. Respondents are asked to indicate how often (rarely, sometimes, most of the time and almost always) they acted in a certain way, depending on the question asked. Lower values indicate more negative outcomes and higher values indicate more frequent instantiations of the behaviour and, by extension, more positive outcomes. The score ranges between 21 and 84 points and for the purposes of the analysis the score was standardised on the scale of 0 to 100. When a girl scored the lowest possible number of points (21) by responding 'rarely' to all questions, the standardised YLI score will take the value of 0%.

The YLI was developed by CARE International to measure a respondent's self-confidence, decision-making, voice, vision and organizational skills (including the ability to motivate others and work with them to address common issues). We refer readers to Section 7.3 for a more detailed introduction of the YLI, its construction, and its constituent questions.

Cohort 4 NFE Girls

In this section, we briefly describe changes in YLI scores over time since the ML1 round, when C4 NFE girls were first interviewed. This set of girls showed marked improvement in their YLI scores from ML1 to ML2. At ML1 (the baseline for this cohort), the mean YLI score was 49.8. By ML2, YLI scores showed a significant increase of 16.6 points to 66.4. These results suggest a significant success for the programme in raising leadership among the C4 NFE cohort. Increases in the proportion of girls meeting the target of 70 on the YLI also mirror those observed for the baseline cohort 1 girls. At ML1, the proportion of C4 NFE girls who attained a score of at least 70 percent on the YLI was 11.3%. At ML2, this had increased to 42.7% (representing a 31.4% increase).

Finally, YLI scores at the state level all rose dramatically as well. Hirshabelle saw the biggest change, increasing from 42.8% at ML1 to 70.7% at ML2 (27.9% increase). Banadir and South West State had higher baseline values, but also saw similarly smaller increases at ML2. Banadir's YLI scores increased from 53.1% to 67.7% (14.6% increase) while South West State's YLI scores increased from 487% to 62.8% (14.1% increase). The table below summarises the changes in YLI scores from ML1 to ML2 for the C4 NFE girls.

TABLE 84: CHANGES IN YLI SCORES FROM ML1 TO ML2 AMONG C4 NFE GIRLS, BY ZONE

C4 NFE Cohort Subgroups	ML1	ML2	Difference (ML2 – ML1)
YLI Score (0-100 standardised scale)			
Overall	49.8	66.4	16.6***
Banadir	53.1%	67.7%	14.6***
Hirshabelle	42.8%	70.7%	27.9***
South West State	48.7%	62.8%	14.1***
Share of Girls Achieving 70% YLI Score Target			
Overall	11.3%	42.7%	31.3***
Banadir	14.8%	43.8%	28.9***
Hirshabelle	17.6%	57.4%	39.8***
South West State	4.4%	34.9%	30.5***

*** significant at 99% level, ** significant at 95% level, * significant at 90% level

Cohort 5 NFE Girls

In addition to tracking the changes in YLI for the baseline cohorts and the C4 NFE cohort, this study established baseline values for a new cohort of girls (C5 NFE girls) recruited at ML2. Interestingly, baseline values for the YLI scores are actually higher for the C5 NFE cohort (67.1) than any other cohort of girls. Likewise, the proportion of girls achieving a minimum score of 70 percent (a programme target) is highest among the C5 NFE cohort (47.0%). These results are surprising because the C5 NFE cohort is not older than the other NFE cohorts, so this higher-than-expected starting YLI score cannot be attributed to age. Qualitative interviews of C5 NFE girls included several mentions of parents, husbands, or siblings supporting girls to enrol in NFE courses by taking care of their children or the housework.[108] While it is possible that the C5 NFE cohort has a higher baseline levels of leadership skills as a result of the programme creating a more supportive environment at the community level, we do not have strong enough evidence to draw conclusions about the high baseline YLI value for the C5 NFE girls. At baseline, the girls in Banadir have a significantly higher baseline YLI score (67.2) compared to Hirshabelle (56.1) or South West State (55.7).

TABLE 85: YLI SCORES, BY ZONE, AMONG C5 NFE GIRLS

C5 NFE Cohort Subgroups	ML2
YLI Score (0-100 standardised scale)	
Overall	67.1

[108] Vignettes FGD with C5NFE Girls, Banadir, Int. 605 & 606

Banadir	69.2
Hirshabelle	71.3
South West State	62.6
Share of Girls Achieving 70% YLI Score Target	
Overall	47.0%
Banadir	50.6%
Hirshabelle	56.5%
South West State	37.8

11.4 School Management and Governance

Cohort 5 NFE

In this section, we establish baseline values for measurements of school management and governance in C5 NFE centres. In particular, for NFE centres, we focus on perceptions of the involvement of community education committees (CECs) in addressing barriers to the enrolment and learning of ultra-marginalised girls, utilizing data from the household survey.[109] These committees are generally intended to identify and resolve challenges to student attendance and learning, including by monitoring student attendance, reaching out to absent or out-of-school students, and finding solutions to mitigate barriers to attendance; they thus play an important role in enhancing inclusion. To evaluate their efficacy, we first assess the main barriers to girls' attendance, followed by analysis of work undertaken by the CEC to address these barriers. We then highlight results of the qualitative analysis in the previous section on school management and governance to better understand the efficacy of CEC initiatives.

Barriers to Education

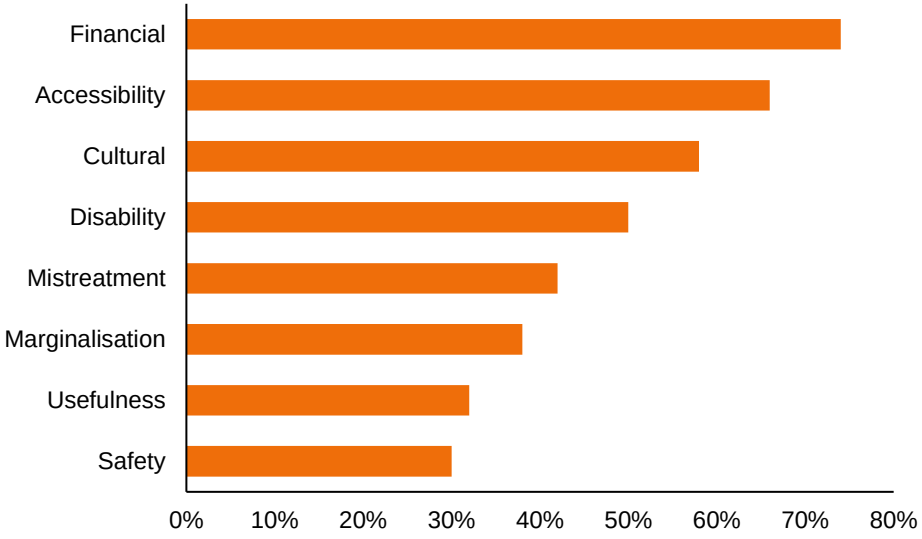
To better understand the efficacy of CECs in addressing barriers to education, we first analyse the most salient barriers facing girls. We categorise these barriers into eight categories: (1) cultural, (2) financial, (3) accessibility, (4) safety, (5) mistreatment, (6) disability, (7) marginalisation, and (8) usefulness. Cultural barriers include conflicts with religious beliefs or the belief that girls' education is not valuable. Financial barriers include insufficient money to pay for school or the need for a girl to work to earn money. Accessibility barriers include long travel distances or insufficient transportation. Safety barriers include lack of safety travelling to or while at school. Mistreatment barriers include poor treatment of girls by teachers or classmates. Disability barriers include insufficient accommodations within schools or from teachers for a variety of cognitive or physical disabilities. Marginalisation barriers include insufficient access to toilets for girls at school, age, marital status, and pregnancy/motherhood status. Lastly, usefulness barriers include perceptions that school is not useful or does not prepare girls for the workforce.

Figure 22 shows the prevalence of various barriers to education for C5 NFE girls; we do not disaggregate results due to low sample size. We first find that NFE girls—in alignment with AGES' intention to target ultra-marginalised girls—face many barriers to education. Indeed, among the 50 caregivers of NFE girls asked

[109] Head teacher surveys were not conducted within NFE centres; as such, we do not utilize data from these surveys in this section.

about barriers to education, only two stated that the girl did not face any barriers, and on average, caregivers stated that the girl faced around four barriers.

FIGURE 22: BARRIERS TO EDUCATION, C5 NFE GIRLS



As shown in the figure, the most prevalent barrier to education was financial, with 74% of caregivers reporting that this had led the girl to not attend school. The second most frequent barrier was accessibility, cited by 66% of caregivers; looking further into this barrier, we find that almost half of caregivers stated that there was no one available to travel with the girl to or from school, again emphasizing the particular challenges faced by girls (rather than boys) to attend school within Somalia. Along these lines, the third most cited barrier was cultural, stated by 58% of caregivers; this includes 32% of caregivers who stated there was insufficient time or resources for both religious and "Western" education, 30% who stated schooling was not important for the girl, and 24% who stated that "Western" education was not religiously or culturally acceptable. These patterns suggest a need for continued sensitization on both the value of education and its acceptability in a Muslim culture.

Disability and marginalisation-related barriers were also cited relatively frequently by caregivers. Looking at components of disability-related barriers, we find that caregivers were most likely to state that improper accommodation in schools was a barrier to education, with 22% of caregivers stating that teachers did not know how to teach the girl and that the school did not have a programme meeting the girl's learning needs. This barrier was echoed in qualitative interviews:

> *[Respondent 1:] There are many challenges faced by girls with disabilities, such as being discriminated against… [Respondent 2:] There are challenges by not having special teachers for students with disabilities.*
>
> - FGD with CEC members, Banadir, Int. 102

For marginalisation, the most frequently cited issue was with insufficient toilets for girls at schools, emphasizing the need for gender-sensitive school infrastructure. Within FGDs, many CEC members also stated that marriage and pregnancy were major barriers for girls to attend school; a CEC member in Bay, for

example, stated that, "One of the challenges is that girls are married off at a young age, which is an obstacle to their education",[110] while another member in Banadir stated the following:

> *A 13-year-old girl who got pregnant, whether she is married or single, may have her parents forbid her from attending school so they won't be subjected to discrimination.*
>
> - FGD with CEC members, Banadir, Int. 101

CEC Initiatives to Address Barriers to Education

With this information in mind, we now analyse the extent to which CECs are addressing barriers of relevance to girls' education. First, we note that only 66% of C5 NFE girls stated that their CEC did anything to support girls' education; involvement ranged from 56% in Hirshabelle to 67% in Banadir and 68% in South West State. This relatively low level of CEC activity points to need to strengthen school management and governance, particularly in Hirshabelle.

The below table shows the specific initiatives undertaken by CECs to support girls' education. Girls most commonly reported that the CEC supported enrolment of out-of-school girls, followed by supporting dropouts to return to school and following up on student absenteeism. However, the frequency of these initiatives was still limited, at only around half of CECs.

TABLE 86: CEC INITIATIVES TO SUPPORT GIRLS' EDUCATION, REPORTED BY C5 NFE GIRLS

Initiative	Overall	Banadir	Hirshabelle	South West State
Enrolment of out-of-school girls	55.4%	48.8%	60.4%	61.9%
Support dropouts to return to school	50.6%	43.8%	43.8%	61.9%
Check when a girl is absent	42.9%	43.2%	41.7%	42.9%
Awareness raising on girls' education	25.6%	27.2%	10.4%	29.4%
Address cases of harassment, abuse, and violence	15.2%	10.5%	8.3%	23.8%
Provide hygiene materials for girls	11.0%	6.2%	8.3%	18.3%
Monitor teaching quality	8.9%	6.8%	6.3%	12.7%
Monitor teacher absenteeism	6.3%	6.2%	2.1%	7.9%
Hire female teachers	3.6%	4.3%	0.0%	4.0%

Furthermore, looking at barriers to education reported by caregivers, we find a mismatch between CEC initiatives and reported barriers. For example, cultural barriers were reported by 58% of caregivers, but a relevant CEC initiative—awareness raising on girls' education—was only undertaken by one quarter of CECs. Similarly, mistreatment was reported by 42% of caregivers, but only 15% of CECs were reported to address cases of harassment, abuse, and violence. Accessibility to toilets was also a salient barrier reported by caregivers, but CECs have little involvement in providing hygiene materials to girls which may help mitigate this barrier.

[110] FGD with CEC members, Bay, Int. 106

Overall, for C5 NFE girls, CECs appear to have a moderate level of activity and, likely, a moderate level of efficacy given some misalignments between reported and addressed barriers. We note also that CEC activity was highest—though still limited—in South West State, and generally most limited in Hirshabelle. There was no major difference, however, in the pattern of initiatives undertaken by CECs across zones.

Data from FGDs, analysed in detail in the FE School Management and Governance section, sheds further light on the activities of CECs and actions taken to address barriers to education for marginalised girls. Despite the infrequency with which CECs were reported to undertake awareness-raising in Table 86, we find that CEC members reported this as a major activity in qualitative interviews. Indeed, several CECs reported that this was one of the only activities they were capable of undertaking due to limited financial resources.[111]

Most CEC members also stated that they were responsible for monitoring student attendance, following up in cases of absent or out-of-school children, and ensuring the quality of instruction. In Bay, for example, CEC activities were summarised as follows: "We check students' attendance and absences, monitor quality of education, new student registration, and contact parents for absent students".[112] In some cases, CECs also described a focus on supporting marginalised students to attend school: "[The CEC] does a lot of outreach to marginalised students, and they try to bring those students to school, and they succeeded".[113]

However, as noted in the previous School Management and Governance section, many CECs described a lack of resources which limited their efficacy to support schools and marginalised students. For example, a CEC member in Banadir stated that they had not done anything to address challenges faced by students with disabilities or pastoralist students because they could not afford anything due to financial circumstances.[114] Overall, this analysis suggests that CECs are moderately active in supporting NFE centres and girls, but that they lack resources and that activities may not align perfectly to the barriers faced by girls and marginalised students.

11.5 Increased Self-Efficacy

Cohort 4 NFE Girls

This section will examine questions and attitudes related to positive youth development as well as access to protection services. The reader should note that the following section focuses exclusively on girls from the NFE Cohort 4 that responded to the questions on self-efficacy in both the ML1 and ML2 evaluations.

Positive Youth Development

During the Midline 1 and Midline 2 evaluation rounds, data on positive youth development was collected from the C4 NFE girls using indicators from the Chinese Positive Youth Development Scale (CPYDS). This globally recognised scale assesses various aspects of youth development such as resilience and confidence. Seven indicators related to self-efficacy were chosen from the CPYDS. The table below lists the seven indicators and tracks the percentage of girls who either agreed or strongly agreed with the statements from ML1 to ML2.

[111] FGD with CEC members, Banadir, Int. 102; FGD with CEC members, Bay, Int. 109; FGD with CEC members, Middle Shabelle, Int. 113
[112] FGD with CEC members, Bay, Int. 108
[113] FGD with CEC members, Lower Shabelle, Int. 12
[114] FGD with CEC members, Banadir, Int. 102

TABLE: C4 NFE CHANGE IN PERCENT AGREEING WITH CPYDS QUESTIONS

CPYDS Question	ML1	ML2	Difference
Lack of control of life	44.6%	68.4%	+23.8
Lack of solutions to problems	54.5%	63.1%	+8.6
Inability to change life	41.5%	64.9%	+23.4
Helplessness	47.8%	62.5%	+14.7
Fate not in hands	57.6%	69.9%	+12.3
Determine own life	56.3%	77.3%	+21.0
Ability to complete tasks	75.2%	87.0%	+11.8

Control of own life

The first indicator measures the girls' perception of control over their lives. When prompted with the statement, "I have little control of things that happen in my life", 44.6% of the C4 NFE girls agreed with the statement during the ML1 round, and this number increased to 68.4% in agreement during the ML2 round. This percentage change of 23.8 points is the largest increase across the seven indicators measured for the C4 NFE girls.

Disaggregating this change by region reveals that the perception of control of one's own life is weakest in the Banadir region with 42.4% of the girls initially agreeing with the statement and 74.3% agreeing with the statement during the ML2 evaluation. Lower Shabelle and Bay likewise saw increases in the percentage of girls agreeing that they have little control of things that happen in their lives. Middle Shabelle, however, saw a decrease in the percentage of girls agreeing with this sentiment (from 84.3% to 79.6%), although the percent who agree in ML2 is still the highest of all the regions.

Interestingly, Maay speakers had a smaller percentage agreeing with the statement in both periods (36.8% and 57.7%) compared to non-Maay speakers (48.9% and 74.2%). Girls with disabilities had higher proportions agreeing that they have a lack of control over their lives (55.2% in ML1 and 70.1% in ML2) compared to girls without disabilities (41.7% in ML1 and 66.8% in ML2), although both showed moderate increases.

Solutions to problems

The second indicator measures the girls' belief in their ability to solve problems. When faced with the statement, "I do not have any solutions for some of the problems I am facing", 54.5% initially agreed in ML1, and this number increased to 63.1% in ML2. These numbers are particularly interesting because a majority of girls *already* believed they did not have solutions for some of their problems in the initial round of CPYDS questioning, and this proportion increased further over time.

Initially, girls in Middle Shabelle were the least likely to feel as though they can solve their own problems as a staggering 89.8% of the girls agreed with the above statement. While this number did decrease in the ML2 evaluation, it only fell to 74.1%. Banadir showed a small increase in the percent of girls agreeing with the statement, rising from 61.2% to 64.5%. Lower Shabelle showed a significant increase in the percentage of girls agreeing with the statement, rising from 13.1% in ML1 to 73.8% in ML2 (eighty-four girls responded from Lower Shabelle for this cohort). Bay was the only region in which the percent agreeing with the statement remained the minority opinion, indicating that most girls still believe they have solutions for the issues they encounter (it did still rise from 40% to 47.9%).

The same trend emerges in terms of Maay versus non-Maay speakers as with the first indicator. Non-Maay speakers increased from 62.1% to 67.2% and Maay speakers from 40.6% to 55.6%, indicating that Maay speakers are more likely to agree that they can solve their own issues. Girls with disabilities did not undergo any significant changes in terms of their perception of their ability to solve problems, but girls without disabilities changed from 50.6% in agreement to 59.5%.

Ability to change

The third indicator measures how the girls perceive their ability to change their own life. When prompted with the statement, "I cannot do much to change things in my life", 41.5% agreed with the sentiment in the ML1 round, while 64.9% agreed during the ML2 round.

Banadir, Lower Shabelle, and Bay all demonstrated increases in the percent agreeing with the statement, while once again Middle Shabelle showed a decrease in agreement. Lower Shabelle experienced a particularly large increase (from 7.1% to 69%). Bay was the only region to remain with a minority in agreement for both periods (from 37% to 47.9%). Girls with and without disabilities both experienced increased percentage changes (+21.7 and +19.3 percentage points, respectively).

Helplessness

The fourth indicator relates to the girls' perceptions of helplessness. When given the statement, "When I face life difficulties, I feel helpless", 47.8% of the girls agreed during the ML1 round while 62.5% agreed during the ML2 round. Once again, there is a transition from a minority in agreement to a majority in agreement with a statement indicative of weaker self-efficacy.

Banadir, Lower Shabelle, and Bay saw increases in the percentage of girls agreeing with the statement such that a minority agreed in ML1, and a majority agreed in ML2. Once again Middle Shabelle showed a decrease in agreement. While Middle Shabelle was the weakest in terms of girls' perception of their helplessness ML1 (84.3%), this percentage fell to 67.6% in ML2.

Disaggregating the results by Maay-language ability and disability status proves to have negligible results for this question.

Fate not in hands

The fifth indicator focuses on how the girls perceive their fate. Overall, 57.6% of the individuals initially agreed with the statement, "I feel my life is determined by others and fate". During the ML2 round, 69.9% of girls agreed with the statement. Once again, the initial ML1 percentage consisted of a majority already agreeing that they feel their life is determined by others, which indicates weakened self-efficacy.

The regions of Banadir, Lower Shabelle and Middle Shabelle all saw most girls agreeing they felt their fate was out of their hands in ML1 and saw this percentage increase in ML2. In Bay, however, the percentage of girls in agreement with feeling out of control of their own fate *decreased* from 47.2% to 46.1% between the two rounds.

Once again following the previous trends, non-Maay individuals had higher rates of agreements in both periods (61.3% and 77.0%) compared to Maay speakers (50.9% and 56.8%). Analysing the results comparing girls with disabilities versus girls without disabilities yields negligible differences for this question.

Determine own life

The sixth indicator measures the girls' perception of their ability to determine their own life, which differs from the preceding indicators because it is a measure of strong self-efficacy. When faced with the question, "I believe things in my life are mostly determined by me", 56.3% agreed in ML1, and this increased to 77.3%

in ML2. This result is important because it demonstrates how positive self-efficacy traits are also increasing; these increases are not limited to indicators related to negative self-efficacy traits such as feeling helpless or unable to solve issues.

The ML1 results from Banadir found that only 46.1% of the girls agreed that they determine most things in their lives, but this number increased to 77.3% in ML2. The other three regions saw increases as well.

Disaggregating the results by language and disability status has negligible results for this question.

Task completion

The seventh and final indicator measures the girls' perception of their ability to finish tasks. The statement provided was, "I can finish almost everything that I am determined to do", which is once again an indicator of positive, strong self-efficacy. The overall percentage of C4 NFE girls agreeing with this statement increased from 75.2% to 87%.

All regions experienced increases in the percentage agreeing that they can finish tasks. Lower Shabelle had the smallest increase, from 70.2% to 75%, while Middle Shabelle experienced a sizeable increase from 78.7% to 90.7%.

Like indicator six, disaggregating the results by language and disability status has negligible results for this question.

Synthesis

As the table above shows, all seven of the statements related to the self-efficacy indicators experienced increases in the percentages of girls agreeing with them from ML1 to ML2 (this is a similar result to the earlier section that analysed the changes in these statements for the C1 NFE girls). These increases were evident with statements that were indicative of a *weaker* perception of self-efficacy and self-confidence (e.g., "When I face life difficulties, I feel helpless"), as well as with statements that were indicative of a *stronger* perception of self-confidence (e.g., "I can finish almost everything that I am determined to do").

When disaggregating the data by region, certain trends emerge. For instance, Lower Shabelle often saw drastic increases in sentiments related to negative self-confidence (see indicators one through four). Additionally, the girls in the Bay region typically had lower, more consistent agreement percentages for questions related to negative self-confidence indicators.

While there are a few anomalies at the sub-group level, these groups and the overall sample all appear to follow the trend of a declining perception of self-confidence for the C4 NFE girls. The rise in proportions agreeing with statements related to helplessness and an inability to change one's life transcend region and all the other variable groups discussed in this section such as the presence of disabilities and language barriers. Due to this transcendence, it is possible that the girls are becoming less optimistic about their capacity to determine their own fate as they become older, more mature, and face certain life challenges. The average age of the C4 NFE girls that were asked the CPYDS questions in the ML1 round was 19.3 years old, with a minimum of 11 and a maximum of 28 years old. It is possible that as these girls become older and encounter obstacles related to life and work changes, their perception of their ability to affect change in their lives may worsen. It is, however, important to note that the last two indicators (determining one's own life and task completion) which are indicative of a positive view of self-efficacy, also saw sizeable increases in the percentage agreeing with the statements.

Access to Protection Services

This section evaluates the C4 NFE girls' access to protection services across ML1 and ML2 in both the school and the community, and the channels through which they can report abuse, harassment, or exploitation.

School

During the ML1 round, 92.9% of the girls responded "Yes" when asked if they have an avenue to report any form of harassment, abuse, or exploitation at the school. During the ML2 round, this percentage decreased to 78.4%, indicating a sizeable drop in the girls' perception of their ability to report abuse. When asked who their top choices are for reporting said abuse, the top choices remained the head teacher or other teachers across ML1 and ML2. The other popular choice was the "Other" category, which included specified responses such as parents, husbands, and sisters.

Community

During the ML1 round, 81.5% of the girls responded "Yes," when asked if they have an avenue to report any form of harassment, abuse, or exploitation in the community. During the ML2 round, this percentage decreased to 73.1%. This decrease in the girls' perceived ability to report community abuse mimics the trend seen above in the school scenario.

Also like the school scenario, the top choices for reporting abuse in the community included teacher and head teacher, but the third most popular option in this case was the police. This differs from the school-focused scenario, presumably because approaching the police is a more natural-seeming method for dealing with abuse/harassment that occurs outside a school setting. At ML2, 8.6 percent of girls felt they could report an incident of this kind, occurring in the community more generally, to the police, though this rate has declined alongside a decline in perceived ability to report to other individuals as well.[115] In other words, the share of girls who feel they could report an incident to a head teacher, a teacher, the CEC, a community leader or clan elder, a religious leader, and police have all declined since ML1 – which is likely a function of the fact that, at the time of the ML1 evaluation, AGES had very recently established reporting mechanisms. These mechanisms were new and well-known at the time; however, as girls have now transitioned out of the NFE centres, these mechanisms may be slightly less helpful, especially for community-based abuse/harassment.

Synthesis

Despite the drop in percentage of girls agreeing that they have avenues to report abuse in both the school and the community settings, a three-fourths majority still agrees that they have options for reporting harassment or abusive situations in both cases. What is particularly interesting in this case is the lack of diversity in who the girls are reporting the incidents to. The options provided included a plethora of options including CEC and GEF representatives, religious leaders, mentors, and female religious scholars. The CEC and GEF options remain underutilised resources in both the ML1 and ML2 rounds. Teachers, police, and family members remain the preferred option for reporting abuse.

Cohort 5 NFE Girls

This section will examine questions related to positive youth development as well as access to protection services. The reader should note that the following section focuses exclusively on Cohort 5 NFE girls. Due to the C5 NFE girls being recruited during the ML2 evaluation round, these answers serve as the baseline for this cohort.

[115] Willingness or ability to report community-based incidents of abuse/harassment to the police has declined from 17.8 percent to 8.6 percent from ML1 to ML2.

Positive Youth Development

During the Midline 2 evaluation rounds, data on positive youth development was collected from the C5 NFE girls using indicators from the Chinese Positive Youth Development Scale (CPYDS). Seven indicators related to self-efficacy were chosen from the CPYDS. The table below lists the seven indicators and the percentage of C5 NFE girls who either agreed or strongly agreed with the statements.

TABLE: C5 NFE PERCENT AGREEING WITH CPYDS QUESTIONS

CPYDS Question	ML2
Lack of control of life	72.0%
Lack of solutions to problems	67.5%
Inability to change life	67.8%
Helplessness	60.2%
Fate not in hands	70.6%
Determine own life	77.3%
Ability to complete tasks	87.3%

Control of own life

When prompted with the statement, "I have little control of things that happen in my life", 72.0% of the C5 NFE girls either agreed or strongly agreed with the sentiment. Perception of having a lack of control over one's life was most prevalent in Banadir, where 77.1% of girls agreed with the statement, and least prevalent in Bay, with only 56.8% of girls agreeing.

When breaking down the effect of language on responses to the first statement, 74.9% of non-Maay speakers agreed with having a lack of control, while 64.9% of Maay speakers agreed with the statement. Girls with a female head of household are more likely to agree with the statement (70.7%) versus girls from a non-female led household (62.5%). Disaggregating the results using the presence of a disability does not yield significant results.

Solutions to problems

When faced with the statement, "I do not have any solutions for some of the problems I am facing", 67.5% of C5 NFE girls either agreed or strongly agreed that they had a lack of solutions. Here, it is Middle Shabelle with the highest percentage in agreement (74.1%) indicating the weakest perception of self-efficacy in this category, followed by Middle Shabelle (72.9%), then by Banadir (67.5%), and once again Bay represents the lowest percentage in favour of the statement (59.1%).

Also similar to the first indicator is the fact that non-Maay speakers agree with the statement more so than Maay speakers (74.9% versus 64.9%). Those with a disability are slightly more likely to agree that they suffer from a lack of solutions (70.5%) compared to those without a disability (65.2%).

Ability to change

When prompted with the statement, "I cannot do much to change things in my life", 67.8% of the girls agreed or strongly agreed with the statement. The belief in an inability to enact life changes was strongest in Middle Shabelle (72.9%) and weakest in Bay (53.4% agree, indicating that almost half believe that they *do* have the ability to change their own life).

Non-Maay speakers had an agreement percentage of 70%, while their Maay speaking counterparts had a percentage of 62%. There was an almost ten-percentage point difference between those without a disability (63.8%) and those with a disability (73.3%).

Helplessness

When given the statement, "When I face life difficulties, I feel helpless", 60.2% of the girls agreed overall, which means that almost 40% do *not* feel helpless when facing difficulties. The lowest percentage in agreement was Middle Shabelle with 52.9%, and the highest percentage in agreement was Lower Shabelle, with 63.9% in agreement.

For this question, the relationship between agreeing with the statement and language status is reversed compared to the previous three indicators in which non-Maay speakers had higher percentages of girls agreeing. In this instance, only 58.2% of non-Maay speakers agree, and 64.9% of Maay speakers feel helpless in the face of difficulties. Additionally, girls with disabilities demonstrate a stronger perception of helplessness (67.7% in agreement) compared to girls without a disability (54.6%).

Fate not in hands

Overall, 70.6% of the girls agreed with the statement, "I feel my life is determined by others and fate". The perception of not having one's fate in their own hands is highest in Lower Shabelle (79.4%) and lowest in Bay (43.2%, which represents the first time a majority *disagrees* with the statement related to weaker self-efficacy perceptions).

Only 57% of Maay speakers agree that their fate is outside of their hands, while 76.3% of non-Maay speakers agree with this statement. The difference in results when disaggregated by disability are negligible.

Determine own life

When faced with the question, "I believe things in my life are mostly determined by me", 77.3% of the girls agreed. This statement associated with the sixth indicator measures the girls' perception of their ability to determine their own life, which differs from the preceding indicators because it is a measure of strong self-efficacy. Therefore, higher percentages of girls who agree with this statement is associated with stronger perceptions related to self-efficacy.

Unlike the first five indicators, differences in percent agreeing with the statement does not differ greatly across region. The lowest percent agreeing that their lives are self-determined was in Bay (75%) and the highest was in Lower Shabelle (79.3%).

The difference in results when the data is disaggregated by disability or language spoken are negligible.

Task completion

The statement provided for the seventh and final indicator was, "I can finish almost everything that I am determined to do", which is once again an indicator of positive, strong self-efficacy, and 87.3% of the girls either agreed or strongly agreed with this statement. In Banadir, 90% of the girls interviewed agreed, in Middle Shabelle 89.4% of the girls agreed, in Bay 85.2% of the girls agreed, and in Lower Shabelle 80.4% of the girls agreed.

Synthesis

When disaggregating the data for the C5 NFE girls based on region, language, and disability existence, certain trends emerge. Generally, girls in Bay appear to have higher levels of positive youth development, as they typically have lower percentages of girls agreeing with statements related to feelings of helplessness and being

stuck or unable to change their lives. On the other end of the spectrum, the Middle Shabelle, Lower Shabelle, and Banadir regions appear to have substantially lower levels of positive development, as high percentages of the girls in these areas report feeling helpless and unable to control their own fate.

One interesting trend relates to language differences. For six out of the seven indicators, Maay speaking girls reported higher perceptions of self-efficacy (the only outlier being indicator four which relates to helplessness) compared to their non-Maay speaking counterparts. This data on the self-perception of Maay speakers seems to challenge the idea that Maay speakers may encounter obstacles that weaken self-confidence, which is a view expressed in certain focus group discussions. For example, in an interview in Banadir, one girl explained, "Since she speaks Maay, it is possible to face insults… [There are] obstacles she may face".[116] The data from these CPYDS questions reveals important insights regarding positive self-perception amongst Maay speakers.

Overall, all seven indicators had a majority of girls agreeing with the statements. This is true for statements that were indicative of a *weaker* perception of self-efficacy and self-confidence (e.g., a majority of girls agreed, "When I face life difficulties, I feel helpless"), as well as with statements that were indicative of a *stronger* perception of self-confidence (e.g., "I can finish almost everything that I am determined to do"). Breaking down the data into sub-groups based on region, disability, and language also yields results in which the majority in each group agree with the statement (the only exception being the girls from Bay for indicator 5, where only 43.2% agree that their fate is not in their own hands).

It is interesting to note that the statements related to the CPYDS indicators focus on the lives of the girls as a whole. When asked about helplessness, the statement refers to "life difficulties" in general. It is possible that splitting up the difficulties into different sectors such as social, school, family, and work difficulties may also show different levels of self-confidence and self-efficacy depending on the sector of the girls' lives under consideration. During a focus group discussion, a group of girls in the Lower Shabelle region were asked what changes they had experienced as result of participating in the NFE courses. The girls were able to cite a multitude of new skills that had been acquired including math, English, the Somali language, and also the ability to make their family understand certain things about their lives and education.[117] In another interview in Middle Shabelle, when presented with a hypothetical scenario centred on a girl who has the opportunity to participate in classes, the interviewees responded that this girl can successfully participate in the classes if, "She shows her ambition and becomes self-sufficient".[118] These interviews demonstrate an acute awareness of the importance of the skills acquired through the NFE courses for self-sufficiency, as well as the connections between educational skills and measures of self-efficacy and self-confidence.

Access to Protection Services

This section evaluates the C5 NFE girls' access to protection services as recorded in ML2 in both the school and the community, and the channels through which they can report abuse, harassment, or exploitation.

School

When the C5 NFE girls were asked if they had an avenue to report a problem should they experience harassment, abuse, or exploitation at school, 73.5% responded, "Yes". In Middle Shabelle, only 54.1% of the girls responded "yes," in Banadir the percentage reached 72.1%, in Lower Shabelle 81.4%, and the maximum percentage reached was in Bay with 87.5% reporting in the affirmative.

[116] Vignettes FGD with Girls, Banadir, 605.
[117] Vignettes FGD with Girls, Lower Shabelle, 610.
[118] Vignettes FGD with Girls, Middle Shabelle, 612.

The girls were then asked who specifically they would report the abuse to. The top choices included the teacher, the head teacher, or "other" which typically referred to the girl's parents, sister, or husband.

Community

When the C5 NFE girls were asked if they had an avenue to report a problem should they experience harassment, abuse, or exploitation in the community, 65.1% responded "Yes". Middle Shabelle (47.1%) and Banadir (61.3%) were once again the two regions with the lowest percentages of girls who responded "Yes". In Lower Shabelle, 77.32% responded affirmatively, and the maximum was reached in Bay with 79.6%.

Also like the school scenario, the top choices for reporting abuse in the community included teachers and the head teacher, but the third most popular option in this case with 24 responses was the police.

Synthesis

In a risk mapping exercise conducted with C5 NFE girls in Banadir, when the girls were asked why they feel happy and safe in the school, one individual replied, "I feel happy throughout the whole school because there is a female teacher who is my friend," while another girl replied, "No one bothers you so everyone at school feels happy and safe".[119] These sentiments encapsulate the importance of the role of the teachers in creating a safe environment for the girls to grow their confidence and skills, which is why a majority feel they can report any abuse to the teachers or head teacher.

In contrast, when asked about feeling safe within the community, the girls reported certain fears. For example, one individual cited a fear of thieves who target girls with mobile phones. Another girl reported being afraid of a dark alleyway that passes by a graveyard behind her house.[120] These descriptions differ greatly from the above feelings of safety and happiness that exist within the school setting, and this is most likely why only 65.1% of the girls felt they knew where to report abuse in the community, as opposed to the 73.5% of girls that knew where to report abuse or harassment should it occur in school.

11.6 Strengthened economic situation of female youth

This section analyses the economic trajectory of girls participating in NFE programming as part of Cohort 4 (recruited at ML1) and Cohort 5 (recruited at ML2). For both cohorts, we analyse patterns in employment and monthly earnings. Employment is captured by sector, while income is a girl's self-reported income over the month prior to the survey. It is important to note that seasonal variation in income should have minimal effects on our findings regarding changes in income across rounds, as the ML1 and ML2 rounds were conducted in a similar period in each year (February-March 2022 for the ML1 round and March-April 2023 for ML2). We refer readers to Section 7.7 for a more detailed discussion of caveats and methodological limitations related to capturing income and employment status in this evaluation.

Cohort 4 NFE Girls

We begin by describing changes in monthly income among the C4 NFE cohort from ML1 to ML2. While our sample of C4 NFE girls includes 661 girls who were initially recruited at ML1 and successfully re-contacted at ML2, the sample whose incomes we analyse is smaller, limited to the girls for whom we have income information in both rounds. Due to the high rate of "don't know" responses to income-focused questions, the sample consists of just 263 girls with data in both rounds.

[119] Risk Mapping with Girls, Banadir, Int. 401.
[120] Risk Mapping with Girls, Banadir, Int. 401

The table below reports changes in income across rounds, disaggregated by geographic zone. The first three columns report results using the "full income panel", which includes all 263 girls for whom we have data in both rounds. Using this sample, we can see that monthly income has increased in all three zones, and that the increases are most substantial in Banadir, where incomes have increased by $84.63 per month – a change of 460% from ML1 to ML2. Gains were smaller – but still substantial – in South West State and Hirshabelle, where incomes more than doubled.

The second set of results differ, in that we have removed particularly extreme values – in terms of change in income over time for the same girl – from the sample. In other words, we removed girls whose changes in income were greater than $500, either decreasing or increasing since ML1. Using this reduced sample dampens the change in income observed in Banadir – while still statistically significant (at the 10 percent level), it is clear that the very large gains in Banadir are driven by a small number of extreme cases. In this case, the more conservative estimate – showing an increase in average income from $12.13 to $22.28 per month – is likely closer to the truth.[121]

TABLE 87: MEAN MONTHLY INCOME AMONG C4 NFE GIRLS, BY ROUND

	Full "Income Panel"			Income Panel, Excluding Outliers		
Zone	**Obs.**	**ML1**	**ML2**	**Obs.**	**ML1**	**ML2**
Overall	263	$18.32	$77.37**	249	$15.06	$24.21**
Banadir	156	$17.97	$102.6**	143	$12.13	$22.28*
South West State	99	$19.88	$41.65	98	$20.09	$26.77
Hirshabelle	8	$6.01	$27.31	8	$6.01	$27.31

Our analysis suggests that gains in income are driven by increased among girls who already had some income at ML1, rather than by broadening the set of girls who earn income. For instance, the share of girls who had non-zero earnings in the last month did not change between ML1 and ML2, but average earnings increased. This is not to say that girls did not enter the labour force or *begin* earning income between ML1 and ML2; indeed, there are many girls who went from no income to earning income and vice versa. In the aggregate, however, 45.2 percent of girls in the panel sample earn non-zero income in ML1, and the same share earn income at ML2 – which girls earn income has changed, but the overall share has not.

Instead, changes in mean income are driven by a change in average income among those who earn at least *some* income. At ML1, mean income among this group was $40.50; the typical (median) girl earned $19.23 per month. At ML2, mean income among girls with a non-zero income had increased to $170.99, and – reflecting the fact that there are some significant outliers in the ML2 income distribution – the typical (median) girl earned $38.46 per month. Even considering only median income, which is conservative in this context, earnings among this set of girls has doubled.

Importantly, the share of girls employed has increased substantially since ML1. Given NFE programming's focus on preparing girls for the labour market, this outcome is encouraging, especially given that girls' NFE programmes would have ended, at most, a few months prior to the ML2 round of data collection. This is not

[121] The positive changes in income from ML1 to ML2 also hold if we move in the opposite direction and include *all* C4 NFE girls, even if they do not have income data available in both rounds. This sample consists of 464 girls at ML1 and 389 girls at ML2, with mean incomes of $17.21 at ML1 and $73.05 at ML2, a difference that is statistically significant at the 1 percent level.

a long job search period, and we would expect higher employment rates when the endline is completed in 2024, once girls have had more time to engage in job search.

The table below reports the share of C4 NFE girls in different occupation categories. The first row makes clear that the share of girls who are unemployed has fallen – by 19 percentage points – since ML1. Note that the set of girls described in this table is more expansive than that considered in our analysis of income, due to the number of respondents who reported they did not know their income.[122] As is true of the C1 NFE girls described in Section 7.7, the biggest gains in employment have come in the form of domestic work, but girls are increasingly engaged in service-sector work, such as retail sales, and even fishing, farming, or pastoralist labor. Outside domestic work, service-sector employment is most common, accounting for a combined 18.6 percent of C4 NFE girls at ML2. Given the urban nature of the sample, the service sectors – retail shops and stalls, restaurants, etc. – are the primary employer, rather than agro-pastoralist labour.

TABLE 88: EMPLOYMENT, BY JOB TYPE, AMONG C4 NFE GIRLS IN THE ML1 AND ML2 ROUNDS

Job Category	Share of Girls. ML1	Share of Girls, ML2
No occupation or job	52.2	33.2
Domestic work	16.4	28.3
Student	15.5	10.1
Unskilled sales/service worker (e.g., hawker, shoe cleaner, domestic helper, cleaner)	9.2	14.6
Farming, fishing, pastoralism	2.7	4.5
Sales/Service worker (e.g., waiter, retailer or clerk at shop, etc.)	0.7	4
Other	1.4	2.7
Trades, craft workers, extractive industries	2	1.8
Professional or managerial positions	0	0.9

Perhaps even more notable than changing occupational status is the shift in business ownership among this cohort. At ML1, 8.1 percent of NFE girls reported having a small business of their own; at ML2, that share has increased to 18.4 percent, a change that is statistically significant at the 1 percent level. The most common form of business is focused on selling agricultural products (vegetables, milk, meat, etc.), with 42.7 percent of businesses falling into this category. The remainder tend to be engaged in retail trade of prepared or

[122] The table captures occupational information from a panel of 446 girls, a much wider set than the income analysis, which may explain a discrepancy between the two analyses: while the share of girls employed has gone up, the share of girls earning non-zero income has stayed the same. This is partially a result of the differences in the sample used. However, even when we limit our analysis to the panel of 263 girls for whom we have income information in both rounds, the share employed has risen by 13 percentage points, while the share earning a non-zero income has not. The two most likely explanations for this gap are that some girls are employed in domestic work that is non-remunerative, and that many girls report very small (under $5) amounts of monthly income that may be unrelated to direct employment. Both of these possibilities would weaken the link between trends in employment and trends in the share of girls reporting a non-zero income.

processed food, clothing, or – much less commonly – charcoal or firewood. One girl reported that she sells henna services as well.

Combined with the findings in Section 10.1 and 10.2, which report transition rates among the C4 NFE cohort, the results in this section paint a relatively optimistic picture of girls' labour market and entrepreneurship outcomes. It is important to note that business ownership and occupational category overlap, so readers should not interpret the gains in business ownership as additional to increased employment. Indeed, many girls report that they are engaged in sales or service-sector occupations *and* own a small business – in many cases, these are likely the same occupation (i.e. girls who own a small retail kiosk are "double-counted" as being employed and owning a business). Despite this, the fact that so many girls have started businesses within a few months of finishing their NFE programme is encouraging.

Cohort 5 NFE Girls

C5 NFE girls are just beginning their NFE programmes, but – given their age – some have already been engaged in the labour market or own small businesses. The table below reports, in the right-most column, the occupational status and categories of C5 NFE girls during ML2. In the middle column, we report the same outcome among C4 NFE girls, as reported during the ML1 round. At that stage, C4 NFE girls were at an equivalent point in their educational programme as C5 NFE girls are currently – in both cases, the girls were beginning their NFE programmes.

Two things are notable about the trends in the table below. First, C5 NFE girls are much more likely to report an occupation than C4 NFE girls were at the same point in their educational pathway.[123] C5 NFE girls are more likely to report being engaged in domestic work, and slightly more likely to be a service-sector worker. Second, C5 NFE girls are also more likely to have a small business than their peers in the C4 NFE cohort were at the outset of their programme. At ML1, 8.2 percent of C4 NFE girls reported having a small business, compared to 11.4 percent among C5 NFE girls currently. The sales of prepared and/or processed food make up a larger share of these businesses than is true for either C4 NFE girls or the C1 NFE cohort supported by FCDO. Overall, 42.5 percent of the small businesses focus on selling prepared food or beverages, while an additional 27.5 percent sell produce, meat, milk, or other agricultural products.

TABLE 89: EMPLOYMENT, BY JOB TYPE, AMONG C4 AND C5 NFE AT THEIR RESPECTIVE BASELINE ROUNDS

Job Category	C4 NFE Girls (at ML1)	C5 NFE Girls (at ML2)
No occupation or job	52.4	34.7
Domestic work	17.3	28.4
Student	15.0	15.6
Unskilled sales/service worker (e.g., hawker, shoe cleaner, domestic helper, cleaner)	8.8	9.1
Farming, fishing, pastoralism	2.4	3.7

[123] Throughout this section, when making comparisons between C5 and C4 NFE girls, we rely on ML1 data for C4 NFE girls (i.e. their baseline round). More importantly, we do not limit the C4 NFE sample to those girls who constitute a panel sample from ML1 to ML2, which is our standard elsewhere when analysing outcomes among this cohort. The reason is that we are interested in a like-for-like comparison between C4 and C5 NFE girls, and the latter cohort has not experienced panel attrition yet. Therefore, the best comparison group is the full C4 NFE, as it was recruited in ML1, prior to the impact of panel attrition.

Sales/Service worker (e.g., waiter, retailer or clerk at shop, etc.)	0.9	3.4
Other	1.3	2.0
Trades, craft workers, extractive industries	1.7	2.6
Professional or managerial positions	0.2	0.6

In line with this difference in occupational status, C5 NFE girls also have higher average monthly incomes than C4 NFE girls did in their own baseline assessment. Among C5 NFE girls, 37.0 percent report a non-zero income, though 11 percent report an income of less than $10 over a month, which is unlikely to be from part-time employment. Their mean income, including those reporting zero income, is $52.15, compared to $19.82 among C4 NFE girls at their baseline. This difference is robust to adjusting for outliers in terms of income by, for instance, capping maximum monthly income at $500.

Incomes among this cohort are not substantially different between Banadir and South West State – the latter has a mean income of $29.52, compared to $37.57 in the former. The sample of girls who self-reported income – as opposed to "don't know" responses – in Hirshabelle was too small to make between-location comparisons.

11.7 Enhanced social support for female youth

Cohort 4 NFE Girls

This section evaluates enhanced social support for ultra-marginalised C4 NFE girls. We will delineate the progression from ML1 to ML2 across the indicators participation in youth groups, local political forums, discussions to improve service provision. We will also evaluate whether girls received humanitarian assistance, while further segmenting the data to scrutinise the influence and degree of variables such as IDP status and household economic status. The data exhibited substantial co-variation between the first three outcomes, implying that many of the same young women were involved in more than one activity.

Participation in all tracked community activities more than doubled, as shown in the table below. All regions witnessed an increase of young women engaging in various development opportunities within their communities. The area related to local governance saw the highest surge in participation among the categories. The increase appeared to be largely driven by young women from Banadir. This group demonstrated by far the highest increase in the rate of participation in every social support activity. For ML1, their participation averaged lower than that of young women's participation in other regions, observing engagement levels between 4 and 8 percent across the categories. However, in ML2, the Banadir participation levels considerably increased to 30 percent for young women indicating participation in discussions concerning local governance; 37 percent for engagement in youth groups, and 42 percent concerning discussions about improving service provision. The only other region with noteworthy increases across all activities is the Bay region, which saw its most substantial increase in participation in the category of local governance, rising from 3 percent to 18 percent.

TABLE 90: PARTICIPATION IN COMMUNITY ACTIVITIES AND DISCUSSIONS, BY ROUND

Community Activity	Participating in ML1	Participating in ML2
Youth groups or networks	14.7%	32.2%
Service delivery improvement discussions	14.5%	37.4%
Local governance discussions	7.9%	26.6%
Received humanitarian assistance	20.0%	28.1%
Any of the above	31.8%	59.2%

The doubling of participation in community activities closely mirrored the participation rates of C4 NFE young women identifying as IDP. The data indicates that participation in discussion regarding service delivery exhibited the most significant increase across all activities, rising from 14 percent at ML1 to 38 percent at ML2. Involvement in local government discussions came a close second, with levels of participation increasing from 8 percent to 29 percent.

When examining data concerning C4 NFE girls with disabilities, we also observe the general increase in participation reflected here. As with IDP young women, service delivery saw the highest increase in activities, with 16 percent participation at ML1 compared to 42 percent at ML2. The increase for the other two categories is almost identical. Youth group participation increased from 14 percent at ML1 to 35 percent at ML2, and involvement in local government discussions from 13 percent at ML1 to 33 percent at ML2.

While at ML1, more girls reporting a mental health disability participated in community activities than girls with physical abilities, figures among both groups somewhat levelled at ML2.

All tracked regions, with the exception of Lower Shabelle, saw an increase in humanitarian assistance received by the C4 NFE girls. However, Banadir saw the largest increase, exactly doubling assistance from 14 percent in ML1 to 28 at ML2. Assistance in Lower Shabelle regressed, with 34 percent indicating having received assistance at ML1, compared to 21 at ML2. Humanitarian assistance provided to IDP C4 NFE girls rose from 22 percent at ML1 to 32 percent at ML2, as did the number of girls receiving humanitarian assistance reporting to have a disability. That figure increased from 19 at ML1 to 28 at ML2. Humanitarian assistance consistently increased among girls with disabilities from ML1 to ML2, with girls with physical disabilities receiving the highest amount at ML2. This figure rose considerably from 13 percent of girls at ML1 to 37 percent at ML1. This represents a shift from girls registering a mental illness having received the highest amounts of humanitarian aid at ML 1 (17 percent vs. 29 percent at ML2).

In drilling down to control for economic status and quality of schooling by applying filters controlling for the roof quality of girls' houses, food availability or the status of their school building and teacher reliability—although schooling information is only collected from C4 NFE girls who attend formal schools and unlike the majority of that group who attend learning centres, thereby rendering the data set smaller, there seems to be little impact on the girls' main interest in service delivery. The data paints an even picture with most girls across all the aforementioned categories maintaining their interest in service delivery with little variation in their in other activities.

Cohort 5 NFE Girls

The newly enlisted C5 NFE girls were the focus of this initial assessment, establishing a baseline of their engagement with several developmental opportunities. Their progress will be measured against these finding of this baseline study during next year's evaluation. As with the c4 NFE girls, enhanced social support for these marginalised girls was identified across a number of indicators, namely involvement in youth networks, political forums, and discussions that foster better service delivery. The evaluation will also cover whether these girls were receiving humanitarian assistance at this primary stage, while establishing how IDP status and domestic financial circumstances impact these interactions.

As with the C4 NFE girls, we observed substantial covariation, signifying a large portion of these girls took part in more than one activity.

Regional variances were evident in the participation levels in various activities. The Lower Shabelle region took the lead in youth group activities participation, with 34 percent of girls taking part, closely pursued by Banadir where 30 percent of local girls partook in youth-centric activities. The Bay region, on the other hand, experienced the lowest participation (12 percent). Discussions geared towards enhancing service delivery noted the highest engagement level in Banadir (42 percent), whilst participation in other regions hovered around one-third of the girls. In the Bay area, girls appeared least engaged or lacked access to local governance dialogues, with only 14 percent participating, as compared to 28 percent in Bandir, 25 percent in Middle Shabelle, and 24 percent in Lower Shabelle. Little variation was observed when delving into the IDP status or when accounting for girls with disabilities. Among girls identified as IDPs, a significant number seemed inclined or capable of partaking in service delivery discussions, with 39 percent participating. Yet, engagement in local governance dialogues and participation in youth groups followed closely behind, with 27 percent and 28 percent respectively. The category of service delivery also scored the highest among girls with disabilities, with 35percent reporting engagement. Local government discussions (26 percent) and youth groups (25 percent) followed suit. Girls indicating physical disabilities reported less activity in youth groups (10 percent) compared to those reporting mental disabilities (27 percent).

When examining economic factors, the category of service delivery saw the broadest engagement. Girls living in homes with poor roofing participated the most in this category (45 percent), followed by girls who reported going hungry for extended periods (30 percent) and girls who reported going without food on most days (25 percent).

The distribution of humanitarian assistance was fairly uniform across regions, averaging around 30 percent for all girls, with 146 girls indicating they received some form of humanitarian aid in the past year. Among the C5 NFE cohort of IDP girls, 31percent reported having received humanitarian aid, a similar figure to girls with a disability, 30percent of whom reported receiving humanitarian assistance. This aid was distributed fairly evenly between girls with mental disabilities (30 percent) and those with physical disabilities (29 percent).

12. Ancillary Analysis

12.1 Menstrual health and hygiene

One goal of the AGES programme is to promote better sexual and reproductive health and hygiene among girls. While this outcome falls outside of the standard outcomes in the broader Girls' Education Challenge programme discussed above, menstruation can have a large impact on girls' attendance and thus learning

outcomes. Menstruation is a common challenge for girls' attendance; girls often stay at home during their period due to shame, lack of menstrual hygiene products, lack of hygienic places at school to manage their periods, for religious reasons, or due to period-related pain.

To better understand these dynamics, the evaluation included a survey module on menstrual health and hygiene administered to girls who had begun menstruating. Girls interviewed by a female enumerator were asked whether they had begun to menstruate; for those who had, they were asked further questions about menstrual health practices, such as use and reuse of pads or rags; and the impact of menstruation on school attendance. The girls were divided into two cohorts: FCDO-supported girls and USAID-supported girls. The following analysis is subdivided into these two cohorts.

Original Baseline Cohorts

This section reports findings related to menstrual health and hygiene among the original baseline cohorts, consisting of FE, ABE, and Cohort 1 NFE girls. Within the FCDO-supported cohort, 355 girls at ML2 stated that they were menstruating and were asked subsequent questions. When carrying out comparisons between BL and ML2, we will use a smaller panel sample that covers the same girl respondents, for a total of 315 girls at ML2 and 233 at BL. These girls were further divided into the three main cohorts targeted by the programme: Accelerated Basic Education girls (ABE girls), Cohort 1 Non-Formal Education girls (C1 NFE), and Formal School girls (FE girls).

The percentage of girls reporting menstruation as a barrier to attendance is 34% at ML2, with a higher rate among C1 NFE girls (37.2 percent) and a lower rate among FE girls (25 percent). Interestingly, this represents a significant increase with respect to the same values reported as BL. At the start of the programme, only 9.4 percent of all girls across cohorts believed that they could not attend school while menstruating. When disaggregating the cohorts, roughly one in ten ABE girls and C1 NFE girls (respectively 11.8 percent and 8.2 percent) felt they could not attend school while on their period. Although a higher percent of FE girls – 16.7 percent – felt they could not attend school while on their period at BL, the number of total girls in this cohort is too small to be significant.

The increase in girls reporting menstruation as a barrier to attend school from BL to ML2 is confirmed in the analysis at the individual cohort level. The most significant increase occurred within the C1 NFE cohort where the percentage of girls that felt they could not attend school while on their period increased of 29 percentage points.

TABLE 91: SHARE OF GIRLS REPORTING MENSTRUATION IS A BARRIER TO ATTENDANCE

Girl Type	Baseline	Midline 2
ABE Girls	11.8% (8)	34.1% (43)
C1 NFE Girls	8.2% (13)	37.2% (51)
FE Girls	16.7% (1)	25% (13)
All Cohorts	9.4% (22)	34% (107)

When asked why menstruation prevents their attendance at school at ML2, over 75 percent of girls identified "pain and discomfort" as the main reason. This is an increase of 35.7 percent since baseline. A similar rate of girls report having experienced menstruation-related illness: overall, at ML2, 75.2 percent of

all FE, ABE, and C1 NFE girls report having experienced such an illness. The share of girls reporting menstruation-related illnesses among these baseline cohorts has not changed appreciably over time, increasing slightly from 74.9 percent at baseline to 75.2 percent at present.[124] As many girls were out of school or had irregular attendance prior to their participation in the AGES programme, the worsening of attitudes towards school attendance while menstruating might be driven by an increase in opportunities to be at school during menstruation and a more frequent experience of its challenges.

On the other hand, lack of sanitary pads was believed to be the main reason for lack of attendance by almost half of the sample – 10 girls across all cohorts – at BL, while this proportion decreased drastically over the next three years, with only 1 girl across all cohorts reporting this at ML2. This finding would suggest that access to menstrual products has likely improved from BL to ML2. This is reinforced by the fact that at ML2 only 11.1 percent of girls across all cohorts reported reusing rags or reusable sanitary products before they are fully dry after being washed, a decrease in 16 percentage points compared to the same figure collected at BL. At the same time, this lower rate of reuse of rags and sanitary pads did not lead to a reduced frequency of changing pads. At ML2 one girl out of four (23.2 percent) across all cohorts reported changing pads less than three times per day, while the frequency of changing pads reported at BL is lower, with almost one girl out of three (33.9 percent) changing pads less than three times per day.

We can observe a difference at ML2 in practices regarding frequency of changing pads between cohorts, with a higher proportion of FE girls changing pads less than three times per day (32.7 percent) compared to 22.2 percent of ABE girls and 20.4 percent of C1 NFE girls. While this is an increase from BL for both ABE girls and C1 NFE girls, only the proportion of FE girls reporting changing their pads infrequently remains consistent from BL to ML2. However, as stated previously, the small sample size for menstruating FE girls at BL reduces the significance of this result.

TABLE 92: GIRLS' ATTITUDES TOWARDS PAD USE

Pads Use	Baseline	Midline 2
Reuses Single-use Pads	27.5%	11.1%
Changes Pad Infrequently (<3 times per day)	33.9%	23.2%

Cohort 4 and Cohort 5 NFE Girls

Within the USAID-supported cohorts (Cohort 4 and Cohort 5 NFE girls), 551 girls at ML2 stated that they were menstruating and were asked subsequent questions. All girls targeted by the USAID-supported programme are part of non-formal education and the following analysis will be divided into the two main cohorts targeted by the programme: Cohort 4 Non-Formal Education girls (C4 NFE), and the new cohort (Cohort 5) Non-Formal Education (C5 NFE). However, as C5 NFE is a "new" cohort that were just enrolled in non-formal education prior to this survey round, so that the data collected serves as baseline data, they will not be included in comparison between ML1 and ML2, which will only cover C4 NFE.

[124] Note, however, that the sample used to assess this shift is not a strict panel sample as used elsewhere in this report. Two factors drive changes in the composition of the sample over time when asking about menstruation: first, girls are only asked these questions if they are being interviewed by a female enumerator, and this changes the set of girls presented with the questions from round to round; second, girls who have not yet started menstruating are not asked follow-up questions, meaning that the set of girls responding changes (increases) with each round.

For comparisons of C4 NFE figures between ML1 and ML2, we will use a smaller panel sample that covers the same girl respondents, for a total of 310 girls at ML2 and 287 at ML1. The sample size referring to C5 NFE specifically is 241.

The prevailing trend among USAID-supported girls regarding beliefs that menstruation represents a barrier preventing girls from attending education programme is roughly in line with the findings of the FCDO-supported cohorts. At ML2, 28.9 percent of girls across cohorts felt they could not attend school while on their period. Considering only C4 NFE for which we have both ML1 and ML2 data, we can observe an increase from 7.3 percent a ML1 to 29.4 percent at ML2.[125] Since the data collected for C5 NFE girls at ML2 could be interpreted as a baseline for that cohort – comparable to C4 NFE at ML1 – the similarity between the two rates at ML2 seems to indicate environmental or other contextual factors influencing this trend, rather than an evolution due to seniority of programme enrolment.

TABLE 93: SHARE OF GIRLS REPORTING MENSTRUATION IS A BARRIER TO ATTENDANCE

Girl Type	Midline 1	Midline 2
C4 NFE (selected sample)	7.3% (21)	29.4% (91)
C5 NFE	-	28.2% (68)
All Cohorts (full sample)	-	28.9% (159)

Note from the project: *The data for this study was collected at the end of the most severe drought recorded in country in 40 years. It is possible that the limited access to water had a negative impact on menstrual hygiene management at both household and school, consequently increasing infections of the urinary tract/reproductive organs, particularly among girls who have undergone type III FGM. This may account for the sharp increase in the % of girls reporting menstruation as a barrier to attendance.*

Among girls that reported that menstruation prevents their attendance at school, pain and discomfort is cited as the main reason by most girls in both cohorts (73.6 percent overall, distinguished between 74.7 percent for C4 NFE and 72 percent for C5 NFE). For C4 NFE girls specifically, there has been a decrease in proportion of girls reporting lack of sanitary pads as the main reason, from 23.8 percent at ML1 to 4.4 percent at ML2, confirming the increase in access to sanitary products hypothesised in the previous FCDO section. These proportions however are calculated based on very small sample sizes and should be only taken as indicative rather than as statistically significant.

Girls' practices relative to pads use is also similar between the USAID-supported participants and the FCDO-supported participants. At ML2, an overall 14.5 percent of girls across cohorts reported reusing rags or reusable sanitary pads before they are fully dry after being washed. While this proportion is slightly higher among C4 NFE girls (17.8 percent) than among C5 NFE girls (11.6 percent), there has been a significant improvement in practices by C4 NFE girls between ML1 and ML2, as evidenced by a decrease of more than 20 percentage points over the course of one year.

Likewise, at ML2 roughly one in four girls (24.9 percent) reports changing pads less than three times per day, with no significant difference between cohorts. When looking at the change over the course of one year in

[125] Notably, this increase is not correlated with a similarly large increase in menstruation-related illness. At ML1, 69.7 percent of C4 NFE girls reported having experienced such an illness; while this has increased to 75.8 percent at ML2 (73.0 percent among C5 NFE girls), this is a much smaller increase than the share who feel they would be unable to attend school while menstruating.

C4 NFE girls' practices, for which we have both ML1 and ML2 data, there was a small reduction (-4.4 percentage points) in the proportion of girls that change pads infrequently.

TABLE 94: GIRLS' ATTITUDES TOWARDS PADS USE

Pads Use	C5 NFE	C4 NFE	
	Midline 2	Midline 1	Midline 2
Reuses Single-use Pads	11.6%	41.5%	17.8%
Changes Pad Infrequently (<3 times per day)	24.5%	29.6%	25.2%

13. Recommendations

Learning Loss and Continuing Education

As Section 4.2 showed, learning scores decline – sometimes sharply – as girls leave school or after they have completed ABE or NFE programmes. Regression in literacy and numeracy reduces the long-term value of the programme's impacts and may reduce the extent to which beneficiaries believe education is worthwhile. A light-touch continuing education intervention for programme graduates, with a focus on application of, and a refresher on, the literacy and numeracy skills they gained, could reduce this documented learning loss. Continuing education could be delivered virtually or remotely; alternatively, in-person sessions, paired with continued engagement with GEFs (mentioned below) could be an effective way to both combat learning loss and maintain networks among the girls and women.

Maintenance of GEFs and Female Social Networks

The evaluation also documented a reduction in the connection between girls and the GEFs in which they participated over time. To some degree, reduced engagement with one's GEF is likely inevitable as girls shift into non-education activities, move away from the area, or otherwise change their focus. The programme should make an effort to promote continued connection between girls who participated in the GEFs together – into adulthood, if possible. These networks could continue to provide a valuable space for women to discuss problems they face, a support system, and a source of continued empowerment. Facilitating occasional meetings of a cohort of GEF participants after they have finished their programmes or left school would be a relatively low-cost method to build women's social networks and – potentially – create a group of adult women with ties to a particular school and an interest in seeing it succeed, similar to alumni networks elsewhere.

Girls' Representation and Participation in CECs

According to the analysis in Section 6, a small but meaningful proportion of GEF members participate engage with their CECs in some form. It is clear there is extensive overlap between the activities of GEFs and CECs, especially around awareness-raising, promoting enrolment, and reaching out to students who have dropped out or stopped attending school consistently. The programme should promote more active engagement of GEF members (and girls more generally) with their CEC. This could take the form of each CEC having a student member or representative that attends meetings, has input into some of the decisions made, etc. This model of shared governance would provide girls a voice in the CEC; it would also provide valuable insight into how the CEC operates, the issues and constraints they face, and how institutions in their community work. Governance models that incorporate students are an important method of socialization or professionalization and could also serve as networking and mentorship opportunities by successful/responsible adult community members.

Coordination and Collaboration between CECs and GEFs

A more systematic relationship between GEFs – or a select set of their members – and CECs could also be maintained post-matriculation. In line with the previous recommendation to help girls maintain ties to their GEFs, building a relationship between the GEFs and CECs oriented around joint awareness-raising and mobilization campaigns could build long-term connections between girls and the school they attended.

Economic Support for Households

The worsening economic situation in southern Somalia is reflected in measures of household well-being among the C4 and C5 NFE cohorts. While the programme already promotes economic self-sufficiency

through VSLAs and other activities, this may be insufficient to maintain enrolment in NFE programmes, especially when one considers the, arguably, higher opportunity cost – in terms of household or paid labour – incurred by the older girls targeted for NFE enrolment. If resource constraints allow, the programme should consider providing a small, potentially targeted, short-term cash transfer or other economic benefit that can ease the way for girls to continue in NFE programmes.

This recommendation could also support continued engagement with VSLAs. As shown in our discussion of VSLAs, just 7.9 per cent of caregivers reportedly participate in a VSLA. Providing short-term cash support through VSLAs could encourage their sustainability as the programme winds down.

School Fees and Increased Economic Support

The evaluation data suggest that there has been a small-to-moderate increase in the share of schools that charge enrolment or tuition fees for attendance. To the extent that schools have increased or begun charging fees, this may need to be reflected in additional bursary or other economic support.

Sustaining Institutions after the Programme's End

As the programme approaches its end, its sustainability is of primary concern. The programme should focus on ensuring increased adoption and effectiveness of interventions related to strengthening economic and financial conditions of programme households. Participation in VSLAs should be an area of focus, together with increased support to GEF activities, potentially also through financial and technical assistance. A specific assessment could shed further light on the underlying barriers to maintaining female networks, like the GEFs, to identify entry points for future programming.

CEC Shift Toward Monitoring of Instructional Quality

In part as a result of resource constraints, CECs are often limited in what they can accomplish. However, within the realm of tasks that require investment of time rather than money – such as enrolment mobilization, monitoring attendance, and monitoring teaching quality – CECs tend to be heavily focused on maintaining or increasing student enrolment and attendance (e.g., following up with students who are absent, encouraging enrolment, etc.). Fewer CECs monitor the quality of teaching, for instance, implicitly prioritising the quantity of a student's time spent in school, rather than the quality or learning value of that time. The programme should make a concerted effort to promote CEC monitoring of teacher attendance, teaching quality, and other aspects of school functioning critical to learning outcomes, such as the application of formative assessments and strategies for teaching to "the right level." Recognising that most CEC members are not education experts, providing them with checklists or simple rubrics for assessing a teacher's lesson planning, use of formative assessments, and so forth, could encourage them to monitor these activities more closely and build a culture of closer oversight over non-student aspects of the school.

Tailoring and Targeting NFE Instructional Levels

Analysis of literacy and numeracy for Cohort 4 NFE girls reveal that there was substantial regional variation in learning outcomes at the outset of their programme (in ML1), with girls in Bay achieving much higher pre-programme learning scores. This variation seems to have impacted the benefits of the programme, as girls in Bay show no tangible gains in learning one year later, while girls in other regions have caught up or surpassed them. This could be a reflection of teaching materials and a programme focus that was ill-targeted for those girls who were comparatively high-achieving at the outset of their learning programme. Developing NFE materials appropriate for girls at perhaps three different "starting points" could allow NFE centres that recruit a high- or low-starting point cohort to adjust their level up or down in response, and generate greater gains in learning.

Promoting Community Prioritisation of Girls' Education

Parental attitudes toward household work continue to be an important determinant of attendance rates. Girls whose caregivers see greater value of schooling relative to a child's other responsibilities or the family's other financial obligations attend learning programmes more consistently. The programme has made important inroads in increasing this type of community support for girls' education, and this progress can be maintained and deepened through relatively low-cost awareness campaigns as the programme winds down.

Sharing Teacher Experience, Expertise, and Materials

The report has documented progress in many aspects of teaching practices. This increased uptake and implementation of inclusive and participatory teaching styles could be maintained and spread through teacher knowledge-sharing platforms similar to those in use in many jurisdictions (e.g., platforms that facilitate peer-to-peer dissemination of lesson plans). Other knowledge-sharing methods could include forums for teacher discussion or training implemented by effective teachers, though these approaches would be more costly and are contingent on accurately identifying high-quality teachers to lead them.

Targeted Focus in Reducing Use of Corporal Punishment

The programme has made strides in reducing the use of corporal punishment in schools, as shown in this report and the ML1 evaluation previously. Based on arguably more-reliable student reports of corporal punishment use in classrooms, its use is relatively concentrated in a handful of schools. Just 17 of 46 C4 NFE centres have more than one student who reported the use of corporal punishment in the last week, but several of these centres had many students (as many as 50 percent or more of the girls interviewed) report it. The programme should consider targeting its efforts to reduce corporal punishment in particular schools where the problem is most pronounced, in order to maximise the programme's impact in a short time period. Students could be cost-effectively surveyed if the assessment was focused exclusively on identifying the use of corporal punishment, which would allow targeting beyond the sample of centres and schools included in the evaluation data here.

More Directly Assessing Sustainability at Endline

The programme's upcoming endline evaluation presents an opportunity to assess the sustainability of many of the institutional structures that were established as part of the programme, especially GEFs and VSLAs. This evaluation and other studies of Somali education have documented the valuable contributions of GEFs to a range of learning and socio-emotional outcomes, and sustaining these groups beyond the life of the programme should be a priority. During the endline, the evaluation team should develop more explicit measures of continued activity and engagement by GEF and VSLA participants; they should also engage with the programme team to develop measures of other aspects of sustainability that are not currently captured in data collection instruments designed at baseline and maintained through the following two rounds of data collection.

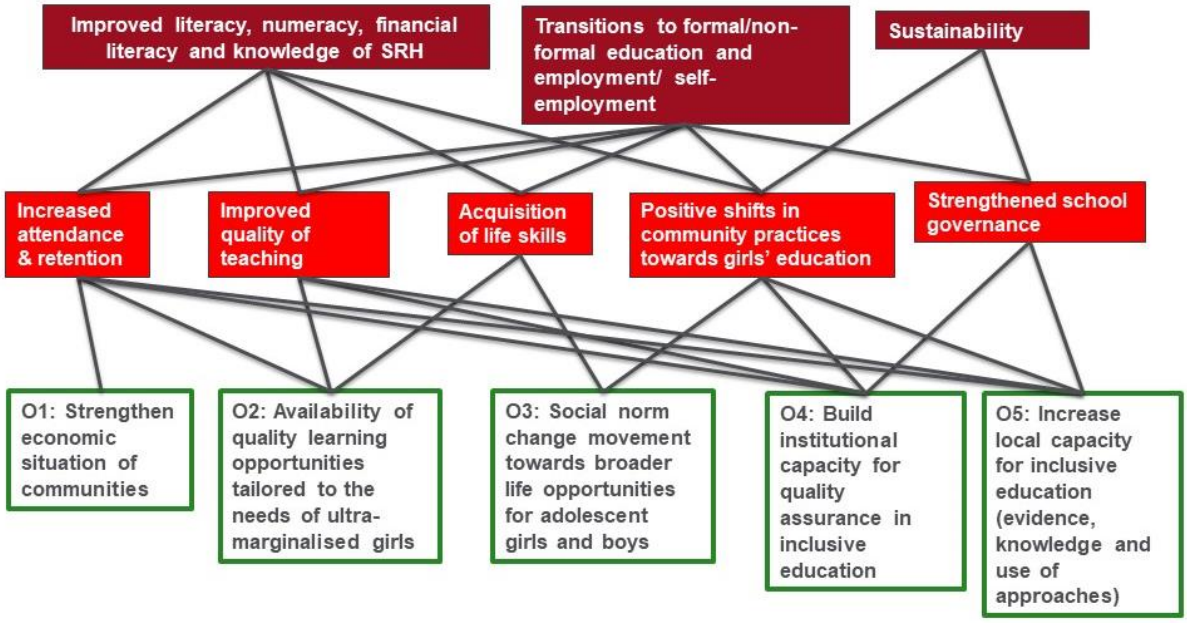
Annexes

Annex 1 - Project Design and Interventions

Since late 2019, AGES has been implemented in areas of south-central Somalia severely affected by the long-running civil war, by al Shabaab's insurgency, and by internecine conflict. The locations selected by the programme include large numbers of IDPs, limited infrastructure and provision of services, and significant security challenges to government actors. The project targets marginalized girls within these communities, where enrolment rates tend to be low and where entire age cohorts of children have missed out on any formal schooling due to conflict.

The project's Theory of Change is outlined in graphical form below. The intermediate outcomes – the middle row, in lighter red – include improved student attendance, teaching quality, acquisition of life skills (leadership, socio-emotional skills, etc.), community attitudes and support for girls' education, and school governance. The primary, overarching outcomes of the programme are expected to emerge from their intermediate outcomes. Improved learning outcomes (defined as improved literacy and numeracy skills; improved financial literacy; and basic knowledge of menstrual hygiene) are expected to arise from the synergy between increased attendance, improved teaching quality, development of girls' agency and positive community practices. Improvements in transition outcomes (continued enrolment in formal school, transitioning from NFE into gainful employment, etc.) are expected to arise from the combination of increased attendance; improved quality of teaching; positive community practices (additional investment in girls, shift in gender and social norms); acquisition of life skills and strengthened school governance (increased capacity for inclusive education and quality assurance). Finally, sustainable change is expected to emerge from shifts in social norms at community and girl level, associated to improved institutional capacity for quality assurance and increased local capacity for inclusive education.

FIGURE 23: AGES THEORY OF CHANGE



The project's approach to improving educational outcomes remains relevant at the time of this writing, largely because the barriers to education identified and targeted by the programme are common and are widely recognised as drivers of limited girls' education in Somalia. We refer readers to the baseline evaluation report, completed 2019-2020 for further discussion of the relevance of the programme's targeting strategy, especially the programme's focus on ultra-marginalised girls. As discussed there, the programme was well-targeted within the context, with a focus on GWDs, linguistic minorities, and adolescent girls who had "missed out" on education previously and, therefore, faced unique disadvantages. Rather than recapitulate the details and our analysis of the programme's targeting strategy, in this section we discuss the theoretical and empirical evidence for the programme's Theory of Change. In the table below, we break down the Theory of Change by intermediate outcome, evaluating the logic and evidence for assumed links between project outputs and the intermediate outcomes.

Output	Intermediate Outcome	Theory and Evidence for Linkage
Strengthening economic circumstances of households and communities	Increased attendance and retention	Strong theoretical linkages, especially given the prevalence of school fees in Somalia and the limited economic capacity of households. Empirically, a majority of girls who drop out of school or are not enrolled cite economic considerations (need to work; cannot afford school fees) as drivers of dropout. In addition, household chore burden and other labor demands of the household are strong predictors of poor attendance among girls, and poor attendance is a predictor of eventually dropping out.
Availability of quality learning opportunities tailored to ultra-marginalised girls	Increased attendance and retention	The strongest theoretical links are between tailored learning opportunities and retention, as girls who have fallen behind in school are more likely to drop out. Alternative education programmes are intended to give girls who missed years of schooling a way to re-enter education. There is good empirical evidence that girls who fall behind early tend to have lower retention rates, including from independent studies with other GEC programmes (e.g., retrospective study of SOMGEP).
	Acquisition of life skills	Alternative learning programmes contribute to life skill acquisition insofar as they provide a venue for girls' clubs, for socio-emotional learning, and for financial skills development. This is supported by the fact participation in girls clubs strongly predicts life skills development, and ABE/NFE programmes make it much more likely women will participate in such activities.

Social norm change toward broader life opportunities for adolescents	Positive shifts in community practices toward girls' education	There is a strong theoretical link between attitudes that support girls' education and downstream pro-education behaviours. The empirical evidence within the context of AGES is comparatively limited. Pro-education attitudes among parents/caregivers do not predict higher transition rates among the baseline cohorts. However, this is likely due to the strength of structural barriers to transition; households may be more supportive of girls' education but these shifts in attitudes may not be sufficient to overcome economic circumstances. Given the deep body of literature from other contexts linking pro-education attitudes to pro-education actions, however, there is good reason to believe this link operates similarly in Somalia.
Building of institutional capacity for quality assurance in inclusive education	Positive shifts in community practices toward girls' education	The theoretical link between official/institutional quality assurance and community support for girls' education is relatively weak. While community support may be higher if they believe the education provided is of high quality, parents and caregivers are poorly placed to assess quality, especially given their limited average educational attainment. Quality assurance provided via government may provide a signal of the importance of education, but this link is comparatively indirect.
	Strengthened school governance	The link between quality assurance and school governance is fairly straightforward, as quality assurance processes should promote consistency in school management and verification of the performance of the head teacher and other management staff. There is limited evidence of this impact as part of the evaluation, in part because most quality assurance is currently provided by CECs embedded in local communities (and which include the head teacher in most cases), rather than from an external educational ministry. As *external* quality assurance becomes more commonplace, school management should benefit as a result.
	Increased attendance and retention	The impact of quality assurance is likely to be limited, given that retention tends to be driven by more substantial causes, such as the economic needs of the household or a change in a girls' life circumstances (e.g., getting married). However, the theoretical link between quality assurance and attendance is clearer, given that a portion of quality assurance

		appears to be follow-up by local CECs with students whose attendance has declined. To the extent that unchecked declines in attendance result in girls dropping out of school, we would expect quality assurance to improve retention, but only indirectly.
	Improved quality of teaching	The link between quality assurance and teaching quality is fairly straightforward, as quality assurance processes should promote more consistent teacher attendance, use of a uniform curriculum, and improved teaching practices. While this evaluation does not provide causal evidence for or against the impact of quality assurance on teaching quality, it is clear that quality assurance by CECs is focused on many aspects of teacher behaviour, which suggests the linkage would be present if studied directly.
Increased local capacity for inclusive education (knowledge and use of inclusive approaches)	Positive shifts in community practices toward girls' education	The link between inclusive education and community attitudes/practices regarding girls' education is not entirely clear. If inclusive education includes the provision of alternative education programmes that are suitable for girls who have fallen behind in school or were never enrolled, this may increase support for girls' enrolment. Similarly, inclusion of the full spectrum of students – e.g., those with disabilities – may have the effect of altering views regarding *who* should be educated, in a positive direction. However, classroom-level changes in education practices are less likely to have a meaningful impact on community viewpoints or practices.
	Strengthened school governance	Improved capacity for inclusive education may not alter the overall quality of school governance, but it is logically linked to improvements in school governance that are specific to inclusivity. This might include CEC activities that promote enrolment of GWDs or other marginalised children.
	Increased attendance and retention	This evaluation does not provide empirical evidence for the link between improvements in inclusive education practices and attendance, though there are good theoretical reasons to expect inclusivity to result in higher attendance rates for many demographic groups. Sections 4.4 and 8.4 both provide evidence – though limited – that improved teaching practices (of certain types, such as highlighting the utility of particular topics) result in improved learning outcomes, which should have downstream impacts on retention.

	Improved quality of teaching	The most direct link between local capacity for inclusive education and any intermediate outcome is with improvements in the quality of teaching. Improvements in the knowledge and use of inclusive approaches should be reflected in classroom teaching practices, such as encouraging adaptation of lessons to the level of different learners from diverse backgrounds, an important form of improved teaching in and of itself.

Turning to the links between the intermediate and primary (learning and transition) outcomes, this report provides additional empirical evidence regarding the programme's Theory of Change. Readers should refer to Sections 4.4 and 8.4 for analysis of the linkage between intermediate outcomes and learning scores for the baseline (FE, ABE, C1 NFE) and C4 NFE cohorts, respectively. Likewise, regression analysis of the linkage between the intermediate outcomes and transition rates are provided in Sections 5.3 and 10.3. Below, we briefly summarise across these findings, but readers interested in understanding additional nuance of the findings can review the full results in the body of the report.

IOs and Learning Outcomes

Perhaps the clearest finding from our analysis of the AGES ToC concerns the link between life skills and GEF participation, on one hand, and learning outcomes, on the other. FE and ABE girls with stronger life skills score significantly higher on both literacy and numeracy assessments. While this finding does not hold for the C1 NFE cohort that was recruited at baseline, it appears again among the C4 NFE cohort that completed their programme immediately prior to this evaluation round. Participation in GEFs or girls clubs are also strongly predictive of learning scores – across four cohorts studied, GEF participation is associated with significant improvements in both literacy and numeracy in three of the four.

It is important to emphasise that neither life skills nor GEF participation are randomly assigned. In practice, more motivated girls may self-select into GEF participation, explaining part or all of the relationship between GEF participation and learning outcomes. Unfortunately, the evaluation design does not lend itself to causal analysis of the impact of the IOs on primary outcomes, though the consistency of the relationship between GEF participation – and its lingering impact even after a girl's ABE or NFE programme has ended – is strongly suggestive evidence.

The link between teaching quality and learning outcomes is much less promising, though this is partially a function of limitations in how we are able to study it. Certain aspects of teaching quality, as reported on by students themselves, are associated with higher literacy and numeracy scores, but the relationship varies markedly from measure to measure and some of the findings run counter to our theoretical expectations (for instance, among C4 NFE girls, negative teaching practices like absenteeism are correlated with higher learning scores). Contradictory findings may stem from the fact that we measure teaching quality – for this analysis – using student reports of teaching practices, rather than direct observation. This is necessary because the analysis requires linking learning scores to teaching practices at an individual level, which is not possible using classroom observations (which cannot be linked to individual students without classroom enrolment lists).

As with teaching quality, community attitudes have a complicated and contradictory relationship to learning outcomes. This is less surprising, because pro-education parental attitudes are likely to predict enrolment and attendance, but learning arises have so many disparate causes – from innate ability to early childhood cognitive

development to test conditions on the day – that the impact of pro-education parental attitudes are unlikely to show up effectively in a non-controlled study of this kind.

IOs and Transition Outcomes

Moving from learning to transition outcomes, our evaluation of the Theory of Change does not reveal a compelling link between pro-education parental attitudes and transition outcomes. It is important to note that our focus is on attitudes among a specific girl's parents, not broader community attitudes. However, despite our focus on this relatively narrow form of community attitudes – where we would expect the most direct impact on transition outcomes – there is no evidence of a causal relationship. As noted in Section 5.3 of the report, this may be the result of the complexity of transition outcomes, especially when we consider that successful transition for many of the cohort girls consists of finding employment – an outcome that is not entirely under their own control.

As with learning outcomes, teaching quality does not have a clear impact on transition rates, but GEF participation does. Among the baseline FCDO cohort girls re-contacted at ML2, those who participated in a GEF in earlier years are 10.6 points more likely to either remain in school or find employment. This result is slightly less stark among the C4 NFE girls, but this may be a function of the fact that their NFE programmes had ended only shortly prior to the ML2 data collection exercise, and positive transition outcomes can take time to materialise.

Annex 2 - Evaluation Methodology

This annex outlines the AGES Midline #2 (ML2) research design. Within this section, we will address the evaluation questions pursued in the assessment, provide details on the tool design, provide a comprehensive account of the fieldwork procedures, elaborate upon the analytical approach utilised, and note the methodological limitations and challenges encountered. More broadly, this section explains how the midline assessment fits within the overarching, multi-year longitudinal evaluation design, which encompasses several rounds of assessment. As was true in the ML1 evaluation round, this round included a baseline evaluation for a new cohort of NFE girls (Cohort 5 or C5 NFE girls) who were enrolled in the programme this year.

The primary objective of the AGES midline assessment is to estimate the changes in key project indicators since each cohort of girls was initiated into the project interventions. Throughout most of the report, our focus is on comparing the present levels of each indicator to the values established during the BL or ML1 rounds. Our choice of temporal comparison – whether to the BL or ML1 – is driven, first, by whether the cohort's initial participation was at the baseline (ABE, FE, and C1 NFE girls) or at ML1 (C4 NFE girls). In the case of girls originally recruited at the baseline, we make comparisons from the current round backward to both the BL and ML1 rounds, but focus our attention primarily on changes since baseline. Beyond straightforward pre-post comparisons, we also disaggregate the findings according to important characteristics of the girls, their households, their communities, and the schools or centres in which they participated in the programme. The goal is to understand differential impacts of the programme across different types or subsets of girls, facilitating project recommendations that aim to improve the inclusion of marginalised groups in education.

Evaluation Design

The AGES assessment employs a longitudinal, pre-post evaluation framework to understand the over-time impact of the project. This evaluation approach incorporates a blend of qualitative and quantitative methods, encompassing various tools for data collection. Qualitative data was obtained through focus group discussions involving multiple respondent groups, interactive exercises specifically designed for girls, and key informant interviews conducted with religious leaders. Quantitative data, on the other hand, was gathered through surveys administered to head teachers, a comprehensive sample of girls and their households selected at random, as well as direct observations of classrooms and attendance headcounts conducted by the field teams. Additionally, learning assessments were conducted with the aforementioned girls.

There are no control or comparison groups in the study. As noted in CARE's MEL Framework at the project's outset, this decision was made in light of the practical and security concerns of using control groups in the volatile and conflict-affected areas of southern and south-central Somalia. Randomization of treatment or intervention distribution would risk unequal allocation of benefits to particular clans, which could contribute to conflict or inter-communal tensions. Importantly, non-random assignment does not circumvent this issue, as communities treated as a comparison group may interpret their exclusion as evidence of bias, increasing risks of conflict and the targeting of programme staff. As such, the evaluation employs a pre-post design, without a difference-in-differences approach; instead, the same cohort of girls and schools that are part of the intervention are being tracked over time, with over-time comparisons – and comparisons to benchmarks established at the baseline, in the case of learning outcomes – being used to understand programme impacts.

In conducting the ML2 evaluation, we tracked the same group of girls who were randomly selected and assessed during earlier rounds (we discuss tracking, re-contact, and the construction of panel samples in more detail below). We administered the same learning assessment and a series of survey questions to these girls.

For girls under 18 years old, we also conducted interviews with their caregivers and heads of household as part of the survey. These girls will continue to be interviewed until the endline assessment to measure changes in learning outcomes over time. Additionally, we collected data to evaluate changes in community attitudes, girls' self-esteem and leadership skills, as well as economic and demographic indicators.

It is important to acknowledge the limitations of a benchmarked pre-post design like that employed in this evaluation. One significant limitation is the inability to account for changes in learning outcomes that would have occurred under the counterfactual condition – the absence of project interventions. Benchmark comparisons do not replicate a true difference-in-differences design because the comparison group is constructed solely from girls who took the learning assessment at baseline. Therefore, the benchmark does not consider broader societal or community-level changes that may impact learning outcomes, such as shifts in enrolment patterns, external shocks (e.g., conflict, drought, or flooding) affecting attendance rates, and various other factors.

To address potential maturation effects, we compare gains in learning from baseline to ML2 with benchmarked differences in learning scores between grade levels at baseline. For instance, for girls aged 11 at baseline, the difference in their performance between ML2 and BL represents the project's naive impact on learning scores. The difference between 11- and 14-year-old girls (reflecting the 3-year gap between BL and ML2) at baseline serves as the benchmark, which is used to assess the naive change in scores. If the gains between rounds (as a girl transitions from 11 to 14 years old) exceed the differences observed between 11- and 14-year-old girls at baseline, the benchmarked impact of the project is considered positive.

Furthermore, changes in attendance and teaching quality are evaluated through classroom observation and headcount tools that involve direct observation of classrooms. Between the baseline and ML1 rounds, the security situation changed, such that schools in Dinsoor were not accessible during ML1 fieldwork. This remained true during the ML2 round. Aside from this change, no further schools were removed or replaced at ML2, and the sample of schools remained the same as at ML1. However, as there may have been changes in teachers over the past two years, the analysis of teaching quality and attendance is conducted cross-sectionally rather than longitudinally.

In addition to examining changes in project outcomes over time, a key objective of the midline evaluation is to assess the validity of the project's Theory of Change. This involves investigating the differential impact of various project interventions based on girls' exposure to each of them, as well as studying the relationship between changes in intermediate outcomes (attendance, teaching quality) and learning outcomes.

Quantitative Methodology

Quantitative data was gathered from a total of 37 formal schools, 32 ABE (Accelerated Basic Education) centres, 34 C1 NFE (Cohort 1 Non-Formal Education – from the BL round) centres, 46 C4 NFE centres (from the ML1 round), and 45 new C5 NFE centres. These numbers take into account the Dinsoor schools and centres that were removed from the sample for security reasons beginning in ML1 and a single school in Hodan (school code HOD05), which was mistakenly included in the BL and ML1 samples by the evaluation team, despite the fact that it was not participating in the programme. It was excluded from the ML2 round and will be excluded going forward as well.

A key distinction between the ML1 and ML2 round is the approach taken to re-contacting girls and maintaining the integrity of the panel sample. Broadly speaking, girls who fell out of the sample during the ML1 round were purposefully brought back into the sample – where they could be located – during ML2.

We discuss this point in further detail below, but note it here because it impacts the sample size targets and sample completion rates we now discuss.

At the start of fieldwork, the sample targets were as listed in the table below. The overall target was of 2,862 girls; this represented a slight increase from the proposal stage, as we increased the target of C5 NFE girls – for internal purposes of being conservative – from 510 to 519 girls. For the purposes of this discussion, the target was the latter. Note that the sample targets for the baseline FCDO cohorts – FE, ABE, and C1 NFE girls – include BL respondents who were not successfully re-contacted during the ML1 round. This explains and increasing sample size target across rounds: at BL, we interviewed 412 FE girls; at ML1, we successfully re-contacted 343 of them and replaced 65; at ML2, our target includes the 343 who were successfully re-contacted at ML1, the 65 selected as replacements, and the 65 who fell out of the sample. This is important to note, because it obscures the successful re-contact rate; indeed, these numbers should not be interpreted as a re-contact rate, because they include replacements and they include girls who were being brought back into the sample after falling out of the sample during ML1. To be clear, these numbers are indicative of the evaluation team's completion of surveys, not of successful re-contact.

The second column in the table indicates the achieved sample size across cohort groups. These are the number of interviews completed, during ML2, with each cohort; the panel samples available for analysis are typically smaller, for the reasons outlined above. The final row in the table is a useful summary of how closely the target for C5 NFE girls was achieved (as noted, the original target was 510 girls).

The bottom panel of the table breaks down the sample targets further, as a function of re-contact status at ML1. In short, we disaggregate the overall sample target among FE girls into the set of FE girls who were successfully interviewed at ML1, and the set who were last interviewed at BL. Again, this should not be read as a statement about re-contact rates and attrition, because the achieved sample numbers include replacements.

TABLE 95: SAMPLE TARGETS AND ACHIEVED SAMPLE, ACROSS COHORTS

Cohort or Girl Type	Total Targeted Girls	Achieved Sample	Completion Rate (Share of Target)
By Cohort			
FE Girls	473	429	90.7
ABE Girls	468	310	66.2
Cohort 1 (C1) NFE Girls	490	302	61.6
Cohort 4 (C4) NFE Girls	916	661	72.2
Cohort 5 (C5) NFE Girls	519	511	98.5
By Cohort and Round of Last Successful Contact			
FE Girls – Baseline	61	36	59.0
FE Girls – Midline	408	393	96.3
ABE Girls – Baseline	104	48	46.2
ABE Girls – Midline	364	262	72.0

C1 NFE Girls – Baseline	122	39	32.0
C1 NFE Girls – Midline	368	263	71.5

The table below expands on the discussion of sample targets by addressing the school-level tools, the achieved sample for each, and their respective targets. The top panel focuses on formal schools, while the bottom panel focuses on C5 NFE centres. Because the earlier cohorts of NFE girls (C1 and C4) and ABE girls participated in learning programmes that have now ended, school-level data collection did not take place in their centres. Of course, many C5 NFE centres also hosted earlier cohorts of NFE girls; however, we label them "C5 NFE centres" because the girls enrolled in them currently are exclusively part of the C5 NFE cohort.

As the table shows, we fell slightly short of our targets – in both types of learning centres – for classroom observations. This was a result of schools either being in the process of holding examinations, or visiting C5 NFE centres that were functional but which had limited students present at the time of our visit. The achieved sample of the head teacher survey – conducted only in formal schools – similarly fell short of the target. Note that attendance headcounts do not have a pre-specified target, because our sampling procedure indicates that field teams should complete headcounts in every functional classroom, from grades 1-5 for formal schools and any NFE classroom in C5 NFE centres. Without knowing, *a priori*, the number of classrooms in each centre/school, it is not possible to pre-specify a sample target.

TABLE 96: SAMPLE TARGETS AND ACHIEVED SAMPLE FOR SCHOOL-LEVEL TOOLS

Cohort or Girl Type	Sample Target	Achieved Sample	Completion Rate (Share of Target)
Formal Schools			
Classroom observation	74	72	97.3%
Headcount	N/A	175	N/A
School survey	37	35	94.6%
C5 NFE Centres			
Classroom observation	90	80	88.9%
Headcount	N/A	129	N/A
School survey	N/A	N/A	N/A

The ML2 assessment, as was true of the ML1 assessment, focused on tracking a panel of girls across rounds. We re-contacted the girls who had participated in either the BL or ML1 assessment, following a set of standardised re-contact procedures, including the use of detailed tracking materials, such as individual tracking sheets for each girl and cohort lists for each school. The recontact procedure consisted of several sequential steps that had to be be completed before the enumerator could mark a girl as "not found." These steps encompassed the following:

- Asking the head teacher and other teachers
- Calling all available phone numbers three times

- Visiting the girl's household twice
- Asking the community, including the girls her age

FE girls who could not be recontacted were substituted with girls from the same school, selected from a list of possible replacement girls. This replacement list was assembled at the BL and has not changed since that time – it is the list of girls who were enrolled *by the programme* immediately prior to the baseline assessment. The purpose of using this list, rather than selecting replacement FE girls directly from classrooms, is that it ensures we make like-for-like replacements; any girl that is replaced is replaced by a girl who was also enrolled in Grades 1 or 2 in late 2019. Within each formal school, a random sample of 20 replacement girls was drawn from this list, excluding girls who were already sampled in previous rounds. The same approach was applied to C5 NFE girls, for whom this evaluation comprises a baseline. Girls were randomly sampled from a list provided by CARE; replacement girls were pre-selected using the same process (random sampling within a centre), and replacements were interviewed only after all the steps outlined above were completed and the original girl still could not be located.

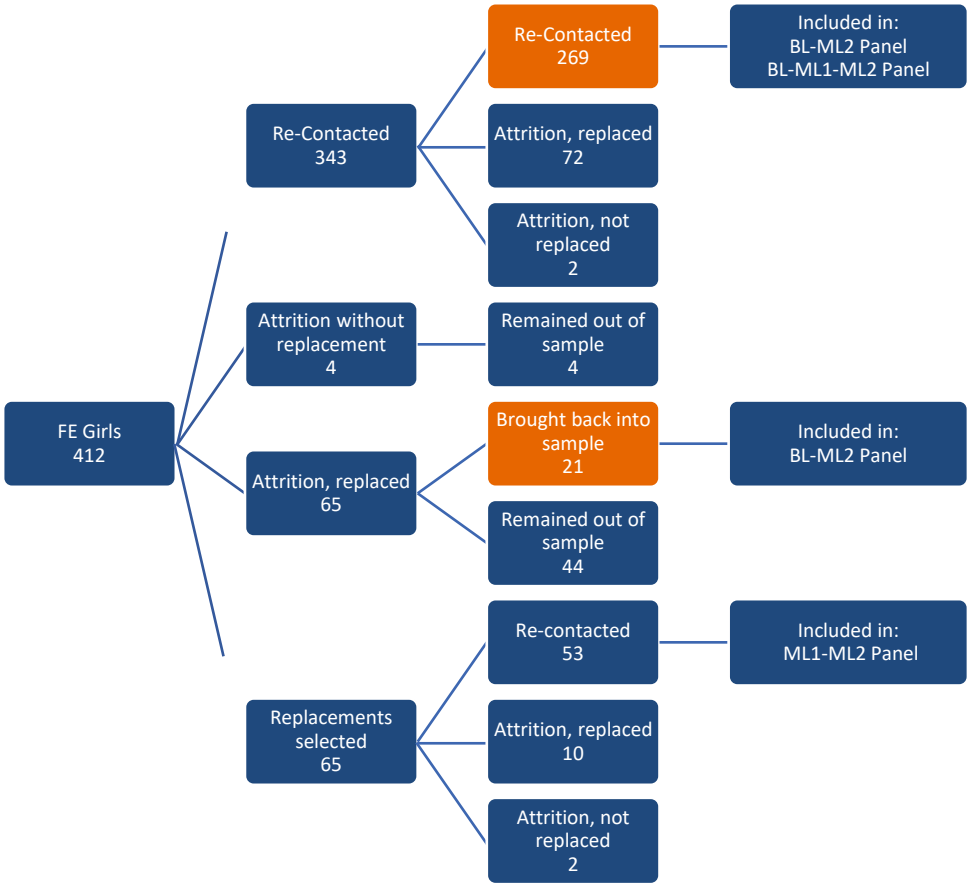
In many instances, girls were replaced due to the following reasons:

- Moved away (abroad or to another district)
- Could not be located in the school/school records and could not be reached even after several attempts
- New (C5) NFE girl had dropped out of the programme between the time the enrolment list was constructed and the time we reached her community.

As alluded to above, there are new complications related to sample construction compared to ML1. As is expected in panel surveys of this kind, a number of girls fell out of the sample during ML1. Our goal is to reduce year-on-year panel attrition through the implementation of rigorous fieldwork procedures, like those listed above. However, we also seek to reduce *aggregate* panel attrition by attempting to bring girls back into the panel wherever possible. For instance, a girl who fell out of the sample during ML1 can often be re-contacted at ML2. While it is obviously not possible to capture data on her from the ML1 round, she can still form a part of a long-run panel of girls from BL to ML2, or from BL to endline. While we prefer having a panel of girls who are included in every round of the evaluation, sample size considerations are always a factor in our analysis, and we generally prefer using a larger BL-to-ML2 panel that excludes data from ML1, rather than relying on a smaller set of girls for whom data is available across all three rounds. This same logic will apply at the endline as well, of course.

The figures below illustrate the re-contact outcomes across rounds, and how respondents move and and out of the panel sample. The most complex situation is illustrated in the first figure, which captures the process for FE girls. As the figure shows, the original sample of FE girls at baseline (first box, far left) included 412 girls. During ML1, these girls were either re-contacted (343), fell out of the sample and were replaced (65), or simply fell out of the sample and were not replaced due to school-specific constraints on available replacements (4). At ML2, we attempted to re-contact every girl who had been successfully interviewed in *either* of the previous two rounds. That is, FE girls who had been interviewed in both rounds; girls who had been interviewed in BL but could not be located in ML1; and girls who had acted as replacements in ML2. The re-contact outcomes for each of these groups are documented in the third column of the figure. A key point from the third column, is that 21 FE girls who had fallen out of the sample at ML1 were successfully brought back into the sample at ML2. These girls are included in our analysis of re-contact rates – they constitute a "success" when considering BL-to-ML2 re-contact rates. And, as the figure shows, they constitute part of the BL-to-ML2 panel sample that is the focus of much of our analysis in this report.

FIGURE 24: SAMPLE ATTRITION, RE-CONTACT AND REPLACEMENT FLOW CHART, FE GIRLS



The next two figures are much simpler than that of the FE girls. The original FCDO cohort of ABE and C1 NFE girls were also contacted, originally, at the baseline evaluation in 2019. In contrast to FE girls, however, no replacements were made at ML1 – if an ABE or C1 NFE girl could not be located, they were not replaced by a newly drawn girl. This fact simplifies things: while we still attempt to re-contact every girl who was originally recruited at the baseline, for these cohorts there is not a new set of (replacement) girls who were brought into the sample during the ML1 round. A different way of viewing this point is that, for the FE cohort, there are girls who appear in the ML1-to-ML2 panel, but not in the BL-to-ML2 panel. For the ABE and C1 NFE cohorts, if a girl appears in the ML1-to-ML2 panel, she is also part of the BL-to-ML2 panel.

FIGURE 25: SAMPLE ATTRITION, RE-CONTACT, AND RE-RECRUITMENT FLOW CHART, ABE GIRLS

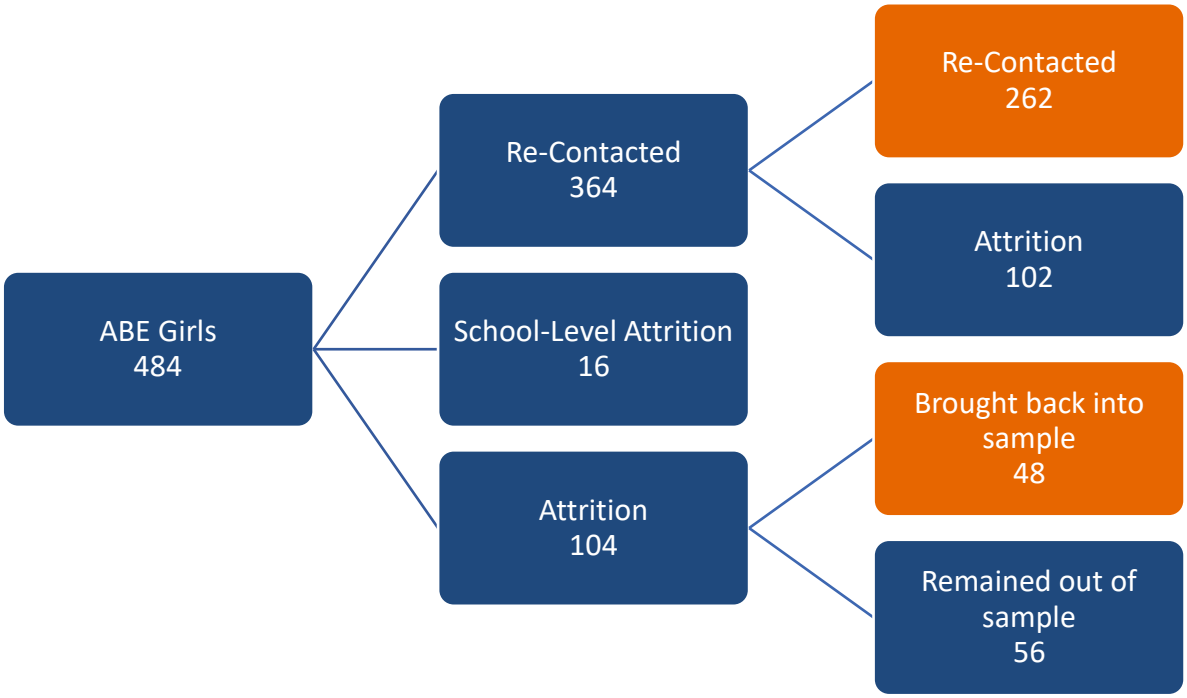
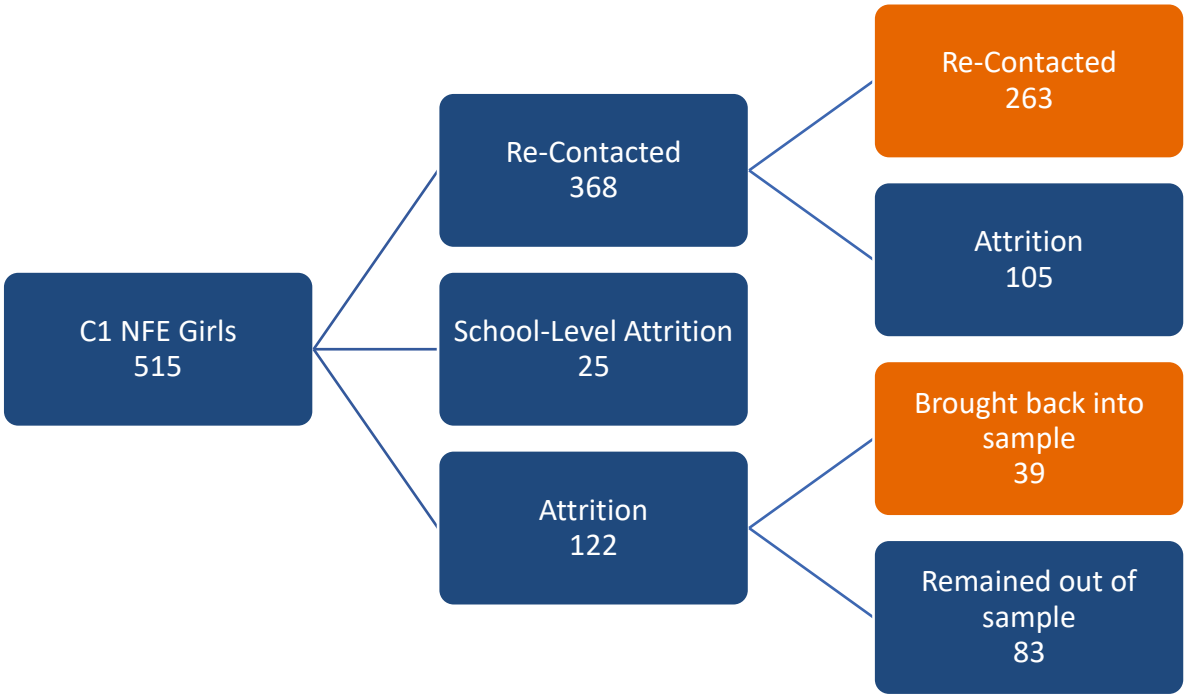


FIGURE 26: SAMPLE ATTRITION, RE-CONTACT, AND RE-RECRUITMENT FLOW CHART, C1 NFE GIRLS



A key point is the inferential value of actively trying to bring girls back into the sample in later rounds. These figures make clear that re-contacting girls who had fallen out of the sample during ML1 increases our available sample size. For instance, if we had not pursued this approach, our sample of ABE girls available for analysis over the BL-to-ML2 period (excluding ML1 data) would have been just 262 girls – those who were successfully re-contacted at both ML1 and, again, at ML2. Including girls (n = 48) who were brought back into the sample after ML1, our sample size for BL-to-ML2 analysis increases to 310. Of course, analyzing this set of girls comes with a drawback: we cannot incorporate ML1 data into the analysis, because this data is only available for 262 girls in the sample. There is a direct trade-off between including ML1 data in our analysis and maximizing the sample size for BL-to-ML2 analysis.

Broadly speaking, our preference is to maximize sample size for BL-to-ML2 analysis. Our interest is primarily in understanding the impact of the programme since baseline, not since ML1. This is especially true in the case of ABE and C1 NFE girls, whose learning programmes generally ended prior to the ML1 round of data collection. It makes more sense to focus on BL-to-ML2 changes, since ML1-to-ML2 changes miss most of the program's actual intervention period, at least for those cohorts. More broadly, our philosophy is that we care most about aggregate impacts, not impacts since the previous round. Therefore, we generally utilise the larger BL-ML2 panels, supplementing our analysis by looking at the "full panel" samples (BL-ML1-ML2 panel) in some cases, where we want to understand something specific about the trends between rounds, rather than the aggregate trend from BL to ML2.

Qualitative Methodology

CARE developed the qualitative interview guides used in this assessment, with inputs from the evaluation team in each round, including the present round. Qualitative interviews included focus group discussions (FGDs) with mothers, teachers, and Community Education Committee (CEC) members, key informant interviews (KIIs) with religious leaders, as well as participatory group discussions with girls, focused on risk mapping and vignette exercises.

The FGD guides for CEC members aimed to gather insights into their experiences with school and CEC management, as well as their attitudes towards girls' education. The FGD guides for teachers aimed to collect information on their attitudes and perceptions regarding gender differences in classrooms, teaching experiences, and interactions with CEC members, school management, and colleagues. The FGD guides for mothers included questions about decision-making processes in households, attitudes towards girls' education, security concerns, and overall community perceptions of girls' education and opportunities.

The risk mapping and vignette exercises were more interactive, with the aim of gaining a deeper understanding of girls' attitudes, perceptions, and experiences related to their educational opportunities, barriers to learning, school environment, and safety. During the risk mapping exercises, girls were asked to draw maps of their community and school grounds, marking the places where they felt safe or happy, as well as places where they felt unsafe or unhappy. They were also asked to explain their reasons for these feelings and whether they believed these places were less safe for girls compared to boys. The vignette exercise aimed to explore girls' perspectives on the value of education, barriers to learning and attending school, and potential solutions. Girls were presented with short stories featuring female characters facing different education-related challenges and were asked to complete the stories and share their opinions on what they believed would happen to the characters and how they could overcome their challenges.

The selection of locations for most qualitative interviews was done randomly, with consideration given to proportionately representing the assessed districts. However, for certain types of respondents, e.g., CEC members, interviews were specifically conducted at schools with verifiably active CECs, as they were expected to provide more comprehensive insights into the committee's work. While this approach sacrificed sample representativeness, this bias was taken into account during the analysis.

During participant selection for FGDs with girls (i.e. risk mapping and vignette exercises), team leaders were instructed to choose girls from the same cohort and from the same broad age group. This facilitated interaction among the girls during the exercises, created a more comfortable environment, and minimized significant age gaps between participants.

To identify teachers for FGD sessions, team leaders selected teachers who taught classes with girls participating in the AGES program, including teaching at the appropriate grade level. The desired number of participants was six, and a list of eligible teachers was provided by the school principal. In cases where an NFE centre had only two or three teachers, which is fewer than the typical number of FGD participants, the sessions included all teachers at the centre. While it would have been possible to reallocate these FGDs to NFE centres with a larger number of teachers, the evaluation opted to complete the FGDs in the originally selected centres, recognizing that smaller NFE centres may face different challenges – such as teacher shortages or larger class sizes – that would be important to capture.

For FGDs with mothers, team leaders received a list of contact details for the mothers of girls who had been interviewed for the household survey. Team leaders individually contacted each mother and invited them to participate in the FGD sessions.

In total, we completed 69 qualitative interviews, with the following breakdown by respondent type:

- 13 FGDs with CEC members
- 12 FGDs with teachers
- 12 FGDs with mothers
- 12 vignette FGDs with girls
- 12 risk mapping FGDs with girls
- 8 KIIs with religious leaders

One extra interview was completed with CEC members due to miscommunication among the evaluation team.

Fieldwork

In this section, we describe critical aspects of data collection and discuss how the data was analysed. With regard to data collection, this section includes details on the number and reasons for school-level replacements and removals (i.e. replacement or removal of entire sampling points or clusters), and how replacements were selected. In terms of data analysis, we describe our general approach to the qualitative and quantitative data, and how their analyses relate to one another. Further details on enumerator selection, training, quality assurance, and data cleaning are also provided.

Enumerator Selection

Nine data collection teams were deployed, each consisting of one team leader and between three and five enumerators. Enumerators and team leaders participated in a 6-day training, which included a 1-day pilot. A total of 54 people were invited to participate in the training, based on their experiences, gender, and language requirements for the fieldwork locations. Following the completion of the training, a quiz was administered to all participants to finalize the hiring decisions. In total, a group of 9 team leaders and 36 enumerators were mobilized to carry out the data collection process. Gender balance was considered when assembling the teams, particularly because of the need to conduct in-depth FGDs with girls. Two of the team leaders were female; as at ML1, this reflects the difficulty of recruiting sufficiently experienced female researchers also able to work in many of the areas of south-central Somalia where the evaluation was implemented.

Certain data collection activities, such as focus group discussions (FGDs) with mothers, risk mapping exercises, and vignette exercises (RMV) with girls, required the presence of female researchers. This was necessary to create a comfortable environment where female respondents would feel more at ease to share their thoughts and experiences. During the FGDs, topics like household decision-making dynamics, community perceptions of girls' education, risks faced by girls on their way to school or at school, and barriers related to sensitive subjects like girls' menstrual health could be discussed more openly with a female interviewer. Similarly, the participatory nature of the risk mapping and vignette tools and the sensitivity of certain questions, such as identifying locations where girls feel insecure and explaining the reasons, necessitated the involvement of female researchers. Because qualitative interviews are generally completed by our team leaders, this presented a problem, as just two team leaders were female. For this reason, each team included at least one experienced female enumerator, to ensure that each team had the capacity to administer gender-sensitive FGD tools and capture data – from the quantitative survey with girls – on menstruation and menstrual hygiene.

Each team participated in a comprehensive six-day training program, which included a dedicated day for conducting a pilot test in schools. Following the pilot test, a feedback session was held, involving the evaluation team and enumerators, to address any issues identified during the pilot, address any errors in the programmed surveys, and provide clarification. The training encompassed the following topics:

- Administration of the learning assessments & mock learning assessments
- Review of all the tools & their administration
- Child protection & safeguarding
- Selection of respondents, fieldwork management & assignment of the team leaders' responsibilities

The training provided each team with a comprehensive understanding of the study's objectives, the tools utilised, and the ODK application employed for data collection. Additionally, team leaders received training on fieldwork management, specifically regarding the completion and maintenance of tracking sheets. A significant portion of the training was dedicated to the administration of the learning assessment, allowing enumerators ample time for practice.

A specific session during the training was devoted to child protection and research ethics, led by a child protection specialist from the CARE Somalia team. This session focused on addressing issues unique to conducting a GEC evaluation, such as obtaining consent from children, ensuring their safety during survey administration, maintaining confidentiality when discussing sensitive topics related to teachers or households, minimizing any undue pressure on children during the learning assessment, and safeguarding the confidentiality of collected information. All enumerators and team leaders signed both CARE's child protection policy and Consilient's internal child protection and research ethics standards. To ensure data confidentiality, any data containing personally identifiable information of individual respondents, including children, community members, teachers, or head teachers, was exclusively accessible to a single designated technical evaluation focal point at CARE and was password-protected.

During the training, female team leaders and experienced female enumerators – those female team members who were selected to complete FGDs with female interviewees – participated in a separate training session conducted by experienced female researchers from Consilient, including the project's Fieldwork Manager. The focus of this training was on the menstrual hygiene section of the household survey and the implementation of the risk mapping and vignette exercises. When male team leaders were appointed, capable female team members were trained to conduct focus group discussions (FGDs) with mothers and administer the risk mapping/vignette exercises with girls, as noted above. Only the most experienced female researchers, who had previously conducted similar participatory exercises and qualitative interviews with girls and mothers for projects of comparable scope to AGES, were selected for the administration of these tools.

Pilot

Before fieldwork commented, a pilot exercise was conducted involving all team leaders and enumerators who would be involved in the data collection. Following the completion of training, a pilot took place in three primary schools located in Hargeisa, Somaliland. During this pilot phase, each enumerator was required to carry out a minimum of one learning assessment and one household survey. Simultaneously, team leaders were tasked with completing one classroom observation, one headcount, and one school survey involving the head teacher.

The evaluation team made a deliberate decision to have team leaders focus on conducting their own data collection during the pilot (i.e. completing attendance headcounts, classroom observations, and other tasks for which they would be responsible during fieldwork), rather than observing their team members performing household surveys and learning assessments. This decision was made to ensure that team leaders would be well-prepared for their own tasks during fieldwork. At the same time, experienced enumerators – often those who have worked on multiple past GEC evaluations with Consilient, but who lack the local networks to serve as a team leader in south-central Somalia – were selected to observe their teammates during the pilot,

providing feedback on their administration of the household survey and learning assessment. This approach ensured that enumerators received specific, tailored feedback during the pilot, without reducing the value of the pilot for the team leaders.

Changes to the School Sample

As noted above, during the ML1 round, two schools that had been included in the baseline sample were replaced due to insecurity. These schools were located in Dinsoor, and were replaced with schools in Afgoye and Baidoa, maintaining the representation of South West State in the sample. At ML2, the same sample was maintained, excluding schools in Dinsoor and keeping the schools that were selected as replacements during ML1.

In the ML1 round, several newly-selected C4 NFE centres were also replaced for accessibility reasons. We refer readers to the ML1 evaluation report (Annex 1) for discussion of the replacements made. As we discussed at the time, the *realised* sample of C4 NFE centres from ML1 was used as the target for ML2. In other words, the actual centres visited during the previous round – which constituted a baseline for the C4 NFE centres – constitute the centre-level sample going forward. No C4 NFE centres required replacement during ML2.

Quality Control

To ensure the accuracy and quality of the quantitative data, various measures were implemented in the survey tools. These included incorporating choice filters, age restrictions, numeric value constraints, and calculations for learning assessment scores. These checks aimed to minimize data-entry errors and ensure that only eligible respondents were interviewed.

During fieldwork, teams were provided with tracking tools to aid in data verification. Individual tracking sheets and sheets specific to each sample point contained identifiers and demographic information for the cohort girls, enabling us to cross-reference and validate the survey data. These tracking sheets will serve as supplementary materials for future evaluations, facilitating tracking purposes.

A dedicated quality control tracking tool was developed and used on a daily basis by the project's research officer and fieldwork manager. This tool tracked the number of submitted surveys and re-contact rates, disaggregated by cohort group, and any changes or information pertaining to the quantitative and qualitative data collection. Additionally, it recorded the time taken to complete the surveys. By incorporating timestamps throughout the survey, we were able to identify enumerators who appeared to be rushing through the survey and provide feedback to team leaders to ensure proper monitoring of survey delivery pace. Daily quality control checks were accompanied by daily data cleaning. Any inconsistencies or mistakes were discussed with the field teams and, if necessary, corrected in the data.

Newly implemented in this round was an effort to audio record interviews for further quality assurance. This approach built on Consilient's years of experience performing quality control on phone-based, remote (CATI) interviews. As part of our CATI approach, we generally record interviews and have a quality control team review the audio files, verifying that responses were filled correctly, identifying any issues in survey administration, and providing a wide range of feedback to enumerators. This round, we implemented this approach in the field, having enumerators audio record their interviews – where respondents provided explicit, supplementary consent to be recorded – and transmitted these recordings regularly, via secure connection, to our home office in Hargeisa. Audio quality control was conducted on a random subset of interviews; in addition, when issues arose with particular enumerators or interviews, we were able to review audio transcripts to correct the data, identify the source of data issues, and so forth.

To ensure the quality of the qualitative data, much of our focus was on the use of experienced qualitative interviewers, with years of experience working on qualitative projects with Consilient. Additionally, during the training, female team leaders and selected female enumerators received guidance from experienced staff on conducting risk mapping and vignette exercises with the girls. Rosters containing participant details, including age and gender, were provided to the teams for effective tracking during qualitative interviews.

During fieldwork, a dedicated team of full-time staff members stationed in Hargeisa was responsible for regularly reviewing the recorded audio files of the FGDs, risk mapping, and vignette exercises. Their role was to identify any instances of inappropriate tool administration from the outset. One team member, who had extensive experience in implementing these tools and a strong understanding of their proper administration, had previously participated in evaluations of CARE's educational projects. Systematic review of all received qualitative interviews was conducted following their reception.

Data Management and Cleaning

For the quantitative data, to ensure secure data management, the evaluation team used an online data management platform – ONA – and all teams were required to submit the surveys to the ONA servers once they were completed. The submitted data were downloaded on a daily basis for regular quality control and data cleaning.

Daily data cleaning focused on general inconsistencies and the duplicate unique ids/observations, age variables, the respondent types, school grade variables, phone numbers, spelling of string variables and learning assessment scores. While household survey and learning assessment data were reviewed daily, the review and cleaning of the data from other surveys were done bi-weekly.

On a weekly/bi-weekly basis, depending on the specific survey data, a more in-depth data cleaning was conducted by our team. All the variables were separately examined and cross-tabulated to identify any possible inconsistencies in the data. If logical inconsistencies were discovered, we contacted team leaders to double-check the answers in case they included typos or accidental mistakes.

As far as the qualitative interviews were concerned, team leaders were required to share audio recordings with our team controlling the quality of the data. Once reviewed, all the qualitative interviews were transcribed and translated by our full-time staff members and externally contracted staff, using specifically developed templates. The process of transcription and translation was supervised by our full-time staff member and the quality of the English translation was reviewed by international full-time staff members. Subsequently, the quality of the translations was reviewed and corrected.

Analysis

The quantitative analysis makes use of clustered standard errors wherever appropriate, but generally does not incorporate survey or sampling weights. Clustered standard errors were used wherever there are multiple observations or respondents in a given learning centre; in short, clustered errors are applied in all cases except when analyzing the head teacher survey. Clustering occurs at the level of learning centre, not sampling point. To recap a discussion from the baseline evaluation, we draw a distinction between FE schools, ABE centres, and NFE centres that may take place in the same schools or buildings. In a strictly physical proximity sense, these learning centres are in the same place. However, the statistical need for clustering is based on logical similarity, not only geographic proximity. There is little reason to believe that ABE girls and FE girls whose classes share the same physical building are necessarily related to one another in a statistical sense. Moreover, the girls selected into the FE, ABE, and NFE cohorts in the same general area were still selected *separately*; in

the context of sampling, they had no relationship to one another. For this reason, we cluster at the centre level – FE and ABE girls in the same "sampling point" are considered distinct clusters. This argument also applies to C4 and C5 NFE centres – they are considered distinct clusters from one another, even if girls from the two cohorts completed their NFE programmes in the same centre.

In contrast to the baseline evaluation, we do not employ survey weights during the analysis. At baseline, survey weights were employed to adjust for unequal cluster sizes within the same institutional category (formal schools, ABE centres, and NFE centres) across sampling points.[126] However, we consider this a secondary concern at this stage of the evaluation. First, cluster sizes have changed since baseline, due to attrition; as a result of differential attrition across clusters, cluster sizes are now quite different from one another. While this means that some clusters will have more influence on our findings (i.e. have greater weight in statistical tests, owing to their larger sample sizes), the large differences in weights that would be necessary to adjust for this would be more problematic, increasing the influence of particular girls or households significantly above those of others in clusters with a larger sample size. It is also the case that – in post-baseline rounds – our interest is less in descriptive findings regarding the nature of girls and education in a particular sampling point, and more in understanding changes over time. This focus means that we focus more on internal validity (valid conclusions drawn regarding the sample) than on external validity (the validity of those conclusions beyond the sample, to a broader population).

Regarding the qualitative data, it was translated, transcribed, and organized in a master spreadsheet for systematic analysis. The aim was to identify insights that supported, contradicted, complemented, or complicated the findings in the quantitative data.

Fieldwork Challenges

This subsection details the challenges encountered during fieldwork, as well as the strategies employed to mitigate them whenever possible. Limitations related to the research design are detailed in the methodology section of the main report body.

Recontact Rates and Replacement Challenges

Despite the team's best efforts, many of the cohort girls could not be located or interviewed. The field teams identified changed or no longer operational phone numbers, moves to another location, and refusals among the key reasons for panel attrition. This issue applied to all cohorts, to varying degrees.

Replacement of FE girls proved especially challenging. If FE girls could not be re-contacted, they were replaced from the original sample frame of FE girls compiled by CARE in late 2019, when FE girls were enrolled into primary schools at Grade 1 or Grade 2. This sample frame is now over three years old, and many of the girls on it could not be located. In four schools, the team ran out of possible replacement FE girls from the sample frame, and therefore were unable to replace some of the targeted FE girls who could not be located.

[126] One exception concerns the classroom observations. As we note in the discussion of fieldwork challenges below, we completed additional classroom observations (4, instead of the expected 2) in several formal primary schools. To ensure that unequal cluster sizes – i.e. overrepresentation of the four schools where we completed more than the required classroom observations – do not affect our conclusions, we test the robustness of our findings regarding teaching quality to the use of survey weights that ensure each cluster/school takes equal weight in the analysis. None of our findings regarding teaching quality are affected by the inclusion or exclusion of survey weights.

Issues with Cooperation of Head Teachers:

After most of the data collection had been completed in their school, one head teacher refused further access to the school for our field team. Despite attempts at negotiation, including the intervention of CARE, the head teacher refused access and prevented further interviews from being completed. At a second school, the head teacher refused to allow school-level data collection; the team was able to complete interviews with the targeted girls outside the school setting, but experienced a high rate of refusal.

Examinations and School Closures:

Examination in several schools in Dhobley limited our team's ability to conduct classroom observations and headcounts, as classes were not in session. Additionally, all school were closed on February 27 for a teachers' exam, causing fieldwork delays.

The team found one school that was not operational at the time of fieldwork. The former students were interviewed, with the goal of maintaining the integrity of the panel sample, but school-level data collection – headcounts, classroom observations, and a head teacher survey – could not be completed. A second school was also found to have been closed for the entire last year.

Low NFE Attendance Rates and Classroom Observations

Several of the C5 NFE centres sampled had only a few students in attendance at the time of fieldwork visits. This made it impossible to conduct classroom observations, as some classes were not in session. In lieu of classroom observations in these centres, the evaluation team completed additional classroom observations in nearby formal schools that are also part of the sample. For this reason, there are four primary schools which, instead of having two classroom observations, have four classroom observations.

Flight Delays and Accessibility Issues:

The team experienced a number of delays during fieldwork due to flight delays. Flight delays impacted the teams working in Jowhar, Bardhere, and Dhobley, often up to several days. A lethal airport attack in Dollow caused a suspension of all flights from that airport also causing delays for one of our field teams.

On the last day of data collection in the December 20th school (WAB01) in Mogadishu security forces arrived at the school and requested that the team provided a permit for data collection. A team member was apprehended and taken to the police station for several hours before CARE, with the assistance from the Ministry of Education, resolved the situation.

Annex 3 - Characteristics and Barriers

Sample Composition of Characteristics and Barriers

The tables in this section describe the characteristics of the girls who comprise the five AGES cohorts, and the barriers to education they face. For each cohort, there are two sets of tables, differentiating between sample characteristics (e.g., the share of the cohort comprised of girls from female-headed households) and barriers to education. In each table, we report the share of the cohort to whom the characteristic or barrier applied at BL and at ML2.

The goal of the tables is to document how the sample's characteristics have changed across rounds, focusing on the impact of sample attrition and replacement. For this reason, we do not focus on changes that occurred *within* the same girl over time, such as a girl who was single at BL but is married at ML2. This type of natural change is interesting, but does not accurately reflect sample composition and how it has changed over time – instead, it reflects natural variation in characteristics over time, including as a girl ages.

The baseline values we report are calculated from the entire baseline sample, including girls and schools who later fell out of the sample or were replaced. The ML2 values we report are the share of girls in the ML2 sample (panel sample) who fell into a given subgroup *at baseline*. To illustrate, consider FE girls who have only one living parent, as reflected in the top row of the table below. For both the BL and ML2 calculation, this outcome is assessed at baseline. Among the baseline sample (n = 421), 10.7 percent of girls had a single living parent. For the ML2 calculation, we assess how many girls in the ML2 sample faced this barrier at BL; in other words, we ask how many girls who continue to appear in the sample at ML2 had a single living parent at BL. This share has fallen to 10.4 percent of the sample (of n = 270) due to attrition. To be clear, the 28 single-parent girls in the ML2 sample were all single-parent girls in the BL sample as well, constituting a portion of the 45 such girls observed in the BL sample. Thus, this table shows how the sample's composition has changed over time in response to attrition; it does not reflect changes in individual girls' characteristics.

Note that a small subset of barriers related to school infrastructure/facilities are available only for the FE girl cohort. This applies to outcomes related to the availability of learning materials, the quality of infrastructure (e.g., availability of electricity), and the provision of school meals. In addition, for the C4 and C5 NFE cohorts, some characteristics and barriers are reported for just a subsample of each cohort; this is due to the structure of the household survey employed at ML1 and ML2, in which caregivers were not interviewed if the girl was 18 years old or older. As a result, girls who were over 18 at the time of their cohort-specific baseline data collection (i.e. ML1 for C4 NFE girls and ML2 for C5 NFE girls) lack information on parental educational attainment, parental presence in the household, and parental attitudes toward girls' education.

As noted in earlier evaluation rounds, the definitions employed for disability status result in very low rates of estimated disability prevalence across all cohorts. In previous reports – and elsewhere in this report – we employed a number of alternative standards for classifying girls' disability status, reflecting our belief that respondents systematically underreported many forms of impairment. In the tables below we generally employ the Washington Group standards for all forms of impairment. The exception are aggregate categories (e.g., physical disability or mental health disability), where we report statistics based on both the standard coding scheme and an alternative, more liberal, coding scheme. We refer readers to the BL evaluation report for additional discussion of this issue and the coding schemes the evaluation team has used.

TABLE 97: CHARACTERISTICS OF FE COHORT AT BASELINE AND ML2

	Baseline	Midline #2	

Subgroup	Pct. of Total	N	Pct. of Total	N	Variable Name & Source
Household Characteristics					
Girl has only one living parent	10.7	45	10.4	28	orphan_single
Girl has no living parents	0.7	3	0.4	1	orphan_double
Girl does not live with either parent in her HH	3.6	15	2.6	7	no_parents
Girl is currently married	0.2	1	0.4	1	married
Girl has ever been married	0.2	1	0.4	1	ever_married
Girl is a mother and is under 16 years of age	0.0	0	0.0	0	mother_16
Girl is a mother and is under 18 years of age	0.0	0	0.0	0	mother_18
Female-headed household	34.9	147	33.0	89	female_hoh
Parental Educational Background					
HoH has no education of any kind (no Quranic)	22.1	93	25.9	70	hoh_noeduc
HoH has no formal education	80.8	340	80.0	216	hoh_noformal
Caregiver has no education of any kind (no Quranic)	29.0	122	33.0	89	cg_noeduc
Caregiver has no formal education	85.7	361	85.6	231	cg_noformal
Household Economic Status					
HH has a poor-quality roof	18.1	76	15.2	41	poor_roof
HH went to sleep hungry most nights, last 12 months	6.9	29	6.3	17	nofood_most
HH went without clean water most days, last 12 months	4.8	20	5.9	16	nowater_most
HH went without medicine most days, last 12 months	14.5	61	14.4	39	nomeds_most
HH went without cash income most days, last 12 months	13.5	57	13.7	37	nocash_most
HH owns lands	38.0	160	40.7	110	owns_land
Caregiver has savings of some form	2.6	11	3.3	9	savings

HH owns a phone	84.8	357	83.7	226	owns_phone
HH owns a smartphone	20.7	87	20.7	56	owns_smartphone
HoH does not earn a regular wage	37.1	156	37.8	102	hoh_nowage
Household head is engaged in pastoralism	1.4	6	0.7	2	pastoral
Disability Status					
Vision disability	0.0	0	0.0	0	disab_vis
Hearing disability	0.0	0	0.0	0	disab_hear
Disability the impedes mobility	0.0	0	0.0	0	disab_mob
Disability of the arms/hands	0.0	0	0.0	0	disab_arms
Disability that impedes self-care	0.0	0	0.0	0	disab_selfcare
Disability that impedes communication	0.0	0	0.0	0	disab_comm
Cognitive disability	0.0	0	0.0	0	disab_cog
Behavioral disability	0.0	0	0.0	0	disab_behavior
Mental health disability	0.0	0	0.0	0	disab_mh
Mental health disability, alternative coding	11.6	49	10.7	29	disab_mh_alt1
Physical disability, any type	0.0	0	0.0	0	disab_phys
Physical disability, any type, alternative coding	0.7	3	0.7	2	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	0.0	0	0.0	0	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	0.5	2	0.4	1	disab_ccb_alt1
Displacement and Language					
Household are IDPs	42.3	178	51.9	140	idp
Household speaks af-Maay	27.8	117	26.7	72	maay

TABLE 98: BARRIERS FACED BY FE COHORT GIRLS, AT BASELINE AND ML2

	Baseline	Midline #2	

Subgroup	Pct. of Total	N	Pct. of Total	N	Variable Name & Source
School Facilities					
Girl will not use drinking facilities at school	17.1	72	16.7	45	wontuse_drinking
Girl will not use toilet facilities at school	8.8	37	10.0	27	wontuse_toilet
No computers available for use at school	94.1	396	94.1	254	no_computers
Girl cannot use books/learning materials at school	29.9	126	31.9	86	no_materials
Not enough seats for every student in class	8.6	36	8.5	23	no_seats
Textbooks are shared between students	39.9	168	45.9	124	shared_textbooks
Girls are able to take textbooks/materials home at night	34.7	146	36.7	99	materials_night
School has reliable electricity	66.7	281	64.1	173	electric
School has water access within 1 km	54.6	230	55.9	151	water
School has only cement floors (no dirt)	67.7	285	67.4	182	cement_floor
School has separate toilets for girls	71.5	301	73.0	197	separate_toilets
School provides at least one meal for students	28.5	120	29.3	79	school_feeding
School Environment and Teacher Behaviors					
Teachers do not make girl feel welcome at school	13.5	57	15.6	42	unwelcome
Teachers are often absent from class	29.7	125	30.4	82	teacher_absent
Teacher rarely or never encourages participation	8.3	35	9.6	26	teacher_participation
Teachers punishes students who give wrong answer in class	78.6	331	81.5	220	punish_wrong

Teacher used corporal punishment in last week	36.1	152	39.3	106	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2.1	9	1.1	3	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	2.4	10	1.5	4	unsafe_journey_cg
Parental and Girl Attitudes toward Schooling					
Girl spends a few hours or more per day doing HH chores	57.7	243	54.8	148	chores
Girls does not feel she can attend school if she is menstruating	0.5	2	0.4	1	absent_mens
Caregiver aspires to university education for girl	88.6	373	87.8	237	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	87.9	370	86.3	233	invest_girlseduc
Caregiver believes work or HH chores are acceptable reason to not attend school	43.5	183	49.3	133	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	58.7	247	60.7	164	school_expensive

TABLE 99: CHARACTERISTICS OF ABE COHORT AT BASELINE AND ML2

	Baseline		Midline #2		
Subgroup	**Pct. of Total**	**N**	**Pct. of Total**	**N**	**Variable Name & Source**
Household Characteristics					
Girl has only one living parent	10.1	49	9.4	29	orphan_single
Girl has no living parents	1.2	6	1.3	4	orphan_double
Girl does not live with either parent in her HH	5.0	24	4.8	15	no_parents
Girl is currently married	0.8	4	0.6	2	married

Girl has ever been married	1.4	7	1.0	3	ever_married
Girl is a mother and is under 16 years of age	0.0	0	0.0	0	mother_16
Girl is a mother and is under 18 years of age	0.4	2	0.0	0	mother_18
Female-headed household	33.7	163	33.5	104	female_hoh
Parental Educational Background					
HoH has no education of any kind (no Quranic)	22.1	107	20.6	64	hoh_noeduc
HoH has no formal education	82.9	401	83.2	258	hoh_noformal
Caregiver has no education of any kind (no Quranic)	27.9	135	27.4	85	cg_noeduc
Caregiver has no formal education	86.0	416	85.2	264	cg_noformal
Household Economic Status					
HH has a poor-quality roof	22.1	107	23.2	72	poor_roof
HH went to sleep hungry most nights, last 12 months	7.9	38	6.5	20	nofood_most
HH went without clean water most days, last 12 months	6.2	30	6.1	19	nowater_most
HH went without medicine most days, last 12 months	16.5	80	14.8	46	nomeds_most
HH went without cash income most days, last 12 months	18.0	87	19.0	59	nocash_most
HH owns lands	31.4	152	32.3	100	owns_land
Caregiver has savings of some form	1.7	8	1.3	4	savings
HH owns a phone	86.4	418	85.8	266	owns_phone
HH owns a smartphone	13.2	64	14.5	45	owns_smartphone
HoH does not earn a regular wage	36.6	177	32.6	101	hoh_nowage
Household head is engaged in pastoralism	2.7	13	1.3	4	pastoral
Disability Status					
Vision disability	0.0	0	0.0	0	disab_vis

Hearing disability	0.2	1	0.3	1	disab_hear
Disability the impedes mobility	0.0	0	0.0	0	disab_mob
Disability of the arms/hands	0.0	0	0.0	0	disab_arms
Disability that impedes self-care	0.0	0	0.0	0	disab_selfcare
Disability that impedes communication	0.0	0	0.0	0	disab_comm
Cognitive disability	0.0	0	0.0	0	disab_cog
Behavioral disability	0.6	3	1.0	3	disab_behavior
Mental health disability	0.0	0	0.0	0	disab_mh
Mental health disability, alternative coding	16.9	82	17.7	55	disab_mh_alt1
Physical disability, any type	0.2	1	0.3	1	disab_phys
Physical disability, any type, alternative coding	0.2	1	0.3	1	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	0.6	3	1.0	3	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	1.0	5	1.6	5	disab_ccb_alt1
Displacement and Language					
Household are IDPs	46.9	227	58.1	180	idp
Household speaks af-Maay	31.2	151	31.9	99	maay

TABLE 100: BARRIERS FACED BY ABE COHORT GIRLS, AT BASELINE AND ML2

	Baseline		Midline #2		
Subgroup	**Pct. of Total**	**N**	**Pct. of Total**	**N**	**Variable Name & Source**
School Facilities					
Girl will not use drinking facilities at school	24.4	118	25.2	78	wontuse_drinking
Girl will not use toilet facilities at school	12.6	61	12.9	40	wontuse_toilet
No computers available for use at school	96.1	465	96.1	298	no_computers

Girl cannot use books/learning materials at school	23.3	113	23.5	73	no_materials
Not enough seats for every student in class	10.3	50	12.3	38	no_seats
School Environment and Teacher Behaviors					
Teachers do not make girl feel welcome at school	12.0	58	11.9	37	unwelcome
Teachers are often absent from class	28.1	136	27.1	84	teacher_absent
Teacher rarely or never encourages participation	13.2	64	12.3	38	teacher_participation
Teachers punishes students who give wrong answer in class	70.0	339	69.7	216	punish_wrong
Teacher used corporal punishment in last week	21.7	105	23.9	74	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	0.8	4	0.6	2	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	0.2	1	0.3	1	unsafe_journey_cg
Parental and Girl Attitudes toward Schooling					
Girl spends a few hours or more per day doing HH chores	63.4	307	63.5	197	chores
Girls does not feel she can attend school if she is menstruating	2.5	12	2.6	8	absent_mens
Caregiver aspires to university education for girl	89.3	432	89.7	278	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	86.6	419	88.4	274	invest_girlseduc
Caregiver believes work or HH chores are acceptable reason to not attend school	38.0	184	36.5	113	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	64.9	314	64.5	200	school_expensive

TABLE 101: CHARACTERISTICS OF C1 NFE COHORT AT BASELINE AND ML2

	Baseline		Midline #2		
Subgroup	Pct. of Total	N	Pct. of Total	N	Variable Name & Source
Household Characteristics					
Girl has only one living parent	17.5	90	17.5	53	orphan_single
Girl has no living parents	1.0	5	1.3	4	orphan_double
Girl does not live with either parent in her HH	8.5	44	7.3	22	no_parents
Girl is currently married	9.3	48	10.9	33	married
Girl has ever been married	17.7	91	21.2	64	ever_married
Girl is a mother and is under 16 years of age	0.0	0	0.0	0	mother_16
Girl is a mother and is under 18 years of age	1.6	8	1.3	4	mother_18
Female-headed household	42.7	220	45.4	137	female_hoh
Parental Educational Background					
HoH has no education of any kind (no Quranic)	19.8	102	18.9	57	hoh_noeduc
HoH has no formal education	81.7	421	82.8	250	hoh_noformal
Caregiver has no education of any kind (no Quranic)	24.5	126	24.8	75	cg_noeduc
Caregiver has no formal education	83.9	432	84.1	254	cg_noformal
Household Economic Status					
HH has a poor-quality roof	19.6	101	20.2	61	poor_roof
HH went to sleep hungry most nights, last 12 months	6.2	32	5.0	15	nofood_most
HH went without clean water most days, last 12 months	4.7	24	4.0	12	nowater_most
HH went without medicine most days, last 12 months	18.8	97	18.9	57	nomeds_most
HH went without cash income most days, last 12 months	12.8	66	10.9	33	nocash_most

HH owns lands	29.1	150	29.1	88	owns_land
Caregiver has savings of some form	1.6	8	1.0	3	savings
HH owns a phone	88.2	454	87.4	264	owns_phone
HH owns a smartphone	20.2	104	19.9	60	owns_smartphone
HoH does not earn a regular wage	43.7	225	40.7	123	hoh_nowage
Household head is engaged in pastoralism	1.2	6	0.7	2	pastoral
Disability Status					
Vision disability	0.2	1	0.3	1	disab_vis
Hearing disability	0.0	0	0.0	0	disab_hear
Disability the impedes mobility	0.0	0	0.0	0	disab_mob
Disability of the arms/hands	0.0	0	0.0	0	disab_arms
Disability that impedes self-care	0.0	0	0.0	0	disab_selfcare
Disability that impedes communication	0.0	0	0.0	0	disab_comm
Cognitive disability	0.2	1	0.3	1	disab_cog
Behavioral disability	0.4	2	0.3	1	disab_behavior
Mental health disability	0.0	0	0.0	0	disab_mh
Mental health disability, alternative coding	16.9	87	17.2	52	disab_mh_alt1
Physical disability, any type	0.2	1	0.3	1	disab_phys
Physical disability, any type, alternative coding	0.8	4	0.7	2	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	0.6	3	0.7	2	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	1.2	6	1.7	5	disab_ccb_alt1
Displacement and Language					
Household are IDPs	37.5	193	50.3	152	idp
Household speaks af-Maay	25.2	130	28.1	85	maay

TABLE 102: BARRIERS FACED BY C1 NFE COHORT GIRLS, AT BASELINE AND ML2

Subgroup	Baseline		Midline #2		
	Pct. of Total	N	Pct. of Total	N	Variable Name & Source
School Facilities					
Girl will not use drinking facilities at school	24.7	127	30.1	91	wontuse_drinking
Girl will not use toilet facilities at school	23.1	119	29.8	90	wontuse_toilet
No computers available for use at school	96.7	498	96.0	290	no_computers
Girl cannot use books/learning materials at school	25.8	133	29.1	88	no_materials
Not enough seats for every student in class	20.4	105	24.5	74	no_seats
School Environment and Teacher Behaviors					
Teachers do not make girl feel welcome at school	11.8	61	10.6	32	unwelcome
Teachers are often absent from class	28.2	145	24.8	75	teacher_absent
Teacher rarely or never encourages participation	8.7	45	10.9	33	teacher_participation
Teachers punishes students who give wrong answer in class	57.9	298	58.9	178	punish_wrong
Teacher used corporal punishment in last week	8.2	42	10.3	31	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2.3	12	1.3	4	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	2.3	12	2.0	6	unsafe_journey_cg
Parental and Girl Attitudes toward Schooling					
Girl spends a few hours or more per day doing HH chores	71.7	369	71.9	217	chores

Girls does not feel she can attend school if she is menstruating	5.0	26	4.3	13	absent_mens
Caregiver aspires to university education for girl	84.7	436	85.8	259	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	88.7	457	87.4	264	invest_girlseduc
Caregiver believes work or HH chores are acceptable reason to not attend school	39.2	202	41.4	125	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	62.7	323	63.2	191	school_expensive

TABLE 103: CHARACTERISTICS OF C4 NFE COHORT AT ML1 (COHORT-SPECIFIC BASELINE) AND ML2

	Midline #1		Midline #2		
Subgroup	Pct. of Total	N	Pct. of Total	N	Variable Name & Source
Household Characteristics					
Girl has only one living parent	5.8	53	6.7	44	orphan_single
Girl has no living parents	0.3	3	0.5	3	orphan_double
Girl does not live with either parent in her HH	3.3	30	3.0	20	no_parents
Girl is currently married	0.8	7	1.1	7	married
Girl has ever been married	1.1	10	1.5	10	ever_married
Girl is a mother and is under 16 years of age	0.1	1	0.2	1	mother_16
Girl is a mother and is under 18 years of age	0.5	5	0.8	5	mother_18
Female-headed household	21.0	192	23.0	152	female_hoh
Parental Educational Background					
HoH has no education of any kind (no Quranic)	5.1	47	5.6	37	hoh_noeduc
HoH has no formal education	25.9	237	26.9	178	hoh_noformal

Caregiver has no education of any kind (no Quranic)	4.7	43	5.0	33	cg_noeduc
Caregiver has no formal education	25.8	236	26.6	176	cg_noformal
Household Economic Status					
HH has a poor-quality roof	21.9	201	22.8	151	poor_roof
HH went to sleep hungry most nights, last 12 months	6.7	61	7.7	51	nofood_most
HH went without clean water most days, last 12 months	8.1	74	8.9	59	nowater_most
HH went without medicine most days, last 12 months	14.1	129	14.4	95	nomeds_most
HH went without cash income most days, last 12 months	20.1	184	21.0	139	nocash_most
HH owns lands	23.0	211	23.6	156	owns_land
Caregiver has savings of some form	1.0	9	1.1	7	savings
HH owns a phone	28.2	258	29.5	195	owns_phone
HH owns a smartphone	3.3	30	3.2	21	owns_smartphone
HoH does not earn a regular wage	14.0	128	15.0	99	hoh_nowage
Household head is engaged in pastoralism	0.3	3	0.5	3	pastoral
Disability Status					
Vision disability	0.4	4	0.5	3	disab_vis
Hearing disability	0.1	1	0.0	0	disab_hear
Disability the impedes mobility	2.4	22	2.9	19	disab_mob
Disability of the arms/hands	0.1	1	0.2	1	disab_arms
Disability that impedes self-care	0.2	2	0.3	2	disab_selfcare
Disability that impedes communication	0.0	0	0.0	0	disab_comm
Cognitive disability	3.1	28	2.7	18	disab_cog
Behavioral disability	1.6	15	1.7	11	disab_behavior
Mental health disability	15.5	142	16.9	112	disab_mh

Mental health disability, alternative coding	17.0	156	18.8	124	disab_mh_alt1
Physical disability, any type	3.2	29	3.6	24	disab_phys
Physical disability, any type, alternative coding	3.4	31	3.9	26	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	4.4	40	4.1	27	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	5.6	51	5.3	35	disab_ccb_alt1
Displacement and Language					
Household are IDPs	54.4	498	61.1	404	idp
Household speaks af-Maay	33.0	302	35.4	234	maay

TABLE 104: BARRIERS FACED BY C4 NFE COHORT GIRLS, AT ML1 (COHORT-SPECIFIC BASELINE) AND ML2

	Midline #1		Midline #2		
Subgroup	**Pct. of Total**	**N**	**Pct. of Total**	**N**	**Variable Name & Source**
School Facilities					
Girl will not use drinking facilities at school	17.1	157	17.9	118	wontuse_drinking
Girl will not use toilet facilities at school	20.9	191	21.2	140	wontuse_toilet
No computers available for use at school	88.6	812	88.8	587	no_computers
Girl cannot use books/learning materials at school	6.4	59	6.4	42	no_materials
Not enough seats for every student in class	3.8	35	3.9	26	no_seats
School Environment and Teacher Behaviors					
Teachers do not make girl feel welcome at school	12.8	117	14.2	94	unwelcome
Teachers are often absent from class	17.4	159	15.9	105	teacher_absent

Teacher rarely or never encourages participation	5.2	48	5.4	36	teacher_participation
Teachers punishes students who give wrong answer in class	28.2	258	26.9	178	punish_wrong
Teacher used corporal punishment in last week	7.6	70	7.0	46	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	1.6	15	1.2	8	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	0.4	4	0.5	3	unsafe_journey_cg
Parental and Girl Attitudes toward Schooling					
Girl spends a few hours or more per day doing HH chores	92.1	844	91.8	607	chores
Girls does not feel she can attend school if she is menstruating	3.2	29	3.2	21	absent_mens
Caregiver aspires to university education for girl	63.0	577	63.8	422	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	80.3	736	79.4	525	invest_girlseduc
Caregiver believes work or HH chores are acceptable reason to not attend school	7.1	65	7.3	48	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	18.3	168	19.1	126	school_expensive

TABLE 105: CHARACTERISTICS OF C5 NFE COHORT AT ML2 (COHORT-SPECIFIC BASELINE)

	Midline #2		
Subgroup	**Pct. of Total**	**N**	**Variable Name & Source**
Household Characteristics			
Girl has only one living parent	20.1	28	orphan_single
Girl has no living parents	0.0	0	orphan_double

Girl does not live with either parent in her HH	8.6	12	no_parents
Girl is currently married	0.8	4	married
Girl has ever been married	2.0	10	ever_married
Girl is a mother and is under 16 years of age	0.2	1	mother_16
Girl is a mother and is under 18 years of age	1.2	6	mother_18
Female-headed household	71.2	99	female_hoh
Parental Educational Background			
HoH has no education of any kind (no Quranic)	18.7	26	hoh_noeduc
HoH has no formal education	74.8	104	hoh_noformal
Caregiver has no education of any kind (no Quranic)	20.9	29	cg_noeduc
Caregiver has no formal education	67.6	94	cg_noformal
Household Economic Status			
HH has a poor-quality roof	19.8	101	poor_roof
HH went to sleep hungry most nights, last 12 months	11.5	59	nofood_most
HH went without clean water most days, last 12 months	17.0	87	nowater_most
HH went without medicine most days, last 12 months	16.8	86	nomeds_most
HH went without cash income most days, last 12 months	33.1	169	nocash_most
HH owns lands	21.3	109	owns_land
Caregiver has savings of some form	2.9	15	savings
HH owns a phone	24.9	127	owns_phone
HH owns a smartphone	6.7	34	owns_smartphone
HoH does not earn a regular wage	11.2	57	hoh_nowage
Household head is engaged in pastoralism	1.0	5	pastoral
Disability Status			
Vision disability	2.3	12	disab_vis

Hearing disability	0.6	3	disab_hear
Disability the impedes mobility	4.1	21	disab_mob
Disability of the arms/hands	0.7	1	disab_arms
Disability that impedes self-care	0.4	2	disab_selfcare
Disability that impedes communication	1.2	6	disab_comm
Cognitive disability	3.5	18	disab_cog
Behavioral disability	4.7	24	disab_behavior
Mental health disability	36.2	185	disab_mh
Mental health disability, alternative coding	42.9	219	disab_mh_alt1
Physical disability, any type	6.1	31	disab_phys
Physical disability, any type, alternative coding	7.2	37	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	7.8	40	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	9.6	49	disab_ccb_alt1
Displacement and Language			
Household are IDPs	37.6	192	idp
Household speaks af-Maay	29.5	151	maay

TABLE 106: BARRIERS FACED BY C5 NFE COHORT GIRLS, AT ML2 (COHORT-SPECIFIC BASELINE)

	Midline #2		
Subgroup	**Pct. of Total**	**N**	**Variable Name & Source**
School Facilities			
Girl will not use drinking facilities at school	16.2	83	wontuse_drinking
Girl will not use toilet facilities at school	24.3	124	wontuse_toilet
No computers available for use at school	92.4	472	no_computers

Girl cannot use books/learning materials at school	6.7	34	no_materials
Not enough seats for every student in class	4.1	21	no_seats
School Environment and Teacher Behaviors			
Teachers do not make girl feel welcome at school	17.4	89	unwelcome
Teachers are often absent from class	23.1	118	teacher_absent
Teacher rarely or never encourages participation	6.8	35	teacher_participation
Teachers punishes students who give wrong answer in class	51.3	262	punish_wrong
Teacher used corporal punishment in last week	16.4	84	corp_pun
Safety of School and Surrounding Area			
Girl does not feel safe traveling to school	2.2	11	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	2.7	14	unsafe_journey_cg
Parental and Girl Attitudes toward Schooling			
Girl spends a few hours or more per day doing HH chores	83.0	424	chores
Girls does not feel she can attend school if she is menstruating	28.2	68	absent_mens
Caregiver aspires to university education for girl	71.2	364	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	55.2	282	invest_girlseduc
Caregiver believes work or HH chores are acceptable reason to not attend school	10.4	53	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	11.4	58	school_expensive

Intersectionality of Subgroups and Barriers

The tables in this section expand on the sample composition statistics provided above by evaluating the intersectionality of particular subgroups and the barriers to education girls face. The goal of this section is to highlight where barriers and subgroups overlap, because this overlap may result in barriers that reinforce one another, contributing to marginalization of particular types of girls.

In light of the large number of characteristics assessed in the previous section, and the number of subgroups/barriers which apply to very few girls, we do not include all subgroups/barriers in the tables below. First, we exclude those – such as specific forms of disability – for which there are very few girls. This is not intended to downplay the importance of intersectionality applied to these characteristics; rather, it reflects the fact that assessing intersectionality with a category into which only 1-2 girls falls is not very informative. Second, we focus on a subset of characteristics and barriers that we expect to have the largest or most direct impact on a girls' educational outcomes. This decision is driven by space considerations, as the tables become unwieldy with too many distinct columns. We report on intersectionality for each cohort separately, given their important underlying differences.

TABLE 107: INTERSECTIONALITY OF BARRIERS TO EDUCATION AMONG FE GIRLS

	Number of Observations	Girl has only one living parent	Female HoH	HoH has no formal education	HH has a poor-quality roof	HH went to sleep hungry most nights, last 12 months	Mental health disability, standard coding	Mental health disability, alternative coding	Physical disability, any type	Cognitive, communicative, or behavioural disability, any type	Household are IDPs	Teachers are not welcoming	Teacher uses corporal punishment	Girl has heavy chore burden	Caregiver values work/HH chores over school attendance
Number of Observations		28	89	216	41	17	0	29	0	0	140	42	106	148	133
Girl has only one living parent	28	10.4	8.9	8.9	2.2	1.5	0.0	1.5	0.0	0.0	5.2	3.3	3.7	6.7	2.6
Female HoH	89	8.9	33.0	26.7	4.4	2.2	0.0	3.0	0.0	0.0	18.1	6.3	14.4	19.3	17.0
HoH has no formal education	216	8.9	26.7	80.0	11.9	5.9	0.0	7.4	0.0	0.0	41.5	11.5	31.9	46.7	39.6
HH has a poor-quality roof	41	2.2	4.4	11.9	15.2	0.7	0.0	3.3	0.0	0.0	9.6	2.2	5.9	9.3	4.4
HH went to sleep hungry most nights, last 12 months	17	1.5	2.2	5.9	0.7	6.3	0.0	0.0	0.0	0.0	4.4	3.7	4.1	1.9	3.0
Mental health disability, standard coding	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mental health disability, alternative coding	29	1.5	3.0	7.4	3.3	0.0	0.0	10.7	0.0	0.0	7.4	1.5	6.3	8.1	2.2
Physical disability, any type	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Cognitive, communicative, or behavioral disability, any type	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Household are IDPs	140	5.2	18.1	41.5	9.6	4.4	0.0	7.4	0.0	0.0	51.9	8.9	22.6	33.0	23.3
Teachers are not welcoming	42	3.3	6.3	11.5	2.2	3.7	0.0	1.5	0.0	0.0	8.9	15.6	9.6	5.6	8.5
Teacher uses corporal punishment	106	3.7	14.4	31.9	5.9	4.1	0.0	6.3	0.0	0.0	22.6	9.6	39.3	24.8	15.2
Girl has heavy chore burden	148	6.7	19.3	46.7	9.3	1.9	0.0	8.1	0.0	0.0	33.0	5.6	24.8	54.8	25.9
Caregiver values work/HH chores over school attendance	133	2.6	17.0	39.6	4.4	3.0	0.0	2.2	0.0	0.0	23.3	8.5	15.2	25.9	49.3

TABLE 108: INTERSECTIONALITY OF BARRIERS TO EDUCATION AMONG ABE GIRLS

	Number of Observations	Girl has only one living parent	Female HoH	HoH has no formal education	HH has a poor-quality roof	HH went to sleep hungry most nights, last 12 months	Mental health disability, standard coding	Mental health disability, alternative coding	Physical disability, any type	Cognitive, communicative, or behavioural disability, any type	Household are IDPs	Teachers are not welcoming	Teacher uses corporal punishment	Girl has heavy chore burden	Caregiver values work/HH chores over school attendance

Number of Observations		29	104	258	72	20	0	55	1	3	180	37	74	197	113
Girl has only one living parent	29	9.4	5.5	8.4	1.3	1.6	0.0	2.6	0.0	0.6	6.1	1.0	2.3	6.5	2.6
Female HoH	104	5.5	33.5	28.1	4.5	2.9	0.0	5.8	0.0	0.3	21.9	3.9	7.7	20.3	14.5
HoH has no formal education	258	8.4	28.1	83.2	18.4	5.5	0.0	14.5	0.3	1.0	49.0	11.6	21.0	57.4	28.4
HH has a poor-quality roof	72	1.3	4.5	18.4	23.2	1.9	0.0	6.5	0.0	0.0	15.2	1.0	5.2	13.5	9.4
HH went to sleep hungry most nights, last 12 months	20	1.6	2.9	5.5	1.9	6.5	0.0	1.0	0.0	0.6	3.9	2.3	1.0	4.2	1.9
Mental health disability, standard coding	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mental health disability, alternative coding	55	2.6	5.8	14.5	6.5	1.0	0.0	17.7	0.0	0.0	12.6	2.6	7.4	11.3	6.5
Physical disability, any type	1	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.3	0.3
Cognitive, communicative, or behavioral disability, any type	3	0.6	0.3	1.0	0.0	0.6	0.0	0.0	0.0	1.0	0.3	1.0	0.0	1.0	0.3
Household are IDPs	180	6.1	21.9	49.0	15.2	3.9	0.0	12.6	0.0	0.3	58.1	7.1	17.1	38.7	20.3
Teachers are not welcoming	37	1.0	3.9	11.6	1.0	2.3	0.0	2.6	0.0	1.0	7.1	11.9	2.6	7.7	5.5
Teacher uses corporal punishment	74	2.3	7.7	21.0	5.2	1.0	0.0	7.4	0.0	0.0	17.1	2.6	23.9	18.7	7.1
Girl has heavy chore burden	197	6.5	20.3	57.4	13.5	4.2	0.0	11.3	0.3	1.0	38.7	7.7	18.7	63.5	21.3

Caregiver values work/HH chores over school attendance	113	2.6	14.5	28.4	9.4	1.9	0.0	6.5	0.3	0.3	20.3	5.5	7.1	21.3	36.5

TABLE 109: INTERSECTIONALITY OF BARRIERS TO EDUCATION AMONG C1 NFE GIRLS

	Number of Observations	Girl has only one living parent	Female HoH	HoH has no formal education	HH has a poor-quality roof	HH went to sleep hungry most nights, last 12 months	Mental health disability, standard coding	Mental health disability, alternative coding	Physical disability, any type	Cognitive, communicative, or behavioural disability, any type	Household are IDPs	Teachers are not welcoming	Teacher uses corporal punishment	Girl has heavy chore burden	Caregiver values work/HH chores over school attendance
Number of Observations		53	137	250	61	15	0	52	1	2	152	32	31	217	125
Girl has only one living parent	53	17.5	11.3	15.2	4.6	0.7	0.0	4.6	0.0	0.3	9.9	2.6	3.6	14.6	4.0
Female HoH	137	11.3	45.4	40.1	9.9	2.0	0.0	7.6	0.3	0.3	24.5	4.3	4.6	35.8	19.9
HoH has no formal education	250	15.2	40.1	82.8	16.6	3.3	0.0	13.2	0.3	0.0	42.4	7.6	9.3	61.3	31.8
HH has a poor-quality roof	61	4.6	9.9	16.6	20.2	2.3	0.0	5.6	0.0	0.0	11.9	1.0	2.3	15.2	8.3
HH went to sleep hungry most nights, last 12 months	15	0.7	2.0	3.3	2.3	5.0	0.0	0.7	0.0	0.0	2.6	1.0	0.0	4.0	2.3

Mental health disability, standard coding	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mental health disability, alternative coding	52	4.6	7.6	13.2	5.6	0.7	0.0	17.2	0.3	0.3	11.6	2.6	4.3	12.9	5.6
Physical disability, any type	1	0.0	0.3	0.3	0.0	0.0	0.0	0.3	0.3	0.0	0.3	0.3	0.0	0.3	0.0
Cognitive, communicative, or behavioral disability, any type	2	0.3	0.3	0.0	0.0	0.0	0.0	0.3	0.0	0.7	0.0	0.7	0.0	0.7	0.3
Household are IDPs	152	9.9	24.5	42.4	11.9	2.6	0.0	11.6	0.3	0.0	50.3	3.6	7.0	40.1	18.9
Teachers are not welcoming	32	2.6	4.3	7.6	1.0	1.0	0.0	2.6	0.3	0.7	3.6	10.6	1.3	8.9	3.3
Teacher uses corporal punishment	31	3.6	4.6	9.3	2.3	0.0	0.0	4.3	0.0	0.0	7.0	1.3	10.3	9.9	4.6
Girl has heavy chore burden	217	14.6	35.8	61.3	15.2	4.0	0.0	12.9	0.3	0.7	40.1	8.9	9.9	71.9	29.5
Caregiver values work/HH chores over school attendance	125	4.0	19.9	31.8	8.3	2.3	0.0	5.6	0.0	0.3	18.9	3.3	4.6	29.5	41.4

TABLE 110: INTERSECTIONALITY OF BARRIERS TO EDUCATION AMONG C4 NFE GIRLS

	Number of Observations	Girl has only one living parent	Female HoH	HoH has no formal education	HH has a poor-quality roof	HH went to sleep hungry most nights, last 12 months	Mental health disability, standard coding	Mental health disability, alternative coding	Physical disability, any type	Cognitive, communicative, or behavioural disability, any type	Household are IDPs	Teachers are not welcoming	Teacher uses corporal punishment	Girl has heavy chore burden	Caregiver values work/HH chores over school attendance
Number of Observations		44	152	178	151	51	112	124	24	27	404	94	46	607	48
Girl has only one living parent	44	6.7	6.2	5.3	1.1	0.3	0.6	1.1	0.2	0.0	3.5	0.5	0.3	5.7	1.2
Female HoH	152	6.2	23.0	19.1	4.1	2.3	2.7	4.5	0.8	0.6	12.6	2.3	2.0	20.4	5.0
HoH has no formal education	178	5.3	19.1	26.9	5.4	2.0	3.5	4.5	1.4	0.5	16.5	2.9	1.8	25.0	5.7
HH has a poor-quality roof	151	1.1	4.1	5.4	22.8	1.5	3.3	3.3	1.1	1.4	18.0	3.9	1.5	21.0	2.3
HH went to sleep hungry most nights, last 12 months	51	0.3	2.3	2.0	1.5	7.7	3.5	3.6	0.6	0.3	4.8	0.6	0.9	7.0	0.3
Mental health disability, standard coding	112	0.6	2.7	3.5	3.3	3.5	16.9	16.9	1.5	1.4	11.8	3.5	1.8	15.4	0.9
Mental health disability, alternative coding	124	1.1	4.5	4.5	3.3	3.6	16.9	18.8	1.7	1.4	12.6	3.6	1.8	17.2	1.5
Physical disability, any type	24	0.2	0.8	1.4	1.1	0.6	1.5	1.7	3.6	0.3	2.3	1.2	0.3	2.7	0.8

Cognitive, communicative, or behavioral disability, any type	27	0.0	0.6	0.5	1.4	0.3	1.4	1.4	0.3	4.1	3.0	0.8	0.3	3.9	0.5
Household are IDPs	404	3.5	12.6	16.5	18.0	4.8	11.8	12.6	2.3	3.0	61.1	8.8	3.6	57.0	4.5
Teachers are not welcoming	94	0.5	2.3	2.9	3.9	0.6	3.5	3.6	1.2	0.8	8.8	14.2	1.2	11.0	0.9
Teacher uses corporal punishment	46	0.3	2.0	1.8	1.5	0.9	1.8	1.8	0.3	0.3	3.6	1.2	7.0	6.5	0.8
Girl has heavy chore burden	607	5.7	20.4	25.0	21.0	7.0	15.4	17.2	2.7	3.9	57.0	11.0	6.5	91.8	7.0
Caregiver values work/HH chores over school attendance	48	1.2	5.0	5.7	2.3	0.3	0.9	1.5	0.8	0.5	4.5	0.9	0.8	7.0	7.3

TABLE 111: INTERSECTIONALITY OF BARRIERS TO EDUCATION AMONG C5 NFE GIRLS

	Number of Observations	Girl has only one living parent	Female HoH	HoH has no formal education	HH has a poor-quality roof	HH went to sleep hungry most nights, last 12 months	Mental health disability, standard coding	Mental health disability, alternative coding	Physical disability, any type	Cognitive, communicative, or behavioural disability, any type	Household are IDPs	Teachers are not welcoming	Teacher uses corporal punishment	Girl has heavy chore burden	Caregiver values work/HH chores over school attendance

Number of Observations		28	89	216	41	17	0	29	0	0	140	42	106	148	133
Girl has only one living parent	28	10.4	8.9	8.9	2.2	1.5	0.0	1.5	0.0	0.0	5.2	3.3	3.7	6.7	2.6
Female HoH	89	8.9	33.0	26.7	4.4	2.2	0.0	3.0	0.0	0.0	18.1	6.3	14.4	19.3	17.0
HoH has no formal education	216	8.9	26.7	80.0	11.9	5.9	0.0	7.4	0.0	0.0	41.5	11.5	31.9	46.7	39.6
HH has a poor-quality roof	41	2.2	4.4	11.9	15.2	0.7	0.0	3.3	0.0	0.0	9.6	2.2	5.9	9.3	4.4
HH went to sleep hungry most nights, last 12 months	17	1.5	2.2	5.9	0.7	6.3	0.0	0.0	0.0	0.0	4.4	3.7	4.1	1.9	3.0
Mental health disability, standard coding	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mental health disability, alternative coding	29	1.5	3.0	7.4	3.3	0.0	0.0	10.7	0.0	0.0	7.4	1.5	6.3	8.1	2.2
Physical disability, any type	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cognitive, communicative, or behavioral disability, any type	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Household are IDPs	140	5.2	18.1	41.5	9.6	4.4	0.0	7.4	0.0	0.0	51.9	8.9	22.6	33.0	23.3
Teachers are not welcoming	42	3.3	6.3	11.5	2.2	3.7	0.0	1.5	0.0	0.0	8.9	15.6	9.6	5.6	8.5
Teacher uses corporal punishment	106	3.7	14.4	31.9	5.9	4.1	0.0	6.3	0.0	0.0	22.6	9.6	39.3	24.8	15.2
Girl has heavy chore burden	148	6.7	19.3	46.7	9.3	1.9	0.0	8.1	0.0	0.0	33.0	5.6	24.8	54.8	25.9

Caregiver values work/HH chores over school attendance	133	2.6	17.0	39.6	4.4	3.0	0.0	2.2	0.0	0.0	23.3	8.5	15.2	25.9	49.3

Disability Status

In Section 3.1 of the report, we described the characteristics of the C4 NFE and C5 NFE cohorts, which were recruited at ML1 and ML2, respectively. In the discussion therein, we opted to aggregate some forms of child functioning difficulty/disability by, for instance, combining all physical disabilities into a single category for reporting purposes. Here we provide a fuller breakdown of disability status, in line with the standard Washington Group classifications. The table below documents disability status among the two newer NFE cohorts, as measured at their respective baselines (ML1 for C4; ML2 for C5).

TABLE 112: DISABILITY STATUS AMONG C4 AND C5 NFE GIRLS

Sample Characteristics	Cohort 4 NFE Girls (ML1)	Cohort 5 NFE Girls (ML2)
Family Structure and Characteristics		
Sight/vision disability	0.4%	2.3%
Sight/vision disability, alternative coding	0.4%	2.3%
Hearing disability	0.1%	0.6%
Hearing disability, alternative coding	0.1%	0.8%
Mobility/walking disability	2.4%	4.1%
Mobility/walking disability, alternative coding	2.6%	5.1%
Disability preventing use of hands/arms	0.4%	0.7%
Disability preventing use of hands/arms, alternative coding	0.4%	0.7%
Disability preventing adequate self-care	0.2%	0.4%
Disability preventing adequate self-care, alternative coding	0.2%	0.6%
Disability inhibiting communication with others	0.0%	1.2%
Disability inhibiting communication with others, alternative coding	0.0%	1.6%
Cognitive disability - trouble remembering, understanding, etc.	3.1%	3.5%
Cognitive disability - trouble remembering, understanding, etc. - alternative coding	4.0%	4.5%
Behavioural disability - trouble making friends or controlling behavior	1.6%	4.7%
Behavioural disability - trouble making friends or controlling behavior - alternative coding	2.2%	5.9%
Mental health disability - frequent anxiety/depression	15.5%	36.2%
Mental health disability - frequent anxiety/depression - alternative coding	17.0%	42.9%
Any type of physical disability	3.2%	6.1%

Any type of physical disability, alternative coding	3.4%	7.2%
Any type of disability	20.4%	42.7%
Any type of disability, alternative coding	22.9%	48.9%

Annex 4 - Learning Tables - Midline Cohorts

In this annex, we provide additional tables documenting learning outcomes among the groups for whom this evaluation constitutes a midline. These include the original baseline cohorts – FE, ABE, and C1 NFE girls – as well as the C4 NFE cohort. The purpose of this annex is to document aspects of learning outcomes that are less relevant, analytically, during a midline, but which may be necessary for internal reporting or other purposes. For instance, we report results on subgroup learning outcomes at ML2 for the FE, ABE, and C1 NFE cohorts in this annex, while the main body of the report includes analysis of subgroup-specific *gains* learning over time. The latter is a measure of change in learning scores from BL to ML2 among subgroups, while the former is a measure of current learning scores, without reference to changes over time. Similarly, we report foundational skill gaps for all midline cohorts; while in the main body of the report we analyse subtask-specific changes in learning scores, the tables in this annex report foundational skill gaps in the cross-section (i.e. current skill gaps, without reference to how they have evolved over time).

The tables below report literacy and numeracy outcomes among specific subgroups of the midline cohorts. Membership in each subgroup is binary, and we report learning outcomes among the members of these subgroups, which can be compared to the "overall" results provided in the first row of the table. We refer to the FE, ABE, C1 NFE, and C4 NFE groups, respectively, as cohorts. For each cohort, we report the size (sample size) of the subgroup within that cohort and the mean learning score for that subgroup. We do not report aggregations across cohorts – e.g., results for all married girls, averaged across all three cohorts – because the three cohorts are so different. Averaging results across the cohorts makes little sense, given their differences in location, age, and intervention exposure. Note that panels within each table clearly delineate the type of subgroup to be reported, such as state, family characteristics, disability status, and so forth.

Subgroup Learning Outcomes

The tables below report subgroup-specific learning scores for the FE, ABE, C1 NFE, and C4 NFE cohorts, respectively (in consecutive tables). For each cohort, we report scores for Somali literacy, numeracy (the EGMA version with 8 subtasks), and the full numeracy score (using all 11 subtasks). Note that we do not restrict the sample to the BL-to-ML2 panel sample or the ML1-to-ML2 panel, because the tables do not report change over time, where consistency of the sample across rounds is essential. Instead, we report learning scores for all girls of a particular cohort who were interviewed at ML2. In the case of FE girls, this includes girls who were selected as replacements during ML1 or ML2, as well as girls who fell out of the sample at ML1 but were successfully re-contacted at ML2. For ABE and C1 NFE girls, the sample analysed here includes girls who fell out of the sample at ML1 but were re-contacted at ML2. For C4 NFE girls, the sample is exactly equivalent to the ML1-to-ML2 panel sample, as it includes every C4 NFE girl interviewed at ML2; because no replacements are selected for this cohort, this is exactly coterminous with the panel sample.

TABLE 113: SUBGROUP LEARNING SCORES AMONG FE COHORT, AT ML2

Subgroup	Sample Size	Somali Literacy	Numeracy (8 subtasks)	Numeracy (11 subtasks)
Overall	429	55.9	72.9	66.0
Geography				
Banadir	189	54.3	70.4	62.8
Jubaland	140	63.5	80.1	74.6

South West State	100	48.4	67.8	60.1
Disability Status				
Any physical disability	27	45.3	68.6	61.0
Any physical disability, alt. coding	32	45.7	69.7	61.8
Any cognitive disability	35	42.6	68.8	60.4
Any cognitive disability, alt. coding	48	45.9	68.2	60.2
Any mental health disability	99	52.7	70.7	62.4
Any mental health disability, alt. coding	188	54.4	72.0	64.3
Any non-mental health disability	57	43.6	69.1	61.1
Any non-mental health disability, alt. coding	73	45.2	68.7	60.7
Any disability	139	50.0	70.0	61.9
Any disability, alt. coding	222	52.5	71.1	63.4
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	98	54.3	73.2	65.2
HoH has no formal education	307	56.4	72.8	66.0
Caregiver has no education of any kind (no Quranic)	120	54.2	71.2	63.7
Caregiver has no formal education	323	56.7	73.3	66.5
Household Economic Characteristics				
HoH does not earn a regular wage	168	58.1	73.4	66.5
HH has a poor-quality roof	58	53.2	72.7	66.7
HH went to sleep hungry most nights, last 12 months	29	43.8	66.2	56.9
HH went without clean water most days, last 12 months	45	50.5	69.7	61.1
HH went without medicine most days, last 12 months	60	48.6	70.4	60.8
HH went without cash income most days, last 12 months	75	52.1	73.5	65.5
HH owns lands	147	58.3	73.8	67.3

Caregiver has savings of some form	22	58.1	73.8	66.8
HH owns a phone	362	56.9	73.7	66.8
HH owns a smartphone	97	61.9	76.0	68.9
Parental Circumstances				
Girl has only one living parent	48	56.4	73.7	66.5
Girl has no living parents	3	84.1	85.6	77.2
Girl does not live with either parent in her HH	22	63.3	72.3	65.4
Female-headed household	186	59.5	74.0	66.9
Linguistic Status				
Household speaks af-Maay	117	46.4	66.8	58.8

TABLE 114: SUBGROUP LEARNING SCORES AMONG ABE COHORT, AT ML2

Subgroup	Sample Size	Somali Literacy	Numeracy (8 subtasks)	Numeracy (11 subtasks)
Overall	310	46.4	69.6	60.5
Geography				
Banadir	134	42.7	65.5	56.8
Jubaland	82	46.2	74.6	65.3
South West State	94	52.1	71.1	61.6
Disability Status				
Any physical disability	26	48.1	73.9	62.2
Any physical disability, alt. coding	27	46.4	73.3	61.6
Any cognitive disability	29	41.6	64.1	53.6
Any cognitive disability, alt. coding	34	42.0	65.4	54.4
Any mental health disability	119	44.0	68.1	58.5
Any mental health disability, alt. coding	141	42.6	67.5	57.8
Any non-mental health disability	50	44.4	68.1	56.9
Any non-mental health disability, alt. coding	53	43.4	68.1	56.7
Any disability	143	43.2	68.4	58.4
Any disability, alt. coding	162	42.4	67.7	57.8
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	64	43.5	68.9	60.5
HoH has no formal education	258	46.2	70.6	61.2
Caregiver has no education of any kind (no Quranic)	85	46.4	70.3	62.1
Caregiver has no formal education	264	46.2	70.5	61.3
Household Economic Characteristics				
HoH does not earn a regular wage	101	47.2	70.9	63.0
HH has a poor-quality roof	72	39.2	66.6	55.8
HH went to sleep hungry most nights, last 12 months	20	38.6	59.4	52.3

HH went without clean water most days, last 12 months	19	38.0	73.7	64.4
HH went without medicine most days, last 12 months	46	38.9	66.6	56.9
HH went without cash income most days, last 12 months	59	46.0	68.6	59.4
HH owns lands	100	51.9	74.4	64.6
Caregiver has savings of some form	4	23.8	55.9	47.7
HH owns a phone	266	46.7	69.4	60.4
HH owns a smartphone	45	43.7	70.0	60.4
Parental Circumstances				
Girl has only one living parent	29	37.9	67.5	57.0
Girl has no living parents	4	26.1	52.9	42.3
Girl does not live with either parent in her HH	15	47.2	64.7	57.4
Female-headed household	104	46.4	67.8	58.7
Linguistic Status				
Household speaks af-Maay	99	47.3	70.7	60.3

TABLE 115: SUBGROUP LEARNING SCORES AMONG C1 NFE COHORT, AT ML2

Subgroup	Sample Size	Somali Literacy	Numeracy (8 subtasks)	Numeracy (11 subtasks)
Overall	302	48.1	67.6	58.1
Geography				
Banadir	143	52.3	69.2	59.3
Jubaland	76	44.2	70.6	61.5
South West State	83	44.6	61.9	53.2
Disability Status				
Any physical disability	21	43.4	65.0	56.5
Any physical disability, alt. coding	21	43.4	65.0	56.5
Any cognitive disability	31	37.1	61.4	50.3
Any cognitive disability, alt. coding	31	37.1	61.4	50.3
Any mental health disability	137	46.1	68.7	58.6
Any mental health disability, alt. coding	138	46.2	68.8	58.7
Any non-mental health disability	46	40.9	63.9	53.9
Any non-mental health disability, alt. coding	46	40.9	63.9	53.9
Any disability	156	45.4	67.0	57.1
Any disability, alt. coding	157	45.5	67.0	57.1
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	57	53.2	71.8	61.6
HoH has no formal education	250	47.2	67.6	58.1
Caregiver has no education of any kind (no Quranic)	75	52.5	70.6	61.3
Caregiver has no formal education	254	47.0	68.2	58.9
Household Economic Characteristics				
HoH does not earn a regular wage	123	49.6	67.6	58.4
HH has a poor-quality roof	61	39.6	65.0	54.6
HH went to sleep hungry most nights, last 12 months	15	40.2	62.2	50.4

HH went without clean water most days, last 12 months	12	46.8	61.9	53.7
HH went without medicine most days, last 12 months	57	47.0	62.0	53.9
HH went without cash income most days, last 12 months	33	48.0	66.7	56.5
HH owns lands	88	50.7	69.5	61.3
Caregiver has savings of some form	3	58.0	65.8	54.6
HH owns a phone	264	48.5	67.4	57.5
HH owns a smartphone	60	58.7	71.8	62.0
Parental Circumstances				
Girl has only one living parent	53	46.7	69.1	58.7
Girl has no living parents	4	54.3	52.5	49.3
Girl does not live with either parent in her HH	22	46.6	62.5	54.0
Female-headed household	137	45.1	66.4	56.2
Linguistic Status				
Household speaks af-Maay	85	43.2	65.9	56.9

TABLE 116: SUBGROUP LEARNING SCORES AMONG C4 NFE COHORT, AT ML2

Subgroup	Sample Size	Somali Literacy	Numeracy (8 subtasks)	Numeracy (11 subtasks)
Overall	661	34.9	59.1	49.6
Geography				
Banadir	304	35.5	62.5	52.6
South West State	249	34.6	57.6	48.9
Hirshabelle	108	33.8	53.0	42.9
Disability Status				
Any physical disability	49	27.1	54.7	44.3
Any physical disability, alt. coding	51	27.2	54.3	44.0
Any cognitive disability	67	23.5	48.9	38.9
Any cognitive disability, alt. coding	81	24.1	50.4	40.4
Any mental health disability	278	31.8	56.4	47.2
Any mental health disability, alt. coding	312	32.9	57.3	48.0
Any non-mental health disability	100	25.0	51.8	41.5
Any non-mental health disability, alt. coding	114	24.9	52.2	42.0
Any disability	318	31.8	56.2	46.7
Any disability, alt. coding	355	32.5	57.0	47.5
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	56	34.1	64.3	52.6
HoH has no formal education	199	37.9	61.4	51.4
Caregiver has no education of any kind (no Quranic)	53	32.3	64.4	52.9
Caregiver has no formal education	198	38.5	61.9	51.9
Household Economic Characteristics				
HoH does not earn a regular wage	99	41.1	59.8	50.3
HH has a poor-quality roof	151	23.1	52.0	42.7
HH went to sleep hungry most nights, last 12 months	51	32.4	51.0	41.7

HH went without clean water most days, last 12 months	59	29.3	51.8	42.2
HH went without medicine most days, last 12 months	95	33.0	55.0	45.2
HH went without cash income most days, last 12 months	139	26.7	54.5	44.8
HH owns lands	156	37.7	61.4	52.2
Caregiver has savings of some form	7	55.6	75.1	60.0
HH owns a phone	195	38.5	62.2	52.0
HH owns a smartphone	21	41.0	67.3	54.8
Parental Circumstances				
Girl has only one living parent	53	41.0	63.3	53.7
Girl has no living parents	4	43.1	56.9	45.9
Girl does not live with either parent in her HH	24	39.6	69.0	57.3
Female-headed household	169	38.5	61.3	51.1
Linguistic Status				
Household speaks af-Maay	234	30.4	56.0	47.3

Foundational Skill Gaps

The tables below report foundational skill gaps – the relative frequency of proficiency across subtasks – for each of the four pre-existing cohorts in the evaluation. As with the subgroup-specific learning scores in the previous section, the sample employed in these tables is the full set of girls interviewed during ML2 for a given cohort, without reference to whether she is part of a particular panel sample.

TABLE 117: FOUNDATIONAL SKILL GAPS IN NUMERACY, FE COHORT AT ML2

Subtask #	1	2	3	4	5	6	7	8	9	10	11
Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	1.2	1.6	10.7	6.1	16.6	13.5	22.8	7.5	35.7	31.5	28.4
Emergent Learner	0.2	0.9	77.2	2.8	7.0	4.2	21.4	2.1	15.4	9.6	21.7
Established Learner	0.7	7.2	9.8	10.7	28.2	9.3	28.0	15.4	27.0	17.7	37.8
Proficient Learner	97.9	90.2	2.3	80.4	48.3	73.0	27.7	75.1	21.9	41.3	12.1

TABLE 118: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, FE COHORT AT ML2

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	4.9	14.5	20.3	27.0	26.8	39.9
Emergent Learner	9.3	11.0	16.1	9.8	6.8	22.4
Established Learner	14.0	28.4	29.1	40.3	31.2	34.3
Proficient Learner	71.8	46.2	34.5	22.8	35.2	3.5

TABLE 119: FOUNDATIONAL SKILL GAPS IN NUMERACY, ABE COHORT AT ML2

Subtask #	1	2	3	4	5	6	7	8	9	10	11
Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	0.6	0.6	22.3	6.5	17.7	15.8	26.8	7.1	51.9	44.2	34.2
Emergent Learner	0.3	3.5	70.0	4.5	11.0	6.1	22.9	2.6	12.6	13.2	26.5
Established Learner	1.9	8.7	6.5	16.1	31.0	13.5	25.8	23.9	22.3	14.8	31.6
Proficient Learner	97.1	87.1	1.3	72.9	40.3	64.5	24.5	66.5	13.2	27.7	7.7

TABLE 120: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, ABE COHORT AT ML2

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	9.0	23.2	30.6	38.1	39.4	51.0
Emergent Learner	11.0	15.8	16.1	12.9	5.8	17.7
Established Learner	13.5	27.4	31.3	30.3	30.3	28.1
Proficient Learner	66.5	33.5	21.9	18.7	24.5	3.2

TABLE 121: FOUNDATIONAL SKILL GAPS IN NUMERACY, C1 NFE COHORT AT ML2

Subtask #	1	2	3	4	5	6	7	8	9	10	11
Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	4.0	4.3	23.2	12.9	24.8	19.2	30.8	10.6	55.6	47.7	37.1
Emergent Learner	1.0	2.6	73.5	2.3	8.6	3.0	22.5	1.3	13.6	11.9	23.8
Established Learner	1.3	7.9	2.3	7.9	22.2	10.6	25.5	15.9	19.9	18.9	32.1
Proficient Learner	93.7	85.1	1.0	76.8	44.4	67.2	21.2	72.2	10.9	21.5	7.0

TABLE 122: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, C1 NFE COHORT AT ML2

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	14.6	24.5	31.5	35.4	37.1	49.3
Emergent Learner	7.3	11.9	14.2	7.0	7.3	19.9
Established Learner	17.2	31.1	28.5	35.4	25.5	27.5
Proficient Learner	60.9	32.5	25.8	22.2	30.1	3.3

TABLE 123: FOUNDATIONAL SKILL GAPS IN NUMERACY, C4 NFE COHORT AT ML2

Subtask #	1	2	3	4	5	6	7	8	9	10	11
Subtask Description	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-Learner	4.8	6.2	28.3	21.6	37.4	33.6	45.7	16.6	69.9	63.2	45.1
Emergent Learner	1.8	3.6	66.3	3.5	11.0	4.8	19.5	2.7	8.0	11.2	22.8
Established Learner	3.6	16.9	3.8	12.0	21.6	11.0	17.4	22.8	11.5	11.5	26.5
Proficient Learner	89.7	73.2	1.7	62.9	30.0	50.5	17.4	57.8	10.6	14.1	5.6

TABLE 124: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, C4 NFE COHORT AT ML2

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	19.4	38.1	47.5	55.1	54.6	62.9
Emergent Learner	16.5	19.4	16.5	5.0	4.2	12.1
Established Learner	17.2	23.1	23.1	26.2	21.2	22.5
Proficient Learner	46.9	19.4	12.9	13.8	20.0	2.4

Annex 5 - Logframe

The project's logframe is included as a separate annex.

Annex 6 - Beneficiaries Tables

Project guidance

Please provide cumulative numbers of direct and indirect beneficiaries by cohort that the project has reached to date. The data for these tables should come from the Total Reach tab on your project's quarterly workplan tracker. Please report the **unique** total beneficiaries reached without double counting beneficiaries reached over multiple quarters. Do not simply add up reach totals across your quarterly workplan reporting.

Tables should include details for all cohorts and cohort types that the project has and is reaching–for example, formal track and nonformal track. Please add on row(s) as needed for disaggregating by cohort *type* and cohort *number*.

Table 6.1: Direct beneficiaries[127]

	Learners			HT/Teachers/other "educators"			MoE/District/ Govn't staff			Parents/ caregivers			Community members		
	Girls	Boys	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total
[Cohort 1, formal]	6,623		6,623	80	183	263	8 REOs/DEOs 8 Gender Focal Points	25REOs/DEOs	33[128] REOs/ DEOs 8[129] Gender Focal points	9,502 mothers		9,502 [130]mothers	431 CEC members 36 community health workers 423 female mentors 193 VSLA facilitators	724 CEC members 186 male mentor 43 VSLA facilitators	1,155[131] CEC members 36 community health workers 609[132] GEF/BEF mentors 236[133] VSLA facilitators
[Cohort 2, formal]	5,994		5,994												
[Cohort 3, formal]	6,479		6,479												
[Cohort 4, formal]	1,444		1,444												
[Cohort 5, formal]	1,405		1,405												
[Cohort 1,SNE]	388		388												
[Cohort 2, SNE]	676		676												
[Cohort 3, SNE]	112		112												
[Cohort 4, SNE]	68		68												
[Cohort 1, ABE]	7,241		7,241	34	82	116									
[Cohort 2, ABE]	6,035		6,035	75	93	168									
[Cohort 1, NFE]	6,604		6,604	88	80	168									
[Cohort 2, NFE]	4,319		4,319	78	26	104									
[Cohort 3, NFE]	4,198		4,198	66	24	90									
[Cohort 4, NFE]	13,439		13,439	175	108	283									
[Cohort 5, NFE]	12,942		12,942	21	17	38									
BEF peer mentors Cohort 1		600													
BEF peer mentors Cohort 4 NFE		950													
BEF peer mentors Cohort 4 formal		190													
Head teachers				11	88	99									
Qur'anic teachers				6	134	140									

[127] Cohort 6 NFE not included as it was enrolled after the evaluation study. Boys engaged in Boys' Empowerment Forums are listed separately as they are not learning beneficiaries (i.e. not reached through learning-focused activities, but just through youth-led action).
[128] MoE staff trained as ToTs and coaches for quarterly coaching of ABE/NFE facilitators
[129] MoE Gender Focal Points staff trained on Girls and Boys Empowerment concept.
[130] Parents of the enrolled girls participating in the VSLA groups
[131] Trained Community Education Committee Members (452 cohort 1, 23 cohort 2, 350 cohort 4, 145 cohort 5 and 185 cohort 6
[132] Trained GEF/BEF mentors (cohort 1; 60 F, 60 M; cohort 4; 283 F, 87 M; cohort 5; 80 F and 39 M)
[133] Trained VSLA facilitators (79 cohort 1, 91 cohort 4 and 66 cohort 5

Table 6.2: Indirect beneficiaries[134]

	Learners			HT/Teachers/other "educators"			MoE/District/ Govn't staff			Parents/ caregivers			Community members		
	Girls	Boys	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total
[Cohort 1-5, formal]	38,335	44,621	82,957												

[134] Total formal student beneficiaries extrapolated from the total enrolment in cohort 1 sample schools at the time of the baseline, extrapolated to the total number of formal schools, minus the students enrolled by AGES. The EMIS does not provide enrolment disaggregated by school; therefore, the enrolment data captured by evaluation was used to determine reach. Reach is calculated conservatively and does not consider impact on new entrants after the end of the intervention in formal schools.

Table 6.3: Direct beneficiaries by intervention/activity

Girls	Intervention/activity						Total (learning track)
	[GEF]	[VSLA]	[Apprenticeship]				
[Cohort 1, formal]							6,623
[Cohort 2, formal]							5,994
[Cohort 3, formal]							6,479
[Cohort 4, formal]							1,444
[Cohort 5, formal]							1,405
[Cohort 1,SNE]							388
[Cohort 2, SNE]							676
[Cohort 3, SNE]							112
[Cohort 4, SNE]							180
[Cohort 1, ABE] benefitting (in co-hosted schools) cohorts 1-3 formal and cohorts 1-3, NFE[135]	600						7,241
[Cohort 2, ABE]							6,035
[Cohort 1, NFE][136]		4250					6,604
[Cohort 2, NFE]		2475					4,319
[Cohort 3, NFE][137]							4,198
[Cohort 4, NFE]	2,830	2,199	586				13,439
[Cohort 5, NFE]	2,830	1,756	0[138]				12,942

[135] GEF was only planned for ABE cohort 1 with the assumption that they were to benefit cohort 2 ABE as well as the formal schools that are hosting the ABEs.

[136] VSLA was planned only for NFE students as this was meant to facilitate their transition to self-employment. ABE and FE were left out as their main transition pathway is formal education.

[137] VSLA was not planned for cohort 3 as this cohort was not in the original plan and was an adaptation due to the impact of COVID-19 on the program.

[138] There are 294 cohort 5 enrolled in apprenticeships between July and August 2023 not included in this table since the enrolment happened after the second midline evaluation.

Annex 7 - Evaluation Inception Report

The external evaluator's inception report is provided as a separate annex to this report.

Annex 8 - Data Collection Tools

The data collection tools – quantitative surveys and qualitative interview guides – are included as a separate annex.

Annex 9 - Qualitative Transcripts

The qualitative data, consisting of verbatim transcripts of all FGDs and KIIs, have been anonymized and provided to CARE's Monitoring & Evaluation team separate from this report.

Annex 10 - Datasets, Codebooks, and Programs

The quantitative datasets, a codebook of key variables, and the Stata .do files necessary for replicating the main results are provided in a separate .zip archive with this submission.

Annex 11 - Quantitative Sampling Framework

The sample frame for preexisting cohorts has been provided in earlier rounds of this evaluation. Specifically, the sample frame for FE girls recruited at baseline, including individual-level data on FE girls, was provided with the submission of the baseline report; the centre/school-level frame for selecting ABE and C1 NFE centres during the baseline and ML1 round were provided alongside the baseline and ML1 evaluation reports, respectively.

The sampling frame for the C5 NFE cohort, which was newly selected in this round of data collection, has been attached separately. This frame is at the individual level, including personally identifying information, and has been provided as a password-protected file to CARE's staff. This submission also includes the program/script (Stata .do file) for selecting C5 NFE centres and the NFE girls selected within each, allowing readers to replicate the sample draw precisely, if desired.

Annex 12 - External Evaluator Declaration

Name of Project: Adolescent Girls' Education in Somalia (AGES)

Name of External Evaluator: Brenton D. Peterson, Consillient Research

Contact Information for External Evaluator: brenton.peterson@consilientresearch.org

Names of all members of the evaluation team: Brenton D. Peterson

I, Brenton D. Peterson, certify that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received.

The following conditions apply to the data collection and analysis presented in the midline/endline report:

- Household surveys, learning assessments, head teacher surveys, classroom observations, classroom headcounts, and qualitative interviews were collected independently by the EE. No analytical data was provided by the project. Initials: BP
- The data analysis was conducted independently by the EE and provides a fair and consistent representation of findings. Initials: BP
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed. Initials: BP
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by Consilient Research. Initials: BP
- All child protection protocols and guidance have been followed. Initials: BP
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols. Initials: BP

Brenton D. Peterson .

(Name)

Consilient Research .

(Company)

November 4, 2023 .

(Date)

Annex 13 - Project Management Response

REFLECTIONS ON KEY findings – Longitudinally tracked cohort 1 participants (FCDO)

Cohort characteristics

The recontact rate stands precisely at 70.5%, matching the program's anticipated 70% target. For cohort 4, recontact rates varied between 70-78%. The likelihood of recontact was higher among the poorest, those experiencing hunger, and households of GwDs; *therefore, if anything, the midline results are skewed towards the worst off.*

Disability: The proportion of girls with physical disabilities among C5 is twice as high as in C4 (6% compared to 3%), while the proportion of girls with mental health disabilities has increased from 15% in C4 to 36% in C5. *The results confirm AGES' trend of progressively refining targeting towards the most marginalized in each cohort.*

The proportion of girls experiencing severe anxiety and/or depression has increased from 17% in 2022 (NFE cohort 4) to 42% (cohort 4) and 43% (cohort 5), highlighting the burden of the prolonged drought and escalating insecurity[139] on girls' wellbeing. Only 40% of the girls with mental health disabilities are enrolled in school compared to 58% of girls with other disabilities, illustrating the particular challenges posed by *mental health issues*.

Impact of the drought on living conditions: As expected, the NFE C5 participants are more affected by extreme hunger (12% sleeping hungry most or all nights) and extremely limited access to water (17% not having sufficient water for home use on most/ all days) than NFE C4 respondents, *reflecting the escalation of the drought.*

Learning

Impact of the drought on learning outcomes: Hunger, dropout, and absenteeism during the 2022 drought have taken a major toll on literacy learning, triggering an overall learning loss of 8 percentage points among girls originally enrolled in formal education. There were no learning losses in numeracy, but in average, progress has stalled among formal education students. South West – the area most affected by the drought – had the worst impact, with the lowest gains since the baseline. While results are significantly higher among girls who remained in school and were promoted year on year, they fail to reach the 2019 benchmark. *The results show how the 2022 drought has eroded learning gains, affecting cognitive development and school attendance.*

Learning patterns: Among formal education girls, the highest gains in literacy were, encouragingly, on reading comprehension, while the highest gains in numeracy were on number identification, quantity discrimination, and addition. Among the girls who stayed in school and were promoted year on year, the highest gains were observed in intermediary level tasks – reading fluency, lower levels of reading comprehension, addition, and subtraction. Among ABE students, the highest gains were on reading fluency and subtraction.

[139] While the security situation in targeted areas has improved since April 2023, the period of data collection was marked by major attacks, including one attack in Mogadishu which killed an AGES student.

Post-ABE learning gains: ABE girls who report, at ML2, that they are enrolled in formal school, have literacy scores at 25.2 points higher than at BL and numeracy scores are 8.5 points higher than at BL, far surpassing the average results of 7 and 4 pp, respectively. *The result suggests that those who transitioned into formal school were well prepared to engage with the primary curriculum.*

Post-NFE learning vs employment: 62% of the C1 NFE girls have since transitioned into employment or self-employment, but also experienced a loss of 9.1 pp in literacy scores since the baseline, and a decline of 6.3pp in numeracy. While those losses are significantly lower than among girls who stayed idle, they suggest that participants are not actively using advanced literacy or numeracy skills in daily life.

Transitions: Over 47% of the ABE students have continued their education, while 25% are employed/ self-employed. Among NFE students, 62% have transitioned into employment or self-employment, while 20% have returned to school. In both cases, the high proportion of older adolescents/ female youth returning to school suggests a major shift in gender norms and household support of education for older girls.

Positive shifts in attitudes towards girls' education: The proportion of caregivers who believe it is acceptable for a girl not to attend school due to the cost of education has fallen from 62.9% to 44.6%. The decline was particularly high in *male headed households – from 67% to 45%.* The decline was also higher among IDPs – from 64% to 46% - and among households facing hunger – from 65% to 40% - thus showing social norms change among the most vulnerable segments of the population. The prioritization of girls' chores has also declined: the proportion of caregivers who believe it is acceptable to withdraw a girl from school due to work decreased from 25.2% to 16.3%. Nonetheless, the proportion of girls spending time on chores has increased for both formal education students and ABE graduates, potentially due to the drought (additional time spent fetching water, caring for the sick), poverty, and girls becoming older. The largest increases in the proportion of households reporting that girls' attendance is affected by chores was observed among IDPs and in male-headed households.

Teaching and school management practices

Formal education, teaching practices: Girls are generally being favored in class, with teachers calling on them more often, providing positive feedback more frequently and using harsh language less often with girls than boys. Nonetheless, a larger proportion of teachers are now providing positive feedback to both boys and girls.

There is *a major decline on the use of corporal punishment*, both in classroom observations (less reliable finding due to potential desirability bias) and in the girl survey (more reliable source), with the proportion of girls affirming that their teacher uses corporal punishment declining by 39 percentage points. The proportion of CECs engaged in child protection activities – potentially linked to the reduction of corporal punishment – has increased by 44 percentage points. Overall, 82% of the girls affirm that they are able to report cases of abuse at school, and 78% are able to report cases of abuse at the community. Both findings confirm monitoring trends. Nonetheless, physical discipline still exists and remains a major barrier for the affected students, particularly for low performers.

The reported use of formative assessments has increased substantially, but the actual proportion of teachers who can demonstrate their use remains low (59%). The finding suggests that a sizable number of low performing students are not being identified or receiving timely support, resulting in learning losses over time (as noted above). The finding highlights the need for public investment in this area and in remedial education, and lends support to the adoption of approaches such as Teaching at the Right Level (included in the upcoming Somalia Partnership Compact). Conversely, there is evidence that *teachers are increasingly adopting more inclusive practices*, such as engaging students who are not participating, with the proportion of teachers doing so jumping from 47% to 81%. This is corroborated by students, with an *increase of 24 percentage points in the proportion of girls who affirm that their teacher often encourages*

participation. The use of traditional teaching practices (copying from the board, repeating after the teacher) has decreased.

Attendance & enrolment: In the targeted formal schools, the share of female students in the classroom has risen to be nearly at parity with boys, from 39% at BL to over 49% now. Girls' enrolment has increased by 91%, compared to 79% among boys. Enrolment has soared in South West (a 149% increase, potentially reflecting displacement towards urban areas); encouraging, dropout rates have also declined by 77% among girls in this area. Among the longitudinally tracked formal education students, 75% have remained in school.

Despite the increased enrolment, girls' attendance rate remains lower than boys – 84% compared to 87% - and has declined since the baseline (from 87% to 84%) while boys' attendance has remained the same. Nonetheless, the presence of an active CEC - particularly where providing specific support to girls – was found to drive attendance rates up, as well as the presence of female teachers. *Schools with higher proportions of students with disabilities also had higher attendance rates*, particularly where there was a larger proportion of students with mobility-related disabilities. The finding suggests that *more inclusive school environments have an overall positive impact on attendance rates*.

The quality of the attendance records has increased over time in Banadir and South West, but not in Jubaland.

External financial support: Overall, 41% of the caregivers reported that their daughter is receiving a scholarship, cash transfer, or other form of financial assistance (not from the project). This is a critical finding as 53% of the schools are charging fees (up from 42% in 2022, potentially reflecting the decline in financial conditions and the end of capitation grants). Private contributions are also crucial to retain teachers, with CECs contributing about 28% of the total teacher salaries.

Social-emotional learning

Leadership skills: Leadership scores have significantly increased – between 14 and 20 percentage points depending on the learning track – with qualitative data highlighting changes in girls' self-confidence and positive perceptions of their own abilities and gender rights. Qualitative data has also linked this to increased participation in entrepreneurship and the ability to voice their opinions.

The massive shocks experienced in 2022 have eroded girls' perceptions of self-efficacy; not surprisingly, the proportion of girls who felt that they lacked control over the events in their lives increased from 41% to 58%. Nonetheless, a surprisingly high proportion of girls believe that they are able to determine their own lives – increasing from 67% to 73%, in a strong demonstration of Somali girls' resilience and life skills.

GEF sustainability: 14% of all girls are still participating in GEFs, out of the 26% who were originally members of a group.[140] Additionally, 28% of the students / former students are members of a youth network, out of whom 68% are not GEF members. Therefore, 14% of the girls are still members of a GEF, while an additional 19% are also members of a youth network and 9% participate in both GEFs and youth networks.

GEF activities are mostly focused on awareness raising (41%) and preventing early marriage (11%).

Economic empowerment

Income and employment: Overall, the average income for girls participating in non-formal education has nearly doubled (96%) while the average income for minority girls has increased 400%. Although there is

[140] 26% of all girls participated in GEFs, out of whom 54% remain active members.

still a gap between the minority girls' income and the general average income of all participants, it has decreased by half in a year's time: at the first midline, minority girls earned 35% of the average income, and now earn 73% of the average income. The proportion of girls in entrepreneurship increased from 8.4% to 13.4%.

VSLA: 27% of the C1 NFE girls are reportedly VSLA members. The proportion of girls and caregivers who report that their group is still active remains similar to the midline, despite the major shocks triggered by the drought. Overall, 69% of the caregivers who are part of a VSLA are still using the savings to support girls' education.

Menstrual hygiene practices

The proportion of girls unable to attend school during menstruation has increased over time, with girls reporting pain and discomfort as the main reason. This may be related to an increased proportion of girls experiencing infections due to the combination of type III FGM and limited availability of water. Conversely, however, menstrual hygiene practices have significantly improved, with the proportion of girls reusing single use pads declining by 16 percentage points and the proportion changing pads less than three times a day reduced by over 10 percentage points.

REFLECTIONS ON KEY FINDINGS – LONGITUDINALLY TRACKED NFE COHORT 4 (USAID)

Learning

C4 NFE participants had major gains in both literacy and numeracy – 18 and 21 percentage points, respectively – far surpassing the performance of the first NFE cohort. The results are a testament to the refinement of approaches to cater to the needs of marginalized students, including the emphasis on remedial education, use of remote learning materials, and multilingual strategies. In the case of literacy, major gains were observed in reading comprehension – a difference of 20 percentage points. In the case of numeracy, the largest gains happened in mid-level tasks (word problems, addition, subtraction) but also, to a lesser extent, among more advanced tasks (multiplication and division). Nonetheless, the performance in the logical reasoning task of identifying missing numbers remained the same, indicating the need for further investment in teacher training in this area.

The highest literacy gains were observed among the most marginalized – girls with disabilities (22 percentage points), female-headed households (24 pp), those who slept hungry many/ most days (22pp), and those without cash income most days (23pp). The results confirm the effectiveness of AGES' targeted approach to support the worst-off girls.

Social-emotional learning and learning outcomes

The participation in GEFs is a strong predictor of improved learning outcomes, predicting a 15-percentage point increase in literacy scores and a 16.5 percentage point increase in numeracy. The participation in GEFs predicts a 12-percentage point increase in the Youth Leadership Index score.

Overall, C4 participants had a 17-percentage point increase in their average Youth Leadership Index score.

Transitions and economic empowerment

Overall, 30% of the C4 NFE graduates are now in employment or self-employment, while 8% have transitioned into formal school and 37% have continued in other forms of education. The highest transition rates were observed in Middle Shabelle, followed by Lower Shabelle, and among the most marginalized: girls who lacked consistent access to water, who experienced hunger, and those living with physical disabilities.

Income: Has increased by 295%.

Safeguarding

Girls' reported access to means of reporting abuse cases at community level has declined – from 82% to 73% - but while previously the main reporting channels were only the head teacher and teacher, girls are now also referring to the police. This may indicate greater confidence in asserting rights and ability to use local services.

Civic engagement

Cohort 4 participants are engaging in broader forms of civic action beyond GEFs. In Banadir, 30% of C4 girls have reported participation in dialogues related to local governance; 37% are involved in youth groups; and 42% involved in discussions to improve service provision.

RESPONSE TO RECOMMENDATIONS

Learning Loss and Continuing Education

As Section 4.2 showed, learning scores decline – sometimes sharply – as girls leave school or after they have completed ABE or NFE programmes. Regression in literacy and numeracy reduces the long-term value of the programme's impacts and may reduce the extent to which beneficiaries believe education is worthwhile. A light-touch continuing education intervention for programme graduates, with a focus on application of, and a refresher on, the literacy and numeracy skills they gained, could reduce this documented learning loss. Continuing education could be delivered virtually or remotely; alternatively, in-person sessions, paired with continued engagement with GEFs (mentioned below) could be an effective way to both combat learning loss and maintain networks among the girls and women.

Project response: The project agrees with the need to integrate measures to mitigate learning losses in future designs. This is an ex-post evaluation for Cohort 1, thus making it difficult to implement new components due to funding limitations.

An analysis of the learning losses by bands of achievement and specific subtasks shows that those have specifically resulted in a reduction of the proportion of participants who were 'proficient' (scoring 81% and above) in more advanced tasks: two-digit subtraction, multiplication, passage reading, and reading comprehension. The proportion of participants scoring 41%-80% ("established" learners) has increased for all of those, suggesting that participants are lost some degree of fluency in those skills.

There are two potential explanations for the learning losses, and the trend is probably the result of a combination of both: (i) the limited utilisation of the skills learnt, particularly among those who have not transitioned into further education/ business and (ii) the impact of malnutrition and mental health issues. A substantial learning loss of over 8 percentage points is also observed among formal education students, among whom 75% continued to attend class; therefore, learning losses are explained by more than non-utilisation of skills.

It is likely that a successful strategy to maintain foundational skills would involve incorporating those in specific follow-up training on financial literacy, access to services, business management, etc – thus aligning it with the interests of older girls and the utilisation of those skills in daily tasks.

Maintenance of GEFs and Female Social Networks

The evaluation also documented a reduction in the connection between girls and the GEFs in which they participated over time. To some degree, reduced engagement with one's GEF is likely inevitable as girls shift into non-education activities, move away from the area, or otherwise change their focus. The programme should make an effort to promote continued connection between girls who participated in the GEFs together – into adulthood, if possible. These networks

could continue to provide a valuable space for women to discuss problems they face, a support system, and a source of continued empowerment. Facilitating occasional meetings of a cohort of GEF participants after they have finished their programmes or left school would be a relatively low-cost method to build women's social networks and – potentially – create a group of adult women with ties to a particular school and an interest in seeing it succeed, similar to alumni networks elsewhere.

Project response: The project has already incorporated a similar approach in its design for cohorts 4-6, which involves (i) coalescing GEFs into district-level networks, intending to facilitate connections and activities beyond one's own school/ neighborhood and higher level advocacy, which may contribute to long-term participation; and (ii) near-total overlay of VSLAs and GEFs. Since VSLAs cater to more "adult" economic interests, there is a likelihood of a more persistent connection between peers. Alumni networks may be difficult to maintain in a context of high mobility/ displacement, and lack the economic/ social support appeal of VSLAs and GEFs.

Girls' Representation and Participation in CECs

According to the analysis in Section 6, a small but meaningful proportion of GEF members participate and engage with their CECs in some form. It is clear there is extensive overlap between the activities of GEFs and CECs, especially around awareness-raising, promoting enrolment, and reaching out to students who have dropped out or stopped attending school consistently. The programme should promote more active engagement of GEF members (and girls more generally) with their CEC. This could take the form of each CEC having a student member or representative that attends meetings, has input into some of the decisions made, etc. This model of shared governance would provide girls a voice in the CEC; it would also provide valuable insight into how the CEC operates, the issues and constraints they face, and how institutions in their community work. Governance models that incorporate students are an important method of socialization or professionalization and could also serve as networking and mentorship opportunities by successful/responsible adult community members.

Coordination and Collaboration between CECs and GEFs

A more systematic relationship between GEFs – or a select set of their members – and CECs could also be maintained post-matriculation. In line with the previous recommendation to help girls maintain ties to their GEFs, building a relationship between the GEFs and CECs oriented around joint awareness-raising and mobilization campaigns could build long-term connections between girls and the school they attended.

Project response: The project agrees with the recommendation. Stronger linkages between CECs and GEFs have already been incorporated in the Cohort 4-6 GEFs. For future designs, the GEF-CEC link could be further expanded to include a formal 'junior membership' of CECs in formal schools, similar to CARE's approach to girls' participation in school shuras in Afghanistan (under community-based education projects such as PACE-A, STAGES, STAGES II, and EEA-III).

Economic Support for Households

The worsening economic situation in southern Somalia is reflected in measures of household well-being among the C4 and C5 NFE cohorts. While the programme already promotes economic self-sufficiency through VSLAs and other activities, this may be insufficient to maintain enrolment in NFE programmes, especially when one considers the, arguably, higher opportunity cost – in terms of household or paid labour – incurred by the older girls targeted for NFE enrolment. If resource constraints allow, the programme should consider providing a small, potentially targeted, short-term cash transfer or other economic benefit that can ease the way for girls to continue in NFE programmes.

Project response: For cohorts 4-6, the project is already facilitating linkages between humanitarian aid (particularly cash and in-kind support) and NFE participants. The project has elected to facilitate linkages and support participants with VSLAs (with a small cash transfer to VSLAs to enable the implementation of business plans) instead of potentially duplicating humanitarian aid.

While the project agrees that there is a need for further direct economic support to ultra-marginalized girls, particularly those who are married/ young mothers, dropout is also being driven by other factors. Overall, 79% of the C4 NFE participants completed the course. A major share of the dropout - 41% - took place in Hirshabelle, corresponding to a dropout rate of 43% (compared to 12% in Banaadir and 22% in South West). The majority of the dropouts left the program between August-October 2022, coinciding with the military conflict in Hirshabelle.

In addition, the project notes the need to go beyond short-term cash assistance and further expand access to long-term economic empowerment solutions – VSLAs, skills training, market linkages. Through additional USAID support, the project is expanding those components, particularly apprenticeships, to boost the likelihood of retention and positive transitions.

School Fees and Increased Economic Support

The evaluation data suggest that there has been a small-to-moderate increase in the share of schools that charge enrolment or tuition fees for attendance. To the extent that schools have increased or begun charging fees, this may need to be reflected in additional bursary or other economic support.

Project response: AGES does not provide bursaries or scholarships to support formal school attendance. This evaluation is an ex-post evaluation for Cohort 1. The project's agreement with the schools hosting C1 students involved the provision of infrastructure improvements, teacher salaries, and technical assistance to schools in exchange for free enrolment in 2019-2022 for enrolled students. The VSLA component with parents was intended to enable ongoing enrolment of participants after the project had been completed. The finding is, however, valuable to shed light on the evolution of economic barriers to access and retention and should be used for advocacy with the MOECHE. It is likely that the increase proportion of schools charging fees is occurring as a result of the end of the ESPIG capitation grants, which supported primary school costs in 2021-22. Additional grants are currently planned under the upcoming Somalia Partnership Compact and World Bank-led Human Capital Development project, which can be informed by the AGES findings.

Sustaining Institutions after the Programme's End

As the programme approaches its end, its sustainability is of primary concern. The programme should focus on ensuring increased adoption and effectiveness of interventions related to strengthening economic and financial conditions of programme households. Participation in VSLAs should be an area of focus, together with increased support to GEF activities, potentially also through financial and technical assistance. A specific assessment could shed further light on the underlying barriers to maintaining female networks, like the GEFs, to identify entry points for future programming.

Project response: The project agrees with the recommendations. For C4-6, the project has increased substantially the provision of technical assistance to VSLAs and GEF (through ongoing coaching informed by gaps identified during quarterly monitoring); the results are clearly reflected on the additional gains for C4, compared to C1. The project has also increased the coverage, linkages, and financial assistance to VSLAs and GEFs (see response above on coalescing GEFs into district-level networks).

CEC Shift Toward Monitoring of Instructional Quality

In part as a result of resource constraints, CECs are often limited in what they can accomplish. However, within the realm of tasks that require investment of time rather than money – such as enrolment mobilization, monitoring attendance, and monitoring teaching quality – CECs tend to be heavily focused on maintaining or increasing student enrolment and attendance (e.g., following up with students who are absent, encouraging enrolment, etc.). Fewer CECs monitor the quality of teaching, for instance, implicitly prioritising the quantity of a student's time spent in school, rather than the quality or learning value of that time. The programme should make a concerted effort to promote CEC monitoring of teacher attendance, teaching quality, and other aspects of school functioning critical to learning outcomes, such as the application of formative assessments and strategies for teaching to "the right level." Recognising

that most CEC members are not education experts, providing them with checklists or simple rubrics for assessing a teacher's lesson planning, use of formative assessments, and so forth, could encourage them to monitor these activities more closely and build a culture of closer oversight over non-student aspects of the school.

Project response: The project agrees that CECs can monitor simple measurements of teaching quality – including the use of formative assessments, positive discipline, and teacher support of non-learners, for instance. Some investments in these areas have been made for C4, particularly on safeguarding and following on non-learners/ GwDs. However, it is unlikely that CECs will be able to effectively monitor if teachers are 'teaching at the right level' – particularly in areas where most CEC members are illiterate themselves. This is a role the project has worked with MOE coaches to provide oversight and support to, liaising with CECs.

Tailoring and Targeting NFE Instructional Levels

Analysis of literacy and numeracy for Cohort 4 NFE girls reveal that there was substantial regional variation in learning outcomes at the outset of their programme (in ML1), with girls in Bay achieving much higher pre-programme learning scores. This variation seems to have impacted the benefits of the programme, as girls in Bay show no tangible gains in learning one year later, while girls in other regions have caught up or surpassed them. This could be a reflection of teaching materials and a programme focus that was ill-targeted for those girls who were comparatively high-achieving at the outset of their learning programme. Developing NFE materials appropriate for girls at perhaps three different "starting points" could allow NFE centres that recruit a high- or low-starting point cohort to adjust their level up or down in response, and generate greater gains in learning.

Project response: The project agrees with the recommendation, while noting that the lack of greater learning gains in Bay is also likely to be associated with greater deprivation of basic needs. The actual ML2 mean literacy scores in Bay were lower than those achieved in other regions (Lower Shabelle, Banaadir) suggesting that there was room for improvement, even if at a lower scale.

The recommendation of 'NFE levels' is also consistent with observations from C1 (where a larger proportion of participants started at higher learning levels). This could be piloted for C6, adapting the current approach to remote learning. That being said, there are potential risks associated with (1) limited capacity of facilitators to handle the additional complexity of multiple learning levels; and (2) enrolling 'better off', younger participants (typically easier to enroll than the ultra-marginalised targeted by AGES).

Promoting Community Prioritisation of Girls' Education

Parental attitudes toward household work continue to be an important determinant of attendance rates. Girls whose caregivers see greater value of schooling relative to a child's other responsibilities or the family's other financial obligations attend learning programmes more consistently. The programme has made important inroads in increasing this type of community support for girls' education, and this progress can be maintained and deepened through relatively low-cost awareness campaigns as the programme winds down.

Project response: While awareness campaigns tend to reduce resistance to girls' education in general, this typically has a higher impact on enrolment vs attendance. The latter is determined by more complex dynamics at the household, including mothers' ability to distribute household workload (which may be limited in female-headed households with a larger number of younger children, or for divorced girls with younger children) and drought-related constraints (in conditions where access to water is extremely limited and disease outbreaks are common, participants may prioritize chores such as fetching water and caring for the sick as a matter of survival). Prioritizing attendance vs household chores often depends on the combination of ongoing gender dialogues to shift norms (in platforms such as VSLAs, CECs) and a shift in economic conditions, particularly for those facing extreme drought. The project agrees that there is a need for further action to shift gender norms – dialogues, working with religious leaders – but also that it should be synchronized with economic empowerment activities and expanding linkages with humanitarian aid to address other causes of absenteeism.

Sharing Teacher Experience, Expertise, and Materials

The report has documented progress in many aspects of teaching practices. This increased uptake and implementation of inclusive and participatory teaching styles could be maintained and spread through teacher knowledge-sharing platforms similar to those in use in many jurisdictions (e.g., platforms that facilitate peer-to-peer dissemination of lesson plans). Other knowledge-sharing methods could include forums for teacher discussion or training implemented by effective teachers, though these approaches would be more costly and are contingent on accurately identifying high-quality teachers to lead them.

Project response: The project agrees with the recommendation. There are good examples of knowledge-sharing strategies used by SOMGEP-T and by the AGES team during COVID, which could be leveraged to further expand support to teachers.

Targeted Focus in Reducing Use of Corporal Punishment

The programme has made strides in reducing the use of corporal punishment in schools, as shown in this report and the ML1 evaluation previously. Based on arguably more-reliable student reports of corporal punishment use in classrooms, its use is relatively concentrated in a handful of schools. Just 17 of 46 C4 NFE centres have more than one student who reported the use of corporal punishment in the last week, but several of these centres had many students (as many as 50 percent or more of the girls interviewed) report it. The programme should consider targeting its efforts to reduce corporal punishment in particular schools where the problem is most pronounced, in order to maximise the programme's impact in a short time period. Students could be cost-effectively surveyed if the assessment was focused exclusively on identifying the use of corporal punishment, which would allow targeting beyond the sample of centres and schools included in the evaluation data here.

Project response: The project agrees with the recommendation and has already operationalized it through quarterly surveys with girls (using cross-sectional, representative samples). The project is also investing in capacity building of MOECHE/ State MOEs staff as investigators for cases of abuse/ SEAH and supporting the development of a policy on the code of conduct, thus creating the capacity for official responses to cases.

More Directly Assessing Sustainability at Endline

The programme's upcoming endline evaluation presents an opportunity to assess the sustainability of many of the institutional structures that were established as part of the programme, especially GEFs and VSLAs. This evaluation and other studies of Somali education have documented the valuable contributions of GEFs to a range of learning and socio-emotional outcomes, and sustaining these groups beyond the life of the programme should be a priority. During the endline, the evaluation team should develop more explicit measures of continued activity and engagement by GEF and VSLA participants; they should also engage with the programme team to develop measures of other aspects of sustainability that are not currently captured in data collection instruments designed at baseline and maintained through the following two rounds of data collection.

Project response: There is room for further review of the GEFs' own perceptions of sustainability and benefits – for example, using the networking tools applied by SOMGEP-T and/or assessing GEF networks' achievements and continued engagement of Cohort 1-4 GEFs vs Cohort 5-6 GEFs. The evaluation will be an ex-post assessment of three GEF cohorts – 1, 4, and 5 – thus lending itself for a thorough assessment of sustainability of individual and collective benefits and actions.

Annex 14 - Learning and Transition Beneficiaries

In this Annex, we estimate the number of learning and transition beneficiaries of the AGES programme. It is important to note that we estimate these numbers for the FCDO-supported cohorts only – those girls

who were enrolled into formal school, ABE programmes, or NFE programmes in late 2019. Moreover, given the important demographic differences between these groups, we provide estimates for each cohort separately, and then provide an aggregate estimate, to ensure the FM has the information needed for annual reporting.

Learning

Learning beneficiaries are calculated using the first line of evidence, "improvement in mean score over a prior timepoint". The prior timepoint we employ is the baseline. For this calculation, we compare literacy and numeracy scores at ML2 to BL, for each of the three tracked cohorts. We restrict our analysis to the successfully tracked sample of girls (i.e. the panel sample).

We do not employ a benchmarked comparison because of the effects of COVID-19. Benchmarks for learning for the FE and ABE cohorts were established at baseline, at a time when girls serving as benchmarks had not been exposed to education during the COVID era. This group experienced fewer school interruptions during the equivalent years of education (Grade 1-4). We do not view the benchmarks established prior to COVID-19 to be a particularly effective standard against which to assess learning progress.

Table 14.1 below simply indicates the line of evidence we use for assessing progress on learning outcomes. In Table 14.2, we report the number of girls achieving learning improvements and the progress toward the target, broken down by learning type (literacy versus numeracy) and cohort.

TABLE 125: LINES OF EVIDENCE EMPLOYED FOR ASSESSING LEARNING OUTCOMES

Evidence line	Description	Comments	Calculating # girls	Calculating project progress for AR	Are you using this evidence line?
1	Improvement in mean score over prior timepoint.	No comparison group	All girls in intervention group counted if difference over prior timepoint is statistically significant. No girls in intervention group counted if not stat. sig	Project meets target if mean score comparison over prior timepoint is stat. sig.	Yes
2	Fewer non-learners	Can use mean score, subtasks of focus Non-learner status must be meaningfully defined (1 word read isn't reading!)	Tracked sample: # girls who meet criteria vs. prior timepoint; non-tracked sample: extrapolated from % reduction across samples	Project meets target if reduction is stat. sig.	No
3	More girls meeting a benchmark	Benchmark must represent a meaningful level of	Same as evidence line #2	Same as evidence line #2	No

		attainment for the girls			
4	Girls mastering a skill	Based on project intervention focus, girls achieving a meaningful score on a skill or task of focus	Same as evidence line #2	Same as evidence line #2	No

The tables below are separated by the learning outcome analysed: Table 14.2A analyses literacy; Table 14.2B analyses numeracy; and Table 14.2C analyses girls' combined (mean) score in literacy and numeracy.[141] In column 2 of each table, we report the coefficient and p-value for a regression that assesses the improvement in learning scores from baseline to midline #2; we also report whether the change from baseline to midline #2 is statistically significant. In column 3, we indicate the percentage and number of girls, in the evaluation sample, meet the target, in line with the criteria in Table 14.1. If improvement since baseline is statistically significant, all girls in the sample are counted.

In column 4, we extrapolate from the evaluation sample to the population. As before, a cohort's attainment of the target is binary, all-or-nothing. For instance, all 13,590 FE girls are counted as meeting the literacy target, because the target was met among the evaluation sample.

We calculate the size of each cohort's population based on cohort sizes present at the baseline. At that time, the programme had enrolled 6,623 FE girls, 7,241 ABE girls, and 6,604 NFE girls (20,468 girls in total). The programme estimated eventual reach of 42,000 girls. The cohort-specific numbers we use in column 4 are derived from distributing the 42,000 total girls targeted across cohorts in line with their distribution at the baseline. For instance, FE girls made up 32.4% of girls enrolled at baseline (6,623 or 20,468); therefore, we assume they made up 32.4% of the 42,000 total girls targeted, or 13,590 FE girls. These assumptions are used to extrapolate to the overall population for the FCDO-supported AGES cohorts.

TABLE 126: LEARNING (LITERACY) OUTCOME REPORTING

Indicator(s)	Calculation for # girls learning	# girls learning - target	# girls learning - actual
FE Cohort Evidence Line 1	+20.5 point change (coefficient); p-value < 0.001 Statistically significant	100% (n = 270; all girls in evaluation sample)	13,590
ABE Cohort Evidence Line 1	+6.8 point change (coefficient); p-value = 0.03 Statistically significant	100% (n = 310; all girls in evaluation sample)	14,858
NFE Cohort (C1 NFE) Evidence Line 1	-6.5 point change (coefficient); p-value > 0.05 Not statistically significant	0% (n = 0; no girls in evaluation sample meet criteria)	0

[141] To be clear, the combined score is, for each girl, the mean of their literacy score (on a 100-point scale) and their numeracy score (on a 100-point scale).

Total	65.8% across 3 cohorts (n = 580 of 882 in evaluation sample)	28,449

TABLE 127: LEARNING (NUMERACY) OUTCOME REPORTING

Indicator(s)	Calculation for # girls learning	# girls learning - target	# girls learning - actual
FE Cohort Evidence Line 1	+15.3 point change (coefficient); p-value < 0.001 Statistically significant	100% (n = 270; all girls in evaluation sample)	13,590
ABE Cohort Evidence Line 1	+4.3 point change (coefficient); p-value = 0.053 Not statistically significant	0% (n = 0; no girls in evaluation sample meet criteria)	0
NFE Cohort (C1 NFE) Evidence Line 1	-6.0 point change (coefficient); p-value > 0.05 Not statistically significant	0% (n = 0; no girls in evaluation sample meet criteria)	0
Total		30.6% across 3 cohorts (n = 270 of 882 in evaluation sample)	13,590

TABLE 128: LEARNING (AGGREGATE LITERACY AND NUMERACY) OUTCOME REPORTING

Indicator(s)	Calculation for # girls learning	# girls learning - target	# girls learning - actual
FE Cohort Evidence Line 1	+17.9 point change (coefficient); p-value < 0.001	100% (n = 270; all girls in evaluation sample)	13,590
ABE Cohort Evidence Line 1	+5.5 point change (coefficient); p-value = 0.02	100% (n = 310; all girls in evaluation sample)	14,858
NFE Cohort (C1 NFE) Evidence Line 1	-6.3 point change (coefficient); p-value > 0.05	0% (n = 0; no girls in evaluation sample meet criteria)	0
Total		65.8% across 3 cohorts (n = 580 of 882 in evaluation sample)	28,449

Transition

To assess the number of girls benefitting in terms of transition, we calculate the share of girls who have successfully transitioned between baseline and midline #2, within the evaluation sample. We then extrapolate this value to the overall population targeted by the programme.

Note that we cannot reasonably compare transition outcomes at midline #2 to transition outcomes at baseline, because all girls, by definition, achieved successful transition at baseline. At that time, all targeted girls were shifted from being out-of-school to being enrolled in FE, ABE, or NFE programmes. Even if the programme achieved a transition rate of 99% between baseline and midline #2, this would be lower than the transition rate at baseline (of 100%).

Our approach recognizes this fact and defines a girl as benefitting from transition if she personally achieved successful transition – in all its forms – between baseline and midline #2. In Table 14.3, below, we report transition beneficiaries among each cohort, with FE girls reported in the first panel, followed by ABE girls and NFE girls. In the table, column 1 describes the transition pathway; columns 3 and 4 provide the percentage and number of girls achieving each transition pathway among the evaluation sample. Note that we only report on successful transition pathways, as our interest is in calculating the share of girls in the sample who achieved success, such that we can extrapolate to the broader population.

The right-most column provides this extrapolation to the overall target beneficiaries (as established at baseline for the FCDO-supported AGES cohorts) of 13,590 FE girls, 14,585 ABE girls, and 13,551 NFE girls, respectively. Using the ABE cohort to illustrate: 18.4 percent of ABE girls transitioned into formal school at the time of midline #2; applying this value to the overall population of ABE girls targeted, we expect 2,734 ABE girls (out of 14,585 targeted in total) to have this outcome. Likewise, 34.5 percent of ABE girls were employed or self-employed (and were of appropriate age to be in employment), which extrapolates to 5,126 total ABE girls following this pathway.

The final row in the table aggregates the total transition beneficiaries across the three cohorts. Of 42,000 girls targeted across the three cohorts, we estimate that AGES produced successful transition outcomes, at midline #2, for 26,533 girls.

TABLE 129: TRANSITION OUTCOME REPORTING, BY COHORT

Pathways[142]	Indicator(s) Disaggregated for transition pathways	Calculation for # girls transitioning	# girls transition- target	# girls transitioning - actual
FE Girls				
Progression through school		69.3%	187 (out of evaluation sample of 270)	9418
Transition into skills or vocational training		0.0%	0 (out of evaluation sample of 270)	0
Total		**69.3%**	**187**	**9418**
ABE Girls				
Transition into formal school		18.4%	57 (out of evaluation sample of 310)	2734
Transition into skills or vocational training		0.0%	0 (out of evaluation sample of 310)	0
Transition into work or self-		34.5%	107 (out of evaluation sample of 310)	5126

[142] Individual project transition pathways may be slightly different than the ones mentioned in this table. Please adapt as applicable for you project and liaise with the FM if required, to determine how to report as per the GEC portfolio pathways.

employment (If of appropriate age)				
Total		**52.9%**	**164**	**7860**
NFE Girls				
Transition into formal school		6.0%	18 (out of evaluation sample of 302)	813
Transition into skills or vocational training		0.0	0 (out of evaluation sample of 302)	0
Transition into work or self-employment (If of appropriate age)		62.3%	188 (out of evaluation sample of 302)	8442
Total		**68.3%**	**206**	**9255**
Aggregate				
Total				**26533**

consilient