Girls' Education Challenge

UKaid from the British people

USAID FROM THE AMERICAN PEOPLE

LEAVE NO GIRL BEHIND

Adolescent Girls' Education in Somalia (AGES)

ENDLINE EVALUATION

PREPARED ON BEHALF OF:

care®

JULY 2024

PREPARED BY:

consilient

Table of Contents

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
EL	Endline
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
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

List of Tables

List of Figures

Executive Summary

About the Evaluation

This evaluation of CARE International's Adolescent Girls Education in Somalia (AGES) programme concludes a four-and-a-half-year effort to track the programme's impact on various cohorts of girl learners. The evaluation is preceded by three previous evaluations: a baseline (BL) round in late 2019, a Midline 1 (ML1) round in early 2022, and a Midline 2 (ML2) round in early 2023. Three key outcomes for girls are measured through this study: 1) learning outcomes, as measured by numeracy and literacy skills; 2) transitions outcomes, measured by girls' trajectories since first surveyed, such as whether a girl has progressed in grade levels, or secured gainful employment after their participation in the AGES programme; and 3) sustainable change emerging from shifts in social norms at the community and individual level, as well as strengthened institutional capacities to support inclusive education.

In addition, the programme's theory of change (ToC) posits that several intermediate outcomes mediate the effect of programme interventions on the primary outcomes. As such, this study also examines progress on several intermediate outcomes: 1) attendance rates, 2) quality of teaching practices, 3) girls' leadership and life skills, 4) school management and governance, 5) community support for girls' education, 6) girl's self-efficacy, 7) strengthened economic circumstances for female youth, and finally, 8) access to social support services.

Importantly, this study comprises evaluations for three separate groups of girls, all recruited into the study at various points in time. The study began with an original baseline cohort of girls participating in the formal education (FE), Alternative Basic Education (ABE), and Non-Formal Education (NFE) programmes. These girls were recruited from schools in the following states, or geographic zone: Banadir, South West State, and Jubaland. The ABE programme focused on accelerated education of basic skills to enable girls to transition into the formal education system, while the NFE programme aimed to equip girls with the skills – both hard and soft – needed to pursue livelihoods, though some NFE girls also transition to the formal education system. These initial groups – whom we collectively refer to as Cohort 1 (C1) or the baseline cohort – were first recruited at BL in 2019, and were funded by the UK's Girls' Education Challenge (GEC) fund.

Two further cohorts, funded by USAID, were later introduced. Cohort 4, or C4, NFE girls were introduced into the study at ML1, while Cohort 5, or C5, NFE girls were introduced to the study at ML2. As such, unlike the C1 girls, we take the ML1 and ML2 evaluation rounds as the cohort-specific baselines for C4 and C5 NFE girls, respectively, and compare changes in main and intermediate outcomes from those rounds to EL. The USAID expansion of the NFE programming also meant the C4 and C5 girls were sampled from an additional state/geographic zone, Hirshabelle, but not from Jubaland.

Learning

Original Baseline Cohorts

Changes in learning outcomes among original baseline cohorts are heterogeneous. While both FE and ABE girls showed significant learning gains in both literacy and numeracy, the scores of C1 NFE girls declined slightly in both subjects. Compared to Midline (ML) 2, results at endline (EL) are reassuring, with both FE and ABE girls' improvements meeting benchmark targets (or being very close to them) and showing a significant improvement between ML2 and EL in both literacy and numeracy. C1 NFE girls also showed a

slight improvement between ML2 and EL, with scores now being closer to BL values after a decline over ML1 and ML2. The new improvement during the last year, after the decline observed between ML1 and ML2, may be explained by the decreasing intensity of the record-breaking drought that occurred between 2020 and 2023.

More specifically, learning scores for FE girls improved by 34.2 points in literacy (compared to a benchmark of 39.1 points) and by 19.9 points in numeracy (compared to a benchmarked gain of 23.1 points). ABE girls gained 22.0 points in literacy, surpassing the expected (benchmarked) achievement by 1.5 points, and 10.7 points in numeracy, missing the benchmark target by 3.7 points.

It is important to note that results were distributed heterogeneously across regions, with improvements in the Banadir region being significantly lower than in other regions, possibly because of the larger presence of girls from IDP families, who face harsher living conditions and improved their scores less than the average.

An important consideration when assessing the improvements among the girls is their varying exposure to schooling across the cohorts. Girls who remained consistently enrolled in FE demonstrated significant progress, but the overall cohort scores were diminished by the dropout rates over time; this trend was also observed among ABE and NFE girls. Regular attendees typically met the improvement benchmarks, while those not enrolled in any learning programs tended to fall behind. These differences underscore the critical importance of continuous enrolment for achieving the programme's learning objectives.

C4 and C5 NFE Cohorts

Both C4 and C5 NFE girls show significant improvements in both literacy and numeracy scores since their respective baselines. C5 girls started the NFE program in 2023 (ML2) and recently concluded their programme, whereas C4 girls started in 2022 (ML1) and ended in 2023 (ML2). The improvements for C4 girls are about 10 points larger than those of C5 girls. Specifically, C4 NFE girls recorded an improvement of 28.2 points in literacy and 28.8 points in numeracy, compared to C5 NFE girls' gains of 18.3 points in literacy and 17.7 points in numeracy. This is attributable to two main factors. First, baseline scores of C4 NFE girls were significantly lower than those of C5 girls in both literacy and numeracy, resulting in more room for improvement. Second, some of the C4 girls who ended their NFE program in 2023 participated in other learning programmes from ML2 to EL and continued improving. C4 girls who remained enrolled in a learning programme showed an improvement twice as large as those who did not. For this latter group, the scores did not improve significantly from ML2 to EL. This reinforces the argument that continuous enrolment is crucial to maintaining improvements in learning.

It is important to note that gains in scores are not evenly distributed across regions, with improvements in the Banadir and Bay regions significantly lower than the average, possibly due to the higher number of IDPs and harsher drought conditions, especially in the Bay region.

Finally, we found evidence that the programme seems to also be effective for some disadvantaged girls, such as girls with disabilities and girls with no educated parents. Furthermore, girls (both C4 and C5) who live with their husbands improved significantly less than the average, possibly due to higher absenteeism, as also confirmed by teachers during focus group discussions.

Transition

Original Baseline Cohorts

A large proportion of FE girls remained enrolled after 4+ years, with 52.9 percent advancing grade levels consecutively. In total, 78.0 percent of FE girls remain enrolled. A sizeable number of the original NFE girls transitioned into non-exploitative employment and education, with 44.3 percent of girls employed, 14.5 percent enrolled in formal education, and 2.3 percent self-employed. Additionally, only 0.4 percent of NFE girls transitioned to vocational training. With the ABE cohort, while 39.9 percent of girls are now enrolled in formal schooling, only 3.4 percent of ABE girls enrolled at a more advanced grade than their ABE level equivalent; additionally only 25.1 percent of ABE girls are employed, 0.4 percent are self-employed, and 0.4 percent are in vocational training.

Notable differences in transition outcome rates include but are not limited to the following: 1) FE girls with uneducated Head of Households (HoHs) had higher transition rates, specifically into employment, likely driven by a higher prioritisation of gaining more educational skills to procure better opportunities compared to her HoH; 2) girls from the South West State had significantly lower success in transition, possibly due to external factors associated with the local political tension affecting economic opportunity for girls in the region; 3) NFE girls with schools lacking school materials had significantly higher transition outcomes into employment, likely due to poor school qualities becoming an inadvertent incentive for NFE girls to shift out of school and seek employment.

Interestingly, no intermediate outcomes related to the Theory of Change had a significant relationship with successful transition outcomes, including caregiver attitudes towards girls' education, teaching quality, Youth Leadership Index scores, and GEF participation.

C4 and C5 NFE Cohorts

C4 NFE girls experienced a relatively high successful transition rate of 59.7 percent. We observed significantly higher rates of successful transition among girls whose house has a poor roof and who went to sleep at least ten night in the last 12 months with hunger. This suggests that, given the vulnerable situation of the girl's household, they are driven to work, probably by necessity. We also observed significantly higher rates of successful transition among girls whose HoH lacked any form of education, which we hypothesise to be driven by higher rates of employment. There were not significant and meaningful associations among the intermediate outcomes outlined by the ToC and successful transition outcomes.

Given the shorter timeframe for evaluating the C5 NFE cohort from enrolment to EL, the overall rate of successful transition – at 54.6 percent – is sizeable, but smaller compared to the other NFE cohorts. Regional differences observed include a higher transition rate among girls in Banadir and a lower transition rate among girls in South West State compared to their respective reference groups. Contrary to the C4 NFE cohort, there was a significant and association of one intermediate outcome outlined by the ToC and successful transition rates among C5 NFE girls. Girls who report that her teacher punishes students who get things wrong during a lesson have a higher rate of successful transition from NFE. This result is counterintuitive, as one would expect improved transitions where teachers do not resort to punishment in class.

Sustainability

Sustainability in the AGES evaluation at EL is assessed for both the FCDO and USAID cohorts. Because the sustainability indicators were previously evaluated for the FCDO cohorts only (and thus can be compared between rounds), the sustainability evaluation of the USAID cohorts was conducted in a cross-sectional manner with the indicator metrics analysed at EL. Additionally, the school-level indicators (i.e., proportion

of teachers implementing inclusive education strategies in class) and GWD retention are FCDO-only indicators since 1) the school-level indicators only focus on FE schools and classrooms and 2) the NFE programmes were not designed for retention, only to provide pathways to either employment or academia.

Original Baseline Cohorts

Caregiver attitudes towards girls' education remains relatively strong, with 59.0 percent of caregivers saying that "education being too costly" is not an acceptable reason for a girl to not attend school. Village savings and loans (VSL) group participation, while being significantly higher than the participation rate at baseline (4.0 percent) still remains relatively low at 10.0 percent; while C1 NFE girl participation (33.1 percent) is notably higher than that of caregivers, only 25.3 percent of those participating reported that their savings group is still active. Additionally, GEF programming was not as expansive for the 26.4 percent of girls participating, with only 9.2 percent of girls in GEFs actually participating in their activities. Moreso, the portfolio of intended programming rolled out by the GEFs (girls education support, business & finance, changing community attitudes) focused heavily on awareness raising (61.3 percent) compared to other activities like supporting savings groups (14.2 percent), enrolling out-of-school girls into an education programme (30.0 percent), or hosting various trainings (14.6 percent).

School-level trainings attendance among female teachers seem to be much lower during the 2023-24 period compared to the 2022-23 period for math and reading/writing; however, gender-sensitive training attendance among male teachers was higher compared to that among female teachers (7.2 vs. 4.3 percent) indicating a larger focus on ensure gender equity practices are instilled among teachers likely to be more resistant to these pedagogical changes. Inclusive training attendance was relatively low compared to the previous year, but the impact of the training on ensuring inclusion for girls from marginalised groups or with disabilities remains inconclusive since teachers and CEC members claim that they do not have any marginalised girls in their schools.

C4 and C5 NFE Cohorts

Because the enrolment of the C4 and C5 NFE girls were much later than their FCDO-counterparts and at different time intervals (2022 for C4 NFE; 2023 for C5 NFE), the rates of participation in the economic and community support components of AGES are expected to not be high. Even so, some of these programmes and their impact on girls' enrolment post-AGES – particularly that of the GEFs – are likely not to continue without additional financial and operational support. GEF sustainability is very likely not to continue due to low participation rates, with only 16.6 percent of C4 NFE girls and 20.5 percent of C5 NFE girls participating. Additionally, out of the girls who participated, only 9.0 and 9.3 percent of C4 and C5 NFE girls, respectively, have ever participated in any GEF activities. As also represented in the FCDO cohorts, the GEFs do not provide a wide array of activities aside from awareness raising (63.1 percent for C4 NFE; 61.8 percent for C5 NFE).

Even so, girls' attitudes towards education is still strong, with 94.5 percent of C4 NFE and 96.2 percent of C5 NFE girls agreeing that education is worth it "even when funds are limited." This is also represented in VSLA participation rates, as while only 28.2 and 23.5 percent of C4 and C5 NFE girls, respectively, participate in one, 42.0 percent of C4 NFE girls and 48.9 percent of C5 NFE girls are still involved in their groups, much higher than the C1 NFE girls' participation rate by comparison. It is likely that these attitudes might persist longer after the conclusion of AGES, but this assessment is limited by the short timeline of evaluation for these cohorts since their relative baselines and their exposure to these support programmes by AGES.

Attendance

Original Baseline Cohorts

The evaluation assessed trends in girls' and boys' attendance in formal education schools, tracking round-to-round changes since baseline of a panel of schools that have been part of the sample in each round. Attendance is assessed through physical headcounts conducted by field team leaders, and through teachers' own reports. Relying first on physical headcounts, the evaluation team found evidence of declining attendance among both boys and girls since baseline (BL). Among girls, declines appear particularly steep for early grades, with attendance rates at EL for Grades 1 to 3 declining by 7.4, 6.0, and 5.9 points, respectively, after controlling for grade level and geographic zone. Boys' attendance similarly saw sizable declines between 5.6 and 1.9 points from BL to EL. Though not significant, these remain sizable changes in attendance rates, much of which driven by declining attendance rates in Baidoa and Banadir for girls, and Banadir for boys. Concurrent to these findings, we also found evidence of teachers' attendance records showing increases in reported attendance rates from BL to EL, possibly due to monitoring and/or capacity building provided by Community Education Committees (CECs), though it is unclear to what extent social desirability bias incentivises teachers to increase students' reported attendance. Nonetheless, we find some indicative evidence of the improving state of attendance record keeping, noting a statistically significant increase of 29.2 points between BL and EL in the proportion of classrooms observed to have "extremely complete" attendance records.

Teaching Quality

Original Baseline Cohorts

The evaluation assesses teaching quality through four components: teacher professionalism, classroom gender equality and equity, use of negative disciplinary practices, and use of positive pedagogical teaching methods. In all, the evidence on progress made is mixed. Progress appears most clear-cut on teacher professionalism, where we observe an aggregate increase of 17.1 points between BL and EL in the share of girls who claimed that teachers were not often absent from class and school. This increase was uniformly observed in Banadir, Jubaland, and South West State, all of whom experienced at least a 12.0 point increase on this measure. On classroom equality, we find evidence that teaching practices had become more gender-equal between BL and EL, as seen, for example, in the significant 8.9-point increase in the share of girls claiming that teachers direct questions at both boys and girls equally. However, when asked more generally about whether they felt that teachers treat girls and boys differently in the classroom, there is a large and significant decline of 24.1 points from BL to EL in the share of girls who disagreed.

On disciplinary practices, data from classroom observations from BL to ML2 indicate a drop in corporal punishment being observed by our field teams. This is seemingly substantiated in a large and significant 43.2 point decline since BL in the proportion of girls who claimed to have witnessed the use of physical punishment in the preceding week. However, the share of girls claiming as much has nearly doubled since ML1 (from 18.6 percent to 32.9 percent), after an initial drop of nearly 60 points from BL to ML1. Among girls who claimed that teachers punish students for providing wrong answers, a larger share of them at EL (69.7 percent) compared to BL (50.0 percent) claimed that that punishment would include physical punishment. These findings suggest a strong need for continuous monitoring to prevent a reversal of progress made during the AGES implementation period. Finally, while EL data shows large and significant increases in the share of FE girls who claimed that teachers use positive teaching practices, such as offering lessons at the right speed, classroom observations data from BL to ML2 fail to substantiate girls' perceptions, suggesting a need for continued improvement in institutionalising positive teaching practices.

Leadership and Life Skills

Original Baseline Cohorts

To measure girls' leadership and life skills, we draw on the Youth Leadership Index (YLI), and 21-item scale that asks girls about the extent to which they agreed with statements related to leadership and self-confidence, including "I can show what is important to me with my actions" and "I am willing to work hard for my dreams." The index score calculates the percentage of the aforementioned items a girl agrees with. Among the baseline cohorts, we see large and significant increases in YLI scores for all cohort types between BL and EL – all increases are in the double digits. These increases become larger when measuring the proportion of girls who achieved at least a 70.0 on the YLI. FE girls experienced a significant 25.5 point increase in their YLI score between BL and EL, as well as a 45.2 point increase in the share of FE girls scoring at least 70.0 percent on the YLI. Indeed, being the youngest girls on average, FE girls started with the lowest YLI scores at BL, and subsequently experienced the largest gains. In terms of location, girls in Jubaland registered the steepest increases, with Cohort 1 girls in Jubaland experiencing a 29.0-point increase in their YLI scores, and with the share of girls with at least 70.0 percent on the YLI growing by 54.3 points between BL to EL.

C4 and C5 NFE Cohorts

Like Cohort 1 girls, C4 NFE girls experience large increases in their YLI scores – in all geographic zones, these increases range between 11.1 and 24.9 points, and all are significant. Similarly, there are sharp increases across zones in the share of girls scoring at least 70.0 percent on the YLI, with an aggregate increase of 25.6 points. South West State's gains are the largest, registering a significant increase of 34.6 points between ML1 and EL. Among C5 girls, we observe small declines in their YLI scores between ML2 and EL, as well as declines in the share of girls scoring at least 70.0 on the YLI. Though these declines are not significant, the plateau among C5 girls is in contrast with C4 girls. Importantly, we note that C5 girls at ML2 had scores that were already higher, on average, than C4 girls, which leads C4 and C5 girls to have similar scores by EL.

School Management and Governance

Original Baseline Cohorts

As with the ML2 evaluation, school directors maintain substantial responsibilities over school management, with school directors being the person most commonly selected by head teachers when asked who had responsibility over a range of school administration tasks. At the same time, Community Education Committees (CECs) play an important role in school governance. The key functions of CECs cited by head teachers are: tracking student attendance (85.7 percent), tracking teacher attendance (77.1 percent), and promoting enrolment of out-of-school children (74.3 percent). CECs also work to address barriers to education. Surveys with girls' caregivers indicate that school materials being culturally or religiously inappropriate, a girl's disability, and financial barriers, hypothetically speaking, would be the primary reasons it would be acceptable for a child to not attend school. CECs in turn play a role in addressing each, such as by raising awareness on the importance of education, constructing of disability-friendly facilities, and raising funds, respectively. Nonetheless, CECs are generally resource-strapped and operate in challenging environments, reducing their ability to fully address these barriers.

Community Attitudes

Original Baseline Cohorts

Community attitudes are collected via household surveys with girls' caregivers, where we specifically asked them whether they believe it is acceptable for girls to miss school for a range of factors. In examining whether caregivers find it acceptable for a girl to stop attending school because school is too costly, we observe an

aggregate decrease in the share of caregivers supporting this notion between BL and EL, though the proportion of caregivers who still support it remains high at EL, at 40.7 percent for FE girls and 43.3 precent for ABE girls. Turning to the measure on whether caregivers believe it is acceptable for a girl to not attend school due to housework, we observe a BL-EL increase in the share of caregivers agreeing with the statement. This corresponds with actual and significant increases in girls' household chore burden, with the share of girls doing housework for the whole day increasing by a significant 13.6 points among ABE girls since ML1, and by 8.9 points among FE girls since ML1. For FE girls, this may be due to girls' increasing age leading them to take on more household responsibilities, while the cause of this trend is less clear for NFE and ABE girls. Overall, then, the findings paint a mixed picture of support for girls' education

Self-efficacy

Original Baseline Cohorts

Focusing exclusively on C1 NFE girls who were asked the relevant questions, we use girls' responses to the Chinese Positive Youth Development Scale (CPYDS) to understand their reported levels of self-efficacy. Our assessment of changes in girls' responses to CPYDS questions paint a mixed picture. On one hand, between BL and EL, girls were more likely to agree to statements such as "I have little control of things that happen in my life" and "I do not have any solutions for some of the problems I am facing". On the other hand, they were also more likely to agree to statements such as "I can finish almost everything that I am determined to do." Qualitative data also indicates that girls, by and large, have gained skills related to resilience and self-confidence as a result of AGES programming.

In terms of accessing protection services, we note decreases in the share of girls who have a channel to report abuses cases occurring in schools, and a channel to report abuses cases occurring in the community. This represented a 12.5-point decreases on both measures from BL to EL.

C4 and C5 NFE Cohorts

Cohort 4 and 5 NFE girls demonstrated opposite patterns in their CPYDS item scores, though both appear to converge to similar levels by EL. Where C4 girls showed large and significant increases in individual CPYDS item scores, indicating worsening self-efficacy, C5 girls registered large and significant decreases on average. The aggregated CPYDS index score did not change between ML1 and EL for C4 cohort girls, but disaggregating by region highlights steep and significant increases in the Lower Shabelle region between ML1 and EL (17.3 points), and similarly steep decreases in the Middle Shabelle region (20.6 points), highlighting the different trajectories of girls in both regions, though their scores converged towards the same levels at EL. Among C5 NFE girls, the declines in their CPYDS scores are more evenly distributed across different geographic zones and subgroups.

In terms of accessing protection services for abuse cases occurring at school and in the community, C4 girls registered a significant decline between ML1 and EL in the proportion of girls who report having at least one channel to report such cases. On the latter measure of reporting abuses occurring in the community, girls with any reported disabilities also registered a significant decrease of 16.2 points, compared to 2.7 points among girls without disabilities. C5 girls registered opposite trends on both measures, with small increases in C5 girls having at least one channel to report abuses in school.

Across both cohorts, the most common parties to turn to – both for abuses at school and in the community – are head teachers, teachers, and family members. This finding highlights the important role that educators can play in child protection and safeguarding, even in regard to incidents occurring outside of school. At the

same time, it is worth noting that there were statistically significant declines between rounds in the share of both still-enrolled C4 and C5 girls who reported having a channel to report cases to and who would turn to head teachers. For example, among C5 girls, the share who would report to a head teacher declined by 27.5 points between ML2 and EL.

Strengthened Economic Circumstances of Female Youth

Original Baseline Cohorts

Among C1 NFE girls, changes in mean monthly income since ML1 do not point to any clear findings. While there is a net increase of $8 per month on aggregate, this was not statistically significant, and only a $12.9 increase for girls in Banadir was statistically significant. Nonetheless, the very small sample size of girls were able to report a specific dollar amount precludes drawing firmer conclusions. In terms of employment, however, there is a significant decline in the share of C1 NFE girls who had no occupation or job, decreasing from 56.7 percent at ML1 to 34.7 percent at EL, which translates to a 22.0 point decline. Between ML1 to EL, the biggest change in occupation category was in the "domestic work category", whose share of girls increased from 19.9 percent at ML1 to 31.3 percent at EL, though the low levels of income among girls taking on domestic work suggests that is it mostly done in their own households, or that it is non-remunerative.

C4 and C5 NFE Cohorts

After adjusting for outliers in the income data, we find moderate aggregate increases in mean monthly incomes for C4 girls, with the mean income rising from $15.65 at ML1 to $40.00 at EL – a statistically significant increase. This trend holds true across Banadir, South West State, and Hirshabelle. For C5 girls, no significant changes in mean monthly income were evident. Importantly, we observe an increase in the share of both C4 and C5 girls who reported earning no income at all. For C4 girls, this was driven by sharp increases in the South West State zone, and more specifically, Afgoye district, though the small sample size necessitates caution in interpreting the results. Among C5 girls, the main driver of this increase lies in the Banadir region.

Both C4 and C5 girls saw decreases in the share unemployed girls. The ML1 – EL change for C4 girls is significant, while the ML2 - EL decrease for C5 girls was not. This is unsurprising, as C4 girls have now had more than a year to seek unemployment, whereas C5 girls' NFE programming had only recently concluded. One encouraging finding across both cohorts is that there is a significant increase in the share of girls in each cohort who claim to be running a business. This figure jumped from 8.3 percent at ML1 to 16.5 percent at EL for C4 girls, and from 12.1 percent at ML2 to 15.2 percent at EL for C5 girls.

Enhanced Social Support for Female Youth

C4 and C5 NFE Cohorts

The share of C4 NFE girls who have benefitted from additional social support – including community engagement and humanitarian assistance – outside of AGES, increased substantially between ML1 and ML2, but exhibited slower growth between ML2 and EL, likely due to difficulties reaching more inaccessible marginalised girls. For instance, youth group participation, while increasing substantially from 14.8 percent to 32.2 percent between ML1 and ML2, only increased to 33.4 percent by EL. The rate of receiving humanitarian assistance from ML2 to EL remained constant, at 30.1 to 30.6 percent, respectively.

While the portion of C5 NFE girls receiving additional support and engagement increased rather marginally between ML2 and EL (except for engagement in service delivery improvement discussions), the rate of

humanitarian assistance reception decreased substantially among the relatively more marginalised girls (i.e., those with disabilities, IDP status, and low SES indicators) in the cohort. For instance, while the overall humanitarian assistance reception rate remained constant from 29.8 percent in ML2 to 30.5 percent in EL, the rate of reception among those with a physical disability declined from 34.6 to 23.1 percent during that time.

Value for Money

The analysis of Value for Money was limited by the lack of a benchmark for education costs in Somalia. In addition, the analysis does not include M&E and CA costs whose data are not divided by donor (FCDO vs USAID).

Original Baseline Cohorts (FCDO-funded)

The direct cost per girl for FCDO-funded activities was £159, rising to £176.2 when indirect costs are included. Most of the budget (55% of direct costs) was allocated to the Output 2 of the ToC, which aimed at enhancing the literacy and numeracy skills of ultra-marginalized girls by providing quality learning opportunities. This was the largest part of the program and included the provision of ABE and NFE programs, along with support for formal schooling. Only the costs under Output 2 can be broken down by type of learning program. The cost per girl for ABE-related activities was £134.1, which is £8.6 higher than for NFE girls (£125.5). This difference mainly stems from the provision of desks for ABE girls, costing £18.3 per girl. Despite higher incentives for NFE teachers and the budget allocation for developing the NFE curriculum (compared to only a curriculum revision for ABE), ABE activities remained more expensive. Assuming that costs of activities related to other outputs are uniformly distributed across all types of girls, the total costs per beneficiary are £204.8 for ABE girls, £196.2 for NFE girls and £107.5 for FE girls.

C4-C6 NFE Cohorts (USAID - funded)

The total direct cost per girl for USAID-funded activities was £177.6 (£201.5 including indirect costs). The largest portion of the budget went towards activities directly associated with providing the NFE curriculum, costing £77.3 per girl and accounting for 43.5% of all direct costs. A significant portion, £55.2 or 31.1%, was spent on partner support costs, mainly covering program staff salaries. When looking at the composition of direct costs, the largest expenses were facilitator incentives, which cost £23.9 per girl and made up 30.9% of the costs directly related to the provision of the NFE curriculum. The provision of NFE classes cost £21.7 per girl and accounted for 28.1%. This includes £12.6 per girl for the construction and rehabilitation of classrooms and £9.1 per girl for the provision of desks. Other significant costs included salaries for support staff at NFE centres (£9.6 per girl), learning materials (£7.9 per girl), and the training of NFE facilitators (£7.6 per girl).

1. Introduction

The Adolescent Girls' Education in Somalia (AGES) was implemented in southern and south-central Somalia, and funded by FCDO UK's Girls' Education Challenge (GEC) programme and USAID. With the mobilization of out-of-school girls and their subsequent enrolment in a range of formal and non-formal learning programs, implementation began in September 2018. Through the allocation of extra donor financing, the programme has grown since its beginning. At baseline, it was implemented in Southwest State, Jubaland, and Banadir; it eventually expanded into Hirshabelle, and its reach increased in the regions where it was already operational.

The project's geographic focus was on the regions of Somalia most impacted by three decades of conflict that began in the late 1980s. These included areas with intense battles, repeated displacements, "clan cleansing," large populations of internally displaced people (IDPs) escaping violence, and regions long controlled by al-Shabaab. The AGES project's implementation was further complicated by a complex governance environment.

For a fuller background on the programme and the context in which it operated, we refer readers to the three previous evaluation reports conducted on behalf of the AGES project. The baseline report describes the project's approach, targeting strategy, and activities in detail; the first midline report (Midline #1, or ML1) describes the political, social, and economic context of the regions where AGES works. The second midline report (Midline #2, or ML2) documented the changes in these contextual factors since the ML1 evaluation took place in early-to-mid 2022. Similarly, the endline report discusses the contextual factors that changed since ML2 took place in spring-summer 2023.

A positive Indian Ocean Dipole and El Niño have intensified the 2023 Deyr (October to December) rainy season, resulting in severe flooding across Somalia, particularly in the states of Puntland, Galmudug, South West, Hirshabelle, and Jubaland. To mitigate the destruction caused by intense rains, flash floods, and riverine floods, humanitarian actors swiftly increased response efforts in collaboration with the government, including the mobilization of boats to rescue stranded individuals. Despite these efforts, at least 2.48 million people have been affected, with 1.1 million displaced and 118 fatalities.[1]

Consequently, at least four million people in Somalia (21 percent of the population) faced crisis or emergency levels of food insecurity between January and March 2024, primarily due to heavy rainfall and flooding late last year, compounded by lingering effects of previous droughts. This is according to the latest Integrated Phase Classification (IPC) findings published by the FAO-managed Food Security and Nutrition Analysis Unit on 15 February. The findings indicate an improvement in food security compared to the same period in 2023, when approximately five million people were classified in IPC Phase 3 or worse due to prolonged drought. The current figure represents a 20 percent reduction in the number of food-insecure people.[2]

Despite the improvements in food security mentioned above, it was projected that between 3.6 million and 4.9 million school-aged children in Somalia would not have access to formal education in 2024.[3] Additionally,

[1]UNOCHA Somalia: 2023 Deyr Season Floods Situation Report No. 4. Available at: https://www.unocha.org/publications/report/somalia/somalia-2023-deyr-season-floods-situation-report-no-4-10-december-2023#:~:text=The%20number%20of%20people%20affected%20by%20the%20heavy,highlands%2C%20signaling%20the%20end%20of%20the%20deyr%20season.

[2]UNOCHA Somalia Situation Report. 17 March 2024. Available at: https://www.unocha.org/publications/report/somalia/somalia-situation-report-17-mar-2024

[3]UNICEF Somalia Humanitarian Situation Report No. 1. January 2024. Available at: https://www.unicef.org/media/153271/file/Somalia-Humanitarian-SitRep-No.-01-31-January-2024.pdf

nearly 2.4 million school-aged children would require humanitarian assistance to begin, return to, or remain in school.[4] As in previous years, the primary reason for school dropout and non-enrolment in Somalia was the cost of education, including related expenses, with displaced communities being significantly and disproportionately affected.

The 2024 Humanitarian Needs and Response Plan for Somalia (HNRP) indicated a decrease in the number of people needing assistance, from 8.3 million in 2023 to 6.9 million in 2024. This reduction is largely attributed to the scaled-up humanitarian response and the significantly above-average Deyr rains (October to December) in 2023, which mitigated the threat of localized famine.[5] However, prices of basic commodities have remained elevated, adversely impacting people's purchasing power. Overall, the situation has stabilized due to the sustained humanitarian response last year and the modest benefits from the rains, although ongoing displacement due to conflict, active hostilities, and insecurity continues to exacerbate the situation.

The political and security contexts of Somalia remain challenging. Since June 2023, the coalition assisting the counterinsurgency against al-Shabaab has been undermined by clan warfare and hostility against the federal authority. Conflicts between the Abgal and Hawadle clans around land ownership and power-sharing agreements in the state of Hirshabelle have weakened the military effort. The tenuous state of order prevailing in the alliance is reflected in the at least six violent conflicts between the Abgal and Hawadle clan militias in Hiiraan and Middle Shabelle that ACLED has documented since June 2023.[6] Consequently, government forces and clan militias began to withdraw from security positions in Hiiraan, slowing down the attack and allowing al-Shabaab to retake lost ground.

These changes within the coalition stopped the counterinsurgency effort from spreading into the Southwest and southern Jubaland. The murder of Hashim Mohamed Shareero, the deputy district commissioner of Wanlaweyn, further exacerbated regional tensions. Additional conflicts between rival militias in Wanlaweyn and Barawe displaced hundreds of individuals and resulted in numerous civilian casualties.[7] Clan conflict also emerged in the Diinsoor district of the Bay region in spring 2024. Despite efforts by elders to negotiate a resolution on the first day of the conflict, their attempts were unsuccessful. Families in Diinsoor began evacuating to nearby towns 10 to 15 kilometers away as the violence escalated. Hostilities and the imminent threat of further violence displaced 1,371 households from their homes in Diinsoor.[8] Recently, Somalia's government has attempted to slow the withdrawal of African peacekeepers, warning of a potential security vacuum should the mission leave prematurely. Neighbouring countries are concerned that resurgent al-Shabaab militants could seize power.[9]

---

[4]UNICEF Education Cluster 2024 HNRP Snapshot. February 2024. Available at: https://reliefweb.int/report/somalia/education-cluster-2024-hnrp-snapshot-feb-2024#:~:text=In%202024%2C%20it%20is%20projected%20that%20between%203.6,to%20begin%2C%20return%20to%2C%20or%20remain%20in%20school.

[5] OCHA Somalia 2024 Humanitarian Needs and Response Plan (HNRP).

[6] ACLED Somalia: Renewed counter-insurgency operations expected as al-Shabaab regroups. May 2024. Available at: https://acleddata.com/2024/05/31/somalia-situation-update-may-2024-renewed-counter-insurgency-operations-expected-as-al-shabaab-regroups/.

[7] Horn Observer Somalia: Wanlaweyn Deputy Commissioner killed in ambush, weapons looted as clan conflict escalates in Lower Shabelle. Available at: https://hornobserver.com/articles/2266/Somalia-Wanla-Weyn-Deputy-Commissioner-killed-in-ambush-weapons-looted-as-clan-conflict-escalates-in-Lower-Shabelle.

[8] UNHCR Somalia: Protection and Return Monitoring. Flash Alert 4. March 2024.

[9] Reuters: Somalia asks peacekeepers to slow withdrawal, fears armed group resurgence. June 20. Available at: https://www.aljazeera.com/amp/news/2024/6/20/somalia-asks-peacekeepers-to-slow-withdrawal-fears-armed-group-resurgence; The East African: Somalia defers ATMIS drawdown plan yet again. June 20. Available at:

Since 2023, children have faced ongoing risks, with the nutritional status of 1.5 million children under the age of 5 remaining concerning, as acute wasting is projected to continue until July 2024. Approximately 8 million people are also experiencing water shortages, leading to a spike in cholera cases. Accessing the most vulnerable children has remained difficult, as many of them live in areas that are inaccessible to UN agencies and their partners. Consequently, the involvement of local partners in delivering services has remained essential.

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 ML2 evaluation round, the present evaluation round actually consists of three distinct evaluations, differing in terms of the cohorts of 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 BL, through the ML2 evaluation round in 2023 to the present 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. Throughout this report, we refer to them collectively as the "baseline cohort" or "C1 cohort"; when referring to the NFE girls from this group, as opposed to NFE girls from later cohorts, we use the term 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

https://www.theeastafrican.co.ke/tea/news/east-africa/somalia-defers-atmis-drawdown-plan-yet-again-4663636?view=htmlamp

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 the AGES 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 BL to EL 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.

Cohort 5 NFE Girls

The third component of the evaluation was introduced last year, involving Cohort 5 (C5) non-formal education (NFE) girls, for whom this evaluation serves as a second and final endline round. Broadly speaking, C5 NFE girls are enrolled in the same learning centres as Cohort 4. They were recruited and enrolled in the learning intervention just before the data collection for the ML2 round, and their learning progress, transition, and other outcomes are now tracked through to the endline, similar to the other cohorts described above.

Panel Samples

Given the pre-post nature of this evaluation, much of our analysis rests on tracking changes over time among the same set of girls in each round. However, attrition across rounds poses a major challenge as it diminishes our sample size available for analysis – a particularly important consideration when conducting subgroup analysis that disaggregates our existing sample even further. To maximise the sample size we use for analysis, much of the following analysis uses cohort-specific panels of girls who were surveyed at both the cohort specific baseline, and at EL. The panel samples are as follows for each cohort of girls:

- Cohort 1 FE, ABE, and NFE girls: BL to EL
- Cohort 4 NFE girls: ML1 to EL
- Cohort 5 NFE girls: ML2 to EL

The key benefit of this approach is that we are able to re-recruit girls who had dropped out of the sample at some point prior to the EL round, and still track their changes over time since the cohort-specific baseline. This in turn allows us to maintain a larger sample than would have been possible had we restricted our analysis to girls who were surveyed in *every* round since their cohort-specific baseline. Though we at times use the "full" panel to track round-to-round changes, much of our analysis utilises the respective baseline-to-EL panel for each cohort.

2.2. Data Collection Tools

The AGES evaluation is a mixed-methods study that seeks to measures changes over time in learning, attendance, girls' life skills, and other outcomes. To accomplish this, the evaluation team employs a range of quantitative and qualitative tools carefully calibrated to capture main outcomes. The tools target a range of respondent types, including girls, their family members, teachers, head teachers, Community Education Committee (CEC) members, and local religious leaders. Our quantitative tools encompass the following:

- Learning assessments testing numeracy (Early Grade Math Assessment, or EGMA) and testing Somali literacy (Early Grade Reading Assessment, or EGRA) administered to the cohort girls
- Surveys with the cohort girls
- Household surveys with the caregivers and heads of household of cohort girls[10]
- School survey with head teachers of formal (FE) schools
- Attendance headcounts of FE centres

During the inception period, the AGES technical staff and the evaluation team reviewed the tools from the prior rounds and made several amendments, based on the experience of previous rounds. Several changes made this round or carried over from prior rounds are worth noting. 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, school-level data collection was only conducted in FE schools. Third, even among these schools, classroom observations were no longer included, as AGES support in FE centres had concluded since 2021.

Finally, changes made to targeting of the household survey at ML1 were maintained to this round. In short, the household module – completed by a girl's head of household and her caregiver – was administered only to girls who are under the age of 18 years *and* have never been married as of the EL round. In contrast, girls 18 and over or under-18 girls who have been married at some point completed a shorter set of questions – capturing some of the same information – themselves. This shorter module itself had the addition of several questions that were previously only asked as part of the household survey. Specifically, a series of questions probing for the girl's household diet, as well as her marital status, were now included as part of the girl's survey, where in previous rounds these questions were only posed to a girl's caregiver. This methodological decision was based on the assumption that girls or women of this age and/or with this life experience are their own caregiver, and the best source of information about themselves.[11]

Similar to ML1 and ML2, this decision impacts our analysis in that the shorter girl module does not capture some of the information on household characteristics that would be captured from the more extensive household survey. Consequently, subgroup analysis of main and intermediate outcomes often requires us to use smaller sample sizes.[12]

[10] For girls who are under 18 and have never been married

[11] Note, however, that we never consider a girl's husband as her caregiver.

[12] Wherever possible, we mitigated this effect by utilising data from the cohort-specific 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 during EL, we did not capture data at EL 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 her cohort baseline to her EL data where possible. Because some girls may have been young enough to require the full household survey in prior rounds, in some cases we are able to extract this information even if no household survey was

Table 1 below documents, at EL, 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 mismatch between the number of girls targeted for the full household survey and the number of girls for whom a household survey was completed stems from the fact that not all household members/interviewees consented to a survey, or were available at the time of data collection.

TABLE 1: SHARE OF COHORT GIRLS, AT EL, COMPLETING GIRL VS. FULL HOUSEHOLD MODULE

Cohort	Completed Abbreviated Module with Girl Only		Targeted for Full HH Survey		Completed Full HH Survey	
	N	Pct of Cohort Sample	N	Pct of Cohort Sample	N	Pct of Cohort Sample
FE Girls	100	27.86	259	72.14	234	65.18
ABE Girls	223	84.79	40	15.21	30	11.41
C1 NFE Girls	263	100.00	0	0.00	0	0.00
C4 NFE Girls	531	92.35	44	7.65	35	6.09
C5 NFE Girls	321	85.83	53	14.17	40	10.70
Total	**1438**	**78.41**	**396**	**21.59**	**339**	**18.48**

For school-level tools, our sample has further narrowed since the ML2 evaluation round, which itself had seen the sample for school-level data collection narrow since ML1. During ML2, field teams conducted headcounts, school surveys, and classroom observations in formal schools and C5 NFE centers, both of which included still-enrolled cohort girls. During the EL, however, school-level tools were administered by field team leaders in formal schools only. Moreover, classroom observations were dropped from the current round of data collection.

TABLE 2: SAMPLE BREAKDOWN OF SCHOOL-LEVEL DATA COLLECTION

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	36	36

administered for them at EL. 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.

Attendance Headcounts	1 per classroom, Grades 1 – 5	37	N/A	33	164

Finally, our quantitative data collection was supplemented by the collection of qualitative data from various key stakeholders, including girl learners themselves, as well as mothers, teachers, CEC members, and religious leaders. Our qualitative sample was heavily represented in Banadir, where 41.7 percent of all qualitative interviews and focus group discussions (FGDs) were held. This reflects our overall sample composition of girls to whom quantitative surveys were administered. This is followed by South West State, where 23 out of 72 qualitative interviews, or 31.9 percent, were conducted. 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

TABLE 3: QUALITATIVE SAMPLE BREAKDOWN

Respondent Type	Banadir	Hirshabelle	Jubaland	Southwest State	Total
CEC Members - FGD	4	0	2	6	12
Mothers - FGD	4	0	3	5	12
Teachers - FGD	6	1	2	3	12
Girls – Vignette Exercise	6	2	1	3	12
Girls – Risk Mapping	7	2	1	2	12
Religious Leaders - KII	3	2	3	4	12
Total	30	7	12	23	72

2.3. Recontact Procedures and Results

During this EL evaluation, our team sought to re-contact girls who were initially recruited during the BL round (for Cohort 1), ML1 round (for Cohort 4 NFE girls), and ML2 round (for Cohort 5 NFE girls). We also attempted to recontact FE girls who were recruited in ML1 or ML2 as replacement girls for girls who were surveyed during the BL but fell out of the sample at either ML1 or ML2. In short, we aimed to re-contact every girl from any of the previous rounds,[13] with the goal of maximizing the available sample for analysis. This section outlines the re-contact procedures used to maximize re-contact rates and briefly analyses 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

[13] The primary exceptions to this are schools in the district of Diinsoor, which our teams had forgone during ML2 due to insecurity. The same decision was made for EL due to continued insecurity and restricted access.

members, pre-mobilising girls with the aid of the CARE AGES team, and utilising 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. Furthermore, enumerators were asked to visit the girl's household at least twice, using location information obtained from prior rounds' 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 (clan elders, religious leaders, etc.) for contact information for the girl or her family. In addition, it is worth noting that the CARE AGES team assisted the evaluation team by helping mobilising girls in the formal schools or learning centres throughout fieldwork, which allowed the evaluation team to survey additional cohort girls who might otherwise be left out of the EL sample due to a condensed evaluation timeframe.

To increase the rate of successful re-contacts, and thus work with a larger sample size, field teams were instructed to travel to a girl's home to survey her if she lived in the same town as the school under which she was registered. If the girl lived outside that range and is out of town, the team leader would consider whether the girl could be visited later in the fieldwork when the team was closer to her home. Alternatively, in a handful of cases, field teams were able to collect information that a girl had moved from one district to another district or town where the evaluation team's enumerators were also present. In some cases, the evaluation team instructed the other field team in a girl's new location to attempt to find and survey her.

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 BL, replacement girls had been selected in each of the prior ML rounds to help mitigate the effects of attrition on our analysis. This selection was done randomly by the evaluation team from the original sample frame provided by CARE in 2019, to ensure unbiased sampling. However, no replacement girls were drawn for the FE cohort at EL. Moreover, ABE, C1 NFE, C4 NFE girls, and C5 NFE girls who could not be located were never replaced, as they had typically completed their participation in the educational centre by the time of the evaluations. Thus, it was not possible to select a comparable replacement girl from within the centre.

Re-Contact Rates

In this sub-section, we present the results of our efforts to re-contact previously surveyed girls. In all, the evaluation team had a total possible sample of 3,013 girls who could be recontacted, with the following breakdown: 594 FE girls, 484 ABE girls, 515 C1 NFE girls, 916 C4 NFE girls, and 504 C5 NFE girls. In short, this represents the total number of girls who had been surveyed in prior rounds. Following careful monitoring of field conditions, however, the evaluation team maintained its decision to not attempt to recontact girls in Diinsoor, and in one school in Hodan district in Banadir, due to continued insecurity and accessibility concerns.

Based on our experience conducting past evaluations for AGES, we posit that the primary drivers of unsuccessful re-contact are likely to be unenrolling from the school and migration. This is because the schools and learning centres are the primary channel through which field teams can access 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 AGES treatment arm (FE, ABE, NFE) focuses on girls with specific demographic characteristics, such as age, while also having differing durations for the treatment's implementation. The FE programme included 10–13-year-old girls at BL and had run for four years, ABE ran for two years before concluding in 2022 and targeted 13-16 year old girls, while the NFE treatment lasted only 11 months for girls aged 17-20. Among the baseline cohort, this final group, the Cohort 1 NFE (C1 NFE) girls, completed their programme in late 2020, while additional cohorts of NFE girls (Cohorts 4 and 5) ended their programmes in 2023 and 2024, respectively.

There may be varying degrees of success in re-contacting girls in different treatment arms because those on the younger end of the spectrum, (namely FE girls), may be less likely to get married and therefore move or drop out of school. The recency of programme conclusion may also affect the likelihood for girls to stay engaged with the schools/learning centres or stay in an area, as they may still anticipate future benefits through participation. To disaggregate these trends, we present the table below on the successful recontact rates per cohort group, while also disaggregating by geographic zone to assess whether context-specific factors may be affecting re-contact. The rightmost column of the table highlights the achieved panel sample, defined as the number of girls, per cohort, who were successfully contacted at both the cohort-specific baseline, and at EL. For example, FE girls who were part of the BL and EL samples would fall into the FE girl BL to EL panel, while C4 NFE girls who were part of the ML1 and EL samples would into the C4 girl ML1 to EL panel.

TABLE 4: RE-CONTACT RATES BY COHORT TYPES AND GEOGRAPHIC ZONE[14]

Cohort	Total Possible Sample	Re-Contact Rate	Achieved Panel Sample
Banadir			
FE Girls	180	51.11	92
ABE Girls	218	46.33	101
C1 NFE Girls	252	45.63	115
C4 NFE Girls	440	61.82	272
C5 NFE Girls	241	68.46	165
Jubaland			
FE Girls	132	61.36	81
ABE Girls	108	55.56	60
C1 NFE Girls	107	67.29	72
South West State			

[14] Note that exclude observations dropped at EL due to ambiguity around a girl's actual identity, in relation to previous surveys.

FE Girls	81	61.73	50
ABE Girls	142	71.83	102
C1 NFE Girls	131	58.02	76
C4 NFE Girls	337	64.99	219
C5 NFE Girls	185	77.30	143
Hirshabelle			
C4 NFE Girls	139	60.43	84
C5 NFE Girls	85	72.94	62
Aggregate			
FE Girls[15]	393	56.74	223
ABE Girls	468	56.20	263
C1 NFE Girls	490	53.67	263
C4 NFE Girls	916	62.77	575
C5 NFE Girls	511	72.41	370

Table 4 confirms the more recent cohorts of girls – C5 and C4 NFE girls – had, in that order, the highest aggregate and zone-specific recontact rates of any cohort. Per our discussion above, this is unsurprising the recency of their programme completion. ABE and C1 NFE girls, on the other hand, experience considerably lower recontact rates, most likely as a function of the longer time elapsed since program completion and the EL. Somewhat surprising is that baseline FE girls also had comparatively lower recontact rates compared to C4 and C5 NFE girls. One methodological point to note is that the FE girls' figures presented in this table represent the FE girls who were originally recruited during the BL evaluation. On the other hand, the evaluation team was able to successfully recontact additional FE girls who had been brought into the sample as replacement girls at ML1 or ML2. These girls have not been included in the table, as they are, by definition, not part of the panel sample of girls surveyed at both the cohort-specific baseline, and the EL. Including them in our calculations brings up the aggregate recontact rate for FE girls from the 52.97 percent indicated in Table 4 to 60.44 percent. This increase is due to a 71.05 successful recontact rate among FE replacement girls recruited at ML1, and 84.54 among FE replacement girls recruited at ML2.

Additionally, we examine other variables that might affect the likelihood of recontact. Per the ML2 evaluation, we posit a suite of potential predictors of re-contact, largely focusing on predictors that might affect a girl's likelihood of dropping out of school, or her (or her household's) likelihood of migrating to

[15] Note that the FE girls' figures presented in this table represent the FE girls who were originally recruited during the BL evaluation. On the other hand, the evaluation team was able to successfully recontact additional FE girls who had been brought into the sample as replacement girls at ML1 or ML2. These girls have not been included in the table, as they are, by definition, not part of the panel sample of girls surveyed at both the cohort-specific baseline, and the EL. Including them in our calculations brings up the aggregate recontact rate for FE girls from the 52.97 percent indicated in Table 4 to 60.44 percent. This increase was due to a 71.05 percent successful recontact rate among FE replacement girls recruited at ML1, and 84.54 percent among FE replacement girls recruited at ML2.

another location. These predictors may affect the likelihood of dropping out or migrating through a variety of mechanisms. For instance, household poverty can hinder migration by making travel financially infeasible but may at concurrently increase the likelihood of dropping out, as children in the household may be required to work to support their families. Land ownership may also decrease migration; without a well-established land ownership system in Somalia, informal and undocumented claims are pervasive, and families who wish to keep their land are effectively rooted to the area to fend off potential claimants to their land. On the other hand, internally displaced households and linguistic minorities may be incentivized to leave their host communities because of social or economic marginalisation.

In the interest of brevity, we use the same battery of variables used in the ML2 evaluation round to determine the extent to which they affect recontact rates between, in most cases, cohort baselines and this present EL round.[16] These were originally derived by conducting balance tests between the baseline C1 cohorts who were successfully recontacted at ML2, and those who were not. The balance tests checked for symmetry in the distribution of subgroups between girls successfully recontacted and those who were not. These subgroups included girls who were from displaced households, girls living with disabilities, and girls whose households went hungry many, most, or all days in the past 12 months. Subgroups that had statistically significant differences in their prevalence among successfully and unsuccessfully recontacted baseline cohort girls were then used as predictors in a multiple linear regression model, with all variables used in a single model. In addition to these variables, we also now include a variable on whether a girl's household had experienced a lack of cash for most days in the past 12 months, as the distribution of girls who answered in the affirmative were significantly different between girls successfully recontacted at EL and those who were not.

In addition to above variables, we also include controls for geographic zone and girl type (FE, ABE, NFE) among the baseline cohorts.[17] We then include two indices generated during the ML2 evaluation: the caregivers' perception index, and the teaching quality index. We do so as we expect that caregivers' attitudes towards girls' education would have meaningful effects on their likelihood of staying in school (and thus their likelihood of being recontacted). A higher score on the caregiver perspective index thus represents a more favourable attitude towards girls' education. Teaching quality may similarly affect a girls' likelihood of staying in school, as a negative learning experience may incentivise some girls to drop out of school. A higher score on this index represents a more negative set of teaching practices, as reported by girls. Finally, we include a binary variable on whether the last time a girl was surveyed for an AGES evaluation was prior to the ML2 round (the most recent round before the EL), as we expect that girls who had been more recently surveyed to have a likelihood of being recontacted again at EL.

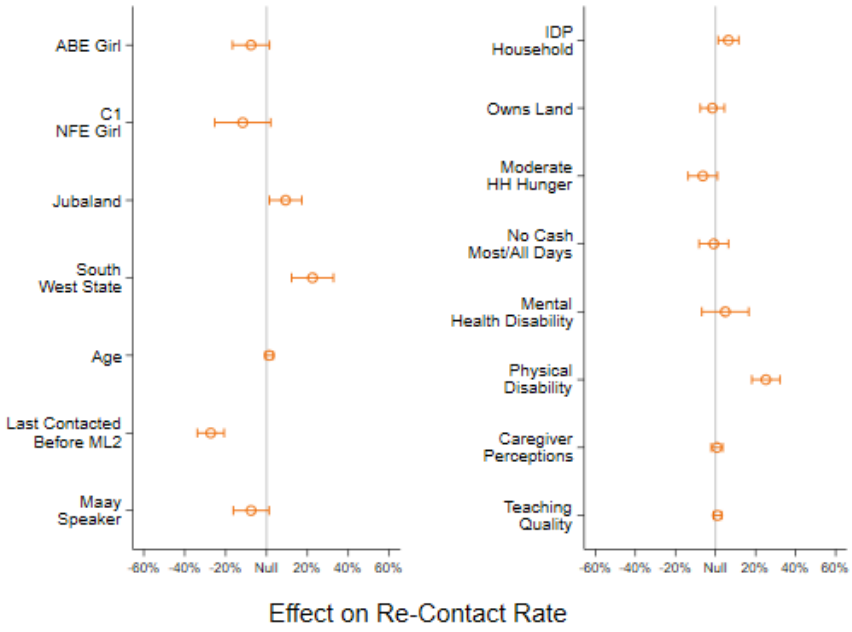
We restrict this analysis to girls in the baseline cohort only, as later cohorts (C4 and C5 girls) had large proportions of girls who did not answer the household surveys, and information on these predictor variables were not collected from them.[18] In addition, we note that subgroup predictors used here to measure the girl's or her household's status as per the earliest round in which data was collected for her.

[16] The exception to this is the FE girls, for whose analysis we include FE replacement girls recruited at ML1 or ML2, and their recontact rates at EL.

[17] These are included as a series of binary variables on whether a girl is included in each geographic zone, and whether she belongs to a particular cohort type (FE, ABE, NFE).

[18] See section 2.2, on Data Collection Tools.

FIGURE 1: PREDICTORS OF SUCCESSFUL RECONTACT AMONG C1 GIRLS



The results of Figure 1 highlight several points. Unsurprisingly, girls who were last contacted prior to ML2 have a significantly lower likelihood of being recontacted again at EL. The coefficient for this variable was -29.54, and the coefficient is significant at the 1 percent level. The coefficients for our two geographic zone binary variables are also significant. In large part, this is driven by comparatively low recontact rates among all cohorts in the Banadir region, as highlighted in Table 4. While it is difficult to ascertain the cause of these differences, one might speculate that because the Banadir region is the most economically developed region under study, it is possible that girls have more opportunities to pursue livelihoods opportunities outside of school, and as a result, are less likely to be recontacted by our field teams. The opposite may be true in South West State, as a dire economic outlook may mean that girls are more inclined to stay in their current communities. Given this finding of heterogeneous recontact rates across geographic zones, much of our analysis in the coming sections will be disaggregated by geographic zone to provide the reader with a more granular view of the trends in key outcomes, and to minimise conflating zone-specific trends with aggregate changes.

As with the ML2 evaluation, after adjusting for all other explanatory variables, we find that the AGES treatment arm and girl's age 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 either. In other words, households who underprioritise girls' education and teachers who are not particularly proficient educators are not associated with failed re-contacts.

2.4. Challenges and Limitations

Over the previous three 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 the last 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 to the greatest extent possible.

Apart from recurring challenges of tracking down girls, an additional challenge faced by evaluation team this round was the significantly condensed timeframe for data collection due to delays with the commissioning of the present EL evaluation. The evaluation team opted to field a much larger field team than in previous rounds to compensate for the shorter timeframe, while the CARE AGES programme team also assisted with mobilizing some girls to be surveyed at the schools or learning centres, to reduce the time field teams would otherwise spend searching for girls in the community. In spite of these mitigation measures, it is possible that higher recontact rates, and thus fuller panels, might have been achieved with a longer timeframe. Nonetheless, the evaluation team maintains confidence in the results of the analysis, and notes where small sample sizes or attrition prevent us from drawing firmer conclusions.

An additional challenge included the complete closure of one of the targeted schools, leading to the displacement of students, and thus lower recontact rates. Additionally, there were two instances of double registration of girls, wherein the same girls registered as beneficiaries under two separate names, and in prior rounds had been doubly surveyed, requiring additional data cleaning in the final dataset. Finally, the evaluation team encountered several cases of alleged fraud, where girls who were not part of the sample to be recontacted were alleged instructed by their teachers to pretend to be the girls who were sought after for this evaluation. While investigations into such instances did not lead to conclusive results, the evaluation team erred on the side of caution and opted to drop the observations where there was a suspected case of identity falsification.

Note from the project: CARE received reports from Consilient of a case of fraud where five girls, all from the same school, reported to the enumerators that they had been asked by a teacher to pretend to be the selected respondents. AGES staff immediately escalated the case through CARE's internal fraud reporting system (EthicsPoint), triggering an investigation process. All financial support to the affected school was halted immediately. The results of the investigation were reported to the Girls' Education Challenge Fund Manager.

FCDO Cohort Results

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

3. Learning

The AGES programme targets three main outcomes. The first is girls' learning in terms of numeracy and literacy (i.e. their proficiency in reading Somali). The second outcome is "transition": this includes promoting girls' progression from one grade to the next, integrating out-of-school girls into formal or alternative educational institutions, and helping older girls transition into gainful employment. The third is sustainability, which entails the creation of sustainable change through shifts in social norms at the community and individual levels, and the strengthening of institutional capacities for supporting inclusive education. In this section, we examine changes in learning outcomes – specifically Somali literacy and numeracy – among the initial groups of girls who joined the programme at 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, looking at what specific components of literacy and numeracy, the programme impacted the most. In the end, the analysis will focus on subgroup-specific programme impacts and on assessing the correlation between the programme's intermediate outcomes and changes in learning scores between BL and EL. This last piece of analysis aims at testing the Theory of Change of the program. 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 due to significant differences in both demographics and programme characteristics across cohorts.

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, the exact questions, such as the stories used in the literacy assessment reading comprehension tasks. 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. As a result, we have no reason to believe changes in the tests across rounds can affect any of the findings in this report.

A notable exception concerns the numeracy assessment, which was supplemented during ML1 by the addition of three more difficult EGMA subtasks. This change brought the number of numeracy subtasks from 8 to 11 and was made to guard against ceiling effects in ML2 and EL. 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 with the FCDO cohort is to make comparisons between baseline and EL, and doing so requires use of the same assessment across rounds.[19]

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 EL; this can include girls who fell out of the sample temporarily in ML1 or in ML2 (or in both) but were brought back

[19] 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 sampled during ML1 and ML2, respectively; thus, their baseline assessments during ML1 and ML2 utilised the full 11-subtask numeracy assessment.

into the sample in this EL round. The BL-to-EL panel sample includes 223 FE girls, 263 ABE girls, and 262 C1 NFE girls.

The second point concerns the utilization of benchmarks obtained 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 compensate for the lack of a comparison group that made it unfeasible to conduct a difference-in-difference analysis of an experimental or quasi-experimental setting. The benchmarks help account for 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 analysis will 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.

3.1. Learning Benchmark

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 expected, benchmark expectations for Grade 1 girls are consistently higher than those for Grade 2 girls: Grade 2 girls are starting from a higher baseline level of learning and their gains are consequently more difficult to achieve.

TABLE 5: BENCHMARK EXPECTED GAINS IN LEARNING SCORES FOR THE FE COHORT

	Benchmark – 3 Grade Progression	
	Grade 1 at BL	Grade 2 at BL
	Somali Literacy	
Aggregate	46.3	29.4
Banadir	57.5	31.7
Jubaland	43.4	27.6
Southwest	35.9	23.8
	Numeracy	
Aggregate	28.6	15.6
Banadir	36.0	17.0
Jubaland	29.9	15.0
Southwest	19.1	9.0

Differently from ML2[20], for each grade level, we have produced only one benchmark score. It captures expected improvements over the course of three years of consistent advancement, e.g., from Grade 2 to 5. If we wanted to follow the methodology used at ML2, we should have measured expected improvement over

[20] In Midline 2, we have produced two benchmark scores for each grade level. 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 course of four years as well. The four-year improvement benchmark would have followed more closely the expected path of girls, given that four years have passed since baseline. However, this was unfeasible due to the fact that benchmark girls assessed at baseline were attending classes between grade 2 and grade 5, as the expectation was for them to be evaluated over the course of three rounds, rather than four. Furthermore, due the context, compounded by the school disruptions caused by COVID-19 during this period, a constant advancement over years represents a relatively high bar for these girls, and for this reason also the previous ML2 evaluation included a more generous 2-year improvement benchmark.

The second panel reports equivalent benchmarks for numeracy. In both panels we calculate state-specific benchmarks for a wide variety of reasons, including differential school quality, that make the expected gains from an additional year of schooling heterogeneous across states. The limited sample size of girls, however, requires caution in interpreting these differences in benchmarks across states. 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.[21] 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 mirror 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 6: 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

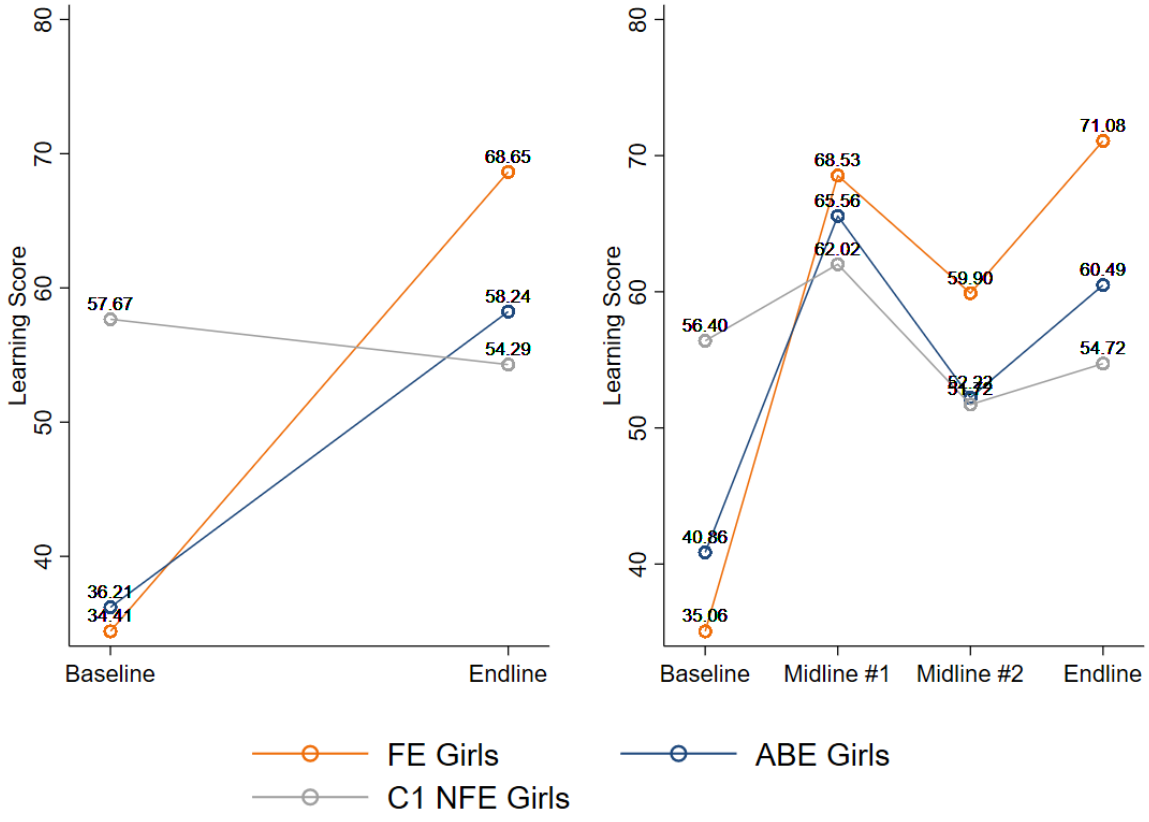
3.2. Aggregate Trends

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

The figure below illustrates the progression of learning outcomes over time for all three cohorts by displaying average Somali literacy scores across different rounds. The left side of the figure uses the BL-to-EL panel sample, which includes all girls successfully interviewed at both BL and EL. In contrast, the right side includes data from the ML1 and ML2 rounds, offering a slightly more detailed view of literacy trends over time. This right panel uses a different sample – girls who were successfully interviewed in all four rounds, resulting in a slightly smaller sample size per round.[22]

The figure reveals various trends. Firstly, both FE and ABE girls have shown improvements in learning since the baseline, with this improvement being more significant among the FE cohort, whose scores doubled from BL to EL. This substantial gain among FE girls is expected, as many have continued their education since 2019, while ABE programs are designed to be shorter. Conversely, the literacy rates of C1 NFE girls have slightly declined since the baseline.

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



Secondly, the right panel displays a distinct trend of decreasing learning levels between ML1 and ML2, followed by a subsequent increase. For FE girls, literacy scores rose more from ML2 to EL than they had fallen from ML1 to ML2, reaching a peak of 71.1 points. However, even with the recent increase, the average

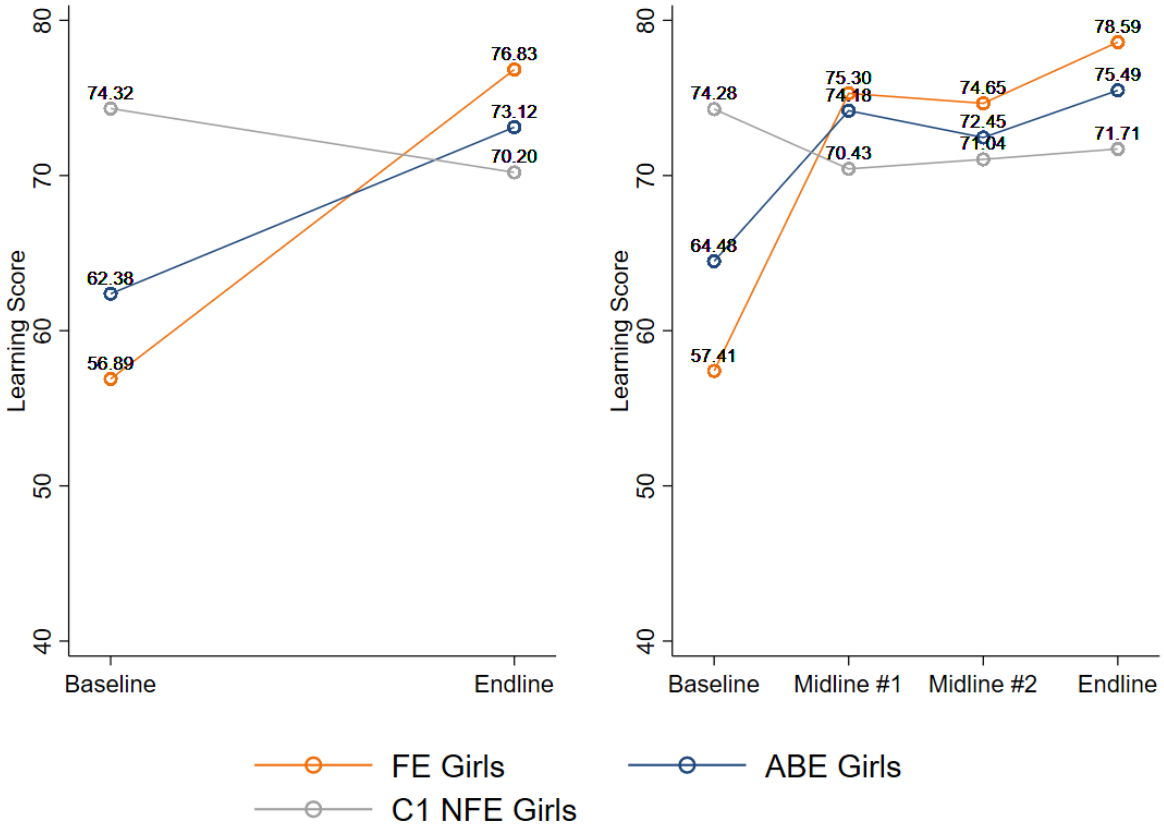
[22] For instance, the sample of FE girls in the left panel (BL-to-EL panel) is 223, while there are 186 girls in the right panel (BL-ML1-ML2-EL panel sample).

literacy score for ABE girls at EL is 60.5, which is still below their highest score of 65.6 at ML1. We delve deeper into this trend in our specific analysis of FE girls, where we observe a predictable disparity in learning gains between those who continued their education and advanced each year and those who did not.

The figure below provides a similar analysis for numeracy outcomes, showing comparable but less pronounced results. In the left panel, it is evident that FE girls have made significant improvements in numeracy (around 20 points), while the ABE cohort has seen smaller, yet notable, improvements of about 10 points. In contrast, the C1 NFE cohort has slightly declined over time. The right panel reveals that these gains for the FE and ABE cohorts are primarily concentrated in the BL-to-ML1 period. Since ML1, conducted in early 2022, both groups have shown minimal numeracy improvements, and only in the ML2-EL period. Notably, the ABE girls' improvements are less than those of FE girls, which is expected since many ABE girls are no longer participating in any learning programs. The modest improvement among FE girls is more concerning and cannot be attributed to ceiling effects: this trend is evident when using the full 11-subtask numeracy assessment introduced at ML1 to mitigate such issues.[23] On the other hand, the C1 NFE girls' numeracy scores decreased between BL and ML1 but slightly increased thereafter.

[23] Looking at the full 11-task EGMA assessment, the scores from ML1 to EL improved from 68.5 to 73.7 points for FE girls, from 66.9 to 69.2 for ABE girls, and from 62.3 to 65.3 for C1 NFE girls.

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



Formal Education Girls

In this subsection, we will delve deeper into the overall findings to better understand changes in learning outcomes within each cohort. We begin with formal school (FE) girls, examining the progression of their literacy and numeracy scores over time. The table below presents the score changes for the group of FE girls successfully re-contacted in all four rounds (top panel, 'BL-ML1-ML2-EL Panel') and for those contacted in both BL and EL ('BL-EL Panel'). As previously mentioned, we generally favour the 'BL-to-EL' panel for its larger sample size, although it might overlook important trends from the ML1 and ML2 rounds. We provide both literacy and numeracy scores here to summarize the overall gains made by FE girls before moving on to the benchmarked analysis.

TABLE 7: LITERACY AND NUMERACY SCORES AMONG FE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score	EL Score
	Literacy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	186	35.1	68.5	59.9	71.1

Banadir	75	33.5	69.2	59.4	70.3
Jubaland	72	36.9	74.0	64.3	76.4
South West	39	34.6	57.2	52.8	62.8
	Literacy Scores – BL-EL Panel				
Aggregate	223	34.4	N/A	N/A	68.6
Banadir	92	34.1	N/A	N/A	66.8
Jubaland	81	36.0	N/A	N/A	75.8
South West	50	32.4	N/A	N/A	60.4
	Numeracy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	186	57.4	75.3	74.7	78.6
Banadir	75	54.8	78.1	75.4	75.1
Jubaland	72	58.3	73.5	78.4	85.1
South West	39	60.8	73.1	66.2	73.2
	Numeracy Scores – BL-EL Panel				
Aggregate	223	56.9	N/A	N/A	76.8
Banadir	92	54.0	N/A	N/A	72.6
Jubaland	81	57.6	N/A	N/A	85.1
South West	50	61.0	N/A	N/A	71.2

Learning gains over time can naturally occur due to maturation effects. Due to the absence of a quasi-experimental setup with a control group, the AGES evaluation design included benchmarks to account for these effects, comparing the literacy gains of girls to the expected gains over the same period, established at baseline. The table below compares the 223 panel FE girls – those baseline girls successfully re-contacted at EL – to the alternative benchmarks described earlier. Notably, learning outcomes have significantly improved over time across all three geographic areas where FE programming is evaluated. Overall, Somali literacy scores have increased by 34.2 points since the baseline for this cohort.

TABLE 8: LITERACY IMPROVEMENTS AMONG FE GIRLS, RELATIVE TO BENCHMARKS, BL-EL PANEL

				Comparison to Benchmark
	Sample Size	**EL Score**	**Gain from BL**	**3 Grade Progression**
Aggregate	223	68.6	34.2	**-4.9***
Banadir	92	66.8	32.7	**-12.5***
Jubaland	81	75.8	39.8	4.4
South West	50	60.4	28.0	-6.2

The last two columns compare the 34.2-point increase – or the region-specific gains listed below – with the expected learning progress from advancing three grades (i.e., from Grades 1 or 2 to Grades 4 or 5). The learning improvements for FE girls are slightly below the benchmark expectations, with variations across regions. The shortfall is mainly due to the Banadir region, where scores are 12.5 points below the benchmark. In contrast, in Jubaland and South West, the difference between the actual gains and the benchmark is minimal. In Jubaland, the literacy score increase even slightly surpasses the benchmark expectations.

The table below presents similar findings for numeracy, comparing FE girls to pre-established benchmarks. Overall, the gains in numeracy align closely with benchmark expectations. However, the Banadir region shows a notable lag of 8.4 points, which is significantly different from zero and higher than in other regions. Conversely, and as seen for literacy, in Jubaland, the numeracy gains exceed the benchmark expectations by 5.2 points, a difference that is statistically significant.

TABLE 9: NUMERACY IMPROVEMENTS AMONG FE GIRLS, RELATIVE TO BENCHMARKS

				Comparison to Benchmark
	Sample Size	**EL Score**	**Gain from BL**	**3 Grade Progression**
Aggregate	223	76.8	19.9	-3.2
Banadir	92	72.6	18.5	**-8.4***
Jubaland	81	85.1	27.5	**5.2***
South West	50	71.2	10.2	-7.4

Compared to results at ML2, when FE girls were overall lagging behind the benchmark expectations, scores at EL are somewhat reassuring. As we noted previously, learning losses occurred between ML1 and ML2 were fully compensated by the gains of the last year.

The improvement in scores may be attributed to different factors. First, the impact of COVID-related school disruptions may be decreasing. As noticed in the ML2 report, 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. These disruptions are likely to have decreased during the last year. A second factor that helped explained the declining literacy scores and the stagnant numeracy scores between ML1 and ML2 was the particularly extreme drought conditions occurred during 2022, especially in some of AGES core programme areas. Even if still severe, the drought conditions have improved in 2023. The analysis by the Famine Early Warning Systems Network (FEWSNET) predicted that famine was no longer the most likely scenario, as it was in 2022.[24] The FEWSNET also reported that 2023 was the last year of a historic drought that brought Somalia on the brink of famine.[25] Above average rainfall in 2023 improved food security across much of rural Somalia from October 2023 to January 2024. Improved conditions are reflected in the coping mechanisms used by AGES households: the share of households that reduced their daily meals in the previous month decreased below 70 percent in EL from the 72 percent peak at ML2. Similarly, the share of girls reporting that they went to bed hungry frequently in the previous year has also decreased, after a sharp increase between ML1 and ML2. 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.[26]

[24] Famine Early Warning Systems Network (FEWS NET) and Food Security and Nutrition Analysis Unit (FSNAU). "Food assistance needs remain high amid ongoing recovery from drought." August 2023. Available at: https://fews.net/sites/default/files/generated-reports/2023/so-fso-2023-08-1694533503.pdf

[25] Famine Early Warning Systems Network (FEWS NET) and Food Security and Nutrition Analysis Unit (FSNAU). "Gradual drought recovery continues, though millions still need assistance". February 2024. Available at: https://fews.net/east-africa/somalia/food-security-outlook/february-2024

[26] At EL we asked girls about whether they consumed any form of protein-rich foods in the last 24 hours. Less than 10% of girls reported no consumption of protein-rich foods. FE and ABE girls who had not consumed any protein-rich foods had significantly

It's important to recognize that the benchmark girls do not serve as an ideal comparison for the FE girls for two main reasons. First, FE girls' learning outcomes depend on their continued enrolment and successful yearly progression, whereas all benchmark Grade 5 girls have, by definition, experienced relatively continuous enrolment. Therefore, the disparity between the gains of FE girls and benchmark expectations partly reflects their comparatively lower exposure to schooling, as some FE girls have not achieved the grade levels represented by our benchmarks.

The second issue is even more significant, as it pertains to the type of girl who reaches Grade 5. Even if we focus our FE girl sample on those who stayed in school and advanced the expected four grade levels – as we do in further analysis – the benchmark Grade 5 girls and FE cohort Grade 5 girls are not necessarily comparable. For example, benchmark Grade 5 girls progressed to Grade 5 without the AGES intervention, achieving this without a specific program. In contrast, when FE cohort girls reach Grade 5, part of their continued enrolment can be attributed to the AGES intervention. It is reasonable to assume that benchmark Grade 5 girls may come from wealthier families, have parents who place a higher value on education, or have a greater personal interest in education. Therefore, these higher-grade benchmark girls do not provide an exact comparison for the FE girls in terms of learning outcomes.

In the left panel of the figure below, we distinguish the Somali literacy growth rates of FE girls who advanced four grades between BL and EL (e.g., from Grade 1 to Grade 5) from the overall FE girl cohort. The figure presents Somali literacy scores over time for FE girls who achieved the expected grade advancement, all FE girls, and the 3-grade benchmarks.

The trends reveal two main points. Firstly, FE girls promoted four times since 2019 show a somewhat steeper increase in literacy compared to the entire FE group. The difference between the trend lines – comparing the orange line to the blue line in the figure – is understated, as the blue line includes all FE girls, even those promoted four times. When we separate these groups, FE girls in the promoted category gained 37.1 points in literacy from BL to EL, while those not promoted four times gained 30.3 points.[27] Essentially, FE girls who stayed enrolled and were consistently promoted came closer to meeting the expected benchmarks than those who dropped out or were not regularly promoted.

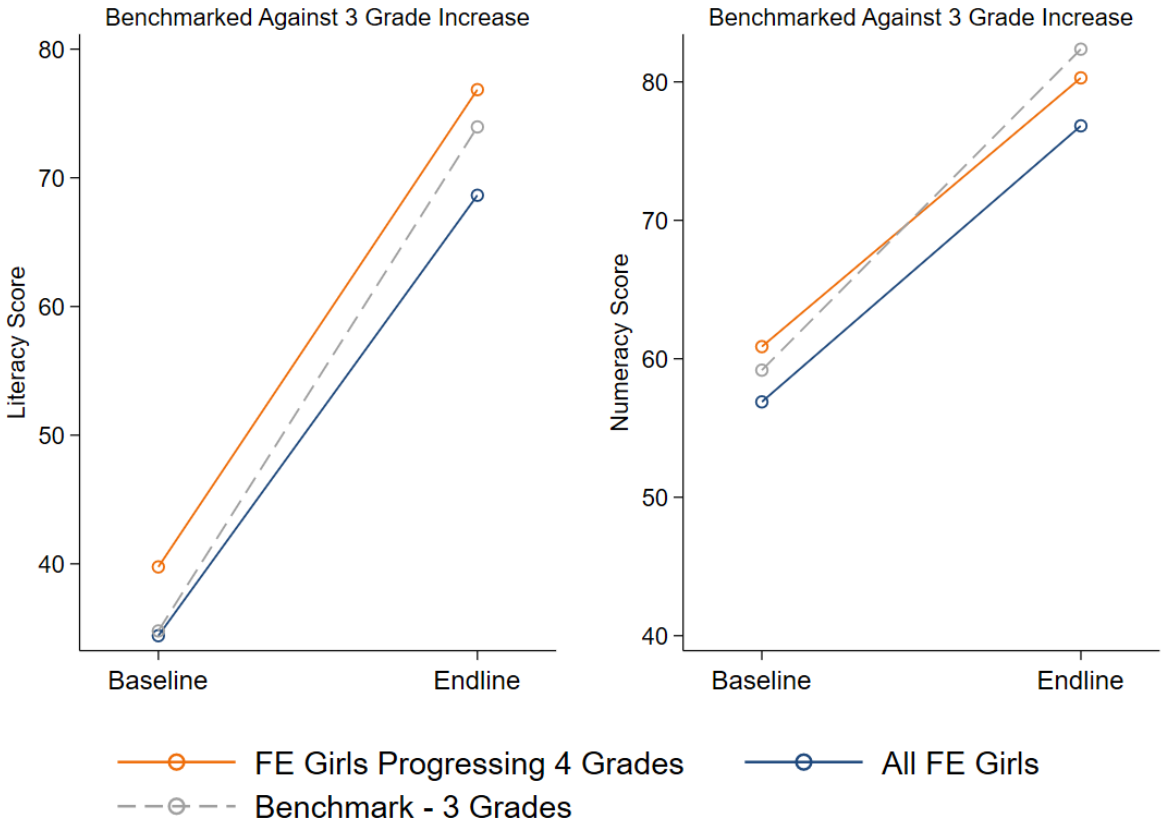
Second, comparing grade-promoted girls to the 3-grade benchmark in the left panel, the grade-promoted FE girls do not achieve the benchmark learning gains, but the difference in slopes between their learning gains (orange line) and the benchmark (grey dotted line) is almost zero.

The right panel of the figure presents the same analysis for numeracy outcomes. In numeracy, the difference between the promoted FE girls – those consistently promoted since BL – and the 3-grade benchmark is similarly small. Additionally, the gap between promoted and non-promoted FE girls is smaller for numeracy than for literacy. Grade-promoted FE girls gained 6.8 points more in literacy from BL to EL compared to other FE girls in the sample. However, in numeracy, grade-promoted FE girls did not improve more the other FE girls (19.4 points of improvement versus 20.7).

lower scores in literacy. However, we found no difference for scores in numeracy nor for C1 NFE girls. Moreover, the impact of protein intake on improvement in scores is not significantly different from zero.

[27] 129 out of 223 FE girls (57.8%) advanced four grades in four years, whereas 94 (42.2%) progressed less than expected.

FIGURE 4: BENCHMARKED IMPROVEMENTS IN LITERACY AND NUMERACY AMONG FE GIRLS FROM BL TO EL



Overall, FE girls recruited at baseline showed significant improvements in both Somali literacy and numeracy from baseline to EL. The gap between their learning improvements and the established benchmark is much smaller than it was at ML2, due to a substantial increase in scores over the last year, following the stagnation and decline between ML1 and ML2. FE girls with greater exposure to schooling experienced more significant gains, meeting the literacy benchmarks.

Accelerated Basic Education (ABE) Girls

Now focusing on the second original baseline cohort, we analyse the learning gains of ABE girls. The table below displays average literacy and numeracy scores by state and overall for ABE girls in each round of the evaluation. This table includes both panel samples previously discussed: the BL-EL panel, which comprises girls recruited at BL and re-contacted at EL regardless of their status during ML1 and ML2, offering the largest sample for observing gains between these two points. The BL-ML1-ML2-EL ("full") panel, instead, includes only those girls who participated in all four rounds, resulting in a smaller but more detailed sample with data from ML1 and ML2.

TABLE 10: LITERACY AND NUMERACY SCORES AMONG ABE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score	EL Score
	Literacy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	179	40.9	65.6	52.2	60.5
Banadir	70	39.4	65.7	48.1	52.9
Jubaland	42	45.5	62.4	49.9	63.4
South West	67	39.5	67.4	58.0	66.6
	Literacy Scores – BL-EL Panel				
Aggregate	263	36.2	N/A	N/A	58.2
Banadir	101	36.2	N/A	N/A	49.4
Jubaland	60	39.8	N/A	N/A	62.0
South West	102	34.1	N/A	N/A	64.8
	Numeracy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	179	64.5	74.2	72.4	75.5
Banadir	70	64.6	76.2	68.6	70.7
Jubaland	42	69.4	64.8	77.5	82.6
South West	67	61.3	78.0	73.3	76.0
	Numeracy Scores – BL-EL Panel				
Aggregate	263	62.4	N/A	N/A	73.1
Banadir	101	63.8	N/A	N/A	66.6
Jubaland	60	66.2	N/A	N/A	82.5
South West	102	58.7	N/A	N/A	74.0

In late 2019, ABE girls were recruited and enrolled in accelerated learning programs aimed at enabling their return to formal schooling after dropping out. Unlike the formal education for FE girls discussed earlier, the ABE program was intended to be short-term, concluding in 2021. The ABE program provided two levels of ABE, each with a duration of a year. At that point, the girls would either move into the formal education system (completion of level 1= into grade 3; completion of level 2 = into grade 5), continue with another educational route, or enter the labour market, among other possibilities. This distinction is significant, as it indicates that our expectations for learning outcomes over a span of three or more years should be more restrained for ABE (and C1 NFE) girls, since their exposure to specific educational interventions is shorter than that of FE girls.

This distinction is somewhat reflected in the benchmarks set for ABE girls, which are based on the expected learning change between Grade 1 and Grade 2 students—a much less ambitious benchmark compared to that for FE girls mentioned earlier. The table below details the gains in Somali literacy among ABE girls from BL to EL, broken down by state and compared to this benchmark. As shown in the table, ABE girls had lower literacy gains than FE girls—up to 30.7 percentage points in Southwest State, and 22.0 points overall. While the literacy gains are generally not significantly different from the benchmark, the results vary across regions. In Southwest State, the literacy improvement was higher, and the benchmark less demanding: here, the ABE girls' literacy score improvement exceeded the benchmark by 21.6 percentage points. Conversely, the gain for ABE girls in Banadir was 14.7 percentage points below the benchmark.

Even though the girls matured one year from ML2 to EL, we retained the same benchmark improvement as in ML2. Thus, the benchmarks used here are even less demanding than those for FE girls, making the results from the Banadir region particularly concerning. Below, we will analyse the performance differences between ABE girls with continuous education and those with intermittent enrolment. Additionally, we will examine the transition outcomes for ABE girls in the following sections.

TABLE 11: LITERACY IMPROVEMENTS AMONG ABE GIRLS, RELATIVE TO BENCHMARKS, BL-EL PANEL

				Comparison to Benchmark
	Sample Size	**EL Score**	**Gain from BL**	**3 Grade Progression**
Aggregate	263	58.2	22.0	1.5
Banadir	101	49.4	13.2	**-14.7***
Jubaland	60	62.0	22.2	-5.6
South West	102	64.8	30.7	**21.6***

The findings from literacy scores are generally similar in the case of numeracy outcomes, as reported by the table below. Compared to benchmarks, gains in numeracy scores were concentrated exclusively in Southwest State (+10.3 percentage points) and the lag in Banadir is significantly large (-19.2 percentage points). Looking at the whole sample, ABE girls fell only 3.5 points below the expected gains in numeracy.

The path of ABE girls' scores across rounds made their analysis and their comparison to the benchmark somewhat more complicated. As already noticed, the ABE girls scores decreased between ML1 and ML2 and improved again during the last year. The difference with the benchmarks then depends on the last year improvement and how this relates to their ability to remain in a learning programme as analysed in the next paragraphs.

TABLE 12: NUMERACY IMPROVEMENTS AMONG ABE GIRLS, RELATIVE TO BENCHMARKS

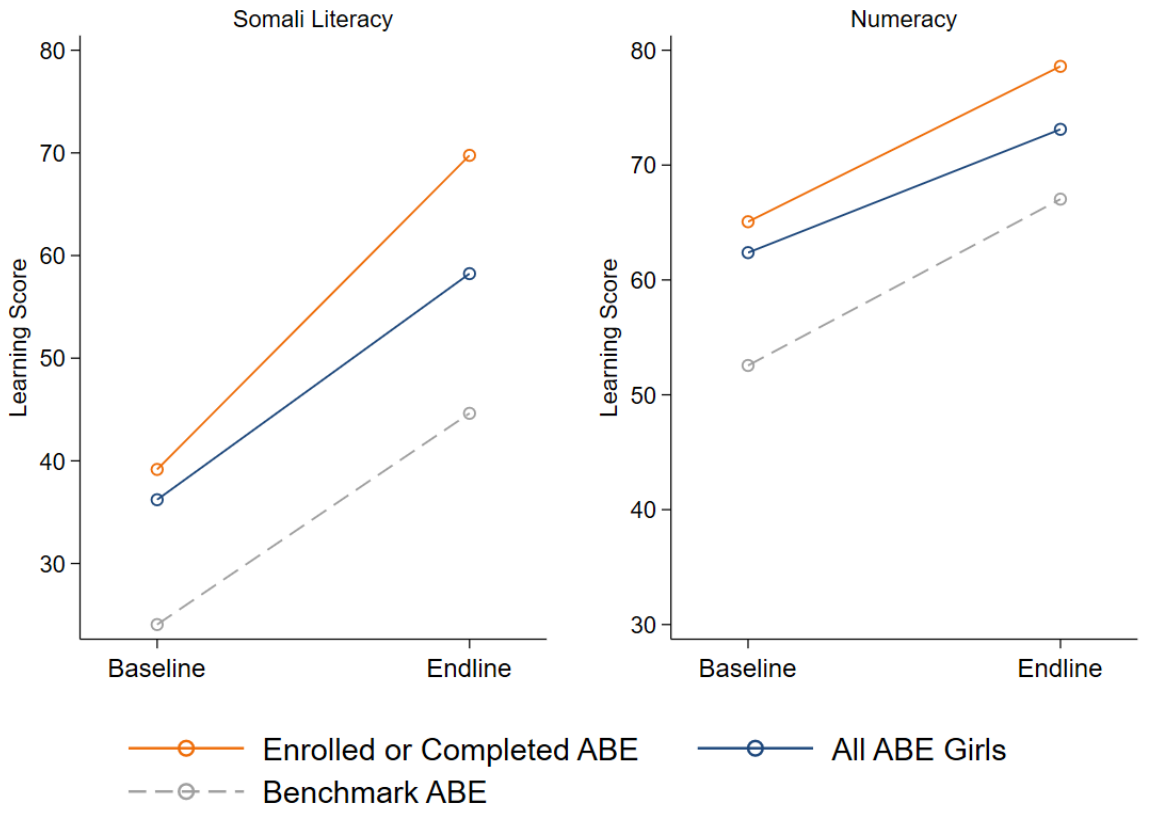
				Comparison to Benchmark
	Sample Size	**EL Score**	**Gain from BL**	**3 Grade Progression**
Aggregate	263	73.1	10.7	-3.7
Banadir	101	66.6	2.8	**-19.2***
Jubaland	60	82.5	16.3	-1.5
South West	102	74.0	15.3	**10.3***

The ability to meet the benchmarks largely depends on the differing outcomes within the ABE cohort. The figure below illustrates the Somali literacy and numeracy scores at BL and EL for two groups of ABE girls. The first group includes those who reported completing an ABE programme or are currently enrolled in ABE, NFE, or formal school. The second group encompasses all ABE girls, including 125 who reported not finishing the ABE programme during the latest data collection. Thus, we can consider two categories of ABE girls in our panel sample: 138 who are either still enrolled or have completed the ABE programme, and 125 who did not complete it.

When we analyse the learning changes separately, girls who are engaged in or have completed the ABE programme exhibit greater learning gains, slightly surpassing the benchmark expectations. In contrast, those who did not complete the ABE programme and are not enrolled in any educational programme show much lower improvements, particularly in literacy. The number of girls who did not complete ABE cohort 1 is 2,317 which is 32% of Cohort 1 ABE. This is based on the number of girls who did not sit for the final level 2 assessment.

As with our analysis of FE girls' learning paths, this figure downplays the difference between the two groups. From BL to EL, ABE completers—those who finished ABE or remain in a learning programme—gained 9.9 points in Somali literacy and fell short by 1.0 points in numeracy compared to the benchmark. On the other hand, ABE non-completers lost 7.7 points in literacy and 6.6 points in numeracy against the expected benchmark gains over the same period.

FIGURE 5: BENCHMARKED IMPROVEMENTS IN LITERACY AND NUMERACY AMONG ABE GIRLS



As mentioned earlier in the context of FE girls, it is clear that girls who fully engage with the intervention programme exhibit learning gains. While this is promising, it underscores that the programme's success relies on the continued participation of the girls. This situation resembles the methodological difference between an "average treatment effect" and an "intent-to-treat" effect. In the latter, an intervention might be effective

for those who fully participate, but its overall impact decreases or becomes negligible when non-compliance is taken into account. This indicates a potential area for further research, as there may be significant variations in outcomes among girls who follow different educational paths after the ABE intervention:

- Dropping out of ABE and not joining any other educational programme
- Completing ABE
- Completing ABE and continuing in higher-level ABE streams
- Completing ABE and transitioning into the formal education system

Due to the small size of the ABE panel sample, our data can only provide indicative evidence about the learning outcomes related to these pathways. This limitation is further compounded by the evaluation team's inadequate documentation of the girls' educational experiences after ABE. Although the data indicates whether a girl is currently enrolled in school, it does not detail the timing of her enrolment or the duration of any gap between completing ABE and starting school.

Despite these limitations, our analysis provides indicative evidence of differences across educational pathways. For ABE girls who, at EL, report being enrolled in formal school, literacy scores have increased by 27.5 points and numeracy scores by 5.6 points since BL. Those who continue to attend an ABE programme at EL show similar progress, with slightly higher gains in both literacy and numeracy. These findings imply that ABE programming can significantly enhance learning, especially when girls use ABE as a bridge to formal schooling or continue in higher-level ABE courses.

Cohort 1 Non-Formal Education (C1 NFE) Girls

The last group of girls from the original baseline cohorts includes those enrolled in non-formal education (C1 NFE). The table below presents their literacy and numeracy scores over time, on aggregate and divided by geographic zone. For both outcomes, we examine two different panel samples, similar to the analysis conducted for FE and ABE girls.

TABLE 13: LITERACY AND NUMERACY SCORES AMONG C1 NFE GIRLS, USING ALTERNATIVE SAMPLES, OVER TIME

	Sample Size	BL Score	ML1 Score	ML2 Score	EL Score
	Literacy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	176	56.4	62.0	51.7	54.7
Banadir	77	62.5	69.5	55.7	55.0
Jubaland	50	50.1	54.3	49.3	57.0
South West	49	53.2	58.1	47.9	52.0
	Literacy Scores – BL-EL Panel				
Aggregate	262	57.7	N/A	N/A	54.3
Banadir	115	63.5	N/A	N/A	54.8
Jubaland	72	51.9	N/A	N/A	54.6
South West	75	54.2	N/A	N/A	53.1
	Numeracy Scores – BL-ML1-ML2-EL ("Full") Panel				
Aggregate	176	74.3	70.4	71.0	71.7
Banadir	77	77.2	78.3	72.2	68.2

Jubaland	50	70.5	54.0	75.5	81.0
South West	49	73.5	74.8	64.7	67.7
	Numeracy Scores – BL-EL Panel				
Aggregate	262	74.3	N/A	N/A	70.2
Banadir	115	76.5	N/A	N/A	65.5
Jubaland	72	71.5	N/A	N/A	79.5
South West	75	73.6	N/A	N/A	68.4

Unlike the FE and ABE girls previously discussed, the C1 NFE girls did not have specific learning improvement benchmarks set at baseline. In this section, we will focus on documenting changes in their learning outcomes since baseline without referring to any particular benchmark.

The absence of a benchmark is less significant here because the learning outcomes for this cohort have generally either worsened or remained unchanged over time. As reported in the table below, in our panel sample of 262 C1 NFE girls, Somali literacy scores have dropped by 3.4 points since BL. This decline is most notable in the Banadir region, where scores fell by 8.7 percentage points. Conversely, girls in Jubaland experienced a slight increase of 2.7 percentage points, while those in the Southwest region remained almost unchanged, with a decrease of 1.1 percentage points. The especially poor performance in the Banadir region is notable because it is the only area where literacy scores did not improve between ML2 and EL.

TABLE 14: LITERACY IMPROVEMENTS AMONG C1 NFE GIRLS SINCE ML1

	Sample Size	Baseline Score	EL Score	Gain from BL
Aggregate	262	57.7	54.3	-3.4
Banadir	115	63.5	54.8	-8.7
Jubaland	72	51.9	54.6	2.7
Southwest	75	54.2	53.1	-1.1

The pattern of numeracy scores is very similar, as shown in the table below. Across all girls, numeracy scores have declined by 4.1 points, with a particularly large decline in Banadir, smaller one in Southwest, and an increase in Jubaland. As for literacy, this reflects the fact that only in the Banadir region scores have decreased between ML2 and EL.

TABLE 15: NUMERACY IMPROVEMENTS AMONG C1 NFE GIRLS SINCE ML1

	Sample Size	Baseline Score	EL Score	Gain from BL
Aggregate	262	74.3	70.2	-4.1
Banadir	115	76.5	65.5	-11.0
Jubaland	72	71.5	79.5	8.0
Southwest	75	73.6	68.4	-5.2

A similar trend to that observed with ABE girls is also seen among C1 NFE girls, though to a lesser extent. C1 NFE girls who remain enrolled in any learning programme at EL (whether FE, ABE, or another NFE programme) exhibit substantial gains in literacy (+11.6 points) and moderate gains in numeracy (+3.2 points) since baseline. It is important to note that continued enrolment in learning programmes was not a primary objective for NFE girls, as they were initially older adolescents or young adults. Despite this, a portion of the sample (38 out of 262 girls, or 14.5 percent) has transitioned into formal schooling, an unexpected result that may suggest that the programme has the potential to increase girls' awareness about the importance of formal education.

Meanwhile, the majority of NFE girls have either moved into employment (122 girls) or are out of school and not employed (76 girls). Among these groups, there has been a notable decline in literacy and numeracy over time. For example, girls who are currently employed or self-employed have seen their literacy scores drop by 11.8 points and numeracy scores by 3.1 points since baseline. This suggests that some form of continued education, even informal, is crucial for maintaining basic literacy and numeracy skills as these girls transition into adulthood. On the other hand, the rare use of these skills in employment may not be sufficient to sustain long-term proficiency.

3.3. Subtask-Specific Changes in Learning

The previous section examined overall changes in literacy and numeracy outcomes. This section shifts the focus to individual subtasks to better understand the specific learning improvements resulting from the programme's educational interventions.[28] For example, although FE girls showed significant improvements in numeracy from BL to EL, these gains might be primarily concentrated in specific numeracy skills, such as performing addition with 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 EL.

The literacy subtasks assess the following skills:

- Letter Sound Identification
- Words Commonly Used

[28] In Annex 4 we present tables with the EL levels of proficiency for each subtask.

- Reading fluency
- Reading Comprehension - Level 1
- Reading Comprehension - Level 2
- Reading Comprehension - Level 3

We begin by examining the progress of FE girls since the baseline evaluation. At baseline, we observed a significant decline in performance between the first and second Somali literacy subtasks, as the assessment moved from letter identification to the identification and reading of common words. This transition marked the sharpest drop in performance on the EGRA for FE girls. After this point, their performance declined more gradually, with a small subgroup performing relatively well across all six subtasks.

The left panel of the figure below shows changes in subtask-specific scores for FE girls in literacy. As we did when looking at aggregate scores, we report results for two groups: those who remained enrolled and advanced through four grade levels between 2019 and 2024, and those who either dropped out or did not progress as expected. For each subtask, we display the mean change in scores from BL to EL, broken down by these two groups. The vertical line labelled "no gain" indicates no change in proficiency since baseline. The mean change in scores is marked by a hollow circle, with horizontal bars around each circle representing the 95 percent confidence interval. If the confidence interval does not intersect the vertical line at 0 ("no gain"), the change from BL to EL is statistically significant.

To clarify, consider subtask 2 of the literacy assessment. For this subtask, FE girls who advanced four grade levels since BL improved their scores by 48.3 points. In contrast, those who did not progress four grade levels saw a smaller increase of 36.6 points. In both cases, the changes since baseline are statistically significant.

Interestingly, the figure shows that both groups improved across all literacy subtasks. Girls who were continuously enrolled and promoted year-on-year showed greater improvements than other FE girls in all subtasks, except for the simplest task of letter identification (subtask 1). Compared to ML2, where learning plateaued between subtasks 5 and 6, all girls now show significant improvements across all subtasks.

The most significant differences in gains between the two groups appeared in subtasks 3 and 4. For example, in subtask 3, which measures reading fluency, girls read as many words as possible from a short story in one minute. Promoted FE girls improved from reading 23 words per minute (29.1 percent of the words in the text) at baseline to 66 words per minute (the 82.4 percent of the words in the text) at EL. Non-promoted girls increased their reading speed from 12 words (16.0 percent) to 39 words per minute (49.2 percent) over the same period.[29]

The right panel of the figure shows less distinct patterns and fewer significant differences between the two groups. However, a positive finding is the overall increase in proficiency in more challenging tasks (subtasks 5 to 8), which involve 2-digit addition, subtraction, and word problems. At baseline, there was a significant drop from proficiency in 1-digit addition (66 percent) to 2-digit addition (23 percent). From BL to EL, girls improved across all addition and subtraction tasks, with the most notable gains in the 2-digit problems.

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.

[29] *Note from the project: The difference in baseline reading fluency for promoted and non-promoted FE girls also reflects the higher vulnerability of the non-promoted, who potentially did not benefit to the same extent from the decoding skills usually acquired during Qur'anic school.*

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 4 grade levels between 2019 and 2024, 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 EL, 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.

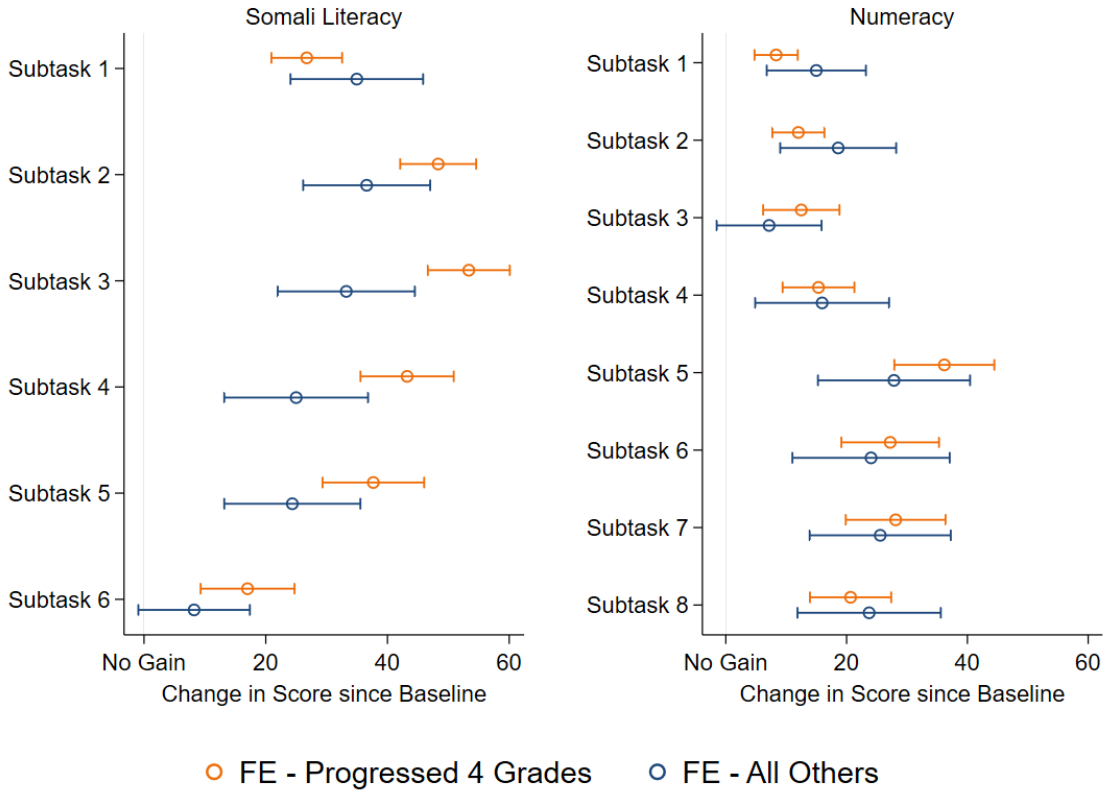
To ensure the interpretation is clear, consider an example using subtask 2 of the literacy assessment. For this subtask, the change in scores among FE girls who progressed 4 grade levels since BL is 48.3 points. In contrast, among the FE girls who did not progress 4 grade levels, the increase in scores on subtask 2 is lower at 36.6 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 all literacy subtasks. FE girls who have been enrolled and promoted year-on-year consistently, improved more than other FE girls in all subtasks, except for the easiest letter identification (subtask 1). Compared to ML2, where a plateau in learning was identified between subtasks 5 and 6, now all girls improved significantly in all subtasks.

The biggest difference in gains between the two types of girls resulted in Subtasks 3 and 4. 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 29.1 percent of the words in the story at baseline; by EL, this had increased to 82.4 percent. 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 16.0 percent to the 49.2 over the same period.

The patterns shown in the right panel of the figure are less clear-cut and with not significant differences in numeracy scores between the two types of girls. The most positive finding concerns the general increases in proficiency in more difficult tasks subtasks (subtasks 5 to 8) that focus on 2-digit addition and subtraction and on the word problem. 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 EL, girls improved across all addition and subtraction subtasks, but these gains were most pronounced in the 2-digit versions of these problems.

FIGURE 6: SUBTASK-SPECIFIC LEARNING IMPROVEMENTS, BL TO EL, AMONG FE GIRLS

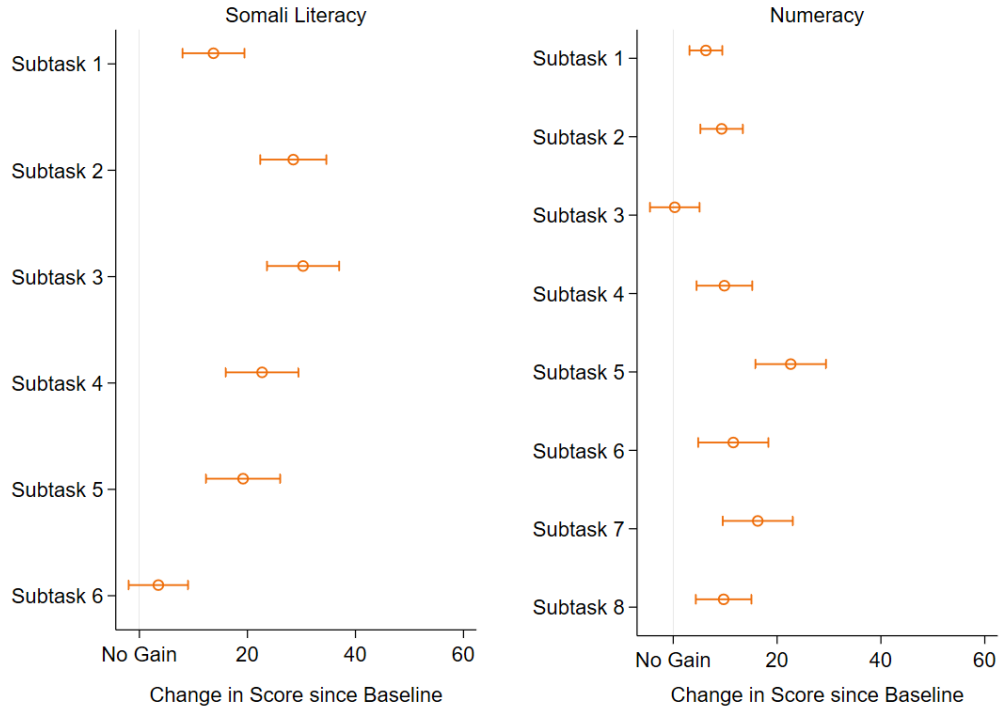


Turning to the ABE girls, we now examine changes in subtask-specific scores for the entire cohort without breaking them down by progression through the programme or transitions to other educational opportunities. Similar to the FE girls, ABE girls show no significant improvement in the most challenging literacy subtask (#6). However, they do exhibit significant improvements in all other literacy subtasks. The greatest gains are seen in subtasks 2 and 3, which involve word identification and reading fluency, respectively, indicating the programme's effectiveness in helping them identify words. While the improvements in reading comprehension (subtasks 4-6) are smaller, they are still significantly better than zero, unlike the results from ML2, where ABE girls showed much lower gains in these areas.

The right panel of the figure shows subtask-specific improvements in numeracy scores among ABE girls. Overall, the results are positive, with improvements in all subtasks except for subtask 3 (Missing Numbers). The most notable progress is in subtask 5, which tests the ability to complete simple (1-digit) subtraction. Girls showed significant improvement in this area and, compared to ML2 results, also improved in slightly more difficult addition and subtraction problems involving two digits.

It is expected that there are fewer substantial gains in subtask-specific literacy or numeracy skills among ABE girls, given their less significant overall improvement on EGRA or EGMA in our earlier analysis.

FIGURE 7: SUBTASK-SPECIFIC LEARNING IMPROVEMENTS, BL TO EL, AMONG ABE GIRLS



3.4. Testing the Theory of Change

Youth Leadership Index

The Youth Leadership Index (YLI), ranging from 0 to 100, estimates girls' self-assessed leadership skills.[30] The index is based on their responses to 21 questions regarding their thoughts on the consequences of their actions, their confidence in clearly expressing their ideas, and their ability to organize others to achieve a common goal. In this section, we test the Theory of Change by examining whether YLI predicts improvements in learning outcomes. We also explore the relationship between the YLI scores of FE, ABE, and C1 NFE girls and the changes in their learning outcomes from BL to EL.

TABLE 16: 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	-0.06	0.732	0.02	0.908

[30] See Section 6.3 for a detailed description of the index.

ABE girls	-0.11	0.469	-0.30	0.054
C1 NFE girls	0.00	0.989	-0.26	0.142

To assess the impact of YLI scores on changes in learning outcomes since BL, we employ a linear regression model, controlling for girls' age and region. As indicated in the table above, an additional point in YLI does not significantly correlate with any increase in numeracy or literacy scores across all groups of girls. The only notable effect is a minor and unexpectedly negative impact on the numeracy score improvement among ABE girls.

GEF Participation

Following the discussion on life skills in the previous section, we now explore girls' participation in the Girls' Empowerment Forum (GEF) and its impact on learning outcomes for FE, ABE, and C1 NFE girls. GEFs are the main avenue through which AGES aims to enhance self-esteem, leadership abilities, and life skills among girls. These after-school programmes provide a peer support network, positive female role models, and opportunities for tutoring and direct mentoring.

Research from both the SOMGEP-T programme and AGES indicates that girls participating in GEFs tend to perform better in learning assessments. Our study reveals similar trends for ABE and C1 NFE girls who participated in the programme at BL and maintained contact with GEF members. Unlike the results from ML2, where the positive effect on learning outcomes persisted roughly two years after their participation in ABE or C1 NFE ended, we did not find any significant impact at EL. The effect remains statistically insignificant among those still in contact with GEFs.

Similarly, FE girls who participated in a GEF did not show significantly greater improvements in learning outcomes compared to their non-participating peers. It's important to note that FE girls are of school age, and the FE programme is designed to help them stay in school. Consequently, FE girls receive support from formal schooling to enhance both their learning outcomes and leadership skills. This consistent support in formal school, regardless of GEF participation, may explain why the effect of their GEF participation appears muted. 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. Overall, it is important to note that the lack of significance in the effect of GEF participation on improvement in scores may be due to the fact that only the 10 percent of FE, ABE and C1 NFE girls reported being in contact with GEF at EL.

TABLE 17: 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	0.87	0.817	2.28	0.569
	ABE girls	4.36	0.207	5.06	0.324
	C1 NFE girls	4.29	0.324	10.02	0.070

GEF Participation Continued	FE girls	-5.03	0.138	1.87	0.654
	ABE girls	3.20	0.375	7.66	0.193
	C1 NFE girls	-5.72	0.115	1.04	0.836

Teaching Quality

The AGES Theory of Change posits that enhanced teaching methods will result in better learning outcomes for girls. The idea is that if teachers are better trained in teaching numeracy and literacy, are consistently present, create a safe and welcoming classroom environment, and actively encourage schooling, students will learn more and have higher school attendance rates. This endline assessment included a series of questions for girls about their teachers' pedagogical practices, as listed below. In this section, we examine the relationship between these student-reported measures of teaching quality and the improvement in learning scores from BL to EL.

For each measure of teaching quality, we use a linear regression to predict changes in literacy and numeracy scores. The teaching quality measure serves as the predictor variable, while age, state, and cohort are included as control variables.[31]

TABLE 18: EFFECT OF TEACHING PRACTICES ON CHANGES IN LEARNING OUTCOMES, BY COHORT

Teacher quality	Cohort	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
My teacher does not make me feel welcome in classroom	FE Girls	2.95	0.597	-4.83	0.463
	ABE Girls	-0.65	0.900	-0.64	0.909
	C1 NFE Girls	5.28	0.349	**20.57***	0.024
My teachers are often absent	FE Girls	3.63	0.280	1.97	0.639
	ABE Girls	8.40	0.077	**17.83***	0.014
	C1 NFE Girls	0.92	0.843	9.21	0.111
My teacher rarely/never encourages participation	FE Girls	6.49	0.225	-2.76	0.751
	ABE Girls	-3.28	0.478	7.82	0.212
	C1 NFE Girls	**16.39***	0.004	11.64	0.243
	FE Girls	**-21.65***	0.072	**-44.11***	0.001

[31] 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 explains how learning things is useful in our lives	ABE Girls	0.51	0.967	6.95	0.585
	C1 NFE Girls	7.93	0.173	3.64	0.814
My teacher's lessons move too fast for me	FE Girls	**-14.16***	0.000	**-15.85***	0.001
	ABE Girls	**-11.00***	0.002	**-13.04***	0.026
	C1 NFE Girls	**-16.17***	0.000	**-16.06***	0.015
My teacher punishes students who get things wrong in a lesson	FE Girls	-0.03	0.994	1.49	0.783
	ABE Girls	4.46	0.205	6.61	0.236
	C1 NFE Girls	-0.85	0.805	-0.56	0.903
My teacher used corporal punishment in last week	FE Girls	8.38	0.069	5.06	0.299
	ABE Girls	8.85	0.188	12.51	0.155
	C1 NFE Girls	-4.53	0.735	-8.97	0.565

Our findings, as shown in the table above, generally do not reveal a strong connection between teaching quality measures and learning outcomes. The only notable result is related to girls who reported that their teachers' lessons moved too quickly. These girls showed significantly smaller improvements in both literacy and numeracy compared to those who did not report this issue, with literacy scores change decreasing by 11.0 percentage points for ABE girls and by 16.2 percentage points for C1 NFE girls.

Some measures of teaching quality have a counterintuitive correlation with scores improvements. For example, the use of corporal punishment[32] is correlated positively with scores improvement of FE and ABE girls, though the coefficients are not significant. One possible explanation for this finding could be that in certain contexts, the immediate compliance and increased focus resulting from the fear of punishment might temporarily enhance academic performance. Additionally, in environments where corporal punishment is culturally normative and widely accepted, students may perceive such discipline as a sign of teachers' investment in their success, thereby motivating them to perform better. However, it is important to note that these short-term gains are achieved at the cost of potential long-term negative effects on students' psychological well-being and the overall learning environment.[33]

Other measures are not consistently related to improvements in 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.

[32] To hit hands of students with a stick was reported by several girls interviewed. We considered this practice as corporal punishment as well.

[33] We also tested whether negative teaching practices reported at BL are related to girls drop-out and to the likelihood of being interviewed at both BL and EL. We do not find any evidence on such relations that could have been explained, at least partly, the counterintuitive results.

Community Attitudes

The Theory of Change suggests that positive shifts in community attitudes toward girls' education will lead to increased school attendance and improved learning outcomes for girls. When caregivers perceive girls' education as valuable compared to other household demands, they are more likely to send their daughters to school and support their educational endeavours instead of prioritizing household chores, marriage, or work. Caregivers who value education are also more likely to provide the necessary support to their daughters, whether financial or otherwise.

In this section, we evaluate whether community attitudes, particularly those of caregivers, toward girls' education predict improvements in learning outcomes from BL to EL. We use a linear regression approach to explore the relationship between community attitudes and changes in learning outcomes, controlling for age and region.

TABLE 19: 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	5.64	0.359	11.96	0,065
	ABE girls	-5.56	0.186	-11.33	0.158
	C1 NFE girls	**-7.79***	0.035	-11.75	0.065
Girls' education is worthwhile investment even if funds limited	FE girls	9.83	0.171	12.49	0.072
	ABE girls	6.69	0.236	14.72	0.028
	C1 NFE girls	-5.98	0.303	-11.71	0.074
Work/chores acceptable reason for non-enrolment	FE girls	4.40	0.191	5.79	0.211
	ABE girls	-0.09	0.978	1.16	0.829
	C1 NFE girls	3.10	0.496	7.85	0.176
Expense acceptable reason for non-enrolment	FE girls	0.40	0.912	3.96	0.291
	ABE girls	-6.11	0.069	-4.93	0.362
	C1 NFE girls	**-8.03***	0.041	-7.13	0.133

We find no clear pattern in correlations between caregiver attitudes and either numeracy or literacy scores for the FE, ABE, and C1 NFE cohorts. Surprisingly, not all the signs of the correlations between some caregivers' attitudes and improvement in scores are also not in line with expectations. For example, except for FE girls, we find that the attitude of a caregiver in which the caregiver aspires for a university education

for their girls is negatively correlated with increases in both numeracy and literacy scores, although the coefficient is significantly different from zero only for numeracy score of C1 NFE girls. On the other hand, some coefficients are as expected and significantly different from zero. This is the case for the positive coefficient from the regression run with ABE girls interacting the caregivers' belief that girls' education is worthwhile investments even when funds are limited and the change in literacy score; and for the negative coefficient from the regression run with C1 NFE girls interacting the caregivers' belief that excessive expenses are an acceptable reason for non-enrolment and numeracy scores.

4. Transition

The second outcome of focus for AGES concerns retention and life pathways, or "transition". As a core outcome of several GEC, GEC-T and LNGB projects, transition aims to capture how the project affects girls' trajectory 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.

The analysis in this section is broken into three parts: aggregate transition outcomes, subgroup transition outcomes, and testing the Theory of Change on transition outcomes. First, 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 in 2019. Second, 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, providing evidence for or against the programme's Theory of Change.

4.1. Aggregate Transition Outcomes

Defining a successful transition outcome is complex, as transition is an inherently complicated, contextual, and multifaceted outcome. In much of our analysis, we will define transition binarily – 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.

Table 20 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] A formal schoolgirl – with a mean age of 11.1 years old at BL – who drops out of school to pursue employment is not considered a success because she is comparatively young and employment after completing just 3-4 years of primary schooling is not sufficient. In contrast, an NFE girl – with a mean age of 17.9 years old at BL – 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.

[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.

TABLE 20: 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 • Drops out but is enrolled in a more advanced level of ABE	• 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, if the grade of enrolment is more advanced than the ABE level. • Transition into a technical or vocational education programme • Retention in ABE with progression to advanced levels • Transition into age-appropriate, non-exploitative employment • Transition into self-employment	• Drop out • Idleness after programme completion • Transition into NFE • Retention in ABE without appropriate level advancement
NFE Girl **Enrolled in NFE at baseline**	• Enrolment in formal school, at any grade level • Transition into ABE, at any 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 • Retention in NFE

In the analysis below, unlike from the previous rounds of evaluation which utilised a liberal and conservative definition of success,[35] we are now considering only one definition of successful transition for ABE and C1 NFE girls. This distinctly includes any transition into vocational training programmes – regardless of cohort type – as a successful outcome. For NFE girls, transition into ABE programming is considered a successful outcome; for ABE girls, level progression in the ABE programme is considered successful while retention without advancement is an unsuccessful transition.[36] Additionally, ABE girls transitioning into formal school

[35] See Consilient Research, *AGES Midline 2 Report* (2023).

[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 (with the exception that the girl is shifting to a vocational programme).

enrolment is only successful if the grade of enrolment is more advanced than the ABE level the girl was originally in (i.e., a girl moving from ABE Level 2 to FE Grade 5 is a successful transition outcome).[37]

Our approach to analysis considers transition *since* baseline; for this evaluation round, we set aside transition outcomes that are specific *between* either ML1 and ML2, ML1 and EL, or ML2 and EL. The reason is hinted at by our discussion of FE girls' grade 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 (as of EL) 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 and ML2 rounds. To ensure our results are not driven by this decision – which excludes 136 FE replacement girls – we also check our main findings related to FE with the expanded sample and its more complicated structure.[38]

In this aggregate analysis, 223 FE girls, 263 ABE girls, and 262 C1 NFE girls were included using the panels of girls tracked with 1) either no attrition since BL or 2) attritted during either/both ML rounds but were brought back into the sample at EL. Using the definition of successful outcomes outlined above, 57.0 percent of FE girls, 32.3 percent of ABE girls, and 61.8 percent of C1 NFE girls reported having a successful transition outcome.

Table 21 reports on the transition outcomes of the FE girls, as well as the proportion of FE girls with a successful transition outcome. Among FE girls, 52.9 percent remain enrolled in school and have progressed in grade level; however, the 22.0 percent remain enrolled but were held back, and 19.7 percent are out-of-school without employment.

TABLE 21: TRANSITION OUTCOMES AMONG FE GIRLS[39]

Transition Outcome	Number of Girls	Share of Sample (%)
n=223		
Positive grade progression	118	52.9
Enrolled, held back	49	22.0
Now OOS	44	19.7
Now employed	8	3.6
Employed, but underage (under 18 years)	2	0.9
Self-employed	1	0.5

[37] ABE programming is accelerated education by design, with each level of ABE being the equivalent of completing two FE grade levels. The four levels of ABE (with their FE equivalent) includes: L1 (Grades 1 & 2), L2 (Grades 3 & 4), L3 (Grades 5 & 6), and L4 (Grades 7 & 8).

[38] This structure is slightly more complicated by the fact that replacement girls were recruited in ML1 and ML2 and we lack information about them at baseline.

[39] The below measures of various transition outcomes may include overlaps. For instance, a girl may claim to both be employed and still enrolled in school. In our classification of transition outcomes, (self-)employment is only considered as a transition outcome if a girl did not concurrently report being enrolled in school.

Now in NFE	1	0.5
Successful Transition Outcome		
Successful Transition	127	57.0

Few FE girls are engaged in employment of any kind. This is not entirely surprising, because FE girls remained relatively young at EL, with the mean age at 15.8 years. More importantly, girls remain enrolled at a very high rate – whether the girl was promoted across grades or not, 78.0 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 still appears encouraging, particularly in the context of the COVID-19 pandemic. It is important to remember that FE girls were enrolled *as part* of the initiative and were not already engaged in schooling. The fact that the programme brought FE girls into primary school and so many remain enrolled 4+ years later, despite COVID-related disruptions, is suggestive evidence of the programme's impact on enrolment outcomes.

Table 22 provides the equivalent results for ABE (top panel) and C1 NFE (bottom panel) girls. The most common outcome among ABE girls, at 35.7 percent at EL, is to be enrolled in formal education at a grade equivalent not higher than the ABE level the girl previously completed; only 3.4 percent of ABE girls enrolled at a grade level more advanced than their ABE level. Since a successful transition for formal education enrolment is defined only if the grade of enrolment is higher than the ABE level previous enrolled at, this heavily accounts for the notably low rate of successful transition among the ABE girls compared to their other cohort-counterparts. However, the high rate of ABE girls moving from the ABE programme to formal schooling does indicate that the girls are more motivated to continue their education after their involvement with the programme.

The second most common outcome was being out of school with no employment at 26.6 percent. However, a sizeable proportion of ABE girls reported being employed in a non-exploitative capacity, with 25.5 percent of them either employed or self-employed.

TABLE 22: TRANSITION OUTCOMES AMONG ABE AND NFE GIRLS

Transition Outcome	Number of Girls	Share of Sample (%)
ABE Girls (n = 263)		
Continued enrolment in ABE for ≤2 years with no advancement	4	1.5
ABE Progression for 3+ years	7	2.7
Out-of-School, Idle	70	26.6
Employed	66	25.1
Enrolled in formal school, but not at a more advanced grade level[40]	94	35.7

[40] For some of the ABE girls, the number of girls who reported transitioning to a more advanced GE grade from ABE is not completely represented here because the ABE level the girl enrolled in at baseline was not recorded. To assess an approximation for the ABE to FE transition, the recorded ABE level from the earliest midline evaluation for each girl was used in place of the baseline level. However, because not all the girls were in the midline samples for ML1 and/or ML2, some of the girls do not have

Enrolled in formal school, but at an advanced grade level	9	3.4
Self-employed	1	0.4
Employed, but underage (under 18 years)	2	0.8
Now in NFE	9	3.4
Now in Vocational Training	1[41]	0.4
Successful Transition Outcome		
Successful Transition	85	32.3
NFE Girls (n = 262)		
Employed	116	44.3
Out-of-School, Idle	73	27.9
Continued enrolment in NFE	27	10.7
Enrolment in ABE	1	0.4
Enrolled in formal school	38	14.5
Self-employed	6	2.3
Now in Vocational Training	1[42]	0.4
Successful Transition Outcome		
Successful Transition	162	61.8

Among the C1 NFE girls, the most common transition outcome was non-exploitative employment, with 46. percent of girls either employed or self-employed. The second-highest outcome was being out-of-school with no employment at 27.9 percent, followed by enrolment in formal schooling at 14.5 percent.

The relatively high employment rates among NFE programme participants suggests that the NFE 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 transition outcomes represent a single moment in time, however. Girls enter and exit employment, as they do with schooling. We would not know whether this level of employment would be sustained among the girls post-AGES, as this endline evaluation might have recorded this outcome at a fortunate time.

Although the lack of a comparison group tempers the strength of our conclusions, the initiative produced the expected outcomes among FE, ABE, and NFE girls. The outstanding question is whether these outcomes – consistent and continued enrolment and grade promotion for most FE girls 4 years on and gainful employment

an earlier ABE level to compare with their endline transition outcome in order to assess whether the FE grade at EL is higher than the ABE level the girl was originally in.

[41] The respondent noted that they were in a technical training; however, the specific details of the training type were not disclosed when asked.

[42] Respondent noted their participation in financial literacy training

for a majority of NFE girls – can be attributed to the programme itself or would have obtained in the absence of the programme. Additionally, the pathways observed most with ABE girls were aligned well with expectations if the definition of successful transition was to move to formal education, regardless of the level of enrolment. However, that a substantial portion of ABE girls who moved on to FE did not enrol at a higher FE level than the equivalent ABE level suggests that an insufficient level of academic preparedness was instilled upon the girls to move up to more advanced grades. Rather, it is possible that they were repeating the lessons they were already taught in a lower grade level during their time with the ABE programme.

4.2. 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 BL, under the assumption of *ex-ante* characteristic lag, or the lagged effect on the transition outcome based on how exogenous subgroup characteristics prior to the intervention's implementation would affect programme outcomes. For many outcomes, this is non-controversial. In other cases, this is not true. 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 BL. 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.

Even though the subgroup analysis is conducted with the characteristics defined and assumed at BL, a set of variables have been denoted with being recorded "at BL" or "at EL." One of the reasons the latter is included is because some of the subgroup characteristics reported at EL deviate exceptionally in transition outcome results compared to that from the BL that its inclusion and explanation would provide some insight into the results of the project. The caveat of this inclusion, however, is that we cannot assume that the samples each round constitute the same respondents and thus cannot rule out the possibility that the observed sub-group transition outcomes were primarily driven by changes in the sample composition, rather than by the actual effects of sub-group characteristics on transition outcomes. As a result, the subgroup analysis will focus primarily on the BL subgroup subtype but will consider the EL subtype for specific variables due to specific changes in the data collection process, such as with marital and maternal status.[43]

In Table 23, 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

[43] During BL, these questions were only asked in the caregiver survey only, which meant that if information from the girls who became 18 and older in the subsequent ML rounds were lost since they did not have the separate survey for caregivers conducted. To address this issue, these questions were added directly to the girls' survey to ensure a more complete dataset at EL by getting these responses from the girls 18 and older directly. As such, the data concerning being ever married, currently married, or having children at EL is more complete.

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 effect of membership in the subgroup on transition rates. Subgroup differences marked with an asterisk are statistically significant at the 0.05 level, after accounting for school-level clustering; however, some subgroups with a remarkably low sample size but statistical significance in subgroup differences will have an additional note about the external validity to the overall CARE AGES population.

The latter three columns disaggregate transition rates by cohort per subgroup. Note that we do not report p-values or statistical significance for these results, as the sample sizes per girl type 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 displayed significantly better transition outcomes compared to the those not in Banadir by 14.4 points, while girls in Southwest State underperformed significantly compared to those not in the state by 21.8 points. Banadir's overperformance was reflected most strongly in the NFE cohort at 70.4 percent; however, Banadir's ABE cohort was the only one to underperform relative to non-Banadir ABE girls, at 37.6 percent. Jubaland's underperformance was evident in all the cohorts, with the weakest performing cohort being the ABE cohort at 35.0 percent.

Interestingly, Banadir was the region with poorer learning outcomes among the BL cohorts. Disaggregating the data, we find that Banadir had the highest share of FE girls making positive grade progressions: Banadir leads the zones at 63.0 percent, followed by Jubaland at 50.6 percent and South West State at 38.0 percent. This implies that while Banadir FE girls were likelier to advance grades, this does not necessarily translate into grade-appropriate numeracy and literacy skills, suggesting that the standards for advancing grades in Banadir schools may be comparatively laxer compared to schools in other zones. Among Banadir NFE girls, the most successful transition outcome in Banadir is attributable to the higher share of NFE girls who are currently employed, which stood at 55.7 during EL, compared to 50.0 for Jubaland and 21.3 for South West State. This finding is unsurprising, given that the Banadir region is the most economically developed and thus likely present girls with the most employment opportunities. At the same time, employment does not necessarily provide girls with adequate opportunities to learn and retain the skills, so a higher employment rate may not lead to improved learning outcomes, despite boosting Banadir girls' successful transition rate.

The second panel showcases the disability status subgroups. Here, the sample of girls with the afflicted conditions in EL who participated in the survey were too low, with sample sizes as low as 2, depending on the coding scheme. This is very much likely due to the sample attrition between ML2 and EL. Due to the small subgroup sample sizes in the disability section, the observed differences in transition outcome likelihood compared to the overall sample are, for the most part, not representative of true transition outcome experience from those girls in the overall AGES cohort population with these disabilities (physical, cognitive, mental, any).

TABLE 23: SUBGROUP-SPECIFIC TRANSITION RATES, IN AGGREGATE AND AMONG COHORTS

					Cohort-Specific Transition Rates		
Subgroup	**Sample Size**	**Aggregate Transition Rates (%)**	**Subgroup Difference (%)**	**S.E. Subgroup Difference**	**FE Cohort (%)**	**ABE Cohort (%)**	**NFE Cohort (%)**
Overall	748	50.0	N/A	N/A	57.0	32.3	61.8
Geography							

Banadir	308	58.4	**14.4***	5.3	66.3	37.6	70.4
Jubaland	213	54.0	5.6	5.7	55.6	35.0	68.1
Southwest State	227	34.8	**-21.8***	4.9	42.0	25.5	42.7
Disability Status							
Any physical disability	2	0.0	**-50.1***	2.7	N/A†	0.0	0.0
Any physical disability, alt. coding	6	33.3	-16.8	19.0	50.0	0.0	33.3
Any cognitive disability	4	0.0	**-50.3***	2.8	N/A†	0.0	0.0
Any cognitive disability, alt. coding	7	28.6	-21.6	19.1	100.0	0.0	33.3
Any mental health disability	0	N/A	0.0	0.0	N/A†	N/A†	N/A†
Any mental health disability, alt. coding 1	113	55.8	6.8	4.4	65.4	42.6	65.0
Any mental health disability, alt. coding 2	247	52.6	3.9	5.1	56.9	35.6	65.3
Any non-mental health disability	6	0.0	**-50.4***	2.8	N/A†	0.0	0.0
Any non-mental health disability, alt. coding	13	30.8	-19.6	15.6	66.7	0.0	33.3
Any disability	6	0.0	**-50.4***	2.8	N/A†	0.0	0.0
Any disability, alt. coding 1	123	53.7	4.4	4.6	66.7	39.2	62.2
Any disability, alt. coding 2	285	51.2	2.0	4.7	60.0	31.3	63.2
Parental Educational Attainment							
HoH has no education of any kind (no Quranic)	154	59.7	**12.4***	5.1	71.7	40.6	75.0
HoH has no formal education	602	49.7	-1.0	5.4	54.2	32.1	63.8
Caregiver has no education of any kind (no Quranic)	192	55.2	7.2	4.9	64.7	34.8	69.1
Caregiver has no formal education	628	50.0	0.8	5.8	55.4	32.1	62.7
Household Economic Characteristics							
HoH does not earn a regular wage	257	51.8	2.7	4.7	61.1	30.3	60.6
HH has a poor-quality roof	142	45.8	-5.2	4.9	52.9	35.6	53.1

HH went to sleep hungry most nights, last 12 months	42	45.2	-5.0	7.5	54.5	38.1	50.0
HH went to sleep hungry for many nights, last 12 months, at BL	128	41.4	**-10.4***	5.2	46.7	32.2	51.3
HH went to sleep hungry for many nights, last 12 months, at EL	101	59.4	**10.4***	5.3	62.5	41.2	72.1
HH went without clean water most days, last 12 months	38	50.0	0.0	8.0	62.5	46.7	46.7
HH went without medicine most days, last 12 months	113	46.9	-3.6	5.2	40.6	40.6	55.1
HH went without cash income most days, last 12 months	95	43.2	-7.8	5.5	62.5	27.1	56.5
HH owns lands	260	43.8	**-9.4***	4.4	45.2	24.4	61.2
HH owns a phone, at BL	652	50.3	2.4	5.9	58.5	31.2	63.0
HH owns a phone, at EL	50.9	2.6	5.1	58.0	29.9	63.5	50.9
HH owns a smartphone, at BL	136	50.0	-0.4	5.7	54.9	32.4	58.3
HH owns a smartphone, at EL	295	51.5	1.8	3.5	67.5	26.9	63.2
Parental Circumstances							
Girl has only one living parent	90	61.1	**12.6***	5.8	76.2	39.3	68.3
Girl has no living parents	6	66.7	16.8	19.4	N/A†	60.0	100.0
Girl does not live with either parent in her HH	36	52.8	2.9	9.4	60.0	33.3	63.2
Female-headed household	267	56.9	**10.8***	4.2	65.2	39.3	65.0
Language							
Linguistic Affiliation							
Household speaks af-Maay, at EL	216	42.1	**-11.4***	4.9	42.4	28.3	58.4
Other Individual Characteristics							

Girl has ever been married (even if divorced now), at BL	53	54.7	5.1	7.0	0.0	33.3	57.1
Girl has ever been married (even if divorced now), at EL	205	55.6	7.7	4.4	43.8	52.0	58.3
Girl is currently married (not divorced/separated), at BL	27	51.9	1.9	10.5	0.0	0.0	56.0
Girl is currently married (not divorced/separated), at EL	136	57.4	9.0	5.6	58.3	53.3	58.5
Girl has a child, at BL	42	59.5	10.1	8.2		0.0	61.0
Girl has a child, at EL	163	60.1	**22.0***	7.9	54.5	60.5	60.5
IDP	306	54.2	6.4	4.7	56.3	39.7	68.9
Consumed any protein in the last 24 hours	646	49.8	-6.3	6.8	58.0	30.9	62.1
School Facilities and Characteristics							
Girl will not use drinking facilities at school	171	52.6	3.4	5.9	50.0	37.3	66.2
Girl will not use toilet facilities at school, at BL	129	56.6	8.0	5.4	56.5	38.2	65.3
Girl will not use toilet facilities at school, at EL	278	47.8	-3.4	3.6	55.2	27.1	66.7
No computers available for use at school	712	49.6	-8.8	9.6	56.3	31.0	62.2
Girl cannot use books/learning materials at school	210	58.6	11.9	5.1	64.4	30.6	76.0
Not enough seats for every student in class	112	57.1	**8.4***	6.1	55.0	31.0	69.8
Girls can take textbooks/materials home at night[44]	79	54.4	-3.9	10.7	54.4	N/A	N/A
School has reliable electricity	140	55.7	-3.3	9.2	55.7	N/A	N/A

[44] Subgroups with a blank cell for the ABE- and NFE-specific cohort rates are due to the fact that these school-specific questions were only conducted from formal education (FE) schools and classrooms, and as such apply only to the FE cohort.

School has water access within 1 km	125	56.8	-0.3	9.5	56.8	N/A	N/A
School distributes sanitary towels[Error! Bookmark not defined.]	67	72.1	**21.7***	9.9	72.1	N/A	N/A
School provides at least one meal for students	68	52.2	-6.7	11.7	52.2	N/A	N/A
School has only cement floors (no dirt)	152	53.3	-11.5	8.8	53.3	N/A	N/A
School has separate toilets for girls	151	58.3	4.1	10.6	58.3	N/A	N/A
No female teachers, either FT or PT	26	42.3	-16.6	11.9	42.3	N/A	N/A
Short instructional time per day	56	46.4	-14.1	12.0	46.4	N/A	N/A
Teachers miss 1-2 days per week, on average	40	65.0	9.8	9.9	65.0	N/A	N/A
Agree: my teachers are often absent	219	42.0	**-11.3***	5.0	50.7	N/A	N/A
Textbooks are shared between students	103	58.3	2.4	9.3	58.3	N/A	N/A
School charges school fees	106	56.6	-0.7	9.6	56.6	N/A	N/A

†Zero observations were present for the aggregate or cohort-specific transition rate to be calculated.

The third panel showcases parental educational attainment and its impact on transition outcomes. Girls whose head of household (HoH) had no education – formal or Quranic – overperformed significantly over those whose HoH or caregiver had at least some form of education by 13.1 points. Interestingly, the cohort with the highest performance was the C1 NFE cohort with a successful transition rate at 75.0 percent. When broken down into rate differences per cohort type, the significantly higher rate was observed among girls from the FE cohort who had an HoH with no education (18.6 points higher, p=0.026).

Looking at the transition outcomes, among girls in the FE cohort, the most frequent outcome among girls whose HoH lack any education and among those with educated HoH was positive FE grade progression; however, the grade progression rate among girls with a HoH without any education – at 65.2 percent – was much higher than the rate among girls with a HoH with some form of education – at 49.7 percent. From this, the plausible reason for this observation is that the girls in formal schooling may have wanted to advance their education after seeing the limitations of opportunities, including financial, within a household where the HoH has no education whatsoever.

The fourth panel looks at the impact of household economic characteristics on successful transition. Among girls whose household owns land, the rate of successful transition is 9.4 points lower compared to those whose household does not own land. When stratified by cohort, this significant relationship is observed only among

the girls in the FE cohort (-20.2 points; p=0.01); this was confirmed in a Chi-square test between land ownership and transition outcomes among FE girls, as only 45.1 percent of girls whose household owns land successfully transitioned compared to 65.4 percent of girls whose household does not own land.[45] One possible explanation may be that land ownership correlates with the propensity for girls to be asked to assist with household work, which may reduce time available for studying. To evaluate if chores – or having to work on household work (i.e., farming obligations) – would be a potential mediator leading FE girls whose household owns land to register significantly lower transition rates, linear regression models evaluating 1) the role of household land ownership as a predictor of household chore burden and 2) the role of household chore burden as a predictor of successful transition outcomes. While significantly positive association was observed in first component model between household land ownership and chore burden (p=0.027), no significant association was observed between household chore burden and the rate of successful transition (p=0.88), providing insufficient evidence to claim that chores is a mediator to explain the negative association between household land ownership and successful transition outcomes.

Additionally, the rate of successful transition among girls who went to sleep hungry for many nights within the past year was 10.4 points higher than among girls who did not. However, when controlled for age and region, the rate difference was insignificant, but age and region were significant confounders.

Most individual characteristics had a more minimal impact on transition outcomes than expected. Marital status, IDP status, and protein consumption did not have a significant relationship with transition outcome at EL. However, girls whose household speaks *af-Maay* as of EL reported an 11.4-point lower rate of successful transition compared to girls whose household do not.

In terms of maternal status, girls who have a child as of EL had a likelihood of successful transition 22.0 points higher than those who are not mothers. Additionally, unlike for that of the group of those who reported having a child at BL, it is also probable that the statistical association observed might be due to another hidden external factor or explanation since temporality cannot be established here (whether the maternity status contributed to successful transition or vice-versa).

In terms of school-specific characteristics, the successful transition rate among girls who could not use their books or materials in their schools was 11.9 points higher than those who could. When evaluated by cohort, on the impact of lacking materials in school on the successful transition rate was significant only among those in the C1 NFE cohort (p=0.01). Looking into the transition outcome types among girls in the C1 NFE cohort who noted the lack of materials in their school, the most common transition outcome was employment (50.7 percent). From this, one might hypothesise that the lack of school materials may have contributed to an unpleasant school experience for the NFE girls, possibly incentivizing them to look for employment, rather than commit to more education through continued NFE programs, or through formal education. Girls in schools that provided meals for students had significantly higher transition rates, 21.7 points higher than those which did not provide school meals. Additionally, girls who agreed that their teacher has been absent often observed a lower transition rate, by 11.3 points, compared to everyone else.

4.3. Testing the Theory of Change

Here, we subject the AGES Theory of Change, as it pertains to transition, to testing. 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

[45] Chi-square=9.0, p=0.003, d.f.=1

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, EL, while setting aside the transition status girls had achieved at ML2. 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 state, 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 Table 24. 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 are denoted by one asterisk.

No characteristics pertaining to teaching quality, caregiver attitude, and household chore burden maintained a significant impact on the successful transition rate.

TABLE 24: RELATIONSHIP BETWEEN KEY INTERMEDIATE OUTCOMES AND TRANSITION RATES AT EL

Independent Variable of Interest	Sample Size (n)	Transition Outcome (%)
Teaching Quality		
Girl Feels Unwelcome by Teacher	86	-0.1
Girl Feels Her Teachers Are Often Absent	219	-8.3
Girl Feels Her Teacher Rarely/Never Encourages Participation	68	-3.1
Girl Feels Her Teacher Rarely/Never Explains How Things Learned Are Useful in Her Life	19	-2.9
Girl Feels Her Teacher's Lessons Move Too Fast for Her	337	-1.8
Girl Believes Her Teacher Treats Boys/Girls Differently in The Classroom	275	-7.8
Girl Reports Her Teacher Punishes Students Who Get Things Wrong During a Lesson	514	2.3
Girl Reports Her Teacher Used Corporal Punishment during the Week of Interview	168	5.7
Caregiver Attitude		
Caregiver does not feel it is safe for girls to travel to the school	5	19.6
Caregiver aspires to send girl to university	660	-5.4

Caregiver believes girls' educ worthwhile, even if funds are limited	657	1.1
Caregiver believes work or HH chores are acceptable reason to not attend school	311	2.8
Caregiver believes cost of education is acceptable reason to not attend school	463	-6.2
Household chore burden at BL	489	0.5
YLI score at BL	748	0.1 (per 1-point increase)
GEF Participation (binary indicator)	378	2.2
GEF Participation (ordinal 0-4 score)	107	-3.0 (per 1-unit increase)

Another intermediate outcome we study is a girls' self-esteem and leadership skills, as measured by the Youth Leadership Index (YLI). Our regression analysis of the role of YLI scores showed no statistically significant impact on the likelihood of successful transition at EL from either results approach.

Finally, looking at the role of Girls' Empowerment Forums or girls' clubs on successful transition outcomes, 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 0-4 scale.[46] We report regressions using these two respective independent variables in the bottom rows of Table 24. Under the binary measure, 50.6 percent of the sample had participated in a GEF in the past. However, there was no significant impact of GEF participation on successful transition outcomes.

5. Sustainability

This section presents a narrative analysis of the sustainability of AGES, based on key sustainability indicators identified and measured at each round since baseline for the FCDO cohorts. This utilises both quantitative data (i.e., household surveys and school-level surveys) and qualitative data (FGDs with teacher, mothers, CEC members) to fully evaluate the long-term impact of AGES via the noted indicators.

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 in school. As part of these economic interventions, AGES established and supported Village Savings and Loans Associations (VSLAs) for FE and ABE girls' caregivers and girls enrolled in NFE. VSLAs are community self-managed groups that often provide the only opportunity for families – especially in rural areas and amongst poor households – to save money and obtain loans. The borrowed money can be used to finance new businesses or

[46] A girl receives one point on this scale if she reports having ever participated in a GEF in either the ML1, ML2, or EL round. She receives an additional point if she reports that she continues to be in contact with the GEF at ML1, an additional point for the same outcome at ML2, and an additional point for the same outcome at EL. Thus, a girl with a score of 4 on this metric self-reported participation in a GEF and reported being in contact with the GEF in the ML1, ML2, and EL rounds. This is a proxy indicator for the depth of engagement a girl has with a GEF.

large household expenses, including expenses related to children's education such as school fees or school materials.

Overall participation in the savings groups – while showcasing some degree of growth since baseline – remains low. Only 10.0 percent of caregivers interviewed reported being a part of a VSLA, which is significantly higher than the baseline participation rate among caregivers at 4.0 percent but is still a low level of engagement considering the 4.5 years since baseline. The growth of participation over time also plateaued among the interviewed caregivers: after baseline, the rate of participation at ML1 and ML2 was 8.3 percent and 7.9 percent, respectively; Chi-square analysis of savings group participation between ML1, ML2 and EL revealed insignificant differences in participation levels, indicating that after a small spike in participation growth from baseline to ML1, VSLA participation stagnated for the remainder of the study period (Chi-square=0.97; p=0.615; d.f.=2).

However, the rate of participation among C1 NFE girls in VSLAs was observed to be significantly higher than that of the caregivers. When asked about their participation in a savings group, 33.1 percent of C1 NFE girls interviewed reported their participation in a VSLA at endline, significantly higher than the caregivers' participation rate. While NFE participation in the VSLA was not recorded at baseline, the rate of participation among C1 NFE girls remained constant between ML1 to endline (27.6 percent at ML1 vs. 26.9 percent at ML2 vs. 33.1 percent at EL).[47]

While C1 NFE participation in the savings group was overall higher compared to that of the other cohorts, the rate of continued active membership among those in the savings group plummeted and remained low compared to the active status of those in the FE and ABE cohorts. When asked about whether their VSLA was still active, 76.9 percent of participating caregivers reported the continued active status while only 25.3 percent of C1 NFE girls reported that their VSLA remained active.[48] This mirrors the trend observed in ML1 (82.7 percent and 38.1 percent, respectively) and ML2 (82.9 percent and 34.6 percent, respectively).

This low rate of participation among the caregivers across rounds, as well as the low rate of active membership among NFE enrolees indicates either a vulnerability in the operation, leadership, and/or mechanisms of the VSLAs or external factors that might influence the girls to not continue their membership, either of which present barriers to the long-term sustainability of these groups. These rates also suggest that very few VSLAs might continue after the conclusion of AGES.

Proportion of parents able to support costs of girls' education

A significant indicator of sustainability of the AGES is the proportion of caregivers whose financial situation allows them to support the costs of girls' education. The participation in VSLAs contributes to this indicator as well as the disbursement of cash support, bursaries, households' livelihood and income generating activities, and general economic context.

Overall, household attitudes towards girls' education despite the costs remain high for both caregivers and girls; however, the economic contexts in these households have been declining enough for the attitude to prioritise schooling to start to fade in both parties. A majority of caregivers (59.0 percent) reported at EL that it was not acceptable for a girl to not attend school due to "education being too costly", significantly higher than the 37.8 percent of caregivers reporting the unacceptability at baseline[49] and consistent with the

[47] Chi-square=3.15; p=0.21; d.f.=2
[48] Caregivers (n=26); C1 NFE girls (n=87)
[49] Chi-square = 41.2; p<0.0001; d.f. = 1

rate among caregivers when asked at ML1 (54.8 percent) and ML2 (56.4 percent). Additionally, 72.4 percent of caregivers reported at EL that it was not acceptable for a girl to not attend school due to the child "having to work", which is statistically similar to the reported rate among caregivers when asked at baseline (75.4 percent). However, the endline rate is significantly lower than the reported rates from ML1 (87.3 percent) and ML2 (85.3 percent), suggesting that between 2023 and 2024, financial circumstances changed for a number of households for the caregiver to prioritise work over education for the girl.[50]

Additionally, 95.7 percent of girls interviewed agreed or strongly agreed that "even when funds are limited it is worth investing in your education". However, when comparing the proportion of girls who strongly agreed at EL with those at ML1 and ML2, there was a clear and sustained decline in support for the statement, from 88.6 percent *strongly* agreeing in ML1 to 64.4 percent in ML2 and finally at 55.1 percent at EL. This is also evident in change in the proportion of girls agreeing (but not strongly) with the statement from ML1 to EL, which increased from 11.0 percent of girls agreeing at ML1 to 32.8 percent at ML2 and finally to 40.6 percent at endline.[51] This suggests that while girls in the household still hold strong beliefs in their education and its importance, household financial dynamics might be gradually eroding that resolve among the girls.

Parental support for girls' participation in GEFs

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 the long-term sustainability, we asked AGES girls whether they participated in GEFs and, if so, whether they were still in contact with the other GEF members. As of endline, when asked if the girl ever participated in a GEF, 26.4 percent of girls reported their past participation with the programme, with 56.1 percent of those participating still maintaining some form of contact with other GEF peers. This rate of GEF participation and continued contact with peers has remained consistent, with no significant changes between each round since BL.[52] Interestingly, only 9.2 percent of girls reported ever participating in activities implemented by the GEF, indicating that the girls' level of engagement with their GEF membership and connections are likely more minimal than expected.

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

Indicator	Total	FE girls	ABE girls	C1 NFE girls
Ever Participated in a GEF	**26.4%**	22.0%	32.2%	26.2%
Still in Contact with GEF Members	**56.1%**	59.4%	55.7%	51.6%

While GEFs are a social platform intended to bring girls together and empower them, this is only one possible option. When we asked C1 NFE girls if they participated in any other youth groups or networks, 44.5 percent

[50] Chi-square = 31.3; p<0.0001; d.f. = 2
[51] Chi-square = 160.0; p<0.0001; d.f. = 8
[52] **Ever Participated in a GEF per round (ML1/ML2/EL):** Chi-square = 3.8; p = 0.152; d.f. = 2; **Still in Contact with GEF Members:** Chi-square = 0.4302; p=0.806; d.f. = 2

of them said that they did, including 41.5 percent of girls who did not participate in the GEFs. On the other hand, among girls who had ever participated in a GEF, 56.25 of them had also participate in other youth groups or networks – an intuitive finding given that girls who are already involved in one forum are more likely to be involved in similar types of activities.

Proportion of GWDs who remain in school

AGES 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.

Because of the low count of GWDs based on the original coding[53], the enrolment rates were evaluated using the second alternative coding structure[54] for GWDs. Overall, under this definition, the continued enrolment of GWDs at endline (57.4 percent) is marginally higher than the enrolment of girls without disabilities (53.6 percent), although that difference is not statistically significant[55]. The rate of enrolment for GWD is consistent with that at ML2, (53.9 percent for GWD enrolment vs. 49.5 percent for girls without disabilities enrolment), indicating that a positive impact of the programme's ability to retain schooling for GWDs.

Table 25 showcases the enrolment rate for each round after baseline based on the specific type of disability. Chi-square analysis was conducted to evaluate the prevalence of any significant changes in enrolment rates by round.

TABLE 25: ENROLMENT RATES OF GWDS, BY DISABILITY TYPE[56]

Disability Type	% Enrolled (ML1)	% Enrolled (ML2)	% Enrolled (EL)	p-value
Sight/Vision Disability, alt. code 2	50.0%	58.1%	54.5%	0.83
Hearing Disability, alt. code 2	77.8%	68.4%	68.4%	0.86
Mobility Disability, alt. code 2	69.6%	82.6%	75.0%	0.44
Self-care Disability, alt. code 2	66.7%	85.7%	66.7%	0.68
Communication Disability, alt. code 2	53.9%	80.0%	69.2%	0.22
Cognitive Disability, alt. code 2	72.6%	60.0%	66.0%	0.33
Any disability, alt. code 2	52.4%	53.9%	57.4%	0.34

GWD enrolment and retention for each and any disability seems to be an overall positive indication of the programme's impact, as there were no significant changes in the enrolment rate since ML1, nor did any of the enrolment rates fall below 50.0 percent. The highest rate of enrolment was observed among girls with

[53] Includes the caregiver response only.
[54] Includes the caregiver and girl response who report "some" difficulty, as opposed to only those reporting a lot or cannot do at all.
[55] Chi-square = 1.26; p = 0.261; d.f. = 1
[56] 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.

mobility disabilities, at 75.0 percent, and the lowest rate of enrolment was observed among girls with visual disabilities, at 54.5 percent.

Among cohort types, GWD enrolment was significantly highest among FE girls - 82.2 percent – while that of ABE girls was at 44.4 percent, and that of C1 NFE girls was at 29.0 percent.[57] Additionally, the enrolment of GWD in each of the three cohorts were not significantly different compared to the enrolment of girls without disabilities, indicating that the barriers to enrolment and retention do not specific impact GWDs.

The distribution of accessibility and teaching accommodations for GWDs was mixed, based on the school. Some teachers noted that no extra accommodations are provided, aside from anti-discrimination training for the other students, while others noted that even if they had GWDs attending, they did not have the resources to be able to fully support them.[58] Interestingly, among teachers who do not have any GWDs in attendance, they noted that they would be able to readily provide or at least make the effort to assist them.[59]

In any case, the sustainability of GWD support and retention is not likely to be threatened after the cessation of the AGES programme, given teachers' and peers' supportive attitudes towards GWDs' learning; however, continued monitoring and resource allocations would be needed to equip the schools with necessary tools to physically accommodate GWDs by a higher administrative body.

Proportion of GEFs implementing community actions to support attendance and retention

AGES focuses on the 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, and through indirect interventions related to economic support and change in community perceptions.

As shown in Table 26, GEFs overall were substantially more active in raising awareness for girls of all cohorts than in any other key activities related to changing community attitudes, supporting girls' education, and mentoring in business and financial support. The implementation of girls' education support was relatively low, with OOS enrolment activities at 30.0 percent, CEC participation at 13.4 percent, and teaching other girls and study groups at 19.8 percent. Additionally, the implementation of business and financial components from the GEFs also remained weak, with only 14.2 percent involved in supporting savings groups and 7.9 percent in joint business support. However, for those who did engage with GEFs, some girls did note the positive impact of their participation in both business, household, and academic engagement, particularly on how the GEFs peer member assist each other.[60]

TABLE 26: PROPORTION OF GEFS ENGAGEED IN ACTIVITIES, BY COHORT

Indicator	Total	FE girls	ABE girls	C1 NFE girls
Girls Education Support				

[57] Chi-square = 85.9; p<0.0001; d.f. = 2
[58] See FGD with teachers, Bay, Int. 504 and FGD with teachers, Banadir, Int. 507
[59] See FGD with teachers, Banadir, Int. 506.
[60] See Vignettes FGD with Girls, Lower Shabelle, Int. 601.

Enrolling out-of-school girls	30.0%	31.7%	28.4%	29.7%
Participation in CECs	13.4%	19.8%	10.2%	7.8%
Teaching other girls/ study groups	19.8%	25.7%	13.6%	18.8%
Business & Finances				
Savings group	14.2%	8.9%	12.5%	25.0%
Joint business	7.9%	8.9%	5.7%	9.4%
Changing Community Attitudes				
Preventing early marriage	23.7%	27.7%	17.1%	26.6%
Awareness raising	61.3%	67.3%	54.6%	60.9%
Trainings (health, girls' rights, gender, etc)	14.6%	20.8%	5.7%	17.2%
Community discussions	10.3%	15.8%	4.6%	9.4%

The range of GEF activities implemented decreases along with the decreased engagement with GEF. It is likely that this decline is attributed to the ceased support towards a sizeable number of GEFs. As mentioned previously, support for GEFs embedded in C1 NFE centres ended in 2021, so the decline in GEF engagement was to be expected. However, this does reinforce the previous conclusion that the GEFs' impact may not be sustainable without continued programmatic support. [61]

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. 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.

Table 27 shows the number of teachers trained on different skills in the targeted schools in the previous year while Table 28 compares the absolute count of teachers trained between ML2 and endline. The training of female teachers is especially important as it can further empower girls to stay in school by providing positive role models and understanding difficulties faced by female students both at school and at home.

TABLE 27: NUMBER AND PROPORTION OF TEACHERS TRAINED IN PAST YEAR

Skill	Total Teachers			Average (per school)	
	Female	Male	TOTAL	Female	Male

[61] Consilient, "Sustainability", *CARE AGES Midline 2 Report* (2023).

	N	%[62]	N	%[63]			
Gender sensitive teaching methods	11	*4.3*	59	*7.2*	70	0.31	1.64
Teaching methods for maths	62	*24.4*	111	*13.6*	173	1.72	3.08
Teaching methods for reading and writing	78	*30.7*	139	*17.0*	217	2.17	6.03
Inclusive education	73	*28.7*	153	*18.8*	226	2.03	4.25
Child protection	84	*33.1*	140	*17.2*	224	2.33	3.89

When we interviewed head teachers in the programme schools, they reported that the proportion of female teachers trained on teaching methods for maths and for reading and writing was 24.4 percent and 30.7 percent, respectively. This is lower than the proportion of female teachers trained in the previous round (40.6 percent and 50.9 percent, respectively), but the proportion of trained female teachers at EL is still higher than for male teachers, among whom 13.6 percent were trained in teaching maths and 17.0 percent in teaching reading/writing.[61] Additionally, gender sensitive training was attended by more male teachers, percentage-wise, compared to female teachers this past year (7.2 percent compared to 4.3 percent), which indicates an emphasis on ensuring that more male teachers understand gender equitable practices in a school seetting.

TABLE 28: NUMBER OF TEACHERS TRAINED AT ML2 AND AT EL

Skill	Trained (ML2)	Trained (EL)
Gender sensitive teaching methods	153	70
Teaching methods for maths	192	173
Teaching methods for reading and writing	259	217
Inclusive education	285	226
Child protection	215	224

In terms of absolute count,[64] fewer teachers attended the trainings for all but child protection in the 2023-24 period compared to the 2022-23 period. This is most prevalent in regard to gender sensitivity training, with half the number of teachers as of EL compared to ML2.

[62] Out of the total female teacher count of 254.
[63] Out of the total male teacher count of 816.
[64] The total count of teachers in the schools at ML2 were not available.

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.

Among teachers trained in the previous 12 months, the rate of completion with the training on inclusive teaching methodologies was much lower than the previous round, with 21.1 percent of teachers trained compared to the 52.0 percent previously.[61] In terms of the number of teachers trained, only 226 teachers were trained in inclusive education strategies compared to the 285 teachers trained the year prior. This does indicate that the rate of training and strategy instillation has been on a decline among the teachers. Additionally, some schools have not received the training for inclusive pedagogical practices, as – in terms of challenges for accommodating GWDs – one teacher noted that they "don't have teachers trained specifically for disability education."[65]

The impact of the implementation of inclusive training is more difficult to discern, largely since teachers cite that 1) there is no discrimination or 2) there are no marginalised girls in attendance. For the most part, almost none of the girls (0.33 percent) who left school prematurely cited mistreatment or discrimination by the teacher as their reason for leaving. However, when asked if girls who speak different languages can keep up with the lessons, some teachers noted that they did not have any other girls who spoke a different language. As one teacher noted: "In this school, the students have the same dialect. It is possible that a few speak another dialect, but they know the Maay language." [66] Additionally, in terms of GWD support, some teachers reported not having any specific accommodations for GWDs but did not indicate any cases of discrimination against them;[66] however, a number of teachers also noted not having any GWD in their school, even if they said that they would be able to support them.[59]

Overall, the impact of the trainings on inclusive teaching methodologies are inconclusive when it comes to how it has affected the teaching styles of the teachers.

6. Intermediate Outcomes

6.1. Attendance

Attendance in class is the first intermediate outcome we measure for the AGES Cohort 1 girls, funded by FCDO. Per the theory of change, we expect that attendance correlates strongly with learning outcomes since regular presence in class is essential for gaining new skills and internalizing lessons, which in turn also contributes to grade advancement and has implications for ongoing school enrolment (as discussed in our transition outcomes chapter).

[65] See FGD with teachers, Lower Juba, Int 509.
[66] See FGD with teachers, Bay, Int. 504.

To monitor attendance numbers, our field teams conducted physical headcounts: enumerators would visit classrooms and physically count how many learners are present in class at that time. These numbers are then later compared to the enrolment records, or records of how many learners *should* have been in class if attendance were perfect. Headcount assessments are carried out per grade, resulting in multiple headcounts within the same school and thus a larger sample size for analysis. In addition to physical headcounts, the field teams also consulted (head) teachers in each school to provide information on the latter's own records of attendance for the day prior our field teams' visits, as well as for the day of. This provides additional, useful data to understand the degree to which teachers' attendance records differ from enumerators' own physical headcounts. For this EL evaluation, headcounts were only conducted for formal education (FE) schools (or FE streams of a school), and our assessment includes separate headcounts and teacher reports for both girls' and boys'.

In assessing changes in attendance from the baseline (BL) round to each of the later evaluation stages—first midline (ML1), second midline (ML2), and endline (EL) — our analysis is restricted to a strict panel of formal education schools, each of which had headcounts taken in each of the four rounds. We do this to avoid potential biases that might be introduced if attendance rate calculations include schools who had at some point dropped out of the sample. Because the sample of FE schools that remain in the full panel by EL might differ from those included in the full panel up to prior evaluation rounds, our calculated attendance rates (and changes therein between rounds) might differ from the figures we had previously reported in the ML1 or Ml2 reports, since the former are based on a slightly different set of schools considered in the analyses.

Two limitations of using headcounts data warrant attention. First, actual attendance of learners might vary on a daily basis. Since headcounts capture attendance on a particular day, our physical headcount measures are a snapshot of a particular point in time. It is possible that had our field teams visited each school on a different day, the results of the headcounts and teachers' reported attendance records might be appreciably different. Second, we use school-reported enrolment numbers as a basis for attendance calculations. Our measure of attendance uses the reported enrolment number as the divisor for the number of learners observed in class (or reported to be in attendance by the teacher). As such, measures of attendance are only as reliable as the school's reported enrolment figures, which the evaluation team regrettably did not have alternative means of verifying.

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 considers attendance in classrooms for grades 1 to 5, as these are the grades where AGES interventions were concentrated.

It is important to note a shortcoming of the EL data collection in particular. During the BL and ML1, we collected attendance and enrolment 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 and EL rounds meant that field teams were generally instructed to complete just one headcount per grade level; if a school had multiple grade 4 classrooms, for example, the team leader would randomly select one classroom for the headcount. This differs from BL and ML1 rounds, during which all classrooms in the target grade ranges were assessed for attendance. This means that our sample size of classrooms is slightly smaller

than expected (n = 164, compared to 260 at ML1) and it includes slightly fewer lower-primary classrooms than we would expect, as these are often the grade levels where there are multiple classrooms. However, this is mitigated by the fact that the smaller sample sizes are 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. Thus, we expect this shortcoming to have relatively minimal impact on the trends we observe in the data.

Table 29 below presents the calculated attendance rates in formal schools, disaggregated by the measurement method: headcounts, teacher's reports for attendance the day of the visit, and the equivalent for the day prior. The table is also disaggregated by the gender of the learners. We note that the results below are based on a "naïve" average of attendance records for each grade and school, without controlling for changes in grade-level compositions across evaluation rounds, which we control for later in the section. In many ways, the findings here echo those previously reported in the ML2 report: while there is a decline in attendance as measured by the headcounts, attendance as measured by teachers' reporting (both on the day of the headcount and the day prior) has seen steady increases for both boys and girls. Between BL and EL, these measures have increased by a margin of 7.1 to 13.7 percentage points. Over the same time interval, attendance rates as measured by headcounts have declined by 3.4 points for girls, and by 4.4 points for boys. Given that we do not see a concomitant in attendance rate increases as measured by headcounts, the growth in teachers' reported student attendance likely stems from changes to their own reporting, rather than actual increases in attendance.[67]

TABLE 29: ATTENDANCE RATES IN FORMAL SCHOOLS, DISAGGREGATED BY METHOD OF MEASUREMENT AND LEARNER GENDER

Attendance Measure	BL	ML1	ML2	EL
Girls				
Attendance – Physical Headcount	86.6	83.2	82.6	83.2
Attendance Today – Teacher-Reported	80.0	83.6	83.4	87.1
Attendance Yesterday – Teacher-Reported	77.7	86.3	85.8	86.5
Boys				
Attendance – Physical Headcount	87.1	84.9	85.0	82.7
Attendance Today – Teacher-Reported	79.9	85.4	85.5	87.6
Attendance Yesterday – Teacher-Reported	74.1	84.4	86.3	87.8

[67] In conducting this analysis of attendance rates by headcounts and teacher reports, we exclude outliers whose attendance rates were above 100.0 percent, which indicates that there were more observed or reported to be in the classroom than registered in the class's enrolment records. An alternative coding scheme instead recoded these observations to be capped at 100.0 percent attendance rate, rather than dropping them entirely. These different coding decisions do not fundamentally alter the results presented in this section.

One possible explanation may stem from teachers' increased awareness of the importance of proper record keeping – and the fact that this is a metric that AGES implementers and community stakeholders monitor. As we later note in the section on school governance, there is large and significant increase in the proportion of surveyed head teachers in formal schools who claim that Community Education Committees (CECs) actively monitor student attendance; in fact, it is the CEC function most commonly cited by head teachers at EL.[68] While less than half of head teachers surveyed at BL claimed the CECs monitor student attendance, 85.7 percent of head teachers claimed as much by EL.[69] This is confirmed by CEC members themselves, many of whom report that monitoring student attendance is one the important roles they play.[70] One teacher also offered the following perspective, explaining how the monitoring by other stakeholders helps the teachers change their attendance record-keeping:

> *The monitoring visits consistently provide teachers with valuable experiences and compliments, proving to be beneficial for them. And yes, we received monitoring visits from the district education authorities, and they normally monitor how we teach the students and their attendance. These efforts encourage us.*
>
> Teacher FGD, Banadir, Int. 511

Recognition of efforts to monitor student attendance may thus incentivise teachers to maintain better attendance records, and/or to report figures closer to full attendance. On one hand, the difference between teacher-reported attendance counts and the third-party headcounts appear to be narrowing slightly over time,[71] which may indicate that teachers and school administrators are indeed improving their systems for attendance record keeping. Our brief analysis on improving attendance records later at the end of this section provides some tentative evidence in support of this interpretation. On the other hand, while teacher-reported attendance counts are uniformly lower than headcounts during BL, they have steadily become higher than actual headcounts since ML1, and are now uniformly higher than headcounts during the EL. Taken together, it is difficult to ascertain whether (and the extent to which) the increasing attendance rates reported by teachers reflect more accurate record keeping, or overestimates stemming from a desire to report better attendance figures.

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

[68] See "School Management and Governance" in the FCDO cohort section.
[69] This change is significant at the 0.01 level.
[70] E.g. FGD with CEC Members, Int. 101; FGD with CEC Members, Int. 104; FGD with CEC Members, Int. 105
[71] For instance, the gap for boys' attendance as measured by headcounts and as measured by teachers' report of boys' attendance the day prior was 13.0 points, which narrowed to 5.1 points by EL, though teachers' records are now overestimates.

The regression estimates broadly reflect trends described in the descriptive analysis above. For girls' attendance, we observe a decrease in the attendance rate since BL. The largest drop was from BL to ML1, where the attendance rate dropped by 5.4 points ($p = 0.043$). We note an increase of 1.6 points between ML1 to ML2, followed by another decrease of 1.0 point from ML2 to EL, though neither of the latter coefficients was statistically significant. Similarly, boys' attendance rate saw an overall decrease since BL, with a decline of 4.5 points, though this was not statistically significant ($p = 0.104$). Regression results on teachers' reported attendance likewise echo the descriptive analysis in a broad sense: for both boys and girls, teacher-reported attendance increased from BL to EL. The results appear starkest among boy learners, where we observe increases that are statistically significant at the 0.1 level for teachers' reports on boys' attendance on the day prior (13.4-point increase from BL to EL, $p = 0.074$) and the day of our field teams' visits (10.6-point increase from BL to ML2, $p = 0.056$).

Geographically, attendance figures for girls and boys remain relatively unchanged for Jubaland.[72] However, larger changes were observed in Banadir and South West State. In a regression model interacting round and geographic zone, girls' attendance in South West State declined by 12.0 points from BL to ML1; while the rate increased slightly between ML1 and EL, there is still a 9.1 point decline since BL in girls' attendance rates. These declines are primarily driven by classrooms in Baidoa: attendance at BL was 91.6 percent, which subsequently fell to 78.1 percent at EL. For the Banadir region, girls' attendance fell by 6.2 points between BL and EL, with the districts of Yaqshid, Hodan, and Waaberi experiencing particularly steep declines. Among boys, the steepest decline was observed in the Banadir region, where we observe a 9.2-point decline in attendance.

In assessing grade-level differences in attendance rates, we estimated additional regression models that incorporated interaction terms between the round of data collection and the grade level of the classroom assessed in the headcount, while also controlling for geographic zone. This model specification enables us to determine which grades in particular experienced cross-round attendance rate changes. In the table below, we report the change in headcount-based attendance rates for a specific grade, between BL and (alternately), ML1, ML2, and EL. For instance, the results show that girls' attendance in Grade 4 classes declined by 8.5 points from BL to ML1. Importantly, all results are *relative to the BL* round; thus, girls' attendance in Grade 4 classes declined by 1.2 points from BL to EL, but this implies that girls' attendance in Grade 4 classes increased between ML1 and EL, as the decline at from BL to EL is smaller than that seen between BL and ML1.

TABLE 30: CHANGES IN GRADE-SPECIFIC ATTENDANCE RATES, BY ROUND

Grade Level of Classroom	Girls' Attendance Relative to BL			Boys' Attendance Relative to BL		
	ML1	ML2	EL	ML1	ML2	EL
Grade 1	-3.9	0.5	-7.4	0.0	2.9	-1.9
Grade 2	0.1	-1.7	-6.0	0.2	-1.0	-4.8

[72] The data did not offer any clear explanation for the different trajectories of different states. However, data and trends across evaluation rounds suggest some possible factors that may have affected attendance rates differentially. First, we note that Jubaland's attendance rate at both BL and EL remains substantially lower than in Banadir and South West State. In the first case, the attendance was 78.0 percent at BL and 77.7 percent at EL, compared to 91.9/85.7 at BL/EL for Banadir, and 91.9/82.8 at BL/EL for South West State. During the BL evaluation, the evaluation noted reports of significant post-election insecurity in Kismayo and Gedo from field teams. This possibly affected the attendance rates of learners at that time period, and afterwards as well, leading Jubaland to maintain consistently lower attendance rates. It is possible that because Jubaland already had low attendance rates due to contextual challenges, additional challenges such as floods or droughts had less of a marginal effect on attendance rates compared to Banadir and South West States schools.

Grade 3	-3.9	-3.3	-5.9	-1.6	-0.9	-4.7
Grade 4	-8.5*	-2.5	-1.2	-6.2	-2.8	-5.5
Grade 5	-5.1	-4.7	2.8	-5.0	-3.9	-5.6

* = statistically significant at the 5 percent level

While only one change (from BL to ML1) was observed to be significant, the numbers in Table 30 above further highlight the general trend of declining attendance rates between BL and EL, as measured by headcounts, across both boy and girl learner groups and across grade levels. Among girl learners, we observe that the largest declines in attendance rate between BL and EL are, in decreasing order, for grades 1 to 3.[73] On the other hand, higher grades (grade 4 and 5) observed larger declines between BL and ML1. With the data available, it is unclear what might be causing this reversal in trends between rounds. It may be the case that the younger learners who at around ML1 were concentrated in lower grades have in general progressed to higher grades. Given that they have done so, one might assume that they are generally more motivated and/or capable of staying in school, which may explain why attendance in higher grades saw slight increases since ML1. However, without individual-level data to verify learners' trajectories across grade levels and across time, this remains largely speculative. Moreover, this does not explain why the more recent cohorts of lower grade learners have experienced sharper declines in attendance rates. It also does not explain why attendance rates for boys have declined at a similar rate across all most grade levels. As such, caution is warranted explaining the observed results.[74]

Finally, in examining some other variables beyond attendance rates, we find some encouraging signs of progress over time. Specifically, the likelihood of attendance records being available appears to have increased over time: when controlling for geographic zone and grade level, we find that the proportion of classrooms with available records increased by 13.8 percentage points between BL and ML1 ($p = 0.098$) and by 12.2 percentage points between BL and ML2 ($p = 0.054$). Using a similar model, we examined how complete the records are over time. We observe an increase of 23.2 percentage points ($p= 0.054$) between BL and ML1, and an increase of 29.2 percentage points ($p = 0.007$) between BL and ML2, in the proportion of classrooms with records marked as "extremely complete".

6.2. Teaching Quality

The second intermediate outcome we examine is teaching quality. Like attendance, teaching quality is expected to have an important impact on future learning and transition outcomes by improving learners' ability to absorb and apply lessons. The AGES programme aims to strengthen teaching quality through a combination of teacher training addressing gaps in literacy and numeracy teaching, improved capacity for

[73] The sizable declines in attendance rates in early grades, particularly for girls, coincide with recent governmental efforts to increase enrolment over the past two years. It may be possible that increased enrolment may contribute to decreasing attendance rates. This may be due to a number of reasons, including increasing class sizes that are not conducive for some students' learning, or the enrolment of students who still lack the means to attend school regularly after enrolment.

[74] The evaluation team also examined dropout rates to determine whether they might explain the declining attendance rates. This was done under the assumption that students who have dropped out, by definition, would not be attending classes, but due to poor or outdated recordkeeping may still be listed as an enrolled student. This would enlarge our denominator when calculating attendance rates, when a reasonable case could be made that students who have dropped out should not count towards the attendance rate. In re-running the above regression with grade and round interaction terms, but adding the dropout rate as a control variable, we find a significant and negative relationship between dropout rates and attendance rates ($p = 0.017$ for girls, $p = 0.042$ for boys). However, the coefficients of the grade and round terms remain relatively unchanged, which might suggest that this cannot fully explain the declines in attendance rates.

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 practices. The use of positive teaching practices may also improve transition (defined as continued engagement in formal education) as girls may be more incentivized to attend and stay in school; in contrast, poor teaching practices may push learners to drop out of school and engage with other out of school activities.

In this section, we discuss four dimensions of teaching quality and practices: professionalism, gender equity, the use of physical punishment, and pedagogical practices. One important caveat to note is that the current EL evaluation differs from prior rounds in that we did not conduct classroom observations in the formal school classrooms in which field teams collected school-level data. This limits our analysis of teaching quality to the use of reports from girls themselves, without addition real-time observations to validate.

However, we believe that this has minimal impact on our presentation of results in this section. First, since classroom observation data was already collected in formal schools during ML2 – just over a year prior to the EL data collection – we summarize the main findings from the ML2 report, where relevant. Second, there is likely a degree of social desirability bias that influences teachers' use of teaching practices when visited by our field teams, particularly as many schools have already been visited numerous times in prior evaluation rounds. As such, the use of girls' surveys may in any case provide a more reliable means of verifying teaching practices.

In the following sections, the results are reported for the full panel of FE girls who were surveyed in each of the four evaluation rounds (BL, ML1, ML2, EL). However, doing so excludes some girls who were contacted in ML2 (and may have been in the BL – ML1 – ML2 panel) but who were not reached at EL. As such, the results from previous evaluation rounds presented here may differ slightly from previous reporting due to attrition of some FE girls.

Teacher Professionalism

We examine teacher professionalism by analysing changes in two measures: whether teachers are reported to make girls feel welcome in the classroom, and the degree of teacher absenteeism. In keeping with trends observed during the ML2 evaluation, we find that the vast majority of girls – over 94 percent in all rounds – claimed that their teachers made them feel welcome. Unsurprisingly, this means there is very little change from BL to EL: while the data reveal a slight decline of 1.6 points in the percentage of girls claiming as much, this change is not statistically significant. The proportion of girls claiming that teachers were not often absent has remained substantially lower compared to the proportion of girls who claimed their teachers create a welcoming learning environment for them. However, these figures are still high in each round. Moreover, these measures saw a significant 17.1 point increase from BL to EL, highlighting an improvement in schools' ability to ensure teacher attendance.[75]

As with the tracking of student attendance discussed in a previous section, this may reflect the strong emphasis which school governance and monitoring programs place on tracking teacher attendance. Indeed, as will be later discussed in our analysis of school management and governance, there is a significant and large increase in the proportion of teachers who claimed the CECs monitor teacher absenteeism: from BL to EL, this figure increased from 55.6 percent to 77.1 percent.[76] However, it is worth noting that the proportion of girls who claimed teachers were rarely absent declined somewhat between ML2 and EL, after making steady and

[75] The results presented below were based on a simple bivariate regression model to examine the changes across rounds. Adding a control for geographic zone, however, does not change the results or our interpretation thereof.
[76] See "School Management and Governance" in this report.

substantial gains in from BL to ML2. While the ML2 to EL decrease is not significant (p = 0.188), it is worth continuous monitoring during successor programs, as it may portend a reversal of the gains made during and immediately after the implementation of the AGES program.

TABLE 31: CHANGE IN REPORTED TEACHER PROFESSIONALISM AMONG FE GIRLS

Outcome	BL	ML1	ML2	EL	Difference (BL to EL)	P-value
Feels welcome	95.7	98.9	96.8	94.1	-1.6	0.60
Teacher not often absent	67.2	77.8	89.8	84.3	**17.1***	0.00

Disaggregating the changes in reported teacher absenteeism by geographic zone revealed that this increase in girls reporting teachers being rarely absent is observed across all zones, and all in the double digits. Jubaland experienced the highest increase, at 18.9 points, followed by Banadir at 17.1 points, and South West State at 12.8 points. Of these, only the change in Banadir was statistically significant. As with the ML2 round, however, the regions that saw the largest increases were located in Jubaland and South West State. In the former case, the Gedo region saw a 69.2-point increase from BL to EL in this measure; in the latter case, the Lower Shabelle region saw an increase of 50.0 points from BL to EL, though we note the EL level for Bay region remains the lowest of all regions, at 66.7 percent.

TABLE 32: PROPORTION OF GIRLS REPORTING THAT TEACHERS ARE RARELY ABSENT, BY GEOGRAPHIC ZONE

Zone	BL	ML1	ML2	EL	Difference (BL to EL)	P-Value
Banadir	63.5	94.7	89.3	81.1	**17.6***	0.02
Jubaland	68.6	50.7	88.9	87.5	18.9	0.14
South West State	71.8	94.9	92.3	84.6	12.8	0.25

Neither the quantitative nor qualitative data revealed clear-cut explanations for why these regions in particular experienced precipitous increases on this measure. However, it may simply a case of "catch-up" growth, where the AGES' (or similar) programme's impact would be most strongly felt in the areas that faced the greatest challenges.

Classroom Gender Equality

In this sub-section, we analyse the degree to which teachers use teaching practices that advance gender equality and 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. The AGES

programme seeks to not simply ensure gender equality, but to be gender transformative through an emphasis on equity. However, since our analysis of the EL data can only include analysis of girls' self-reported experiences, we are unable to examine gender equity per the EL round. As such, we draw on the EL girls' survey data to address the equality component.

To begin, we use a simple metric of classroom gender equality, as experienced by girl learners themselves: their response to the statement, "My teachers treat boys and girls differently in the classroom." We consider disagreement with this statement as a belief that the classroom is gender equal. Disagreement with this measure has seen consecutive round-to-round decreases since BL; by EL, the proportion of girls disagreeing with the statement had declined by 24.1 points, from a high of 63.6 percent at BL to 39.5 percent at EL. This change is also significant at the 5 percent level. Even after controlling for geographic zone and region in separate regression models, we still observe a significant 24.0-point decline from BL to EL.[77] [78]

In contrast, girls' responses to additional questions about specific teaching practices would imply improving gender equality in the classroom. Specifically, we ask girls 1) whether teachers directed more questions to girls, to boys, or equally to both, and 2) whether teachers directed more difficult questions to boys, to girls, or equally to both. Table 33 below outlines the results across all rounds for the above two indicators, as well as the proportion of girls who claimed that teachers asked more questions and asked harder questions, respectively, to girls. The results all highlight increasing equality among boys and girls, defined as the equal treatment of both genders. Nearly all changes from BL to EL are statistically significant, while one (on whether questions to boys and girls are equally difficult) has a p-value marginally above 0.05.

TABLE 33: CHANGE IN GENDER EQUALITY PRACTICES IN THE CLASSROOM[79]

Outcome	BL	ML1	ML2	EL	Difference (BL to EL)	P-Value
Questions equally difficult	87.6	98.7	92.5	96.0	8.4	0.05
Questions harder for girls	11.6	0.6	3.1	2.9	**-8.8***	0.03
Questions directed equally	87.7	96.2	97.5	96.6	8.9*	0.03
More questions directed at girls	12.3	3.8	2.5	2.8	**-9.5***	0.02

One reason for these contrasting findings may simply be that the indicators highlighted in Table 33 do not adequately capture all elements of gender equality in teaching practices. These questions capture teaching practices occurring inside the classroom, but may not adequately capture the full spectrum of additional practices that may not be equally applied to both genders. In this sense, while it is possible that the AGES

[77] Disaggregating by region, however, highlighted interesting trends in the data. Specifically, both Gedo and Lower Shabelle regions saw increases in this measure while other regions saw large declines, though both Gedo and Lower Shabelle still had the two lowest scores at EL.

[78] We also control for enrolment status (i.e. whether a surveyed girl was enrolled in school at the time of the survey). The coefficient for the control variable was not significant. While the addition of this control variable reduced the magnitude of the BL to EL change to 17.0 points, it remains a large and significant coefficient at the 5 percent level.

[79] Note that the ML2 table presented results from the classroom observations, whereas this report presents results from the girls' survey.

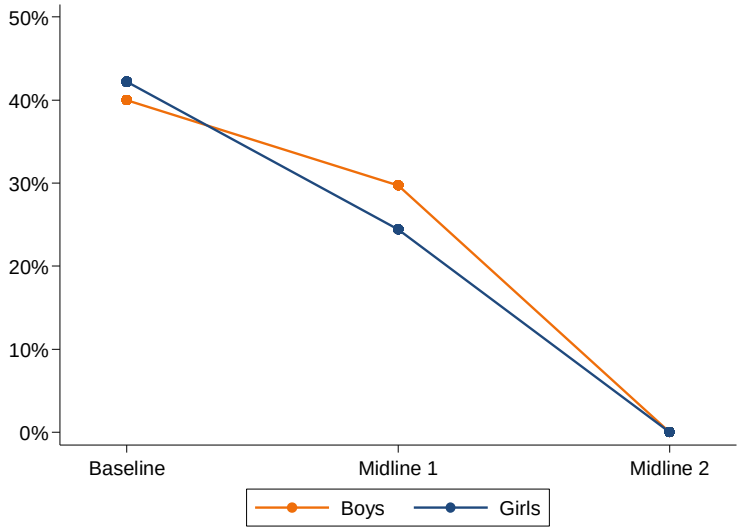
programme has helped improve some practices related to gender equality, more time and resources may be needed to more comprehensively implement a gender equality across all teaching and practices. Taken together then, this analysis shows both encouraging signs of progress in AGES-supported school, as well as hints of additional teaching practice areas that may need improvements.[80]

Disciplinary Practices

We now turn to examining negative disciplinary practices used by teachers, including punishing students for wrong answers and the use of corporal punishment. Such practices may deter learning by making students afraid to take part in class or to attend school, and may even push students to drop out or seek out other activities outside of school. We first briefly summarise the findings from the ML2 report, which had included analysis based on classroom observations. However, bearing in mind 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 also validate our findings with data from the girls' survey, which includes data from the EL.

In brief, the analysis of classroom observation data up to ML2 highlighted a sharp drop in the proportion of classrooms where our monitoring teams observed physical punishment used by teachers. The rate of decline was nearly identical for both boys and girls, who were observed to be physically punished in 40.0 percent and 42.0 percent of classrooms, respectively at baseline, only to have both figures drop to zero by the time of the ML2 classroom observations.

TABLE 34: OBSERVED USE OF PHYSICAL PUNISHMENTS, FROM BL TO ML2



[80] One other possibility may lie in the conceptual difference between equality and equity. Where the former implies the same treatment of girls and boys, the latter implies the empowerment girls, who are traditionally marginalised in the Somali education system, and may entail different treatment of boys and girls in order to offset the additional barriers girls face in pursuing an education. However, given the continued prevalence of conservative norms in many of the programme locations, it would seem unlikely that girls' substantial increase in agreeing that girls and boys are treated differently is driven by such rapid gains in teachers using gender equitable practices that may lead girls to be treated differently than boys.

Reports of the use physical and negative disciplinary practices by girls largely mirror that finding based on classroom observations. Specifically, there is a sharp decrease in the proportion of girls who claimed to have witnessed their teacher use corporal punishment of any kind in the preceding week. At BL, this figure stood at 76.1 percent, but had declined to 32.9 percent by the EL. This represents a significant and sharp decrease of 43.2 points. The proportion of girls who reported that teachers punished students for wrong answers had also declined, albeit much less sharply, and the decline is not significant.

TABLE 35: TEACHERS' REPORTED USE OF NEGATIVE DISCIPLINARY PRACTICES, AS REPORTED BY GIRLS

Outcome	BL	ML1	ML2	EL	Difference (BL to EL)	P-Value
Use of physical punishment[81]	76.1	18.6	32.9	32.9	**-43.2***	0.00
Punishment for wrong answer	81.7	60.2	74.2	78.0	-3.8	0.42

While these findings are generally encouraging, there are two trends that warrant caution, and potential action in successor programmes to AGES. First, while the decline in the use of physical punishment, as reported by girls, is large and significant, there is a clear trend of girls increasingly reporting the use of corporal punishment between ML1 and EL. Indeed, the proportion of girls has nearly doubled between ML1 and EL. For the indicator on teachers punishing students for wrong answers, the figures have similarly crept back up since ML1, representing a 17.8-point increase. This might suggest that while the use of negative disciplinary practices initially plummeted during the AGES programme's period of implementation in formal schools, negative practices may gradually return in the absence of continued monitoring.

Second, the small decline in punishing students for wrong answers masks a more worrying trend: of those girls who claimed that teachers punish students for wrong answers, there is a consistent, round-to-round increase in the proportion who claim that teachers use some kind of physical punishment to address wrong answers. At BL, this figure stood at 50.0 percent among the girls who reported that teachers punish wrong answers. Even during the ML1, we note an increase up to 55.3 percent. By EL, it had increased to 69.7 percent. The increase was significant at the one percent level. This is largely consistent with the aforementioned observation that the use of corporal punishment appears to be slowly creeping back up since ML1.[82]

[81] In the ML2 round, this variable was measured only if a girl had witnessed physical punishment used in the past week. The EL coded this variable to also include girls who claimed that they themselves had received physical punishment from teachers.

[82] As previously noted, the set of questions from which the teaching quality analysis draws it data is posed to girls who are currently enrolled in a learning program, and those who are not. One important robustness check is to analyse this variable only among the subset of FE girls who are currently enrolled in a learning program. Doing so leads to substantively similar results: the figure remains at 50.0 at BL (as all girls are, by definition, enrolled in a learning program), 54.9 percent at ML1, and 67.6 percent by EL. The change from BL to EL also remains statistically significant, albeit at the five percent level.

The qualitative data, while providing more details on respondents' perspectives, is similarly ambivalent in terms of explaining the trends around physical and negative punishments. On one hand, some respondents, including teachers, express their growing recognition that the use of physical punishment should be reduced:

> *To instil discipline in adult students, we have shifted our approach from physical punishment to verbal communication. We emphasize the importance of adhering to certain rules and provide clear instructions. Previously, punishments like sitting under the sun were used, but now we focus on engaging in dialogue with the students to address any issues.*
>
> FGD with Teachers, Banadir, Int. 507[83]

On the other hand, other respondents provided responses that further emphasised that the use of negative disciplinary practices remains a considerable problem in schools. For instance, one respondent offered the following view, which indicates a degree of acceptance that some physical punishment is acceptable (and may not be considered punishment at all):

> *There is no punishment, if two people fight over something. If we see there is a wrong punishment given to the students, we stop it. Only the small beating for when the students are disciplined is allowed, but the heavy punishments are not allowed. So we monitor it and we are always around are the school.*
>
> - FGD with CEC members, Lower Shabelle, Int. 101

That the former view came from a CEC member, and in response to a question on child protection, further highlights the continued problem of negative disciplinary practices, in spite of significant early progress made between BL and ML1.

Pedagogical Practices

The last component of teaching quality that we examine is the use positive, pedagogical practices by teachers. The ML2 evaluation round utilised both classroom observations and girls' self-reports to cross-reference the findings. As the EL did not employ classroom observations, we first briefly summarise the classroom observation findings from ML2, before proceeding to outlining the findings from the girls survey across rounds.

In short, the classroom observations analysis from ML2 paint a decidedly mixed picture of pedagogical practices. In examining the use of formative assessments (FA), the evaluation found that teachers' stated use of FAs increased much more sharply (49 percent at BL to 96 percent at EL) than teachers' available records of actual FA use (55 percent at BL to 59 percent at EL), which suggests either an overreporting by teachers, or poor record keeping. Classroom observations also sought to document how often teachers used the

[83] This particular quotation comes from an NFE teacher, but as the school itself also has formal education classes, changes in disciplinary approaches may also have occurred among formal education groups in the same school.

following practices, among others: 1) use of student-centred activities or games, 2) open-ended questions, 3) group work, 4) solicitation of student opinions, 5) students instructing each other, and 6) involvement of students who are not participating. Of these, only the sixth indicator saw a significant increase between BL to ML2, while two of these indicators saw decreases from BL to ML2, though these were not statistically significant. Finally, the observations also revealed the continued use of ineffective teaching practices, such as students spending most of their time copying from the board or repeating after the teacher. [84]

To triangulate these observation findings, we turn to girls' reports of positive teaching practices. As with the ML2 evaluation, we find that the reported use of teaching practices has significantly increased between BL and EL. The table below highlights that all but one of the six indicators of positive teaching practices have a BL to EL difference that is between 14.0 and 28.5 points, and that they are significant at the 5 percent level.

TABLE 36: CHANGE IN USE OF POSITIVE TEACHING PRACTICES, AS REPORTED BY FE GIRLS

Outcome	BL	ML1	ML2	EL	Difference (BL to EL)	P-Value
Teacher often explains use of subjects	69.9	84.4	80.1	86.0	**16.1***	0.01
Teachers often gives ideas to learn outside of class	47.8	66.7	63.8	61.8	**14.0***	0.04
Teachers gives ideas to study outside of class	92.5	90.3	93.5	91.4	-1.1	0.76
Lessons move at the right speed	39.8	58.6	67.2	57.5	**17.7***	0.00
Teacher often uses different ways of explaining	49.5	60.2	61.8	68.8	**19.4***	0.00
Teacher often encourages participation	52.7	68.3	81.7	81.2	**28.5***	0.00

The overall improvements in reported teaching practices is also supported by the qualitative data, where some teachers have provided account of changing their practices. One teacher, for example, explains that "The subject that I teach now has questions at the end of the lesson, and I used to ask those questions. But now I get the questions [from the students] about the lesson, and that is an experience that I have gained."[85] Other teachers provided the following perspectives:

> *Yes, I have incorporated group discussion on the daily class activity and stopped thwacking my students.*
>
> - FGD with Teachers, Lower Juba, Int. 509

[84] For more details, refer to the AGES Midline 2 Report.
[85] FGD with Teachers, Bay, Int.504

> *I have changed something. The subject that I teach now has questions at the end of the lesson, and I used to ask those questions. But now I get the questions from the lesson, and that is an experience that I have gained*
>
> - FGD with Teachers, Bay, Int. 504

Still, it is important to note that the classroom observations and the girls' reported experience uses indicators that do not overlap entirely. For example, it is possible that teachers lack any use of participatory or student-centric activities in their lessons, but lecture students at the right speed such that students believe that they have a good grasp over the class material. It is also possible that girls' definitions of the positive teaching practices may differ from what the classroom observations are measuring. Where the classroom observation exercise considers pedagogy as participatory if girls in the classroom are actively contributing to class discussions or activities, girls' themselves may consider their teachers' attempts at getting them to follow along a lesson as encouraging them to participate, even if the girls are mostly passively receiving information from an instructor. As such, we interpret the results of this sub-section's analysis as indicative evidence of substantial improvements that are sustained across rounds, while still noting the room for improvement, as identified in the classroom observation data.

6.3. Leadership and Life Skills

This section aims to evaluate the advancements in leadership skills, self-confidence, and life skills gained throughout the program. These elements are measured through self-perception, primarily utilizing the Youth Leadership Index (YLI) as the primary tool. Besides reporting the overall scores, the analysis will break down the changes from the beginning to the end of the program based on various factors. This will help identify the main influences on self-confidence and leadership, while offering a comprehensive understanding of the program's achievements and areas needing improvement.

The program's theory of change suggests that better-quality learning opportunities tailored to the needs of ultra-marginalised girls, along with shifting social norms towards broader life opportunities, will affect the development of life skills. Improved life skills are expected to enhance literacy, numeracy, financial literacy, and support the transition to further education, employment, or self-employment.

The significance of girls' self-reported leadership skills in enhancing learning and transitional outcomes is supported by broader findings from evaluations and lessons learned from projects like SOMGEP-T and Phase 1 of the GEC initiative. The GEC thematic review on self-esteem shows that interventions aimed at building girls' self-esteem result in positive changes in their attitudes, including increased motivation to attend school, new aspirations, and a sense of belonging at school. These changes positively impact attendance, class participation, and overall learning outcomes.

The Youth Leadership Index is a composite measure based on a set of 21 questions, developed by CARE International, and successfully piloted and used in several countries. This indicator evaluates a respondent's self-confidence, decision-making, voice, vision, and organizational skills (including the ability to motivate others and collaborate on common issues). The table below lists the questions used to construct the YLI.

TABLE 37: 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.
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 endline 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.[86]

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.[87]

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 48.5. This rose to 52.3 at ML1 and had a tremendous jump to 66.4 at ML2. By the endline the YLI score reached 68.8, in total being a statistically significant 20.3 points increase from the baseline (or a 41 percent increase from baseline). It is a very meaningful success for the programme in increasing leadership skills from a relatively low base and falls short by only 1.2 points from the 70 points target.

TABLE 38: CHANGE IN YLI SCORES BY ROUND AND SCHOOL TYPE

Outcome	BL	ML1	ML2	EL	Difference (Earliest Round - Latest Round)
FE girls	45.1	49.0	66.5	70.6	**25.5***
ABE girls	48.1	53.0	66.1	68.1	**19.9***
C1 NFE girls	52.3	55.1	66.7	67.5	**15.2***
Total	48.5	52.3	66.4	68.8	**20.3***

The proportion of girls meeting the target of 70 percent 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 12.4 percent and far from the endline target of 80 percent of girls reaching this target. At ML1 the figure stood at 14.2 percent, an increase of 1.8 percentage points or nearly. But progress has truly been made since between ML1 and ML2, rising 29.1 percentage points to 43.2 percent. From ML2 to the EL progress deaccelerated and it only increased in 3.1 percentage points, to 46.4 percent.

The gains are particularly notable among FE girls, of whom only 5.4 percent at baseline scored above 70 percent on the YLI. This points to some success in the programme of increasing youth leadership skills- a positive finding for the intervention. While the new score at the endline still falls far short of 80 percent and

[86] 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.

[87] 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.

the growth deaccelerated from ML2 to endline, the YLI increased four times the baseline value, and ten times for the formal education centres ten times, which still are considerable numbers.

TABLE 39: CHANGE IN PROPORTION OF GIRLS WITH A YLI ABOVE 70 BY ROUND AND SCHOOL TYPE

Outcome	BL	ML1	ML2	EL	Difference (Earliest Round - Latest Round)
FE Centres	5.4%	14.0%	44.6%	50.5%	**45.2***
ABE Centres	12.3%	14.0%	41.9%	46.4%	**34.1***
NFE Centres (Cohort 1)	19.9%	14.8%	43.2%	42.0%	**22.2***
Total	12.4%	14.2%	43.3%	46.4%	**30.9***

For both the aggregate YLI score 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. We run a regression to estimate what is the coefficient of age at the baseline. This should tell us what the approximate effect of age on the YLI score without the AGES program is. If we multiply the coefficient by 3 (one for each round) we generate a benchmark of what would be the progress of the YLI score due to age without the program. We then compare the estimated increase in score with the actual increase observed at the endline we can observe, what, approximately would be the effect of the AGES program, without the effect of girls growing up.

While the ML2 report shows a significant coefficient of age for girls of FE and ABE centres, for this endline report the same analysis with girls that are present in all rounds yield no significant results at the 5% level. For the FE centres the regression coefficient was 1.96 (significant at the 10% confidence level), meaning we would expect a 7.84-point increase in the YLI index over 4 years. For girls in ABE centres, we measure a regression coefficient of 2.71 at the 10% confidence value, therefore we would expect a 10.84-point increase in the YLI index value. Similar to the ML2 report, the coefficient value for the girls on NFE centres is slightly negative and non-significant at the 10% level (p-value =0.334) and as such it 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, two girls from Banadir expressed: "Yes, I've mastered

tasks I once found challenging, such as using a mobile device. Now, I can confidently articulate my thoughts and opinions, safeguarding my rights from infringement." And "Yes, I have developed abilities such as making phone calls, reading books, and storing contacts, which I didn't possess previously."[88] In the same spirit, many groups of girls declare learning how to read, write and do mathematics, for which some girls declare that they have gained autonomy and are more able to participate in the community[89].

The impact of the programme on the development of leadership skills is underscored by the relationship between ever participating in GEFs and YLI scores. Looking only at YLI score differences at the EL, participation in GEF has no statistically significant effect on the score. But when we consider *changes* in YLI scores over time, we see a positive, statistically and substantively significant effect of GEF on the score. On average, girls who participated at least once in GEF score 2.8 points higher on the YLI score across the whole sample.

In this model, GEF participation is associated with slightly greater *gains* in YLI scores between BL and EL among girls in our panel sample. Ever participating in GEF translates to statistically significant 2.8 points more in the YLI score, after controlling for round, region, and demographic characteristics [90]. Girls who did not participate in a GEF experience a substantial gain – 19.8 points – in YLI scores from BL to EL; however, GEF participants experience a slightly larger gain of 20.3 points over the same period. The impact of GEF participation is most concentrated among the FE cohort, with a statistically significant coefficient of 4.7 points in the regression, while for ABE and NFE girls the coefficient is non-significant. This difference 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 and South West State– reveal distinct impacts of the programme across each region. The most notable change occurred in Banadir, where there was a substantial increase from 49.1 to 59.1 between BL and ML1, followed by another significant jump to 69.1 by ML2, but ended with a small decrease of 1.5 points to 67.6 by the EL. Still, this represents a statistically significant increase of 18.4 from BL to EL, corresponding to a percentage gain of 38% in the YLI score since BL. Jubaland, on the other hand, witnessed a modest rise to ML1, but grew to the same level as Banadir by ML2 with a score of 68. It ended up surpassing Banadir with a further increase to 73 points by EL. 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 18.3 points, a statistically significant net gain of 13.7 points since baseline.

TABLE 40: CHANGE IN YLI SCORES BY ROUND AND AREA

Outcome	BL	ML1	ML2	EL	Difference (Earliest Round - Latest Round)
Total Points					
Banadir	49.1	59.1	69.1	67.6	**18.4***

[88] ABE Participant of Vignettes, Banadir, Int. 1103
[89] Participant of Vignettes, Banadir, Int. 1601; Participant of Vignettes, Banadir, Int. 204.
[90] This finding is robust to regression-based control variables for age, region, and other demographic characteristics.

Jubaland	44.0	47.3	68.0	73.0	**29.0***
South West State	52.4	47.8	61.0	66.1	**13.7***
Proportion of girls reaching 70 points					
Banadir	14.9%	24.3%	50.5%	41.4%	**26.6***
Jubaland	3.0%	11.6%	44.5%	57.3%	**54.3***
South West State	18.7%	2.6%	31.6%	41.9%	**23.2***

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 14.9 percent at BL to 41.4 percent at EL, and Jubaland from 3 percent to 57.3 percent. 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 18.7 percent to 2.6 percent, it has since rebounded to 41.9 percent by EL. In the case of South West State, the high score at BL was driven by C1 NFE girls, who had 36.7 percent of girls passing the exam. These results tumbled by ML1 and for ML2 and EL, the proportion of NFE girls reaching the 70% mark remained lower than for FE ad ABE girls. On the other hand, for Jubaland, the increase between ML1 and ML2 was mainly driven by the sudden increase in the proportion of girls reaching the score in the region of Lower Juba. This region went from 3.5 percent of girls reaching the target score in ML1 to 45.1 percent in ML2.

To summarise, there were increases in scores for the Youth Leadership Index over the period from BL to EL and in the proportion meeting the target of at least 70. The target set at the baseline – 80 percent of girls reaching the target – has not been met by the end of the project, given a deacceleration of the YLI between ML2 and EL. Still, it begs the question if, with more time or resources the target could be met. In general, the point rise occurred across each of the original cohorts – FE, ABE, and C1-NFE schools, and between zones, it was largely driven by improvements in Jubaland, though all regions saw large and statistically significant gains since baseline.

6.4. School Management and Governance

In this section we discuss the quality of school management and governance. Effective school management offers numerous benefits for student learning, enrolment, retention, and motivation. Well-managed schools enhance learning by hiring qualified teachers, providing them with training opportunities, monitoring their teaching practices and attendance, and ensuring they receive sufficient and timely pay. Additionally, effective management may involve engaging with communities to encourage parents to enrol their children, monitor student attendance, promote community support for education, and address barriers to enrolment and retention for marginalised students. Furthermore, well-managed schools can effectively utilise financial resources to improve infrastructure, develop and implement school development plans, ensure child protection, promote inclusivity, and manage responses to crises.

To improve school management and governance, AGES focuses on the activities of Community Education Committees (CECs) and Ministry of Education representatives working with program schools. CECs receive capacity-building support to enhance their ability to reach and assist marginalised girls, particularly those with disabilities, to enrol and remain in school. Specifically, AGES collaborates with CECs to strengthen their community mobilization skills and adopt inclusive practices. AGES also aids government officials in increasing their understanding of the needs and rights of marginalised girls and in designing and implementing policies for inclusive and special needs education. Additionally, AGES supports government officials in improving quality assurance and school monitoring procedures.

In this evaluation, we first examine the management dynamics captured in the head teacher survey. We then analyse indicators relevant to AGES' work with CECs. To assess the efficacy of CECs, we evaluate their efforts to address barriers to girls' education and the extent to which these efforts align with the actual challenges faced by girls in accessing education.

School Management

We begin by briefly discussing the management modalities of schools. At EL, all surveyed head teachers (100%) reported that their school had a management plan, maintaining the same percentage as in ML2 and increasing from 91.2% at BL. The table below displays head teachers as having authority over six school-related tasks at EL. The table indicates that, as expected, school directors have significant influence over school management. Respondents most frequently reported that directors have authority over the school budget, purchase of supplies and equipment, school calendar, student discipline policies, and personnel and hiring. Community Education Committees (CECs), however, also show substantial involvement, particularly in matters related to the budget, personnel and hiring, student discipline policies, and the purchase of supplies and equipment. Finally, it is noteworthy that the national government (and to a lesser extent, the FMS government) has considerable involvement in the school curriculum and school personnel decisions, as well as some influence over the budget. This dynamic highlights the importance of ongoing engagement with the government.

TABLE 41: 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	29.4	23.5	0.0	14.7	0.0	17.7
School director	52.9	55.9	73.5	67.7	20.6	35.3
Teachers	2.9	0.0	8.8	8.8	0.0	0.0
NGO or charity	2.9	2.9	0.0	0.0	0.0	2.9
FMS government	0.0	5.9	2.9	2.9	11.8	26.5
National government	11.8	8.8	14.7	5.9	64.71	14.7
Other	0.0	2.9	0.0	0.0	2.9	2.9

One key responsibility not included in the above table is supporting the provision of training to teachers. The table below illustrates the change in the average number of female and male teachers receiving training on various topics within FE schools. The results are mixed: half of the training types saw a decline in the number of trained teachers (including gender-sensitive teaching methods, literacy teaching methods, and inclusive education), while the other half saw an increase (including math teaching methods, child protection, and others) for both female and male teachers. The increases and decreases in training were not statistically significant, possibly due to an insufficient number of surveyed schools. These findings differ from the ML2 report, where almost all types of training saw increases compared to the baseline (see ML2 report).

Note from the project: This is an ex-post evaluation, as AGES stopped providing support to the sampled formal education schools in 2021 (other than for the provision of NFE, in the case of schools overlapping with C4-C6). Therefore, the provision of training since 2021 is largely related to the support received from other initiatives.

Regarding the types of training received by teachers, at EL, the highest average number of female and male teachers attended training on child protection (shifting from inclusive education in ML2), followed by literacy teaching methods. Head teachers were not asked who provided the training, making it difficult to determine why certain trainings were prioritized.

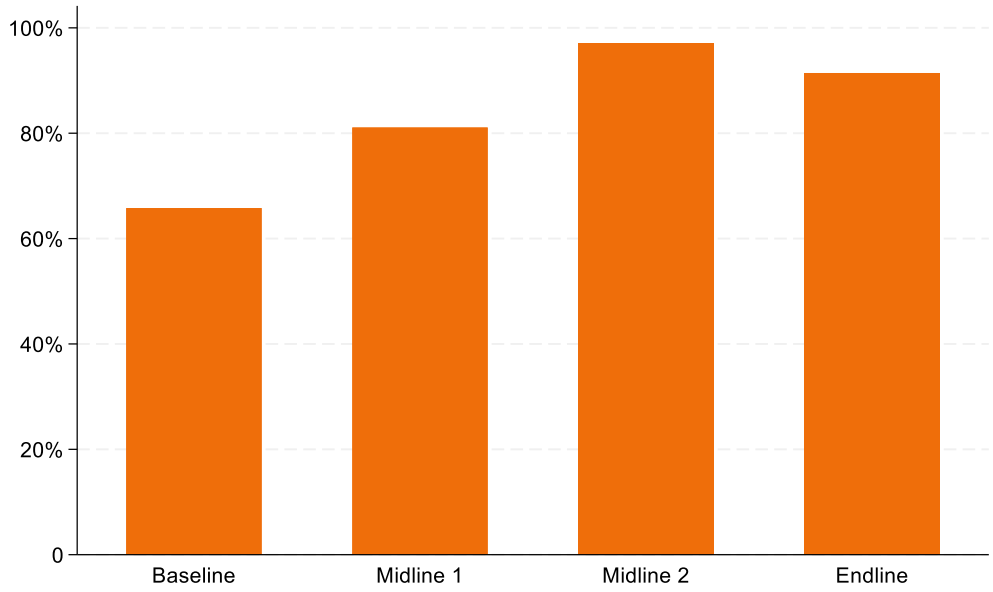
Finally, it is notable that fewer female teachers received training than male teachers. However, schools generally have far fewer female teachers than male teachers; at EL, only about one-quarter of teachers in FE schools were female. Therefore, it appears that female teachers receive training at a higher rate than male teachers. Also, comparatively, there has been an increase in the number of male teachers of 41.0% comparing to an increment of only 9.2% of female teachers.

TABLE 42: AVERAGE NUMBER OF TEACHERS RECEIVING TRAININGS, FE SCHOOLS

	Female teachers			Male teachers		
	BL	EL	Diff.	BL	EL	Diff.
Average number of teachers	6.5	7.1	0.6	16.1	22.7	6.6
Gender sensitive teaching methods	1.1	0.5	-0.6	2.3	1.7	-0.4
Maths teaching methods	1.2	1.7	0.5	2.8	2.1	-0.7
Literacy teaching methods	2.4	2.2	-0.2	3.7	3.9	0.2
Inclusive education	2.6	2.1	-0.5	5.1	4.4	-0.7
Child protection	1.4	2.5	1.1	2.9	4.2	1.3
Other	0.7	1.4	0.7	0.6	1.6	1

We now discuss involvement of CECs in school management as a key indicator of interest for AGES. At EL, 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 is the same number as ML2, and 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 EL, 91.4% of CECs had come to the school during the past year to monitor facilities, a small decline from the previous 97% of schools in the ML2. Still, it means a 25.6 percentage points increase from BL to EL baseline to endline.

FIGURE 8: 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 endline. For data from FE girls, questions on CEC activities were not asked at baseline; as such, the table shows the change from ML1 to endline.

For both head teachers and FE girls at the BL, the most reported activity of CECs was the enrolment of out-of-school girls. But by endline, the most important activity declared by headteachers for the CEC became tracking student attendance, with 85.7% of head teachers mentioning it. On the contrary, 'checking when a girl is absent' was cited by only 29.3% the girls, placing fourth among the activities that girls claimed CECs were undertaking. 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. While asked about their activities, FGD with CEC members do mention their labour of monitoring the schools, and responsibilities in tracking girls' attendance to take awareness actions with the girls and their families. A member of a CEC in Banadir gives the following example: *"For instance, there was a case where a girl was absent from school for a while, but we contacted her parents to understand the situation through our attendance records."*[91] This could also be declared by CEC members due to

[91] FGD with CEC members, Banadir, Int. 104

social desirability bias, but the head teachers report points to the knowledge of the work "behind the scenes" by the CECs.

Head teachers also reported CECs' frequent involvement in tracking teacher attendance and promoting enrolment of out-of-school children. For tracking of student and teacher attendance, reported 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 EL compared to ML1, although there was an increase from ML2 to the EL. Strangely enough, we see an increase of girls stating that CEC support schools between ML1 and EL, but there is also an important decrease in the average number of activities a girl cites the CECs as doing.

TABLE 43: INITIATIVES UNDERTAKEN BY CECS IN FE SCHOOLS

Initiative	BL	ML1	ML2	EL	Difference
Reported by head teachers:					**BL – EL (percentage point)**
Promoting enrolment of out-of-school children	63.9	73.0	82.9	74.3	10.4
Tracking student attendance	47.2	62.2	77.1	85.7	38.5***
Tracking teacher attendance	55.6	59.5	65.7	77.1	21.6*
Following up on cases of dropout	50.0	70.3	65.7	65.7	15.7
Child protection activities	19.4	29.7	60.0	45.7	26.3**
Raising funds for the school	22.2	45.9	37.1	20.0	-2.2
Hiring teachers	11.1	35.1	11.4	20.0	8.9
Reported by FE girls:	**BL**	**ML1**	**ML2**	**EL**	**ML1-EL**
CEC supports school	-	63.6	70.4	76.3	12.8***
Support dropouts to return to school	-	66.6	33.1	49.6	-16.9***
Enrolment of out-of-school girls	-	78.2	33.6	55.4	-22.7***
Awareness raising on girls' education	-	24.4	24.8	36.8	12.4***
Check when a girl is absent	-	49.1	23.8	29.3	-19.8***

Address cases of harassment, abuse, and violence	-	20.6	11.5	17.4	-3.2
Provide hygiene materials for girls	-	13.1	11.6	17.7	4.6
Monitor teaching quality	-	15.4	10.7	12.1	-3.3
Monitor teacher absenteeism	-	8.7	5.4	9.9	1.2
Hire female teachers	-	6.1	5.5	13.1	7.0***

In qualitative interviews, CEC members also described activities conducted by their groups. The amount of times[92] the CEC members mention raising awareness of girls' education as one of its goals or functions coincide with the FE girls' growing perspective of it being one of the CEC initiatives.

The board's role is to raise awareness, engage with parents, and raise funds for the board to assist students [and encourage] girls to study.

FGD with CEC members, Bay, Int. 102

Closely related to raising awareness, some CECs deal with finding drop-outs and encouraging them to come back to school if they are dropping out due to work. Among other roles mentioned by the CEC members is monitoring the schools. Most CEC members say they do it twice a month, while others do it on a weekly basis, several times a week. In their monitoring work they address issues such as the overall cleanliness of the school, student attendance and overall progress of the students' education.

Yes, we do monitoring and we inspect if the work is going on as planned and if the hygiene equipment is complete. And there are people who work for us for free and they keep the hygiene of the school.

FGD with CEC members, Banadir, Int. 109

Some people have left the school without the teacher knowing. The committee looks for them to know why they are absent and, if there is any problem, they solve it. There are also people who left because of money and the committee brings them back and solve their issue.

FGD with CEC members, Lower Shabelle, Int. 101

[92] FGD with CEC, South West State, Int.1003; FGD with CEC, Jubaland, Int. 705; FGD with CEC, South West State, Int. 1202; FGD with CEC, South West State, Int. 1102; FGD with CEC, South West State, Int.1503; FGD with CEC, South West State, Int.1501; FGD with CEC, South West State, Int.1501; ; FGD with CEC, Banadir, Int.104.

While not reported as a primary activity in the table above, several CEC members report some sort of conflict resolution as part of their activities. These can be disputes between the students, teachers and/or school administrators and, in some cases, involves child protection actions, like protecting children from physical abuse from teachers.

> *We conduct monitoring activities every two weeks to ensure smooth operations and address any disputes that may arise.*
>
> FGD with CEC members, Banadir, Int. 108

> *Yes, we monitor the safety of children or students such as problems that occur among students, whether it is conflicts, etc.*
>
> FGD with CEC members, Bay, Int. 103

> *We do monitor the way the teachers are delivering the lessons and other situations that students encounter, such as if there is a student who is beaten or if the methodsused to punish the students which all the schools uses are wrong. We stop them if they beat the students, which is not right.*
>
> FGD with CEC members, Lower Shabelle, Int. 101

The below table provides additional information on CECs' monetary contributions to schools, showing the percent of CECs reported to contribute to teachers' salaries. At EL, around 40.6% of head teachers reported that CECs contribute to teacher salaries, a substantial increase since baseline, but a 7.9 point reduction from ML2. The average contribution amount had also increased substantially since baseline, to around US$42 per month at EL. However, we should note that although sample size is limited, so results should only be taken as indicative. As a result, at EL among schools where CECs contributed to teacher salaries, these contributions made up around 8.6% to 11.2% of teacher salaries.

TABLE 44: CEC CONTRIBUTIONS TO TEACHER SALARIES, FE SCHOOLS

Outcome	BL	ML2	EL	Difference BL - EL
Percent contributing to teachers' salaries	20.7%	48.5%	40.6%	18.9pp
Of whom, average contribution (USD per month)*	4.6	38.2	42.1	37.5
Of whom, average percent of female teacher's salaries contributed by CEC	2.1%	10.3%	11.2%	9.1pp
Of whom, average percent of male teacher's salaries contributed by CEC	3.8%	10.9%	8.6%	4.8pp

*Eliminating extreme values above US$ 1000

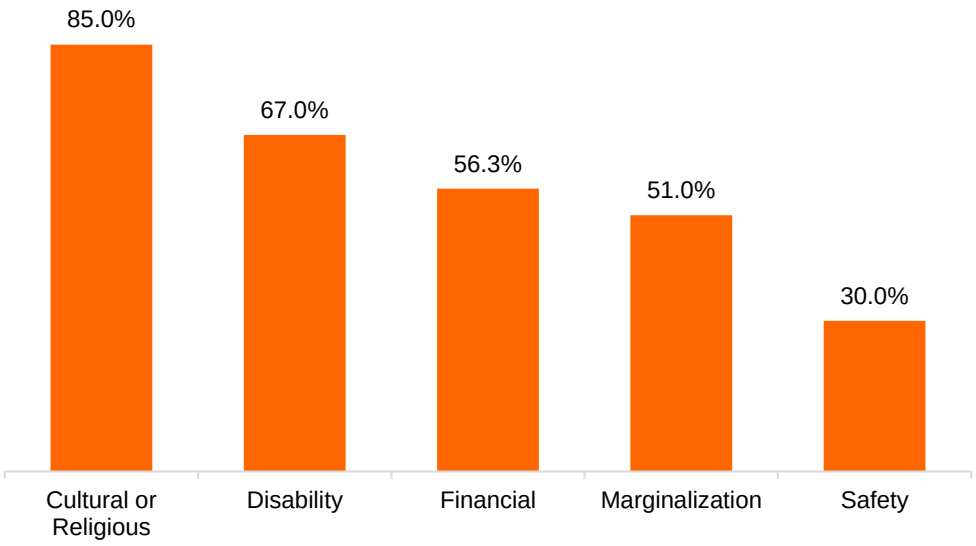
Overall, these results, though mixed, suggest an active role for CECs in school management. However, CEC activity is insufficient to ensure positive learning and transition outcomes; rather, CECs must also be addressing issues of relevance to the challenges faced by marginalised girls. To better understand the relevancy of CEC activities, in Figure 9 below, we show the frequency with which caregivers stated it is acceptable for a child not to attend school in various scenarios.[93] We categorise these scenarios into five groups: (1) 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.

Looking at the Figure 9 from the EL, we find that caregivers were most likely to state that cultural and religious reasons[94] (85%) or disabilities (67%) were acceptable reasons to not attend school. About half of the caregivers said that it was acceptable to not attend school due to financial reasons (56.3%) and other marginalisation-related scenarios. 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 (32.8%) or if she has to do house chores (28.3%). In contrast, caregivers were less likely to express acceptance for students to miss school due to safety-related factors.

[93] 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. We followed the same set of questions for the endline to allow for consistency between rounds. 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.

[94] Caregivers were asked "Under which of the following conditions do you think it is acceptable for a child to not attend school?". For 'cultural and religious reasons', caregivers were asked if it was acceptable to for the child to not attend school if "School teachings are culturally/ religiously inappropriate" and the response options were "Yes (acceptable to not attend)" and "No (not acceptable to not attend)".

FIGURE 9: SITUATIONS WHEN IT IS ACCEPTABLE FOR A CHILD TO NOT ATTEND SCHOOL



As mentioned, Cultural or Religious reasons is the category with the most acceptability for caregiver to not send their children to a given school. Reviewing the qualitative data, the topic of cultural or religious differences or problems with schools is not mentioned by the mother at any FGD, so we do not have a clear picture of what those cultural or religious differences might be, and which ones could represent a barrier to girls' education. It is important to mention that the CEC, with their girls' education awareness actions could be important to address this cultural or religious barriers or misconceptions mothers may have towards their daughters' education. Additionally, the participants in all FGDs with CEC members referred to the support to girls' education by the religious leaders of their zones, and the positive role they take in awareness and promotion of it.

The second category with most acceptability for caregiver to not send their children to a given school was disability, for which people also suffer marginalization, a category where more than half of caregivers consider acceptable for a child to not attend school. CEC members reported awareness raising on the need to educate marginalised girls, including girls with disabilities, pastoralists, IDPs, and married girls, and pregnant girls was important. 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:

> *"One of the issues is that the school was not designed for the disabled, the roads are broken, and the classrooms are not designed for them."*
>
> *"Challenges [for students with disabilities] include being bullied because of their disability and receiving poor care in school, such as bathrooms, which makes learning difficult and leads to dropping out."*
>
> FGD with CEC members, Bay, Int. 102

[On students from agropastoralist families] When the time of plantation comes the parents take their children from the school to help these works, the student has to obey his parents because their lives depend on the animals and the fields. So, this causes a delay to the student's studies.

FGD with CEC members, Bay, Int. 103

These [IDP] students exist and they don't have educational materials when they enroll in the school. They don't have uniforms, and some of them can't afford to buy pens or books and we support them as much as we can. These students need support, sometimes it happens that they don't eat anything, some of them are brought from places that sell Qat and we tell them that education is free. But some of them refuse to attend school, and say that in selling Qat [he receives his] daily income and he or she believes that they will get money from selling Qat [therefore making it a better option].

Member of CEC - FGD, South West State, Int. 109

A variety of activities were described to address these challenges for marginalised students. A couple of CECs described initiatives to help students with disabilities, for example, with one CEC in Banadir stating that they provided a wheelchair (with help from an NGO)[95], or another constructing "berkeds" (small underground water storage/ cisterns) to store drinking water and a few small toilets for people with disabilities for students with special needs.[96] In the case of pastoralist students, CECs mostly mention awareness raising initiatives to attract them to attend school:

Yes, we have done them. When the animals are being run out from them, they came to the city and we enroll them to schools and encourage them to learn. the committee achieved to bring students into the schools by using awareness.

FGD with CEC members, Lower Juba, Int. 110

The third category where caregivers consider acceptable for a child to not attend school was financial. Regarding financial barriers, it is important to note that at EL, 44.1% of schools reported that no fees were charged for enrolment, a decrease since baseline when 52.9% of the same schools charged no fees. Among schools that did charge a fee, the average fee was around US$3.5 per month, which represents a USD$3.5 decrease in the average monthly fees recorded during the BL evaluation, though this figure might remain a substantial amount given the prevalence of extreme poverty in Somalia. Furthermore, 11.8% of schools at EL reported that families had to pay other costs for their child to attend school, the same number of schools as the baseline, considering only schools that were found in both rounds. These results suggest that a large portion of students may face financial barriers to enrolment and continued attendance in school.

[95] CEC - FGD, Banadir 104
[96] Teachers - FGD, Southwest 1203

Note from the project: The average amount charged by schools - $3.5/month – is actually less than half of the average in South-Central Somalia, which stood at $7.7 in 2022 according to the Education Management Information System (EMIS).[97]

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 the following declarations by different CEC members:

> *The parents cannot help as they are farmers. So, the challenge is the parents cannot pay fees, yet they want their children to study.*
>
> FGD with CEC members, Lower Shabelle, Int. 101

> *There are also poor people who cannot afford education, and boys are prioritized over girls, so the girls need support.*
>
> FGD with CEC members, Banadir, Int. 109

> *The presence of fee-charging schools in our vicinity poses challenges for both the community and our school. We are already experiencing high enrolment, and we cannot accept additional students due to the financial limitations faced by the local population. Many people are unable to afford education when fees are involved.*
>
> FGD with CEC members, Banadir, Int. 108

Unfortunately, many CECs member expressed limited ability to help students with financial challenges. A CEC member in Afgoye, for example, stated that, "You know the CEC are normal people with no external financial support, so we cannot do anything."[98] This lack of resources transcends to, for example, the need to pay teachers a salary or provide them and students with transportation to the schools, as discussed in the FGD in Baidoa: "The teacher does not have enough resources to provide education; for the example of books and also the teachers should receive transportation."[99]

Nevertheless, when asked if they do fundraising activities, in almost all of the FGD with CEC members it was stated that the committees collect money for different purposes, such as providing books and learning materials, paying teachers salaries, providing school maintenance, buying water, or assisting students with special needs. However, most of the fundraising is only carried out for specific circumstances.

> *Yes, the committee does fundraising only when there is a need, the committee has a cashier and the money is collected when there is a need, for example, if there is a need for water or that teachers need*

[97] Federal Ministry of Education, Culture and Higher Education, *Annual Education Statistics Report 2021-2022*, pg.12
[98] FGD with CEC, South West State, Int. 1003
[99] FGD with CEC, South West State, Int. 1501

a small amount of money or the school needs maintenance and the money is collected from the committee, teachers and parents.

FGD with CEC members, Bay, Int. 104

It is a little what we have collected, and we give the student based on their needs as support and we collect it from the community.

FGD with CEC members, Banadir, Int. 109

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, the precarity of the working conditions seems to limit the ability of the CECs to reach the more vulnerable children.

6.5. Community Attitudes

In this section, we examine the shifts in community attitudes towards girls' education, as these can significantly help marginalised girls overcome barriers to attending and completing school. The perceived benefits of education may affect caregivers' choices to send their daughters to school instead of keeping them home for household chores, early marriage, or income-generating activities. This perception also impacts the level of financial or other support communities provide to help girls remain in school. Given the high poverty levels and limited resources in the areas where AGES is implemented, investing in education might require households to forgo essential needs. Thus, caregivers' decisions to keep girls in school are influenced not only by their views on the importance of education but also by their belief that the economic and social benefits of education outweigh the costs.

We first analyse changes in caregivers' views on when they would allow their children to not go to school, primarily given the cost of education of the need for the children to work at home or otherwise. We then discuss caregivers' aspirations for the level of education their daughter should achieve. Towards the end, we analyse girls' housework responsibilities and impact on school attendance considering the amount of housework they must carry daily. Finally, we provide some insight from the FGDs of the contrasting views of mothers on girls' education.

Caregiver Attitude 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 EL, 99.3 percent 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 BL and EL, we find that most caregivers aspire for their girls to attend college or university—around 92.7 percent at EL for FE and ABE girls.

For FE girls, caregivers' desire for girls to reach college or university increased slightly from baseline to EL, at the significance level of 90%. There was no significant change in caregivers' aspirations for their girl's education for ABE girls, although it decreased by 2.6 percentage points for college or university.

TABLE 45: CAREGIVERS' ASPIRATIONS FOR GIRLS' EDUCATION

	FE			ABE		
Education level	BL	EL	Difference	BL	EL	Difference
Primary or less	0.6%	0.7%	0.1	0.6%	3.3%	2.7
Secondary	10.0%	6.1%	-3.6	10.1%	10.0%	-0.1
College or university	88.6%	93.5%	**4.9***	89.3%	86.7%	-2.6

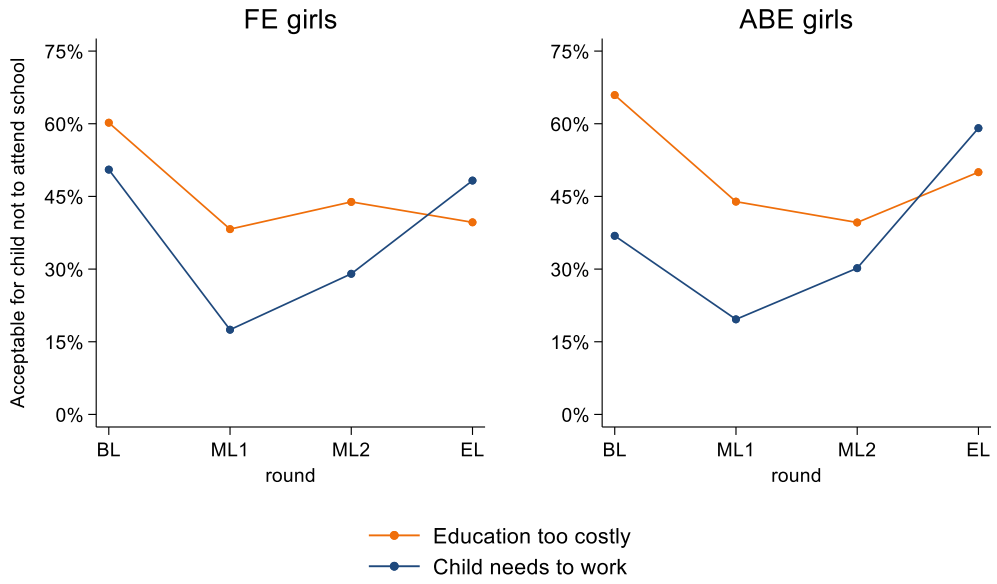
While these results suggest a high level of support for girls' education, it is important to note that this support may be susceptible to social desirability bias and is also phrased very abstractly. Caregivers may wish that their girl be highly educated in the absence of any constraints to do so; however, with the presence of financial, cultural, security, and other limitations, support for girl's schooling may decline.

Investments in Girls' Education

To better understand the value caregivers stance on girls' education despite limited financial resources, we analyse the responses to the statement, "Even when funds are limited it is worth investing in [my daughter's] education".[100] We find high rates of agreement with this statement across all evaluation rounds, at 98.3 percent at baseline, 97.8 percent at ML1, 99.3 percent at ML2 and 99.1 percent at EL for FE girls and 97.8 percent at baseline, 99.1 percent at ML1, 96.2 percent at ML2 and 100 percent at EL for ABE girls. Nevertheless, this result likely reflects social desirability bias, as caregivers are likely aware that opposition to girls' education is seen as undesirable, that leads us to explore other variables to assess the importance caregivers assign to their child's education.

[100] 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 and EL.

FIGURE 10: 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 or do housework. We find a relatively high level of acceptability for children to not attend school due to financial challenges: At EL, 39.7 percent of FE caregivers and 50 percent of ABE caregivers stated it was acceptable for a child not to attend if education was too costly and 48.3 percent of FE caregivers and 59.1 percent of ABE caregivers stated it was acceptable if the child needed to work.

This result is puzzling because, as stated in the previous section, a key component of the work by the CECs has been awareness towards girl's education. An interesting feature is that the acceptance of the caregivers of the girls not attending school due to the need to do (house)work steeply decreased between baseline and ML1, but then bounced back by ML2 and EL. In the case of ABE girls, the endline is almost 23 percentage points higher than the baseline for the acceptability of missing school for housework. For FE girls, the endline is only 2.3 points less than the baseline. In the case of not going to school due to costs, for FE girls the level of acceptability remains at around 20.0 points less throughout ML1, ML2 and EL comparing to baseline. For ABE girls, from BL to ML1 it decreased by around 16.0 points but bounced back by ML2 and ends up at 59.1 percent acceptance at endline, around 23.0 points more than baseline.

To better understand these dynamics, which may be highly contextual, in the below table we disaggregate 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.[101] We first find that the reported level of acceptability has declined from BL to EL for every demographic group analysed; in most cases, the decline was substantial.

Comparing within the analysed demographic groups, we find stronger declines from BL to EL for households headed by men versus households headed by women; for IDP households versus non IDP households; for

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

households where the caregiver has at least primary education versus households where caregiver has none or only madrasa education; and households that have been hungry many or most days in past year versus those which have never been or rarely been hungry in past year. Still, for all measurements except for households headed by women, we see substantial decreases in the acceptability to not attend school due to education costs.

Encouragingly, this pattern suggests an improvement in attitudes towards investment in girls' education among subgroups who might be expected to have less support—and fewer 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 BL. However, the significant and large decrease for male HoHs means that at EL, their level of purported financial support for girls' education is now slightly better than of female HoHs. Similarly, households in a dire and vulnerable situation such as hunger, where one could expect them to prioritize work over education, have lower levels of support of girl's education at baseline, but by endline this trend is reversed. Also, 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. This trend persists, however, and levels of support for girl's education improved more on IDP than non-IDP households.

TABLE 46: ACCEPTABILITY FOR FE CHILDREN TO NOT ATTEND SCHOOL IF TOO COSTLY, BY DEMOGRAPHIC CHARACTERISTICS, IN PERCENTAGE

Subgroup	BL	ML1	ML2	EL	Difference BL – EL
Total	60.2	38.3	43.9	39.7	-20.6***
Zone					
Banadir	49.3	29.3	43.7	43.1	-6.2
Jubaland	66.7	35.2	47.3	43.9	**-22.8***
South West State	69.2	62.2	37.9	25.0	-31.3
Female HoH	43.9	35.7	38.0	42.1	-1.8
Male HoH	67.4	39.4	46.7	38.5	**-20.8***
IDP	70.8	33.8	54.1	46.5	**-24.3***
Non-IDP	53.5	41.1	37.2	35.6	-17.9
Caregiver has no education or Madrassa only	57.2	38.2	45.4	41.0	-16.2
Caregiver has some primary education or above	77.8	38.5	36.0	31.2	**-46.5***

Household has been hungry many or most days in past year	70.0	30.0	60.0	28.6	-41.4
Household has never or rarely been hungry in past year	59.7	38.7	42.8	40.4	**-19.3***
Household speaks Maay	63.8	56.5	42.9	29.0	-34.8
Household speaks Af-Mahatiri, or other (not Maay)	59.0	32.1	44.2	43.5	**-15.5***

Interestingly, at BL, 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 for education among 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 EL the level of support was almost 10 percentage points lower for caregivers with some primary education or above than the caregivers without education.

Looking at results for ABE girls in Table 47 below, we find uneven/irregular results, with high disparities and lack of significant results for the difference between baseline and endline. The main issue for these disparities is the low numbers of respondents for the ML2 and EL whose caregiver responded to the question and who are also part of the survey on every round. This limitation does not allow us to draw firmer conclusions, but, as Figure 10 suggests, there is a general decrease in the acceptance of the child missing school due to education being too costly. In the case of IDP households, we see a significant decrease in the mentioned indicator, which suggests that the AGES program had a positive incidence in this vulnerable group of families. The drop in acceptability in Jubaland, although significant, could be driven by the small amount of observations at EL. For that specific subgroup there were only five responses.

TABLE 47: ACCEPTABILITY FOR ABE CHILDREN TO NOT ATTEND SCHOOL IF TOO COSTLY, BY DEMOGRAPHIC CHARACTERISTICS, IN PERCENTAGE

Subgroup	BL	ML1	ML2	EL	Difference BL – EL
Total	65.9	43.9	39.6	50.0	-15.9
Zone					
Banadir	72.9	47.6	36.4	66.7	-6.2
Jubaland	57.1	41.7	38.5	0.0	**-57.1***
South West State	64.2	41.4	44.4	62.5	-1.7
Caregiver/household characteristics					
Female HoH	67.3	50.0	33.3	66.7	-0.6

Male HoH	65.3	41.1	42.9	43.8	-21.6
IDP	64.4	56.9	43.5	25.0	**-39.4***
Non-IDP	67.0	28.6	36.7	64.3	-2.7
Caregiver has no education or Madrassa only	61.0	40.2	39.0	47.1	-13.9
Caregiver has some primary education or above	87.9	60.0	41.7	60.0	-27.9
Household has been hungry many or most days in past year	93.3	16.7	60.0	100	6.7
Household has never or rarely been hungry in past year	63.4	45.5	37.5	38.9	-24.5
Household speaks Maay	61.4	26.9	45.5	57.1	-4.3
Household speaks Af Mahatiri, or other (not Maay)	68.0	49.4	38.1	46.7	-21.4

Despite these positive results, it is again worth noting that for both FE and ABE girls, 40.7 percent to 43.3 percent of caregivers still 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. This might reflect the very real financial limitations faced by marginalised households in Somalia, despite some notable improvements in levels of support among subgroups that can be considered marginalised, as described in the preceding analysis.

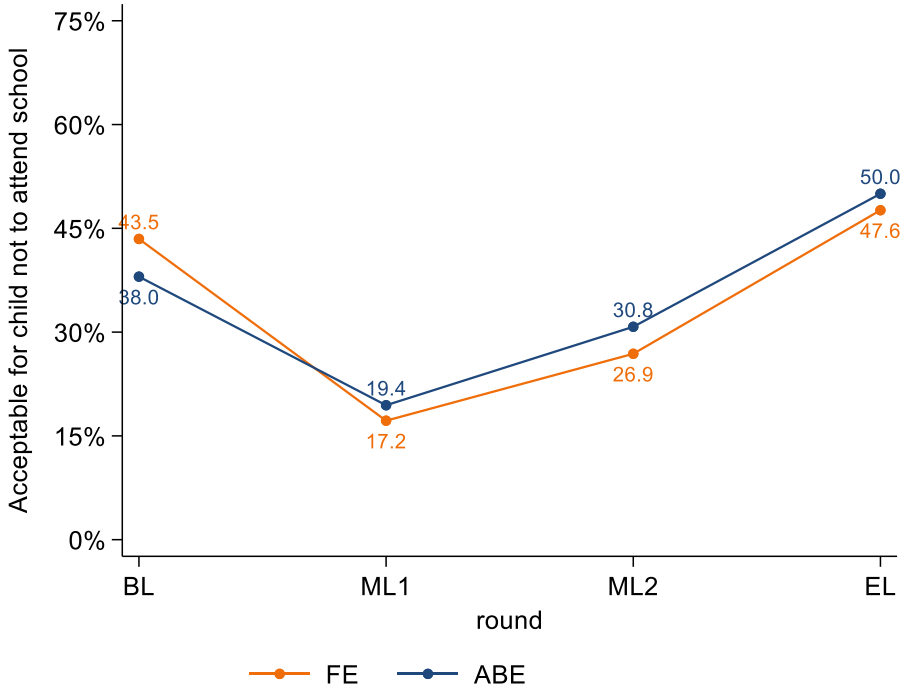
This point is further underscored by the low levels of caregivers who reported having any savings, with only 3.4 percent (21) of 591 reported to having any savings[102] and only 10.0 percent (26) of 235 caregivers participating in saving groups. Of the 26 caregivers participating in savings groups, 14 of them declared they use it for education purposes. Of those 14, 13 say they use those savings for both daughters and sons, and only one says for neither of them. Mind that this response could also be impacted by desirability bias, as respondents might be conscious that giving preference for boys over girls is frowned upon. The most common use for savings was education, followed by food expenses by just a one-person difference.

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 work or do house chores. Figure 11 shows that FE and ABE caregivers' perceptions of acceptability significantly decreased from

[102] We report results for EL only due to limited sample size. We pool all respondents of the endline, equally from their panel completeness to allow for more observations.

BL to ML1, reaching 17.2% and 19.4% respectively, but since, increasing till surpassing the BL level at the EL with 47.6% for FE and 50% for ABE households.

FIGURE 11: ACCEPTABILITY FOR CHILDREN TO MISS SCHOOL DUE TO HOUSE CHORES OR WORK RESPONSIBILITIES



To better understand this evolution of preferences, we analyse changes in the work responsibilities of girls and in the frequency with which girls have been unable to attend school due to work or chores. This analysis will evaluate the difference between the ML1, the lowest point of acceptability of not going to school, with the EL, to derive reason of this important difference.

The results in Table 48 suggest that girls' housework[103] burdens may have substantially increased since baseline. For FE and ABE girls, we find that significantly more caregivers and girls reported that their girls or themselves spend a whole day doing housework, while the proportion for working a quarter a day or for a few hours substantially decreased. This changes on working hours affect the reason for stopping attending school for FE and ABE girls differently. In the case of FE respondents, there is no significant changes on how often housework prevents girl from attending school. For ABE respondents, on the other hand, ABE respondents show a significant increase of declaring that housework *often* prevents girl from attending school, from 5.3 to 14.0 percent of the answers.

TABLE 48: FE AND ABE GIRLS' HOUSEWORK RESPONSIBILITIES AND IMPACT ON SCHOOL ATTENDANCE

	FE	ABE

[103] Work associated with the household include: caring for younger or older family members, cooking or cleaning, fetching water, agricultural work, helping at the family business or working outside home on non-agricultural related activities.

	ML1	EL	Difference	ML1	EL	Difference
Average time spent by girl doing housework						
Whole day	1.7%	10.6%	**8.9***	10.3%	23.8%	**13.6***
Half day	37.9%	37.6%	-0.2	40.0%	43.0%	3.0
Quarter day/a few hours	42.9%	29.4%	-13.5	37.7%	16.9%	**-20.9***
Little time/an hour	17.5%	22.4%	4.8	12.0%	16.3%	4.3
Housework prevents girl from attending school						
Yes, not enrolled because of this	1.2%	1.2%	0.0	5.3%	3.5%	-1.8
Yes, stops her often	2.9%	5.8%	2.9	5.3%	14.0%	**8.7***
Yes, stops her sometimes	19.2%	16.3%	-2.9	22.9%	20.9%	-2.0
No, does not stop her	76.7%	76.7%	0.0	66.5%	61.6%	-4.8

For these variables there are enough observations to include the C1 NFE girls in the analysis. First, we note in Table 49 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 the whole day working by the EL. At the same time, there is a higher proportion of girls at ML1 who say that they do not attend school because of housework. By EL, this amount drops significantly by 9.3 percentage points, also significantly shifting to more girls answering that housework often stops them from attending school.

TABLE 49: C1 NFE GIRLS' HOUSEWORK RESPONSIBILITIES AND IMPACT ON SCHOOL ATTENDANCE

	ML1	EL	Difference
Average time spent by girl doing housework			
Whole day	28.9%	34.1%	5.2
Half day	43.9%	43.2%	-0.7
Quarter day/a few hours	23.1%	16.5%	-6.6
Little time/an hour	4.0%	6.2%	2.2
Housework prevents girl from attending school			
Yes, not enrolled because of this	11.6%	2.3%	**-9.3***
Yes, stops her often	11.6%	22.7%	**11.2***
Yes, stops her sometimes	30.1%	22.7%	-7.3
No, does not stop her	46.8%	52.3%	5.5

The increase in time doing housework could be due to several factors. First, there might have been contextual factors such as the impoverishment of the household due to external reasons like droughts or economic

recession. This could force girls to do housework to meet the household's basic needs. Another reason could be that, given that there has been around 2 years between the ML1 and the endline data collections, girls are no longer in school and with older age, their household responsibilities might have increased. This might be why, for FE girls, housework has increased without significantly changing being the reason that prevents the girl from attending school. To test this idea, we run a regression that controls for age and its interaction with the survey rounds. For FE girls, there is a significant effect of age on spending more time doing housework. The coefficient of the variable 'age' is significant and of a magnitude of 14.7. When analysing how age and time spent on housework affects attending school, the regression shows that working half a day negatively impacts attending school, at least sometimes, but age does not play a role.

For ABE girls, the regression analysis tells a different story. For them, age has no significant relation with the time spent on housework. On the other hand, the time spent on housework does have a positive significant impact on preventing girls from attending school. More time spent on housework reduces the likelihood of the girl's attending to school. Like the ABE girls, for C1 NFE girls age has no significant relation with the amount of daily housework the girls do. In the same fashion, the time on housework has a significant positive relationship with at least being sometimes prevented to attend education. This difference between FE and ABE might be due to different girls' profiles, especially age, as ABE and C1 NFE girls are older girls that use alternatives to formal education.

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 most of the 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 in Jubaland, where caregivers were less likely to report an impact of housework on attendance, although not significantly. In contrast, ABE girls also had the highest burden of housework responsibilities in Southwest State, with a significant increase of 17 percentage points.

Looking at caregiver characteristics, for FE and ABE girls, Table 50 shows contradicting results between the two cohorts. While the effect of housework on school attendance is positive for female headed households and negative for male headed households for FE girls, it is the other way around for ABE girls. This same contradiction is seen for girls in households where the caregiver has at least primary education. Whereas for FE girls the negative effect of housework on school attendance is significantly reduced by 21.7 percentage points, for ABE girls it significantly increases in 24.8 percentage points from ML1 to EL. 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.

TABLE 50: IMPACT OF HOUSEWORK ON FE AND ABE GIRLS' REPORTED SCHOOL ATTENDANCE, BY SUBGROUP

	FE			ABE		
Subgroup	**ML1**	**EL**	**Difference**	**ML1**	**EL**	**Difference**
Zone						
Banadir	19.4%	20.3%	0.9	31.3%	36.4%	5.0
Jubaland	31.9%	26.1%	-5.8	50.0%	35.0%	-15.0

South West State	13.9%	23.5%	9.6	25.4%	42.4%	**17.0***
Caregiver/household characteristics						
Female HoH	13.5%	21.8%	8.4	40.7%	39.6%	-1.1
Male HoH	27.5%	23.9%	-3.6	30.2%	37.8%	7.6
IDP	34.3%	25.4%	-9.0	45.7%	44.1%	-1.6
Non-IDP	16.2%	21.9%	5.7	25.0%	34.6%	9.6
Caregiver has no education or Madrassa only	21.5%	24.8%	3.4	36.2%	36.6%	0.4
Caregiver has some primary education or above	34.8%	13.0%	**-21.7***	21.9%	46.7%	**24.8***

Table 51 shows results for C1 NFE girls. In line with results in Table 50, we find that the impact of housework on school attendance declined—though not significantly—for most subgroups, except for Banadir and non-IDP households. Despite these declines, however, at EL, 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 households in South West State, similar to results in Table 50, showing how the interaction of various dimensions of marginalisation may limit girls' ability to attend school and successfully learn.

TABLE 51: IMPACT OF HOUSEWORK ON C1 NFE GIRLS' REPORTED SCHOOL ATTENDANCE, BY SUBGROUP

Subgroup	ML1	EL	Difference
Zone			
Banadir	50.6%	53.2%	2.6
Jubaland	44.7%	40.0%	-4.7
South West State	65.3%	46.9%	-18.4
Girl characteristics			
Female HoH	59.8%	51.2%	-8.6
Male HoH	47.3%	44.6	-2.7
IDP	62.1%	42.6%	**-19.5***
Non-IDP	47.7%	50.9%	3.3
Caregiver has no education or Madrassa only	54.1%	45.6%	-8.5

Caregiver has some primary education or above	48.1%	59.3%	11.1

Overall, unlike the findings related to financial support for girls' education, these results indicate a potentially worrisome decline in caregiver support for girls' education, as well as a significant housework burden that affects school attendance for FE, ABE, and NFE girls. Acknowledging that a combination contextual changes, poverty, and increasing age for some girls – factors outside of the AGES programme's control – will likely affect some girls' capacity to attend school regularly, future learning programmes might benefit from ensuring that course materials can be adequately adaptive to girls' changing abilities to physically attend school in order to sustain and increase learning gains. This might mean including more modules that are conducive to self-study, as well as more flexible policies for keep learning materials at home (for ABE and NFE girls, as well as older FE girls).

The qualitative data shows that some mothers admit having held their daughters home for chores when the housework is overwhelming. For example:

> *Yes, there were instances when I stopped my daughter from going to school, particularly when there was an abundance of household chores that required her assistance.*
>
> *Yes, there were occasions when I prevented my daughter from attending school, especially when the workload at home became overwhelming for me to manage alone, and I needed her support.*
>
> FGD with Mothers, Bay, Int. 204

Also, mothers point out to financial struggles, and the distance from the house to the education centre as reasons other girls in the community struggle to attend school.

> *Yes, financial constraints pose significant challenges for some girls attending this school, impacting their ability to focus on their studies.*
>
> *Certainly, girls residing in remote areas face obstacles in studying at this school because their families may struggle to afford transportation, resulting in difficulties arriving at school on time.*
>
> FGD with Mothers, Bay, Int. 204
>
> *Yes, some of them come from very far places, and sometimes they face challenges with road closures. They arrive very late, so they miss class because of the road closures.*
>
> FGD with Mothers, Banadir, Int. 207

Nonetheless, a considerable number of interviewed mothers are conscious of the importance of education. Mothers admit to learning to consider girls' education as equally important as the boys'. This signals a change that could have been driven by the AGES project and by different agents that raise awareness about girls' education.

> *Yes, there has been a significant change [on my perspective about girls' education]. Initially, I didn't believe that girls could learn, but now I understand the value of education for girls.*
>
> FGD with Mothers, Banadir, Int. 212

Still, the impact of social desirability bias on these responses must be considered. Consequently, this section reveals mixed community attitudes toward the value of girls' education. A key insight is that, although many people may express support for girls' education, numerous barriers persist that limit actual support. These obstacles appear to affect girls facing various types of marginalisation, such as IDP girls, posing a significant challenge to the success of AGES programming.

6.6. 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 ML1, ML2 and EL evaluations.

Positive Youth Development

During the ML1, ML2, and EL evaluation rounds, data on positive youth development were gathered from the C1 NFE girls using the Chinese Positive Youth Development Scale (CPYDS). This internationally recognized scale measures various aspects of youth development, including resilience and confidence. Seven indicators related to self-efficacy were selected from the CPYDS. The table below shows these seven indicators and tracks the percentage of girls who agreed or strongly agreed with the statements from ML1, ML2 and EL.

TABLE 52: CHANGE IN AGREEMENT WITH CPYDS QUESTIONS OVER TIME

CPYDS Question	ML1 (%)	ML2 (%)	EL (%)	Difference ML1 - EL
1. Lack of control of life	40.3	60.0	63.1	**22.7***
2. Lack of solutions to problems	39.2	54.3	46.6	7.4
3. Inability to change life	28.4	56.6	51.7	**23.3***
4. Helplessness	41.5	49.1	50.0	8.5
5. Fate not in hands	54.0	65.1	60.2	6.2

6. Determine own life	67.6	71.4	67.6	0.0
7. Ability to complete tasks	84.7	81.1	86.4	1.7

Across almost all variables we see an increase in the rates of positive response, three of which are statistically significant. It is only the sixth variable, which asks "I believe things happening in my life are mostly determined by me" that does not grow but stays at the same level as ML1. For the five first variables the wording of the question means that the higher the positive response rate, the worse are the subjects on youth development. The last two are worded the other way around, where higher positive responses mean a positive evolution of youth development.

Control of own life

The first indicator measures the girls' perception of control over their lives, specifically the lack of thereof. When faced with the statement, "I have little control of things that happen in my life," 40.3 percent agreed or strongly agreed during the ML1 round, while 63.1 percent agreed or strongly agreed in the EL round. This increase of 22.7 percentage points is statistically significant and contrary to most of the other CPYDS variables, it also shows an increase from ML2 to EL.

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 39.2 to 46.6 between rounds ML1 and EL. As most variables in the CPYDS scale, the increase was initially produced between ML1 and ML2, where the indicator reached 54.3 percent. Then, from ML2 to the EL the number of girls responding positively decreased by about 7.7 percentage points.

Ability to change

The third indicator measures the extent to which girls think they can 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 28.4. During the ML2 evaluation, this percentage had increased to 56.6 agreeing they lacked the ability to change, and then it slightly reduced to 51.7 by EL. This is the largest and statistically significant increase across all 7 indicators, with 23.3 percentage points from ML1 to EL, although it decreased between ML2 and EL.

Helplessness

The fourth indicator measures the girls' perceptions of helplessness. When prompted with the statement, "When I face life difficulties, I feel helpless", 41.5 percent initially agreed during the ML1 round, and that response significantly increased to 50.0 percent by the endline. As the first CPYDS indicator ('Lack of control of life'), it is the second only response that sees increases of positive responses from ML1, to ML2, to EL.

Fate not in hands

The fifth indicator evaluates how the girls perceive their fate or lack thereof. Overall, 54.0 percent of the individuals initially agreed with the statement, "I feel my life is determined by others and fate". For the ML2 agreement with the statement increased 11.1 percentage points, but ultimately was reduced to 60.2 percent of positive responses by EL. Still, it means a (non-significant) increase of 6.2 percentage points from ML1 to EL.

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 asked the question, "I believe things in my life are mostly determined by me", 67.6 percent initially agreed, while 71.4 percent agreed in ML2. It ultimately returned to the initial level of agreement by EL, negating the small but positive development of this indicator. This result is important because it shows that positive self-efficacy traits have stalled. This indicator reminds us that self-efficacy measures should not be restricted to negative traits, like that one's fate is beyond their control asked in the first question.

Task completion

The seventh and final indicator measures the girls' belief in their ability to finish tasks. The statement of positive self-efficacy provided was, "I can finish almost everything that I am determined to do. There was a small increase in the overall percentage of girls agreeing with this statement from 84.7 in the ML1 round to 86.4 in the EL round. However, this increase was not statistically significant. In between, there was a decrease of 3.6 percentage points from ML1 to ML2, which was recovered for EL. Across all the seven measures, this indicator shows the best result on youth development.

Synthesis

Six out of seven statements related to the self-efficacy indicators experienced increases in the percentages of girls agreeing with them from ML1 to EL. 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"), though accompanied by high levels of agreement 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").

Given this contradictory result of increases in negative youth development indicators and high (but stagnant) rates of positive responses in positive youth development indicators, it is important to contrast the data with the girls' own words across the qualitative data collection from the vignettes exercises and the statements from mothers in FGDs about the development of their daughters' education.

First, we find that almost all girls declare that they have learned how to read and write, and do maths, which allows them to help in their houses and participate in the community, by, for example, helping their mothers to use mobile money. Through education, and the acquired knowledge, the girls also state that they are more involved in the community and, in general, express a gain in confidence.

> *I have become more involved in the community. Previously, I used to stay at home, but now I have learned how to read and do math. I am grateful for this opportunity.*
>
> Vignettes FGD with Girls, Banadir, Int. 606

> *I have transformed my attitudes within the community. Initially, I used to feel apprehensive about interacting with students my age, but now it feels natural to me.*
>
> Vignettes FGD with Girls, Banadir, Int. 612

Mothers, on the other hand, give mixed answers when asked if they think their daughters can express their ideas and opinions at the community and school. Some say that girls express their ideas and needs freely and

stating that they are active on that. Some mothers point out to different systems where the girls can voice their opinions, like girls' club,[104] teachers specifically asking them, or through "Daabuurka", where they stand in line before going to class and have time to express themselves or encourage others to do so.[105] However, other mothers point out that their daughters cannot express their ideas openly, but only to people close to them, like friends or family. One mother attributes it to the girl's shyness.[106]

> *No, she may not be able to express herself to people outside her immediate circle, but she can convey her needs to her siblings and parents, who will assist her accordingly.*
>
> FGD with Mothers, Banadir, Int. 212

It is also important to highlight that the CPYDS indicators address the girls' lives in a general manner. The statement on helplessness, for instance, refers to "life difficulties" in a general sense. Dividing these difficulties into distinct categories such as social, school, family, and work may reveal varying levels of self-confidence and self-efficacy based on the specific aspect of the girls' lives being considered. Furthermore, distinguishing between changes in sentiment due to external environmental changes versus internal emotions can be challenging. This is a broad issue, as short-term events or problems can lead to recency biases in respondents' answers. It is especially pertinent in Somalia, where the environment is highly dynamic, with short-term changes potentially impacting personal safety and food security. These factors are likely to influence perceptions of control, helplessness, and fate.

Access to Protection Services

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

School

During the ML1 round, 95.5 percent of this cohort answered "Yes" when asked if they have someone to report any form of harassment, abuse, or exploitation at the school. For ML2 the percentage decreased 12.5 points to 83.0 percent. For the EL, the percentage remains practically unchanged, staying at 82.4. Still, the decrease between ML1 and ML2 is important to consider. When asked who their top choices are for reporting said abuse, the top choice remained the head teacher or other teachers across ML1, ML2 and EL. The other common choice was the "Other" category, with the vast majority of respondents selecting this option subsequently citing a family member as the party to whom they would report cases of abuse.[107]

Asked about what the places in school were where they felt unsafe, most girls answered the toilets for different reason, such as cleanliness, lack of privacy, and harassment by boys.

> *I feel scared in the bathroom because you can't go in alone unless someone is with you because it doesn't have a door so boys can enter, because it happened in the past that a girl*

[104] FGD with Mothers, Jubaland, Int. 702
[105] FGD with Mothers, Banadir, Int. 101
[106] FGD with Mothers, South West State, Int. 1301
[107] Family members were not included in the close-ended answer choices for this survey question.

> *who was using the bathroom and a boy entered the bathroom the girl was using. So it's scary.*
>
> Risk Mapping with Girls, Banadir, Int. 404

For this kind of cases, the girls rely on the teachers for security. That is why seven out of eleven focus groups girls state that a happy place in school is where the teacher is. It also explains why the teachers and head teachers were marked as people where the girls look for assistance in case of harassment or abuse in the school.

Community

During the ML1 round, 88.1 percent of the girls responded "Yes" when asked if they have an avenue to report any form of harassment, abuse, or exploitation in the community. Like the case of reporting abuse in school, there was an important drop of almost 10 percentage points to the response for the ML2. By EL it decreased further, to 75.6 percent of girls.

In this case, the top choice in the EL to report the incident was the parents or other types of relatives, expressed in the 'others' option, with around 39.1 percent of the respondents opting for it. The next two options were the head teacher and the teachers with 30.8 percent and 23.3 percent respectively. Also different from the case of schools, the percentage of girls choosing to report to the police was higher, at 21.8 percent.

For girls, their main security concern in the community occurs when going to the school. Most of them claim that there is a part of the way home from school where there are groups of boys or men that are drunk and harass them. These insecure places include, for example, a big tree that is located near their house, unoccupied places, like alleys or dilapidated buildings, and livestock market or farms. Another insecure place mentioned by the girls are main roads and highways due to the traffic, and cars passing by at high speeds. To improve safety, the girls propose or ask for more government action against the groups of boys or men that loiter on some community spaces the girls must go through to get to school. And, although a group of girls mention that former places have been made safer due to police presence, they also represent a source of insecurity due to harassment or intimidation.

> *On my way to school, there used to be men on the streets who harass us, but now the presence of security police has greatly reduced such incidents. However, there's still an ongoing issue with some soldiers disturbing students as they pass by. While some soldiers are understanding, others can be intimidating, raising their guns openly, which is very frightening, and it is good that the police should stay in those deserted places to avoid the boys that harass girls.*
>
> Risk Mapping with Girls, Banadir, Int. 406

Synthesis

Despite the drop in percentage of girls agreeing that they have avenues to report abuse occurring in both 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., reporting the issues to the police, head teacher, or their parents to deal with the root cause). The

interviews and FGDs with girls also demonstrate the importance of engaging with boys and men when dealing with girls' education, because, as the qualitative evaluation shows, it is in the interaction with boys at school or men on the street that girls face increased challenges to attending education. Factors that such as lack of resources for transportation, far distances, or poor sanitary infrastructure, that might deter girls to go to school, are exacerbated by the harassment and insecurity created by boys and men. Engaging with this problem would ease girls' life beyond education.

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.

6.7. Strengthened Economic Circumstances

The final intermediate outcome for the original baseline group of girls supported by the FCDO deals with their economic and employment outcomes, particularly for those who completed the NFE programme. In this section, we assess the employment levels among C1 NFE girls, their reported monthly income, and how these metrics have changed from ML1 (the first measurement) through ML2 and the EL.

Measuring employment and income presents certain challenges. Firstly, it's unclear if girls involved in domestic work are compensated and whether this compensation is in cash or in-kind[108]. Additionally, it's ambiguous whether respondents interpreted domestic work (defined as "non-agricultural domestic chores inside the home, such as child-raising and cooking") as tasks performed in their own homes or exclusively in the homes of other families. These issues are significant limitations across the three rounds of evaluation, which encompass a total of 521 C1 NFE respondents.

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, the share of C1 NFE girls who do not have a job has declined significantly between ML1 and EL. Domestic work remains the most common form of work for girls, and from ML1 to endline the share of girls engaged in domestic work increased by 11.4 percentage points, after a spike in ML2 where it reached 38.5 percent. Between ML1 and EL, the growth of domestic work labourers accounts for a bit more than half of all the new working girls.

TABLE 53: EMPLOYMENT, BY JOB TYPE, AMONG C1 NFE GIRLS IN THE ML1 AND ML2 ROUNDS

Job Category	Share of Girls. ML1	Share of Girls, ML2	Share of Girls, EL
No occupation or job	56.7%	21.8%	34.7%
Domestic work	19.9%	38.5%	31.3%
Unskilled sales/service worker (e.g., hawker, shoe cleaner, domestic helper, cleaner)	11.1%	14.4%	8.5%
Student	6.4%	5.2%	6.8%

[108] 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.

Sales/Service worker (e.g., waiter, retailer or clerk at shop, etc.)	1.8%	9.8%	5.1%
Other	0.6%	6.9%	5.7%
Trades, craft workers, extractive industries	2.3%	2.3%	5.1%
Farming, fishing, pastoralism	1.8%	5.8%	1.1%
Professional or managerial positions	0.0%	5.8%	1.7%

Outside domestic work, gains in employment numbers were smaller but still important – there are increasing numbers of girls engaged in service-sector work, and in a broad category of physical trades, crafts, etc. Combining the two service-sector categories –ambiguously distinguished by their degree of formality – 20.8 percent of girls are engaged in this sector. This aligns with the qualitative data from the EL round, in which some mothers indicated that their daughters work in shops or vendor stalls, mostly family run, assisting their mother in parts of the business.[109] These mothers describe having their daughters helping them run the business, helping selling products and, in one case, helping to manage the collected money.[110]

The nature of these jobs raises questions about the nature of domestic work, as previously mentioned. The average income over the past month for girls employed in domestic work is lower than that for girls in other types of employment but higher than for unemployed girls. This indicates that domestic work might generally be compensated. However, 41.8 percent of the girls employed in domestic work reported no income in the past month, compared to 59.0 percent of unemployed girls and 21.7 percent of girls with other types of jobs. Analysing the distribution of earnings reveals that outliers, who earn significantly more than average, may skew the values. For both unemployed girls and those in domestic work, earnings are close to zero or zero, suggesting many girls engaged in domestic work are either working in their own households or not receiving monetary compensation.

Another trend in income-generating activities, not shown in the table above, involves the proportion of NFE girls who have their own small businesses. At ML1, 11.1 percent of NFE girls had their own small business, increasing to 15.5 percent at ML2 and 18.8 percent at EL. These businesses typically involve selling produce, prepared food, and other retail vending activities.

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 and volatility[111] in income (especially as we recorded information of a single month of income) are well-known impediments to accurate data collection. A bigger problem is that a large proportion of girls did not know their income over the past month – out of 1212 girls asked about their income in this evaluation round (all NFE girls, of all cohorts), one third did not know their income. Among the C1 NFE cohort analysed here, just 193 girls (of 263) reported their income, including girls who reported earning no income at all. This issue is worsened by the fact that we seek

[109] FGD with Mothers, Banadir, Int. 201; FGD with Mothers, Banadir, Int. 403; FGD with Mothers, Banadir, Int. 101; FGD with Mothers, South West State, Int. 1405.
[110] FGD with Mothers, South West State, Int. 1405
[111] E.g., for agropastoralist, income is seasonal.

to compare changes in income over time, meaning that we need to compare girls' income from ML1 to EL. The problem is, not all girls who participated in every evaluation round answered the question of income. That would limit our sample of C1 NFE girls with complete income information to 57 observations. Given the small number of observations with the complete data across the 3 rounds, we decided to pool all C1 NFE for the analysis.

Another limitation is that, for the endline evaluation data collection, we asked girls to provide their monthly income in Somali shillings, but many girls and enumerators wrote really small amounts that suggest they responded in US dollars. To avoid extreme values that might have been expressed in Somali shillings, we convert high values of the variable to US dollar exchange rate. This also allows us to standardise the variable to the observations from previous rounds an make them compatible. We present a table with the results of the full "income panel", for observations with income data across all three rounds. We also present the data of the "income panel" without outliers, where observations where the income changes more than US$ 500 between rounds are excluded.

TABLE 54: MEAN MONTHLY INCOME AMONG C1 NFE GIRLS, BY ROUND

Full "Income Panel"					
Zone	**Obs.**	**ML1-US$**	**Obs.**	**EL**	**Diff. EL – ML1**
Overall	105	58.8	105	61.4	2.6
Banadir	48	13.9	48	26.8	12.9**
Jubaland	32	158.5	32	76.6	-81.9
South West State	25	17.3	25	108.4	91.1
"Income Panel" without outliers					
Overall	100	21.6	100	29.6	8.0
Banadir	48	13.9	48	26.8	12.9**
Jubaland	28	38.2	28	27.1	-11.0
South West State	24	17.6	24	37.9	20.3

As shown in the table, income varies between rounds and zones in an inconsistent manner. While in the previous ML2 evaluation report we showed an increase on income on varying degrees in all three geographic zones from ML1 to ML2, in the endline the pooled data of the monthly income shows a decrease from ML1 to EL for Jubaland State and an increase for Banadir and South West State. Though, only the increase for Banadir is statistically significant at the 95% level. We urge caution when relying on those results, given their inclusion of distortions arising from measurement error in the data collection. To avoid the disparities between rounds, we performed some analysis only with data from the EL to see if the was any mediating effect of being part of a pastoral household, being married, enrolment in education, and holding a job. None

of these factors seems to be related to increased income. This result points out to the reliability and difficulty of collecting data about income.

The irregular reports on income and an increasing but swinging number of employed girls are signals of the difficulties they still face to generate income. The most important, according to qualitative interviewees, is a lack of financial support for opening new businesses.[112] Without savings or credit, it is not possible to start a small business. As a mother in Gedo puts it:

> *There is a major issue that exists when they want to establish a business, and they need financing. Most parents do not have money to support their children.*
>
> FGD with Mothers, Gedo, Int. 206

A mother also noted that community attitudes can affect the girls, as working girls may in some cases be subject to mockery by other community members:

> It is a nice thing if a girl establishes a business, but girls do not have the means to do so. Additionally, if they open a business, young people make fun of them, as I have witnessed since my daughter is selling milk with me. So, if the girl opens a place, she will be mocked, and no one will help her.
>
> FGD with Mothers, Banadir, Int. 211

Negative community attitudes such as these might be the reason why many mothers do not only talk about a lack of financial means to start a business, but also the need of *support*, many without specifying support from whom or what kind.[113]

[112] 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.

[113] FGD with Mothers, Southwest, Int. 1301; FGD with Mothers, Southwest, Int. 1404; FGD with Mothers, Southwest, Int. 1405; FGD with Mothers Jubaland, Int. 702; FGD with Mothers, Banadir 101

7. Value for Money

Previous sections have evaluated the impact of the AGES program in meeting its original objectives, such as improving girls' learning and life skills, enhancing the practices and attitudes of teachers and caregivers, and fostering better school management. While the data indicate that the program was generally successful in achieving many of its goals, it remains unclear whether it provided value for money by being cost-effective: Did the intervention deliver the expected results while minimizing costs?[114]

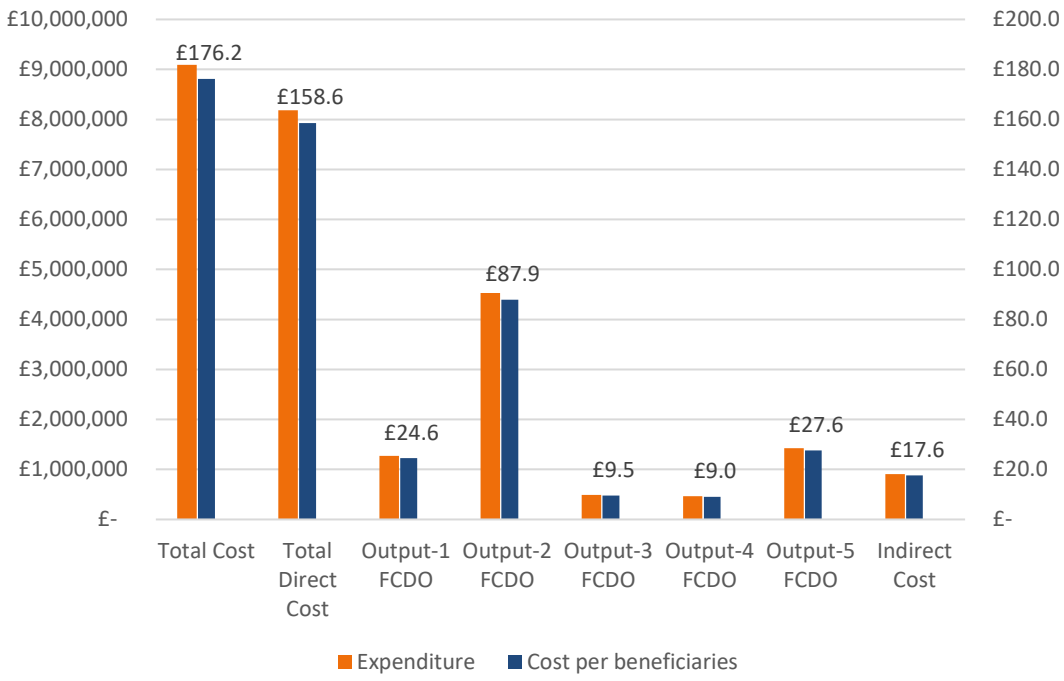
As outlined in the AGES Theory of Change, the FCDO funded five different types of outputs and activities:

- Output 1: Strengthened economic situation of communities
- Output 2: Availability of quality learning opportunities tailored to the needs of ultra-marginalised girls
- Output 3: Social change movement towards broader life opportunities for adolescent girls and boys
- Output 4: Build institutional capacity for quality assurance in inclusive education
- Output 5: Support to girls with disabilities

As a first step, we calculated the total cost per beneficiary. Figure 12 presents the total expenditure and expenditure per girl reached for the whole program and disaggregated by type of output.

[114] The VfM analysis provided in this section does not include M&E and CA costs as they are funded by both FCDO and USAID and their source cannot be derived from the budget data available. It is also important to note that the lack of quality data on cost per student in Somalia does not allow for a comparison with national benchmarks.

FIGURE 12: TOTAL AND PER BENEFICIARY EXPENDITURES



As the figure illustrates, the direct cost per beneficiary was £159 per girl. However, not all costs are uniformly distributed across beneficiary girls. Except for activities related to the Output 2 and Output 5 (targeting Girls with Disabilities), it is not possible to disaggregate cost by type of learning program. Therefore, we consider all costs as uniformly distributed across all types of girls, except for Output 2 costs that can be disaggregated and Output 5 costs that are only related to Girls with Disabilities. We started from the analysis of Output 2 costs. The program allocated most of the budget (55% of direct costs) to activities under this Output, which focused on providing quality learning opportunities to improve the literacy and numeracy skills of ultra-marginalized girls. These activities comprised the largest part of the program and included the provision of ABE and NFE programs, as well as support for formal schooling.

The table below details the breakdown of costs for activities related to this output.

TABLE 55: OUTPUT 2 ACTIVITIES DIRECT COSTS BREAKDOWN BY TYPE OF BENEFICIARY

Activity	Total Expenditure (£)	Cost per beneficiary (£)
ABE Related activities (#ABE girls=13,276)		
Center Rehabs	100,134.0	7.5
Curriculum/Learning Materials	482,038.9	36.3
Teacher Incentives/Training	687,874.7	51.8
Community Mobilization	11,406.9	0.9

Activity	Total Expenditure (£)	Cost per beneficiary (£)
Monitoring with MOE	10,007.4	0.8
Total	**1,291,461.9**	**97.3**
NFE Related activities (#NFE girls=15,121)		
Center Rehabs	11,123.2	0.7
Curriculum/Learning Materials	76,031.0	5.0
Teacher Incentives/Training	342,467.5	22.6
Total	**429,621.8**	**28.4**
FE Related Activities (#FE girls= 21,945)		
Center Rehabs	82,407.3	3.8
Curriculum/Learning Materials	86,982.9	4.0
Teacher Incentives/Training	696,004.5	31.7
Community Mobilization	46,634.4	2.1
Total	**911,029.1**	**41.5**

The table indicates that the direct cost per beneficiary varies depending on the learning program. ABE-related activities had a cost per girl of £97.3, which is significantly higher than the unitary costs for NFE related activities – only £28.4 – as well as the unitary cost for FE activities, equal to £41.5. This difference is primarily due to the fact that the project funded more activities related to ABE programs compared to NFE and FE. Moreover, the cost for teachers' incentives was significantly higher for the ABE school than for other learning programs (£51.8 per beneficiary, compared to £22.6 in NFE programs and £31.7 in Formal Schooling). The cost of reviewing the curriculum and providing learning materials was also significantly higher for ABE schools – £36.3 per girl vs £5.0 and £4.0 per NFE and FE girls respectively. This is largely attributable to the provision of desks, specific to ABE girls, which cost £18.3 per girl.

Some other costs for activities related to the Output 2 were not specific to one education path, but equally distributed across all beneficiaries. These costs – summarized in the table below – primarily comprising program staff salaries, also include travel cost, IT and office equipment and partners support costs, amounting to £36.8 per girl. Adding these costs to those specifically related to FE, NFE and ABE activities, the final direct cost per girl of Output 2 related activities is £135.0 for ABE activities, £66.1 for NFE activities, and £79.2 for FE supporting activities.

TABLE 56: OUTPUT 2 ACTIVITIES DIRECT COSTS DISTRIBUTED ACROSS ALL BENEFICIARY

Activity	Total Expenditure (£)	Cost per beneficiary (£)
Program Local Staff Salaries	962,986.3	19.1

Activity	Total Expenditure (£)	Cost per beneficiary (£)
Program International Staff Salaries	601,340.2	11.9
Travel cost	80,937.2	1.6
IT and Office Equipment	12,335.9	0.2
Partner Support Costs	242,228.0	4.8
Total	**1,899,827.6**	**37.7**

Other activities targeting specific groups of girls are those related to the Output 5: support to girls with disabilities. The program benefitted 1244 girls with disabilities who were enrolled in formal schooling. The table below presents a breakdown of costs for activities related to this output, and the cost per GwD.

The project allocated slightly more than 1.4 million pounds to this Output, which represents 17.7% of all direct costs, excluding M&E and Central Administration costs. Girls with disabilities accounted for the 2.4% of all beneficiaries. For this reason, the cost per GwD is particularly high, at £1145.6. This is mainly due to the high cost per beneficiary associated with fees (£662.6, more than half of the total), and costs specific to the disability condition of the girls. For example, the project provided an average grant of £157.7 to their parents and spent on average £66.7 on transportation for each girl.

TABLE 57: OUTPUT 5 COST BREAKDOWN - GIRLS WITH SPECIAL NEEDS (N = 1244)

Activity	Total Expenditure (£)	Cost per beneficiary (£)
Partial grants to parents of GwD	196,119.5	157.7
Transportation for GwD	82,991.8	66.7
Teacher Training Materials	48,503.2	39.0
Fees (Specific to SNE activities)	824,251.0	662.6
Devices for teachers with disabilities	20,568.4	16.5
Travel (include Hotel Accommodation)	189,519.2	152.3
Partner Support Costs	55,771.5	44.8
IT and Office equipment	7,460.4	6.0
Total	**1,425,185.0**	**1145.6**

Costs related to other outputs are assumed to be equally distributed across all beneficiaries. With this assumption in mind, if we sum cost per beneficiary for activities related to Output 1, 3 and 4 (see Figure 12) to those calculated above for Output 2 and Output 5 activities, we will get the following direct cost per beneficiary[115], for each girl type:

- ABE girls: **£178.1**
- NFE girls: **£109.2**
- FE girls: **£122.3**
- Girls with disabilities: **£1188.7**

After calculating the costs per beneficiary, we need to understand how they translated in improvement in the three outcomes targeted by the program: learning outcomes, positive transitions, and sustainability of interventions.

Regarding learning outcomes, our evaluation, detailed in Section 3, shows that the improvement in both literacy and numeracy between BL and EL has been significant, especially for FE and ABE girls, whose scores in literacy improved by 34.2 and 22.0 points respectively and scores in numeracy improved by 19.9 and 9.7 points. On the other hand, older NFE girls did not improve, mainly for two reasons: the NFE curriculum has lower focus on literacy and numeracy and more on life skills and financial literacy, and their baseline scores were higher than those of FE and ABE girls.

Despite the lack of improvement in learning outcomes among NFE girls, the program succeeded in enhancing skills related to resilience and self-confidence, as well as life skills, as measured by the YLI.

[115] Excluding M&E and Central Administration costs

Moreover, the program was successful in supporting positive transition of NFE girls to formal education or to employment with 14.5% of NFE girls who transitioned to formal schooling and 46.9% of girls who are now either employed or self-employed.

Regarding girls with disabilities, the evidence is mixed, despite the significant investment per beneficiary. Although their enrolment and retention for each and any disability suggest an overall positive impact of the program, the declining rate of training and strategy instillation focused on girls with disabilities raises some concerns about the cost-effectiveness of the activities related to Output 5.

Finally, we found mixed evidence of the program impact on sustainability in terms of strengthened school governance and driving positive shifts in community practices towards girl education. Regarding school governance, school directors maintain substantial responsibilities over school management, however, CECs have limited ability to fully address barriers for girls with disability, being generally resource-strapped and operating in challenging environments.

USAID Cohort Results

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

8. Learning - USAID

8.1. Aggregate Learning Outcomes

One of the main objectives of the AGES programme for the C4 and C5 NFE cohorts is to enhance the girls' numeracy and literacy skills, particularly their ability to read in Somali. The 11-month NFE programme aimed to equip these girls with basic literacy and numeracy skills that could be beneficial in the labour market. In this section, we present findings on the overall changes in learning outcomes observed among C4 NFE girls from ML1 (baseline for C4) to EL, and among C5 NFE girls from ML2 (baseline for C5) to EL. We then detail changes in specific numeracy and literacy assessment subtasks. For clarity, ML1 serves as the baseline for C4 girls, who began the programme immediately after the ML1 evaluation and completed it just before ML2; C5 girls started right after the ML2 evaluation and finished just before this EL evaluation. Next, we explore trends among girls with disabilities and other demographic sub-groups. Finally, we examine the extent to which intermediate outcomes predict changes in numeracy and literacy.

Before presenting the findings on learning outcomes, we describe the composition of the two NFE cohorts in terms of region and age. The table below shows the distribution of 575 C4 NFE girls and 370 C5 NFE girls interviewed at EL across regions and age groups. As illustrated in the table, most girls in both cohorts are older than school age, with the majority being over 18 years old. As anticipated, C5 NFE girls are slightly younger than C4 NFE girls, as they enrolled in the NFE programme one year later.

TABLE 58: DISTRIBUTION OF C4 AND C5 NFE GIRLS BY REGION AND AGE

Characteristic	Nr (%) of C4 NFE girls	Nr (%) of C5 NFE girls
State – Region		
Banadir -- Banadir	272 (47.3%)	165 (44.6%)
South West State -- Bay	135 (23.5%)	59 (15.9%)
South West State -- Lower Shabelle	84 (14.6%)	84 (22.7%)
Hirshabelle -- Middle Shabelle	84 (14.6%)	62 (16.8%)
Age groups		
Age 15 years and under	17 (3.0%)	19 (5.1%)
16-17 years	33 (5.7%)	40 (10.8%)
18-19 years	182 (31.7%)	140 (37.8%)
20-22 years	153 (26.6%)	82 (22.2%)
23+ years	190 (33.0%)	89 (24.1%)
Total	**575**	**370**

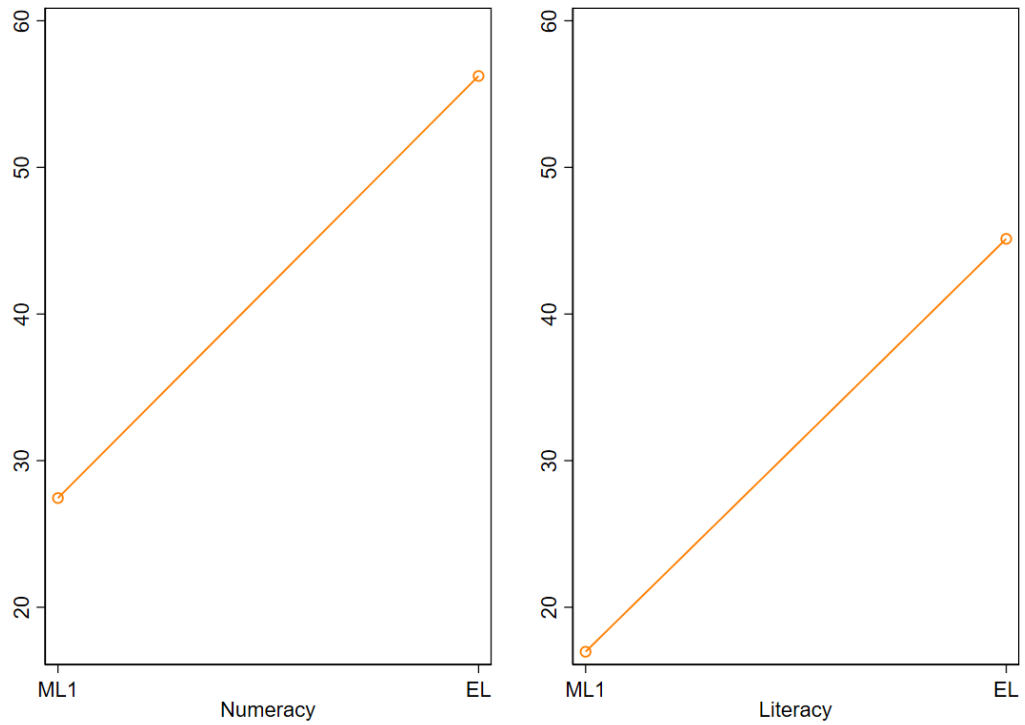
To accurately interpret the changes in learning outcomes, it is crucial to acknowledge that there are no benchmarks or counterfactual cases for C4 NFE girls, making it difficult to estimate the 11-month NFE programme's impact against a comparison group. Since these girls were not enrolled in school, we would expect their numeracy and literacy skills to have only slightly improved or stagnated over time. With the assumption that their skills would have remained unchanged without the programme, we can attribute the observed variation in aggregate numeracy and literacy to their participation in the NFE programme.

The graph below provides a broad overview of the numeracy and literacy test scores of C4 NFE girls from ML1 to EL rounds, showing an overall improvement in both areas.

It is important to note that the increase in scores was significantly larger between ML1 and ML2, due to two main reasons. Firstly, scores had already doubled by ML2 from ML1, so expecting the same level of improvement at EL was unrealistic. Secondly, the 11-month NFE programme ended by the ML2 evaluation, meaning the improvements at ML2 were observed right after the programme's completion. Nevertheless, it is reassuring to see that the girls' scores continued to improve a year after the programme ended.

Additionally, as discussed later in this section, some C4 NFE girls continued their education in some form, either non-formal or formal, after completing the programme. This ongoing education may also have contributed to the rising learning scores documented here.

FIGURE 13: CHANGES IN AGGREGATE NUMERACY AND LITERACY SCORES FOR C4 NFE GIRLS



The graph below provides the same results but for improvement of C5 NFE girls between ML2 and EL. The C5 NFE cohort is the latest group of NFE girls to be brought into NFE learning centres, following on the C4

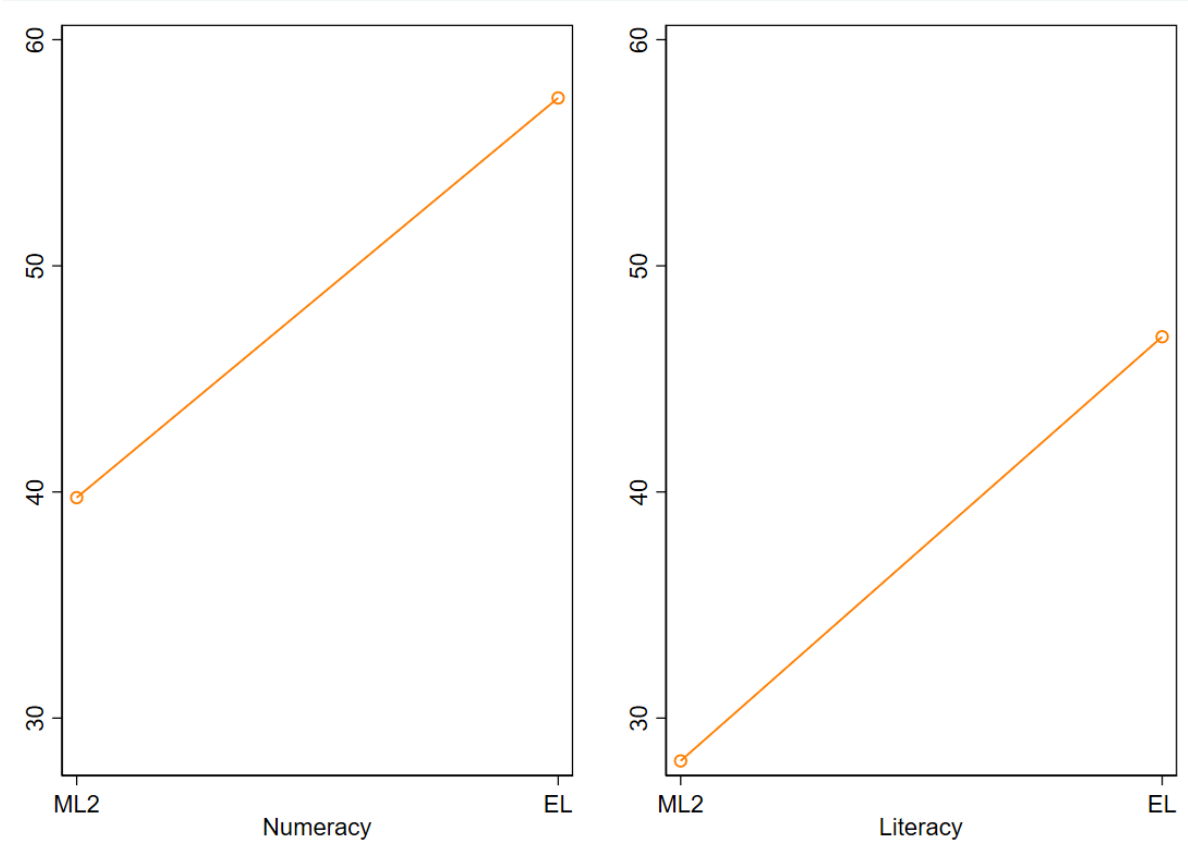
NFE girls who were recruited and enrolled in early 2022. The C5 NFE group was enrolled into the learning centres immediately prior to the ML2 data collection.

As seen with C4 NFE girls, there are no benchmarks or counterfactual cases to estimate the impact of the 11-month NFE programme against a comparison group. As a result, the improvements of C5 NFE girls should be interpreted similarly as those of the C4 NFE girls, with the caveat that unlike this latter group, C5 NFE girls were enrolled to the programme in 2023 that ended in 2024, and their exposure to the programme have been more intense between ML2 and EL rounds of the AGES evaluation, on average, compared to C4 girls over the same time period.

In both literacy and numeracy scores, their improvement has been significant, even if slightly lower than the C4 NFE girls' improvement from ML1 to ML2.[116] The lower gains can be explained by the fact that C5 NFE girls started participating in the programme with significantly higher scores than C4 NFE girls, whose scores at ML1 in both numeracy and literacy were around 10 percentage points lower than the scores of C5 NFE girls at ML2.

[116] Between ML1 and ML2 average numeracy score improved from 27.7 to 49.7 points (Δ=+22.0), whereas from ML2 to EL it improved from 49.7 to 56.1 points (Δ=+6.4). A similar pattern can be seen for the average literacy score that improved from 16.8 to 36.2 points between ML1 and ML2 (Δ=+19.4) and from 36.2 to 44.2 points from ML2 to EL (Δ=+8.0).

FIGURE 14: CHANGES IN AGGREGATE NUMERACY AND LITERACY SCORES FOR C5 NFE GIRLS



The table below provides more detail on the extent to which C4 and C5 NFE cohort achieved improvements on the numeracy and literacy tests from their respective baselines (ML1 for C4 NFE girls, and ML2 for C5 NFE girls) to EL. The observed improvement among both groups of girls in numeracy and literacy was both substantially large and statistically significant. As already noted, C4 NFE girls started from lower scores at BL, and have improved more than C5 NFE girls. As a result, scores at EL of C4 and C5 girls are very similar in both numeracy and literacy.

TABLE 59: DIFFERENCE IN LEARNING OUTCOMES

Type	Cohort Baseline Mean	Cohort Baseline SD	EL Mean	EL SD	Diff.	P-Value
	C4 NFE Girls (BL=ML1)					
Numeracy scores	27.5	25.1	56.2	26.9	28.8	0.000
Literacy scores	17.0	27.2	45.1	35.5	28.2	0.000

	C5 NFE Girls (BL=ML2)					
Numeracy scores	39.7	24.6	57.4	26.0	17.7	0.000
Literacy scores	26.9	29.5	45.3	34.2	18.3	0.000

To further analyse the evolution of C4 NFE girls across different rounds and how it relates to continuous participation in learning programs, the table below reports how the gain in scores from ML1 and EL varies with participation in a learning programme between ML2 and EL. Results are striking, with girls who are enrolled improving their scores in both numeracy and literacy by around 40 percentage points, doubling the improvement of girls who are not enrolled to any learning program. For both numeracy and literacy scores the improvement in even higher for girls who are enrolled in FE schooling, than for girls enrolled in either ABE or NFE programs. For girls not enrolled in any programme, the improvement is similar to the one-year improvement recorded at ML2. This reinforce the assumption that without any learning program, the performance in both learning and numeracy tests is likely to stagnate or even regress if girls drop out of education. It is important to note that other factors may affect this result, with girls who stayed in a learning programme already more motivated, as also suggested by their higher ML1 scores, especially in literacy.

TABLE 60: DIFFERENCE IN LEARNING OUTCOMES BY CURRENT ENROLMENT STATUS – C4 NFE GIRLS

Type	ML1 Mean	ML1 SD	EL Mean	EL SD	Diff.	P-Value
	Girls still enrolled in a FE learning programme					
Numeracy scores	32.4	29.4	73.7	20.1	41.3	0.000
Literacy scores	25.7	32.7	69.6	26.0	44.0	0.000
	Girls still enrolled in an ABE or a NFE learning programme					
Numeracy scores	24.4	24.6	64.3	23.2	39.9	0.000
Literacy scores	15.3	26.7	53.3	33.9	38.0	0.000
	Girls not enrolled in any learning programme					
Numeracy scores	27.4	24.3	48.6	26.9	21.2	0.000
Literacy scores	15.1	25.2	35.9	34.8	20.7	0.000

The table above also highlights that girls who are not enrolled in any learning program are also those with lower literacy scores at the cohort baseline. This may suggest that, despite their significant gains (around 20

points), their initial low literacy level have negatively influenced their ability to transition into formal learning program or to stay in non formal paths. Therefore, a more tailored approach to those girls who perform particularly poorly could be especially beneficial in allowing them to stay in educational programs and in reducing the likelihood of dropping-out.

Geography

In this section, we examine the trends in learning outcome changes for C4 and C5 NFE girls by region, highlighting geographic patterns. As shown in the table below, C4 NFE girls experienced the greatest numeracy improvement in Lower Shabelle (+48.2 points), followed by Middle Shabelle (+38.9 points) and Banadir (+28.4 points). In contrast, girls in the Bay region showed a more modest increase of 11.2 points on average. It is noteworthy that Bay also had the lowest improvement between ML1 and ML2. This smaller increase can be partly attributed to the higher initial numeracy skills of C4 NFE girls in the Bay region at ML1, where their average score was nearly double that of other regions. By EL, this gap in numeracy scores had closed, with girls in Lower Shabelle now outperforming those in the Bay region.

TABLE 61: DIFFERENCE IN NUMERACY OUTCOMES BY REGION

Type	BL Mean	BL SD	EL Mean	EL SD	Difference	P-Value
	C4 NFE Girls (BL=ML1)					
Banadir	25.9	23.2	54.2	27.4	28.4	0.000
Lower Shabelle	18.5	19.9	66.7	22.2	48.2	0.000
Bay	43.7	28.1	55.0	28.4	11.2	0.036
Middle Shabelle	15.3	16.2	54.3	24.8	38.9	0.001
Aggregate	27.5	25.1	56.2	26.9	28.8	0.000
	C5 NFE Girls (BL=ML2)					
Banadir	46.5	24.0	54.1	26.4	7.7	0.000
Lower Shabelle	31.0	21.5	63.3	21.9	32.2	0.000
Bay	30.9	25.7	55.3	30.9	24.4	0.003
Middle Shabelle	42.1	23.5	60.3	23.8	18.2	0.006
Aggregate	39.7	24.6	57.4	26.0	17.7	0.000

For C5 NFE girls, the regional pattern is quite different, with the lowest improvement observed in Banadir, at only 7.7 points. This low improvement can be partly explained by the high literacy scores in Banadir at ML2 (46.7 points versus the regional average of 39.9 points), but it remains concerning as Banadir has the lowest average numeracy scores at EL.

A similar geographic trend is observed in literacy outcomes among C4 NFE girls, with improvements in regions other than Bay being even more pronounced. Initially, at ML1, C4 NFE girls in Bay outperformed those in other regions. However, by EL, girls from other regions significantly boosted their literacy scores

and surpassed the average scores of those from Bay. Meanwhile, literacy scores for C4 NFE girls in Bay showed only minimal improvement, a trend already evident at ML2. This pattern can be partially explained by the severe drought conditions in the Bay region, as noted in the ML2 report.

These harsh conditions persisted between ML2 and EL, affecting the results for C5 NFE girls as well. Girls in Bay not only had the lowest average literacy score (17.9 points compared to the overall average of 28.1 points) but also recorded the smallest improvement (+7.8 points). In contrast, Lower Shabelle saw the highest improvement in literacy scores among C5 NFE girls (+39.9 points). Similar to numeracy scores, C5 NFE girls in Banadir showed significantly less improvement than the average (+10.3 points).

The lower gains in Banadir and Bay can be also explained by the higher presence of girls from internally displaced families in these regions. IDP communities have been the most hit by food shortages during the drought and more generally faced harsher living conditions. This is also reflected by their lower average scores in both literacy and numeracy.[117]

TABLE 62: DIFFERENCE IN LITERACY OUTCOMES BY REGION

Type	BL Mean	BL SD	EL Mean	EL SD	Difference	P-Value
	C4 NFE Girls (BL=ML1)					
Banadir	14.9	26.8	43.6	35.4	28.7	0.000
Lower Shabelle	14.1	22.2	63.5	29.3	49.4	0.000
Bay	29.0	32.1	36.9	36.8	7.8	0.009
Middle Shabelle	7.2	15.3	45.1	33.4	37.9	0.001
Aggregate	16.7	27.2	45.1	35.5	28.2	0.000
	C5 NFE Girls (BL=ML2)					
Banadir	31.6	30.8	41.9	34.5	10.3	0.000
Lower Shabelle	21.7	25.8	61.2	28.8	39.9	0.000
Bay	17.9	25.4	25.8	32.6	7.8	0.033
Middle Shabelle	37.2	30.6	60.3	26.5	23.1	0.011
Aggregate	28.1	29.5	46.9	34.0	18.8	0.000

The aggregate findings above paint a picture of substantive improvements in the literacy and numeracy among both C4 and C5 NFE girls after completion of the 11-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. However, results from C4 NFE girls suggest that the continuous

[117] Average scores at EL of C4 NFE girls coming from IDP families are 34.8 points in literacy (vs 53.8 of non-IDP girls), and 50.2 points in numeracy (vs 61.3 of non-IDP girls). Very similar results are observed among C5 NFE girls.

participation in learning programme is needed to sustain improvements. While NFE girls in Lower Shabelle and Middle Shabelle appear to have all experienced double digits improvements in literacy and numeracy skills, C4 NFE girls in Bay passed from being the best at ML1 in both numeracy and literacy to be the lowest scoring in literacy, and below the overall average in numeracy. The reason for this may be found in higher scores at ML1 for C4 NFE girls and in particularly harsh conditions in Bay due to a more severe drought.[118]

8.2. Foundational Skill Gaps

In this section, we shift our focus from aggregate scores to detailed analysis of numeracy and literacy scores by their individual subtasks. Our aim is to understand the patterns of skill achievement among the C4 and C5 NFE girls. We begin by identifying performance gaps in specific learning subtasks at EL. In the next section, we will examine how girls' scores on each subtask have changed, essentially identifying where improvements have or have not occurred since they began the NFE program.

The tables below present the proportion of girls who meet different achievement thresholds for each of the 11 numeracy subtasks at baseline (ML1 for C4 NFE girls and ML2 for C5 NFE girls) and at endline. These subtasks on EGMA and EGRA typically increase in difficulty throughout the assessments. However, performance on each subtask can vary as they require discrete skills that can be learned independently.

We have defined four bands of achievement: non-learners, who scored 0 on a subtask and were unable to answer any of the test items correctly; emergent learners, who scored between 1 and 40 percent; established learners, with scores between 40 and 80 percent; and proficient learners, who scored above 80 percent on a given subtask.

TABLE 63: FOUNDATIONAL SKILLS GAPS IN NUMERACY, C4-C5 NFE COHORT AT THEIR COHORT BASELINES

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)
	C4 NFE Girls – ML1										
Non-Learner	20.3	27.1	52.7	58.7	76.4	68.4	76.2	56.4	91.5	87.2	72.2
Emergent Learner	10.3	8.4	34.6	2.6	6.1	2.3	9.8	0.8	4	2.1	12.8
Established Learner	8.7	19.3	10.6	4.8	7.1	4.1	6.9	10.6	2.4	3.6	10.7
Proficient Learner	60.7	45.2	2.1	33.8	10.4	25.1	7.1	32.2	2.1	7.1	4.4

[118] Famine Early Warning Systems Network (FEWS NET) and Food Security and Nutrition Analysis Unit (FSNAU). "Nearly 6.6 million people in Somalia still face Crisis (IPC Phase 3) or worse acute food insecurity outcomes despite relative improvement in rainfall forecast and decline in food prices". April 2023. Available at: https://fsnau.org/downloads/Somalia-Multi-Partner-Technical-Release-on-the-March-2023-Follow-up-Assessment-Results-25-Apr-2023.pdf

	C5 NFE Girls – ML2										
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

TABLE 64: FOUNDATIONAL SKILL GAPS IN NUMERACY, C4-C5 NFE COHORT AT EL

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)
	C4 NFE Girls										
Non-Learner	5.0	5.4	28.7	17.7	31.3	25.0	38.6	14.6	56.3	49.7	35.5
Emergent Learner	1.6	4.7	58.4	2.8	8.3	3.7	16.3	1.7	7.5	6.8	25.4
Established Learner	2.6	13.9	5.2	12.2	25.4	9.9	23.3	23.8	15.1	18.6	25.9
Proficient Learner	90.8	76.0	7.7	67.3	35.0	61.4	21.7	59.8	21.0	24.9	13.2
	C5 NFE Girls										
Non-Learner	3.5	6.1	32.6	15.8	28.9	21.7	36.4	11.8	52.7	48.4	34.0
Emergent Learner	3.2	2.9	53.2	2.1	6.7	5.1	18.4	2.4	10.7	8.0	25.1
Established Learner	1.9	11.2	8.6	12.3	27.3	11.5	26.5	19.5	16.6	20.9	29.4
Proficient Learner	91.4	79.7	5.6	69.8	37.2	61.8	18.7	66.3	20.1	22.7	11.5

The table indicates that C4 and C5 NFE girls show very similar performances across various tasks. Their numeracy skills do not decline sharply across subtasks. NFE girls excel in number identification and number discrimination (selecting the larger of two numbers) tasks, and also perform reasonably well in 1-digit

addition, 1-digit subtraction, and simple word problems involving addition and subtraction (subtask 8). However, their performance drops significantly when tasks involve 2-digit numbers (subtasks 5 and 7), especially in comparison to their 1-digit counterparts. That being said, the decline in performance between 1-digit and 2-digit tasks is not absolute. For instance, C4 NFE girls see their proficiency in addition decrease from 67.3 percent in 1-digit tasks to 35.0 percent in 2-digit tasks. This shift largely moves girls from the proficient category to the emergent/established learner categories. The same trend is observed in C5 NFE girls. This indicates that many girls proficient in 1-digit addition can still add 2-digit numbers with some accuracy, though they may struggle with more complex procedures like carrying numbers within subtask 5.

The most challenging task for the girls is the missing number exercise (subtask 3), with only 7.7 percent of C4 NFE girls and 5.6 percent of C5 NFE girls achieving proficiency. This difficulty can be attributed to the nature of the exercise, which requires identifying patterns in sequences of numbers rather than performing calculations, a skill that is rarely practiced during lessons.

In the table below, we report proficiency levels for Somali literacy, across the EGRA's six subtasks, using the same classification scheme.

TABLE 65: FOUNDATIONAL SKILL GAPS IN LITERACY, C4-C5 NFE COHORT AT THEIR COHORT BASELINE

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 2	Reading comp. 3
	C4 NFE Girls – ML1					
Non-Learner	44.5	66.8	76.5	78.9	79.6	81.2
Emergent Learner	18.1	15.7	9.5	3.7	1.6	6.6
Established Learner	16.5	9.6	8.1	10.8	12.2	9.4
Proficient Learner	20.9	7.9	5.9	6.6	6.6	2.8
	C5 NFE Girls – ML2					
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

TABLE 66: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, C4-C5 NFE COHORT AT EL

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 2	Reading comp. 3
	C4 NFE Girls					

Non-Learner	20.9	32.5	40.5	44.2	44.0	52.2
Emergent Learner	8.0	10.6	8.0	2.4	6.3	27.5
Established Learner	11.1	19.5	11.8	35.0	32.2	16.5
Proficient Learner	60.0	37.4	39.7	18.4	17.6	3.8
	C5 NFE Girls					
Non-Learner	17.6	27.8	38.0	40.9	42.5	48.7
Emergent Learner	5.6	9.9	8.6	2.1	6.7	28.9
Established Learner	14.7	21.9	14.2	39.3	32.9	20.1
Proficient Learner	62.0	40.4	39.3	17.6	17.9	2.4

The scenario is less optimistic, with roughly 20 percent of girls unable to recognize at least 40 percent of letters and almost 30 percent unable to identify at least 40 percent of common words. Predictably, girls who struggle with word recognition also have low reading fluency. The word recognition task seems to be the task where critical gaps between girls emerged. Reading comprehension instead remains particularly difficult for both C4 and C5 NFE girls, with more than 40 percent categorized as non-learners in these tasks.

8.3. Subtask-Specific Gains in Learning

After the analysis of subtask-specific learning outcomes at EL, we now evaluate changes in performance on specific subtasks between rounds. For C4 NFE girls, our analysis focuses on changes in girls scores in each subtask from ML1 to EL and evaluate trends in those changes, whereas for C5 NFE girls we look at changes from ML2 to EL.

Numeracy subtasks

In analysing the numeracy subtasks, we first examine the changes in scores for C4 and C5 NFE girls across the numeracy assessments administered at ML1 (for C4 girls), ML2 (for C5 girls), and EL. Each assessment consists of 11 numeracy subtasks. These subtasks are the same described in the previous section on foundational skills.

The table below details each subtask with a description, along with the mean scores achieved by C4 NFE girls at ML1 and EL, the change in scores between these rounds, and the P-value of that change.

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

Subtask Number	Description	ML1 Mean	EL Mean	Difference	P-value
1	# Identification	68.7	92.3	23.6	0.000

2	Quant. Discrimination	61.6	85.7	24.1	0.000
3	Missing #	12.2	18.6	6.4	0.003
4	Addition (1 digit)	37.4	75.1	37.6	0.000
5	Addition (2 digit)	15.4	52.8	37.4	0.000
6	Subtraction (1 digit)	27.1	67.7	40.6	0.000
7	Subtraction (2 digit)	12.6	40.2	27.6	0.000
8	Word Problem (Add. & Sub.)	40.7	76.7	36.0	0.000
9	Multiplication	4.5	34.1	29.6	0.000
10	Division	8.5	40.3	31.8	0.000
11	Word Problem (Mult. & Div.)	13.2	35.1	22.0	0.000

The table above reveals several key findings. Firstly, the C4 NFE girls showed significant improvements in their mean scores across all 11 subtasks, with each increase being statistically significant. Except for subtask 3, all improvements exceeded 20 points, and scores increased by more than 30 points in subtasks 4, 5, 6, 8, and 10. Notably, the addition subtasks (4 and 5) saw some of the highest gains, second only to the 40-point improvement in the 1-digit subtraction task. This indicates that the NFE programme was effective in enhancing the girls' basic calculation skills. Furthermore, the substantial improvement in subtask 8 (+36.0 points) suggests that the programme also succeeded in helping the girls apply mathematical principles to real-world problems.

However, the limited progress in the "Missing #" subtask continues a trend observed in previous evaluations, where NFE girls consistently scored low and showed minimal improvement. This persistent lack of significant gains in subtask 3 may reflect the NFE programme's focus on practical numeracy and literacy skills beneficial for the labour market, rather than on fostering an abstract understanding of numbers and patterns.

The table below presents the results for the C5 NFE girls, showing a markedly different pattern, largely due to their much higher baseline scores at ML2 compared to the ML1 scores of the C4 NFE girls. For instance, in subtasks 1 and 2, the C5 NFE girls had average scores of 86.2 and 79.9 points at ML2, nearly 20 points higher than the C4 NFE girls' scores at ML1. Consequently, due to the ceiling effect, the improvements for C5 NFE girls in these two subtasks, although significant, were less than 10 points.

In other subtasks, the improvements were more uniform, ranging from 17.3 points (for the word problem using multiplications and divisions) to 26.1 points (for the 1-digit subtraction task). The only exception was the "Missing number" subtask, reinforcing the earlier observation about the NFE programme's limited emphasis on developing an understanding of numbers and patterns.

TABLE 68: CHANGES IN NUMERACY SUBTASK SCORES FOR C5 NFE GIRLS

Subtask Number	Description	ML2 Mean	EL Mean	Difference	P-value
1	# Identification	86.2	92.4	6.2	0.005
2	Quant. Discrimination	79.9	86.8	6.9	0.006
3	Missing #	8.7	18.3	9.6	0.000
4	Addition (1 digit)	59.4	77.0	17.6	0.000
5	Addition (2 digit)	34.2	56.1	21.9	0.000
6	Subtraction (1 digit)	43.9	70.0	26.1	0.000
7	Subtraction (2 digit)	20.4	40.0	19.6	0.000
8	Word Problem (Add. & Sub.)	62.1	80.2	18.1	0.000
9	Multiplication	9.7	35.2	25.6	0.000
10	Division	14.4	40.0	25.6	0.000
11	Word Problem (Mult. & Div.)	18.3	35.6	17.3	0.000

Literacy subtasks

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 for numeracy subtasks, the literacy scores of C4 and C5 NFE girls improved significantly, and the increase for C4 girls was larger, mainly de to lower scores at their baseline assessment.

The table below shows the changes in literacy scores for C4 NFE girls. The improvements were meaningful, and the C4 NFE girls improved most with the subtasks assessing fundamentals such as "Letter Sound Identification" in which there was 32-point improvement and reading "Words Commonly Used" which improved by 34.5 points. This translated in a 35.4 points improvement in the "Reading Fluency" which is clearly dependant on the first two. The C4 NFE girls also made substantial but slightly smaller improvements with more challenging subtasks concerning reading comprehension. Except for the most difficult reading passage, where the improvement was limited to 11.3 points, in the other two reading comprehension subtasks the improvement was larger than 25 points. The improvements in reading comprehension indicate that the improvement in fundamental literacy skills translated to better understanding of reading materials.

TABLE 69: CHANGES IN LITERACY SUBTASK SCORES FOR C4 NFE GIRLS

Subtask Number	Description	ML1 Mean	EL Mean	Difference	P-value
1	Letter Sound Identification	36.3	68.4	32.1	0.000
2	Words Commonly Used	17.0	51.5	34.5	0.000
3	Reading fluency	11.8	47.2	35.4	0.000
4	Reading Comprehension 1	13.3	43.4	30.1	0.000
5	Reading Comprehension 2	14.2	39.7	25.5	0.000
6	Reading Comprehension 3	9.2	20.6	11.3	0.000

Improvements of C5 NFE girls follow a similar pattern and are all significant. However, their magnitude is smaller, due to higher baseline scores at ML2. As a result, EL scores in each subtask look very similar when comparing C4 and C5 girls. This suggests the presence of some kind of ceiling in learning for NFE girls, having difficulties to improve their subtasks scores over certain thresholds.

TABLE 70: CHANGES IN LITERACY SUBTASK SCORES FOR C5 NFE GIRLS

Subtask Number	Description	ML2 Mean	EL Mean	Difference	P-value
1	Letter Sound Identification	57.1	72.8	15.7	0.000
2	Words Commonly Used	32.2	55.0	22.8	0.000
3	Reading fluency	21.4	47.8	26.5	0.000
4	Reading Comprehension 1	22.7	44.7	22.0	0.000
5	Reading Comprehension 2	23.8	39.9	16.1	0.000
6	Reading Comprehension 3	11.5	20.9	9.5	0.000

8.4. Subgroup Programme Impact

In this subsection, we present key findings on differences the changes observed in learning outcomes from ML1 to EL (for C4 NFE girls) and from ML2 to EL (for C5 NFE girls), analysing differences among subgroups in terms of aggregate differences of numeracy and literacy scores. Given the many subgroups of the NFE panel cohorts, we have chosen subgroups which have demonstrated impacts on learning trajectories in the

literature. 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 alongside the significance of that difference.

The table below shows ML1-EL changes in learning scores of subgroups of C4 NFE girls. The substantive and significant improvements in literacy and numeracy are also observed among all subgroups.

Maay speaking girls improved significantly less than the average, especially in literacy, where their improvement was almost the half of the average (+15.9 points vs +28.2). Maay speakers improved less than the average also in numeracy, suggesting that the wide use of the Mahatiri dialect as language of instruction can limit their improvements.

The difficulties of girls not speaking fluently the Mahatiri dialect has been also highlighted by teachers:

> *There are students who have difficulties in understanding the language and if they say something that is difficult for them other students laugh at them and discriminate them. We raise awareness about that.*
>
> - FGD with Teachers, Middle Shabelle, Int. 502

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. It is important to note that EL scores of girls with disabilities are still lower than the average, and their gains can be attributable to their particularly low scores at ML1. However, their significant improvements can be considered a success of the programme at eliminating barriers for learning of disabled girls.

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 attended schools which lacked resources we also observed increases in literacy and numeracy skills following the NFE programme. On the other hand, girls living with their husband improved significantly less than the average in both literacy and numeracy, suggesting that marital obligation within the HH limited their participation to the programme and their improvement.

TABLE 71: 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	**28.2**	**0.000**	**28.8**	**0.000**	**575**
Language					

Maay speaker	15.9	0.001	18.5	0.000	207
Disability Status					
Mental health disability (main)	22.8	0.000	26.2	0.000	104
Non-mental health disability	31.8	0.000	30.0	0.000	46
Any disability	26.8	0.000	28.7	0.000	136
Household and Demographic Characteristics					
No parents	55.2	0.000	47.5	0.000	18
Female-headed household	33.8	0.000	33.0	0.000	137
HoH no education	41.7	0.000	31.6	0.000	32
HoH no formal education	34.8	0.000	30.2	0.000	169
Caregiver no education	44.2	0.000	32.3	0.000	28
Caregiver no formal education	35.2	0.000	30.2	0.000	166
Neither HoH nor caregiver has education	46.1	0.000	32.8	0.000	27
Neither HoH nor caregiver has formal education	34.8	0.000	29.9	0.000	162
HoH does not earn wage	35.1	0.000	30.8	0.000	89
HH has poor roof	20.1	0.000	25.1	0.000	138
Went to sleep hungry most/all nights	24.8	0.000	33.9	0.000	40
Went without water at home most/all days	28.3	0.000	29.1	0.000	51
Went without meds most/all days	29.6	0.000	34.7	0.000	85

Went without cash income most/all days	26.8	0.000	33.1	0.000	126
HH owns land	25.2	0.000	22.5	0.000	124
IDP	21.7	0.000	24.9	0.000	262
Girls' Characteristics					
Girl owns phone	28.6	0.000	28.9	0.000	517
Girl owns smartphone	27.3	0.000	25.8	0.000	74
Girl lives with her husband (proxy for marriage)	10.3	0.004	19.2	0.000	144
Girl spends a few hours or more on chores	26.9	0.000	27.6	0.000	532
School Resources and Infrastructure					
Girl won't use drinking water at school	22.1	0.000	20.9	0.000	105
Girl won't use toilet at school	23.8	0.000	24.0	0.000	126
No computers at school	27.9	0.000	27.6	0.000	511
Girl cannot use books/learning materials at school	31.1	0.000	35.6	0.000	36
Not enough seats in class	17.4	0.008	22.8	0.000	24

Changes in learning between ML2 and EL of C5 NFE girls were more homogeneous across subgroups. It is important to note that the overall improvement of these girls was lower, and that the sample size of some groups is particularly low, given that the total sample size of C5 NFE girls assessed at EL is 370. The table below reports in detail results for all subgroups analysed.

TABLE 72: CHANGES IN LEARNING SCORES FOR C5 NFE GIRLS BY SUBGROUP

Subtask Number	Literacy Difference	Literacy P-value	Numeracy Difference	Numeracy P-value	Sample Size (n)

Overall	**18.3**	**0.000**	**17.7**	**0.000**	**370**
Maay speaker	19.1	0.000	23.0	0.000	112
Disability Status					
Mental health disability (main)	17.3	0.000	13.0	0.000	140
Non-mental health disability	25.0	0.002	21.4	0.000	41
Any disability	18.3	0.000	14.8	0.000	160
Household and Demographic Characteristics					
No parents	16.7	0.037	6.0	0.451	8
Female-headed household	16.4	0.000	17.3	0.000	83
HoH no education	10.6	0.018	12.1	0.048	20
HoH no formal education	16.4	0.000	17.5	0.000	82
Caregiver no education	6.0	0.391	5.2	0.346	21
Caregiver no formal education	14.6	0.002	16.6	0.000	72
Neither HoH nor caregiver has education	11.5	0.046	6.5	0.253	15
Neither HoH nor caregiver has formal education	15.1	0.002	17.2	0.000	70
HoH does not earn wage	15.0	0.001	13.4	0.001	44
HH has poor roof	11.0	0.001	12.3	0.001	83
Went to sleep hungry most/all nights	25.1	0.000	25.2	0.000	46
Went without water at home most/all days	12.9	0.002	23.8	0.000	61

Went without meds most/all days	24.0	0.000	22.4	0.000	62
Went without cash income most/all days	18.4	0.000	19.7	0.000	117
HH owns land	23.1	0.000	25.0	0.000	82
IDP	13.4	0.000	13.3	0.000	147
Girls' Characteristics					
Girl owns phone	18.7	0.000	17.8	0.000	310
Girl owns smartphone	16.0	0.000	16.2	0.000	99
Girl lives with her husband (proxy for marriage)	13.4	0.000	14.1	0.000	75
Girl spends a few hours or more on chores	17.0	0.000	16.3	0.000	304
School Resources and Infrastructure					
Girl won't use drinking water at school	12.3	0.011	14.2	0.002	55
Girl won't use toilet at school	14.8	0.001	12.0	0.001	88
No computers at school	18.8	0.000	16.9	0.000	339
Girl cannot use books/learning materials at school	27.6	0.013	24.0	0.010	18
Not enough seats in class	22.3	0.025	17.5	0.058	15

Among C5 NFE girls, some subgroups did not improve significantly their learning scores, or did so significantly less than the average. The subgroups for which the gain is not significantly higher than zero are also those whose sample size was particularly limited (<20 girls) and hence of difficult interpretation. Among subgroups with significant improvements, but lower than the average, C5 NFE girls who live with their husbands improved less than the average, confirming the result seen above for C4 NFE girls. The lower learning of girls who get married has been also confirmed during FGDs with teachers, with the issues flagged in four of the twelve FGDs. As one teacher said:

> *Girls who get married while studying are not learning well. They are more absent and when they come back to school, they forgot everything.*
>
> - FGD with Teachers, Lower Juba, Int. 509

Other groups underperforming in terms of ML2-EL gains include girls with caregivers without formal educations, and girls living in house with roof made of poor materials, suggesting that poor living conditions may affect girls' cognitive development.

8.5. Testing the Theory of Change

The Theory of Change suggests that programme outputs result in changes in various intermediate outcomes, which subsequently positively influence learning outcomes. For C4 and C5 NFE girls, these intermediate outcomes include acquiring life skills and improved teaching quality. In this section, we evaluate how well these intermediate outcomes predict the girls' learning trajectories by examining the relationship between each intermediate outcome and the changes in learning outcomes within the sample.

Youth Leadership Index

The Youth Leadership Index (YLI) is a score ranging from 0 to 100 that measures girls' self-assessed leadership skills in a school context. This index is derived from responses to 21 questions that ask girls about their awareness of the consequences of their actions, their confidence in expressing their thoughts clearly, and their ability to organize their peers to achieve a common goal. To determine if the YLI is a predictor of improvements in learning outcomes, we examine the relationship between the YLI scores of C4 and C5 NFE girls and the improvement in their learning outcomes.

TABLE 73: EFFECT OF 1 POINT INCREASE IN YLI SCORE AND CHANGES IN LEARNING OUTCOMES

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
	C4 NFE			
YLI score variation (EL-ML1)	-0.03	0.747	-0.05	0.104
	C5 NFE			
YLI score variation (EL-ML2)	-0.13	0.152	-0.04	0.766

Using a linear regression model with girls' age, cohort type, and region as control variables, we find no significant relation between YLI scores and change in neither literacy nor numeracy score. It is important to note that YLI score and learning scores in both numeracy and literacy are positively correlated. However, girls with higher scores at YLI also had higher scores at their respective BL (ML1 for C4 NFE, ML2 for C5 NFE girls). Consequently, their room for improvement was little compared to that of girls with a lower baseline YLI score.

GEF Participation

In connection with the life skills discussed in the previous section, this section explores the participation of C4 and C5 NFE girls in the Girls' Empowerment Forum (GEF) and its impact on improving learning outcomes over time. The Girls' Empowerment Forum is an after-school programme where girls build a peer support network, interact with positive female role models, and receive tutoring and mentoring, among other benefits. Previous research on GEFs in the SOMGEP-T programme and AGES has shown that participating girls perform better on learning assessments.

Our study also finds that participation in GEF is linked to higher literacy and numeracy scores, but this effect is significant only for C5 NFE girls. Using a regression framework that controls for region and age, we find that for C5 NFE girls, both GEF participation and ongoing contact with GEF members are associated with higher gains, particularly in literacy scores. In contrast, for C4 NFE girls, we did not observe any significant impact of GEF participation on gains in either numeracy or literacy scores.

TABLE 74: EFFECT OF GEF PARTICIPATION AND CHANGES IN LEARNING OUTCOMES AT EL

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
	C4 NFE			
GEF participation	2.2	0.483	2.5	0.459
GEF participation continued	-0.2	0.956	0.4	0.911
	C5 NFE			
GEF participation	3.3	0.098	**8.8***	0.014
GEF participation continued	2.3	0.078	6.0	0.140

Teaching Quality

The AGES Theory of Change posits that better teaching methods lead to improved learning outcomes. By training teachers in inclusive pedagogy, the aim is to create a classroom environment where girls feel comfortable and engaged, ultimately enhancing their learning. In this section, we explore the relationship between teaching quality and the changes in learning outcomes experienced by C4 and C5 NFE girls.

We use a linear regression model, controlling for age and region, to analyse the effect of various teaching quality indicators on changes in learning outcomes, as presented in the table below.

The table shows results for C4 NFE girls, indicating that the impact of teaching quality on learning improvements is inconsistent. Few of the analysed impacts are significant, and some findings are unexpected. For instance, higher teacher absenteeism is significantly associated with greater score improvements, particularly in literacy. Similarly, when teachers are perceived as moving through lessons too quickly, there is also a significant positive effect on learning scores. These inconclusive results may be due to the fact that many C4 NFE girls did not participate in any learning programme over the past year, and not all those who did were part of an NFE programme.

TABLE 75: EFFECT OF TEACHING PRACTICES ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE – C4 NFE

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
My teacher does not make me feel welcome in classroom	4.6	0.375	0.5	0.926
My teachers are often absent	7.0	0.110	**7.6***	0.046
My teacher rarely/never encourages participation	0.6	0.916	2.1	0.719
My teacher explains how learning things is useful in our lives	8.2	0.431	9.4	0.318
My teacher's lessons move too fast for me	**10.4***	0.011	**13.0***	0.003
My teacher punishes students who get things wrong in a lesson	1.1	0.770	-0.7	0.855
My teacher used corporal punishment in last week	-7.1	0.259	**-12.8***	0.040

The table below shows instead results for C5 NFE girls, who have just ended their NFE programmes. For these girls results are more in line with expectations, even if many teaching practices do not seem related to scores improvement. Teacher absenteeism is now significantly related to lower improvement (-8.5 points) in both literacy and numeracy, and the lack of teachers' encouragement for participation is associated with a significant lower increase in literacy (-18.0 points than the average increase). This reinforces the suspect that increase in scores for C5 NFE girls are more strictly related to reported teaching practices, since their participation to the programme is more recent.

TABLE 76: EFFECT OF TEACHING PRACTICES ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE – C5 NFE

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value

My teacher does not make me feel welcome in classroom	5.7	0.210	**-19.2***	0.000
My teachers are often absent	**-8.4***	0.032	**-8.5***	0.038
My teacher rarely/never encourages participation	-2.7	0.603	**-18.0***	0.000
My teacher explains how learning things is useful in our lives	-3.2	0.582	**-9.5***	0.029
My teacher's lessons move too fast for me	-5.5	0.108	6.3	0.092
My teacher punishes students who get things wrong in a lesson	0.6	0.815	3.3	0.355
My teacher used corporal punishment in last week	2.3	0.475	-3.9	0.357

Nutrition - Protein Intake

Despite not being directly addressed in the programme Theory of Change, nutrition, and more specifically protein intake has been extensively related to cognitive development and learning outcomes. The table below shows the effect of reported protein intake on changes in scores.

TABLE 77: EFFECT OF PROTEIN INTAKE ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE – C4 AND C5 NFE

Participation - Cohort type	Cohort	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
All type of protein	C4 NFE girls	8.67	0.148	9.82	0.089
	C5 NFE girls	5.70	0.297	5.14	0.349

Using a linear regression model with girls' age, cohort type, and region as control variables, we find a positive relation between protein intake and change in neither literacy nor numeracy score. However, the effect is significant only for change in literacy scores of C4 NFE girls and at the 10% level. It is important to note that the 90 percent of girls reported consumption of protein in the 24 hours before the assessment, and that this question was asked consistently only at EL. For this reason, the robustness of this result is limited. On the

other hand, the signs of the impacts are as expected, and suggest that nutrition indeed plays a role on cognitive development.

9. Transition - C4 & C5 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 and C5 NFE cohort's life pathways 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 and C5 NFE programming and estimate the transition rates for girls who were recruited into the NFE programme in 2022 before ML1 and in 2023 before ML2, respectively. 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.

9.1. Aggregate Transition Outcomes

As discussed earlier within this study,[119] 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, as well as unsuccessful. As shown in Table 78, enrolment into formal education or employment are regarded as successful transitions; dropping out, staying at a similar educational level, and non-gainful employment is considered an unsuccessful transition.

TABLE 78: TRANSITION PATHWAYS, ACCORDING TO STARTING POINT OR COHORT

Starting Point	Successful Transition	Unsuccessful Transition
C4/C5 NFE Girl **Enrolled in NFE at ML1/ML2**	• Enrolment in formal school, at any grade level • Transition into ABE[120] • Transition into a technical or vocational education programme • Transition into age-appropriate, non-exploitative employment • Transition into self-employment	• Drop out • Retention in NFE • Non-gainful employment,[121] or employment if girl is under 18 years of age

The C4 (started in 2022) and C5 (started in 2023) NFE programmes lasted approximately 11 months and have concluded. However, the C4 and C5 NFE girls can stay in the programmes longer than the 11 months

[119] See Section 4.1 *Aggregate Transition Outcomes* with FE, ABE, and C1 NFE cohorts
[120] Accelerated Basic Education programme.
[121] Any type of employment situation where the employee does not receive steady work or payment and is not sustainable (i.e., temporary, unconventional, and/or informal work opportunities)

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. While the ML2 evaluation considered girls retention in the NFE programme as a successful transition, the EL evaluation considers NFE retention as an 'unsuccessful' transition,[122] while a transition into a vocational training programme would be considered a successful transition.

The transition outcomes of the C4 and C5 NFE girls are shown in Table 79. A plurality of the C4 (15.5 percent) and C5 (21.9 percent) NFE girls are retained in their respective NFE programmes, which is now considered an unsuccessful transition. 35.1 percent of C4 NFE girls and 28.1 percent of C5 NFE girls are now employed in an age-appropriate, non-exploitative manner. Being in vocational training is considered a positive transition outcome; in total, just 1 (0.2 percent) of girls in the C4 NFE and 5 (1.4 percent) of girls in the C5 NFE cohorts fall into this category. Only 2.6 percent and 1.9 percent of C4 and C5 NFE girls, respectively, reported being self-employed, and only 0.5 of C4 cohort girls and 1.1 of C5 cohort girls in ABE education.

TABLE 79: TRANSITION OUTCOMES FOR C4 AND C5 NFE GIRLS[123]

Transition Outcomes	C4 NFE		C5 NFE	
	N=575		N=370	
	Number of Girls	Share of Sample (%)	Number of Girls	Share of Sample (%)
Retention in NFE	89	15.5	81	21.9
Age-appropriate, non-exploitative employment	202	35.1	104	28.1
Out-of-School, Idle	146	25.4	87	23.5
Enrolled in formal school	119	20.7	82	22.2
Enrolled in ABE	3	0.5	4	1.1
Self-employed	15	2.6	7	1.9
In vocational training	1[124]	0.2	5[125]	1.4
Overall Outcome				
Successful Transition	**340**	**59.1**	**202**	**54.6**

[122] Consilient Research, *CARE AGES Midline 2 Report (2023),* p.146

[123] Here, a girl is only considered to be in (self-)employment as a transition outcome if she did not concurrently report being enrolled in a learning program. See also footnote 39 in our discussion of the FCDO cohort transition outcomes.

[124] The respondent specified being trained in Henna.

[125] Vocational trainings for the participants included Specific Trade Training (5), Business Management (1), Financial Literacy (3), and Language skills (English, Arabic, etc.) (4)

25.4 percent of C4 NFE girls and 23.5 percent of C5 NFE girls said that they were now out of school but not employed. This outcome for both cohorts is regarded as an unsuccessful transition as they are neither employed, self-employed, transferred to formal schooling, nor doing vocational training. Because of the EL definition of successful transition, the aggregate transition outcome for C4 NFE and C5 NFE cohorts is 59.1 and 54.6 percent, respectively, as shown in the bottom-most row of Table 79. The most common unsuccessful transition outcome was being out of school and idle, with a quarter of the C4 NFE and slightly below a quarter of the C5 NFE samples reporting that status.

For the C4 NFE cohort, the transition rate as reported at ML2 was 74.0 percent compared to 59.1 percent at EL. This represents a drop of 14.9 points, primarily because of the change in criteria for defining successful transition at the EL (i.e., retention in an NFE programme is now an unsuccessful transition outcome at EL).

9.2. Subgroup Transition Rates

The previous section explored the specific transition pathways that the C4 and C5 NFE girls followed as a cohort. Here, we examine differences in transition rates among groups within the C4 and C5 NFE cohorts 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 and C5 NFE cohort. To do so for each cohort, we conducted a simple regression model utilising the subgroup characteristics as the predictor with the outcome of successful transition, accounting for school-level clustering.

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 formal schooling, has acquired employment, or is doing vocational training. Any other outcome is defined as an unsuccessful transition. The transition rate is thereby defined as the proportion of C4 and C5 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 the most substantially. The table below compares the transition rates of cohort subgroups of interest relative to the remainder of the sample.

TABLE 80: SUBGROUP TRANSITION OUTCOMES AMONG THE C4/C5 NFE COHORT

Sample Characteristics	C4 NFE			C5 NFE		
	N	Transition Rate (%)	Difference from Aggregate	N	Transition Rate (%)	Difference from Aggregate
Overall	575	59.1	-	370	54.6	-
Geography						
Banadir	272	64.7	10.6	165	66.1	**20.7***
South West State	219	52.5	-10.7	143	44.1	**-17.2***
Hirshabelle	84	58.3	-0.9	62	48.3	-7.5

Family Characteristics						
Girl has only one living parent	39	56.41	2.3	13	63.6	10.2
Girl has no living parents	1	0.0	**-54.8***	22	0.0	0.0
Girl does not live with either parent in her HH	18	77.8	**25.6***	0	75.0	21.1
Female-headed household	137	57.7	10.1	8	51.8	-14.9
HoH has no education of any kind (no Quranic)	32	68.8	**16.9***	20	45.0	-12.8
HoH has no formal education (may have Quranic)	169	55.6	7.4	82	59.8	16.9
Caregiver has no education of any kind (no Quranic)	28	67.9	15.5	21	47.6	-9.9
Caregiver has no formal education (may have Quranic)	166	56.0	9.2	72	59.7	12.5
Household Wealth and Socioeconomics						
HH has poor roof, at EL	110	69.1	**11.7***	53	67.9	**14.5***
Went to sleep hungry many nights (10+), last 12 months, at EL	107	70.1	**13.0***	56	71.4	**18.8***
Went to sleep hungry most/all nights, last 12 months, at EL	26	50.0	-10.0	25	48.0	-8.2
Went without water for home use most/all days, last 12 months, at EL	63	54.0	-6.3	32	59.4	4.1
Went without medicine most/all days, last 12 months, at EL	84	57.1	-2.8	54	64.8	10.8
HH owns land, either solely or jointly, at EL	176	59.1	-2.9	109	55.1	-2.2
Household owns a phone, at EL	31	38.7	13.7	35	51.4	26.4
Household owns a smartphone, at EL (alt. coding)	199	65.3	8.9	136	54.4	-2.0

HoH does not have an occupation or does not earn a wage, at EL	13	30.8	-10.1	14	64.3	25.8
Disability Status						
Girl has physical disability	20	70.0	11.3	20	60.0	5.7
Girl has physical disability, alternative coding	22	68.2	9.4	26	57.7	3.3
Girl has cognitive, behavioural, or communicative disability	27	55.6	-3.8	26	53.9	-0.8
Girl has cognitive, behavioural, or communicative disability, alternative coding	34	52.9	-6.6	34	58.8	4.7
Girl has mental health disability	104	64.4	6.5	140	59.3	7.6
Girl has mental health disability, alternative coding 1	117	64.1	6.2	166	60.8	**11.3***
Girl has mental health disability, alternative coding 2	202	63.9	7.3	194	59.8	**10.9***
Girl has non-mental health disability	46	63.0	4.3	41	58.5	4.4
Girl has non-mental health disability, alternative coding 1	53	60.4	1.4	49	57.1	2.9
Girl has any disability	136	62.5	4.4	160	58.1	6.2
Girl has any disability, alternative coding 1	155	61.9	3.8	185	58.9	8.7
Girl has any disability, alternative coding 2	293	63.5	**8.9***	250	58.4	**11.7***
Girl's Marital Status						
Ever married, at BL	5	60.0	5.6	10	60.0	4.9
Ever married, at EL	276	66.3	**13.6***	155	57.4	5.1
Currently married, at BL	3	66.7	12.3	4	50.0	-5.8
Currently married, at EL	211	64.5	8.3	112	52.7	-2.8

Is a mother, at BL	3	66.7	12.3	6	66.7	16.7
Is a mother, at EL	240	66.7	2.8	127	59.1	7.3
Indicators of Household Marginalisation						
Pastoral Household	7	71.4	12.5	13	46.2	-9.1
IDP Household	262	59.2	0.1	147	56.5	3.1
Household speaks af-Maay	207	54.6	-7.1	112	52.7	-2.8
Other Individual Characteristics						
Consumed any protein in the last 24 hours	506	59.5	-0.5	331	54.7	-13.3

State-wise, for the C4 NFE cohort, disaggregating the subgroups by states did not reveal any significant differences in successful transition rates compared to the remainder of the sample. For the C5 NFE cohort, the girls from Banadir displayed a significantly higher successful transition rate – by 20.7 points – compared to girls not in Banadir; however, the girls from South West State observed a significantly lower rate – by 17.2 points – compared to girls not in South West State. These findings are interesting given that learning outcomes in the Banadir zone were poorer for both C4 and C5 girls when compared to other zones. The contrast between more successful transition outcomes and comparatively poorer learning outcomes might be attributable to the fact that Banadir girls are likelier to be employed, given that Banadir is arguably the most economically developed region of Somalia. Indeed, among C4 girls in Banadir, 43.4 percent are employed, compared to 28.8 in South West State and 25.0 in Hirshabelle. For C5 girls in Banadir, this figure stood at 42.4 percent at EL, compared to 18.9 for girls in South West State, and 11.3 in Hirshabelle state. While employment contributes to successful transition outcomes, it may not necessarily lead to learning gains, as employment may mean comparatively fewer opportunities for girls to study and/or use numeracy and literacy skills.

In terms of family characteristics, for the C4 NFE cohort, girls who have a head of household (HoH) with no education registered a transition rate that is 16.7 points higher compared to girls who do not. This difference is significant. As this runs counter to our expectations on transition outcomes, additional analysis was conducted to identify the source of the association. When adjusted for region, the adjusted rate difference was no longer significant at the five percent level. This higher rate of transition is likely due to increased motivations for the girl to find employment, as when looking at the specific transition outcomes, 40.6 percent of C4 NFE girls with a HoH with no education, respectively, ended up now employed, compared to 22.9 percent of girls with a HoH with some form of education.[126] Also for the C4 cohort, there is a significant difference of 25.6 points in the successful transition rate for girls who do not live with either parents in her HH. The result is mostly driven by a majority of girls transitioning from NFE to formal education (44.4 percent) and to being employed (33.3 percent), but this result needs to be interpreted with caution, as the

[126] It is important to note that the sample size of girls with HoH with no education is low (32), so translation of results to the overall CARE AGES population should be taken with caution.

number of girls living without either parent is low (18). No statistical significance was found with any other family characteristics in the C4 NFE cohort, and none are found for the C5 NFE cohort.[127]

In terms of household socioeconomic status, for both the C4 NFE and C5 NFE cohorts, we find a significant difference in the rate of successful transition from NFE schools among girls from households with poor roofing. For each cohort the difference is 11.7 and 14.5 points respectively. In the same fashion, we also find a significant difference in the rate of successful transition from NFE schools among girls who went to sleep hungry ten nights or more for the last 12 months, as measured at EL. The difference is 13.0 points for C4 cohort girls and 18.8 points for C5 cohort girls. The differences in the rates of successful transition are mostly explained by the larger proportion of girls who transitioned from NFE to work. 44.6 percent of the C4 girls from households with poor roof are now employed, compared to 34.1 percent who are not. Similarly, 52.3 percent of C4 girls who had 10 or more days of sleeping while hungry are now employed, compared to 32.2 percent who did not. For the C5 cohort girls we see the same pattern: of the girls living in houses with poor roofing, 43.4 percent transitioned from NFE to employment, whereas only 26.7 percent of girls with proper roofing did the same. For girls sleeping with hunger at least 10 nights in the past year, 50.0 percent transitioned from NFE to employment, whereas only 25.3 percent of girls without that level of hunger did the same. This suggests that girls from economically vulnerable households are more likely to experience successful transitions because they are more likely to work, possibly due to the necessity to help their households earn income.

No significant difference in the rate of successful transitions was identified among girls with physical or cognitive/behavioural/communicative disabilities compared to those without the respective disabilities for both C4 and C5 NFE cohorts. For C5 cohort girls we observe a significant difference of successful transition rates for the two different definitions of mental health disability (related to anxiety and depression). The points difference is very similar, with 11.3 points for the first coding and 10.9 points for the second. We also see a significant rise in successful transition from NFE in the aggregate for the C4 and C5 cohorts if we consider girls with any disability under the second, more lenient, alternative coding. As with the previous subgroups, the successful transition is explained by a higher transition to employment (37.5 percent for girls with any disability versus 33.3 percent for girls without disabilities, among the C4 girls; and 31.6 percent versus 20.8 percent among the C5 counterparts), and by a lower rate of girls staying in NFE (17.7 percent versus 13.3 percent in C4; 18.8 percent versus 28.3 percent in C5)..

In terms of marital and maternal status, for the C4 NFE cohort, girls who reported ever being married – including those who are currently divorced or widowed – as of EL had a significantly higher successful transition rate compared to those who were never married, at 13.6 points higher. Checking for the composition of transition outcomes via tabular analysis and a Chi-square test, there was a significant association between being ever married and transition outcomes, with the most frequent outcome among those ever married being having any gainful employment – including self-employment – at 52.8 percent.[128] However, when the regression model controls for age, the rate of transition is observed to not be significantly different, though age was a significant and positive predictor. The findings[129] are logical, in that older girls

[127] It is important to note that while statistical significance was found when assessing the impact of having no living parents on transition rates, the subgroup sample size of 1 negates the validity of the reported subgroup difference.

[128] Chi-square=270.0, $p<0.0001$, d.f.=6

[129] Unpaired t-test to compare mean ages between C4 NFE girls who ever married versus not – with both groups having successful transition outcomes (diff= -4.7, t=-14.6, $p<0.0001$, d.f.=338).

would have a higher likelihood of ever being married,[130] while also more likely to have searched for and found employment.[131] It is also important to note that while there was a reported statistical significance among those currently married at BL for the C4 NFE cohort, the sample size of 3 is too small to draw valid inferences from this finding.

In terms of household marginalisation, all but one of the indicators in this set of subgroups did not display any statistically significant differences to the aggregate transition outcome rate for both C4 and C5 NFE cohorts. In the C4 NFE cohort sample, only girls from pastoralist households showed better successful transition rates compared to those from non-pastoralist households, with a difference of 28.2 points. Finally, in terms of nutrition, no significant difference in successful transition rates compared to the reference group was found among girls whose household maintains consistent protein consumption.

9.3. Testing the Theory of Change

In the same manner as assessed with the FE, ABE, and C1 NFE Cohorts, 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 continue with the linear regression approach to assess whether the intermediate outcomes significantly impact transition rates, controlling for state and age, as well as accounting for school-level clustering.

In Table 76: Effect of teaching practices on gains in numeracy and literacy scores since baseline – C5 NFE, the results of the role of the intermediate outcomes as predictors for transition rates are listed, with the left-most column detailing the type of intermediate outcome evaluated under the Theory of Change. The right two columns list the differences in transition outcomes from the reference group (i.e. girls who do not belong in under the subgroup listed in the first column) in C4 and C5 NFE cohort, separately. Unlike with the FE, ABE, and C1 NFE girls, who entered our sample in the same round (BL), the C4 and C5 NFE cohorts' samples were first surveyed at ML1 and ML2, respectively. As such, we do not compare the two cohorts together, as each is in a different point in time in their post-NFE programme trajectories.

In terms of the teaching quality outcomes, we only see a significant relationship in the C5 cohort girls between the rate of successful transition and girls reporting her teacher punishes students who get things wrong during a lesson. This means, for girls who report the abuse in the survey, there is a 5.2-point increase in the possibility of having done a successful transition from NFE. The most salient difference in transition rates is that the rate of girls transitioning from NFE to being unemployed and not studying (an unsuccessful transition) is higher for girls who do not report teacher's punishment than those who do (29.5 percent versus 18.3 percent). This result is counterintuitive, as one would expect that punishment in class would be a negative factor for the girl's education and later successful transition from NFE. For C4 cohort girls there is no significant factor on the girls' successful transition.

While there is a significantly higher level of successful transition among C4 NFE girls whose caregiver is concerned about travel safety to school, this section only had a sample size of 3, which is too small to generalise the finding to their respective cohort population. Furthermore, the rates of successful transition

[130] Unpaired t-test to compare mean ages between C4 NFE girls who ever married versus not (diff= -4.6, t=-19.0, p<0.0001, d.f.=572).

[131] Unpaired t-test to compare mean ages between C4 NFE girls who successfully transitioned versus not (diff=-0.1, t=-3.0, p=0.0029, d.f.=573).

among the girls in both C4 NFE and C5 NFE cohorts who felt unsafe were not significantly different.[132] Household chore burden at their respective baselines also did not correlate with higher transition outcomes for either cohort.

TABLE 81: EFFECT OF KEY PREDICTORS OF TRANSITION RATES, BY COHORT

Independent Variable of Interest	C4 NFE Cohort Sample Size (n)	Effect of Predictor on Transition Rate (%)	C5 NFE Cohort Sample Size (n)	Effect of Predictor on Transition Rate (%)
Teaching Quality				
Girl Feels Unwelcome by Teacher	80	6.9	61	3.2
Girl Feels Her Teachers Are Often Absent	100	3.8	85	2.2
Girl Feels Her Teacher Rarely/Never Encourages Participation	34	1.4	24	4.2
Girl Feels Her Teacher Rarely/Never Explains How Things Learned Are Useful in Her Life	10	3.7	24	4.5
Girl Feels Her Teacher's Lessons Move Too Fast for Her	174	-1.7	92	-1.1
Girl Reports Her Teacher Punishes Students Who Get Things Wrong During a Lesson	151	0.7	197	**10.4***
Girl Reports Her Teacher Used Corporal Punishment during the Week of Interview	38	3.8	65	11.7
Caregiver Attitude				
Caregiver does not feel it is safe for girls to travel to the school	3	**38.7***	9	20.0
Caregiver aspires to send girl to university	359	0.7	262	-1.4
Caregiver believes girls' education is worthwhile, even if funds are limited	458	2.5	201	5.8

[132] The sample size for girls reporting being unsafe traveling to school were also low for both cohorts (9 for C4 NFE; 6 for C5 NFE).

Caregiver believes work or HH chores are acceptable reason to not attend school	46	-0.5	48	7.4
Caregiver believes cost of education is acceptable reason to not attend school	119	3.6	52	2.7
Household chore burden at BL	532	-12.1	304	-6.2
YLI score at BL	575	0.0	370	-0.1
GEF Participation (binary indicator)	234	2.8	142	2.6
GEF Participation (ordinal 0-4 score)	67	3.8 (per 1-unit increase)	52	0.5 (per 1-unit increase)

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.[133] Overall, the YLI score and GEF participation are not significant predictors of transition outcomes for both C4 and C5 cohorts.

Indeed, we mostly did not find evidence of the intermediate outcomes related to Theory of Change acting as significant predictors to successful transition outcomes. One factor that shows a statistically significant relation with successful transition for C4 girls – caretakers' views that it is unsafe for girls to travel to school – is counterintuitive and does not fit the Theory of Change. However, the small sample size of three C4 girls falling into this category prevents us from drawing valid inferences from this finding. For C5 girls, we find that the use of corporal punishment by teachers is associated with a higher transition rate.

10. Sustainability - C4 & C5 NFE

Unlike with the sustainability indicators highlighted with the FCDO cohorts, school-level sustainability indicators and the GWD retention levels were not included for the USAID cohorts for the following reasons:

1. The school-level indicators were applicable to formal schools only.
2. NFE programmes were designed to last for one year, so retention is not a goal for NFE enrolees.[134]

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

Overall participation in VSLAs among caregivers is low, with 14.3 percent of caregivers from C4 NFE and 10.3 percent of caregivers from C5 NFE participating in a savings group. In terms of continued active membership in the savings group, 40.0 percent of C4 NFE caregivers and 75.0 percent of C5 NFE caregivers reported they remained active; however, the low sample count for each cohort for this response indicates that these rates are likely not to translate to the actual rate of continued active membership among these

[133] This is similar to the analysis done in Section 4 with FE, ABE and C1 NFE girls.
[134] See *Section 7.2. Subgroup Transition* to see the impact of the NFE programme on transition outcomes for GWDs.

caregivers in the AGES population.[135] Among NFE girls, the participation rates are comparatively higher per cohort at 28.2 percent for girls in C4 NFE and 23.5 percent for girls in C5 NFE. Additionally, active membership in the VSLAs among the girls was substantially higher, with 42.0 percent of C4 NFE girls and 48.9 percent of C5 NFE girls reporting their continued involvement.

The low rate of VSLA participation and retention among caregivers, but higher rates among the girls is very likely due to the older demographics of the NFE girls, as the girls are likely the primary target of recruitment to engage with the VSLAs and be able to build up their wealth rather than rely on their caregiver to fill that role. However, even with that, the relatively low participation rate among the girls suggests that these savings groups are not as likely to expand after the conclusion of AGES. It is likely that the girls from the USAID may continue their involvement in the savings groups independently for longer than their FCDO counterparts due to their higher rates of active membership compared to that of the C1 NFE girls (25.3 percent) (See Section 5*: Sustainability*).

Proportion of parents able to support costs of girls' education

Among caregivers in the C4 NFE cohort, 68.6 percent of them said that education being "too costly" is an acceptable reason to keep a child out of school; similarly, for those in the C5 NFE cohort, 61.5 percent of caregivers believe this. 57.1 percent of C4 NFE caregivers said it was acceptable to pull a girl from school to prioritise household work and chores while 41.0 percent of C5 NFE cohort caregivers said was unacceptable. This suggests that caregivers are still unable to support the costs related to sending their girl to school, and that the termination of AGES programme support in this component may prevent more caregivers from continuing their child's education. Qualitative interviews with CEC members reinforce this notion, highlighting the role of financial challenges and limited community resources on impacting students' academic continuation and success.[136]

Among girls in the USAID cohorts, education is still a high priority for them despite any experienced financial challenges. When asked if "even when funds are limited it is worth investing in your education", 94.5 percent of C4 NFE girls and 96.2 percent of C5 NFE girls agreed or strongly agreed[137] with that statement. However, vignette interviews with the girls reveal that for some, perceptions of education-related expenses still prevent girls from going to school, even if free education opportunities are available. As one girl noted, "I know several underprivileged girls who believe that school requires money they cannot afford. It is crucial to raise awareness among them about the availability of free education." [138]

Parental support for girls' participation in GEFs

For the C4 NFE cohort, while only 16.6 percent of girls reported ever participating in a GEF, 59.6 percent of them did report continuing their contact with other GEF members. However, only 9.0 percent of girls reported every participating in any GEF activities. Similarly, for the C5 NFE cohort, only 20.5 percent of girls participated in a GEF, but 55.9 percent of those who participated were still in contact with their fellow GEF members. In line with the C4 NFE cohort, only 9.3 percent of C5 NFE girls reported ever participating in any GEF activities.

[135] **C4 NFE Cohort:** N=5; **C5 NFE Cohort:** N=4
[136] See FGD with CEC members, Banadir, Int. 107.
[137] Includes both "agree" and "strongly agree"
[138] See Vignettes FGD with Girls, Banadir, Int. 606.

TABLE 82: PARTICIPATION RATES AND ENGAGEMENT IN GEFS, BY COHORT

Indicator	C4 NFE girls	C5 NFE girls
Ever Participated in a GEF	16.6%	20.5%
Still in Contact with GEF Members	59.6%	55.9%

Looking at whether these girls also participated in other youth group and networking opportunities, only 33.4 percent of C4 NFE girls and 34.0 percent of C5 NFE girls said they have. This includes 30.0 percent of C4 NFE girls and 29.4 percent of C5 NFE girls who reported not ever participating in a GEF. The overall low participation in *any* forums or youth groups indicate that recruitment and retention efforts in the latter periods of AGES (from 2022 onwards) was relatively lower than that effort in 2019 with the baseline FCDO cohorts.

Proportion of GEFs implementing community actions to support attendance and retention

As shown in Table 83, GEFs among C4 and C5 NFE girls were much more active in raising awareness for girls- at 63.1 percent and 61.8 percent, respectively. In contrast, the rates of implementation from the other activities were substantially lower, ranging from 2.8 to 18.4 percent among the C4 NFE cohort and from 4.9 to 24.8 percent for the C5 NFE cohort. These observations indicate that the breadth of programming from the GEFs have been small (similar with findings among the FCDO cohort). As GEFs were designed to provide community-based interventions that would result in downstream improvement in girls' school enrolment, attendance, and retention, the low rates of GEF activity implementation indicates there is room for GEFs to increase their involvement in a wider ranges of community activities.

TABLE 83: PROPORTION OF GEFS ENGAGAGED IN ACTIVITIES, BY COHORT

Indicator	C4 NFE girls	C5 NFE girls
Girls Education Support		
Enrolling out-of-school girls	16.7%	24.8%
Participation in CECs	10.6%	6.9%
Teaching other girls/ study groups	14.2%	12.8%
Business & Finances		
Savings group	10.6%	10.8%
Joint business	2.8%	7.8%
Changing Community Attitudes		
Preventing early marriage	18.4%	6.9%
Awareness raising	63.1%	61.8%

Trainings (health, girls' rights, gender, etc)	8.5%	4.9%
Community discussions	9.2%	4.9%

11. Intermediate Outcomes

11.1. Leadership and Life Skills

As in section 5.3, this section aims to assess the progress in leadership skills, self-confidence, and life skills gained throughout the program by using the Youth Leadership Index (YLI) for girls in USAID Cohort 4 and Cohort 5 of this evaluation. The Youth Leadership Index is a composite measure based on 21 questions, each evaluated on a 4-point Likert scale. Respondents indicate how often (rarely, sometimes, most of the time, almost always) they behaved in certain ways, depending on the question. Lower scores indicate more negative outcomes, while higher scores suggest more frequent positive behaviours. The YLI score ranges from 21 to 84 points and has been standardized on a 0 to 100 scale for analysis. If a girl scores the lowest possible points (21) by answering 'rarely' to all questions, her standardized YLI score will be 0%. CARE International developed the YLI to measure self-confidence, decision-making, voice, vision, and organizational skills (including the ability to motivate and collaborate with others to address common issues). For a detailed introduction to the YLI, its construction, and its constituent questions, please refer to Section 7.3.

Cohort 4 NFE Girls

In this section, we describe changes in YLI scores over time since the ML1 round, when C4 NFE girls were first interviewed. This set of girls showed statistically significant improvements in YLI scores considering both from ML1 to ML2 and from ML1 to EL. Although results show a shortfall in scores between ML2 and EL, there was an overall 15.9 percentage points increase in YLI between ML1 and EL. The proportion of girls who reach at least 70% of the points show a similar evolution as the average YLI scores. Between ML1 and EL there is a statistically significant increase of 25.6 percentage points of the proportion of girls with 70 or more points, though the level shortly falls from ML2 to EL.

At the state level, the three regions analysed, Banadir, Hirshabelle and South West State, show similar evolution of their scores and proportion of girls getting above the 70% mark. The special case is the share of girls achieving 70% in South West state, where it started with 3.8% in ML1 and increased by 34.6 percentage points, more than the other two states.

Analysing for girls' characteristics like IDP status, age, participation in GEF, pastoralist household, Maay speaking household and household led by a female, we only find statistically significant differences between Maay and non-Maay speakers. Controlling for the interaction with rounds, the regression shows a difference of -5.2 points at the 99% significance level in the YLI score for Maay speaking girls, meaning, on average, Maay-speaking girls score 5.2 points less than non-Maay speakers.

TABLE 84: CHANGES IN YLI SCORES FROM ML1 TO ML2 AMONG C4 NFE GIRLS, BY ZONE

C4 NFE Cohort Subgroups	ML1	ML2	EL	Difference (ML1 - EL)
	YLI Score (0-100 standardised scale)			
Overall	49.0	67.2	64.9	15.9***
Banadir	52.4	68.9	63.4	11.1***
Hirshabelle	41.9	70.0	66.8	24.9***
South West State	47.7	64.2	65.8	18.1***
	Share of Girls Achieving 70% YLI Score Target			
Overall	10.4%	43.9%	36.0%	25.6pp***
Banadir	13.0%	46.3%	31.0%	18.1pp***
Hirshabelle	19.7%	54.9%	45.1%	25.4pp*
South West State	3.8%	36.8%	38.5%	34.6pp***

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

Cohort 5 NFE Girls

We now analyse the evolution of YLI results for girls from the Cohort 5 NFE, for which the ML2 report stablished their baseline. Overall, girls of the Cohort 5 NFE show slight and statistically non-significant declines in both their average YLI scores and the share of girls achieving the 70% score target. Although it is not a positive result, it is in line with the decline in scores from ML2 to EL of all previous cohorts analysed. Also, Cohort 5 NFE girls show the highest scores for the ML2 round across all cohorts analysed. It is possible that the C5 NFE cohort has a higher baseline levels of leadership skills because of the AGES programme having already created a more supportive environment at the community level by the time the C5 NFE girls joined the programme, but we do not have strong enough evidence to fully explain the high baseline YLI value for the C5 NFE girls. Nonetheless, C5 NFE girls are later surpassed in the EL by FE and ABE girls in both YLI scores and proportion of with at least 70% points.

Across states, we only see a statistically significant decline of 5.2 points in scores for the Banadir zone. South West state is the only state that sees an increase of 4 points in their scores and 6.3 percentage points in the proportion of girls with at least 70 points, but both are not statistically significant results.

Note from the project: The decline in the YLI at the second evaluation round (endline for C5) is a common pattern; it is associated with the overreporting of leadership skills at the baseline and the realization of barriers to rights during the participation in girl-led action through the program. Gains in the YLI are typically seen from the third evaluation round onwards, as observed for C1 and C4.

Comparing girls' characteristics like we did for Cohort 4 NFE girls (IDP status, age, participation in GEF, pastoralist household, Maay speaking household and household lead by a female), again we find statistically

significant results for the difference between Maay and non-Maay speakers. Like Cohort 4 girls, Maay speakers of Cohort 5 had a statistically significant difference in YLI scores compared to non-Maay speakers, with Maay speakers registering an average score that is lower by4.8 points. Nonetheless, they show a higher, statistically significant improvement in scores than non-Maay speakers between ML2 and EL, with an improvement of 7.9 points between rounds. The same dynamic appears in the proportion of girls reaching the 70% points level, where there is a statistically significant gain of 17.8 percentage points by the EL for Maay speakers. Unlike Cohort 4 girls, we see a larger proportion of girls with 70 points or higher for girls in households led by a female, but we do not see a difference in the change in proportions proportion between rounds, compared to households led by males.

TABLE 85: YLI SCORES, BY ZONE, AMONG C5 NFE GIRLS

C5 NFE Cohort Subgroups	ML2	EL	Difference ML2 - EL
YLI Score (0-100 standardised scale)			
Overall	68.2	66.9	-1.3
Banadir	70.6	65.4	-5.2**
Hirshabelle	72.4	69.2	-3.3
South West State	63.7	67.7	4.0
Share of Girls Achieving 70% YLI Score Target			
Overall	48.1%	44.6%	-3.5pp
Banadir	52.1%	42.4%	-9.7pp
Hirshabelle	58.1%	48.4%	-9.7pp
South West State	39.2%	45.5%	6.3pp

11.2. Increased Self-Efficacy

This section examines the intermediate outcome of increasing the self-efficacy of girl learners. This comprises two components: 1) positive youth development, and 2) access to protection services. This section examines these components jointly for C4 and C5 cohort girls. Unless otherwise noted, we use the sample of C4 and C5 girls who were in both their respective cohort-specific baselines (ML1 and ML2, respectively) and in the EL.

Positive Youth Development

During ML1, ML2, and the EL, the evaluation team collected data on positive youth development from NFE cohort girls, which includes the USAID-supported C4 and C5 NFE cohorts. To measure positive youth development, the evaluation team collected data using indicators from Chinese Positive Youth Development

Scale, or CPYDS, which is a widely recognised and tested scale that measures several aspects of youth development, including resilience and self-confidence. For the purposes of this evaluation, and in conjunction with the CARE AGES team, selected seven items related to self-efficacy from the larger CPYDS pool of indicators. For each CPYDS indicator, respondents were presented with a statement, and asked to select an answer from a four-point scale, with answer choices ranging from "Strongly agree" to "Strongly disagree". The table below presents the percentage of respondents, in each round, who selected "Agree" or "Strongly agree" for each CPYDS question.

Note that indicators one to five are scored "negatively", meaning that respondents who agree or strongly agree with these items are considered to be less self-efficacious. All p-values were derived from bivariate linear regressions with round as the predictor.[139]

TABLE 86: PERCENTAGE OF RESPONDENTS (STRONGLY) AGREEING WITH CPYDS ITEMS, BY COHORT

CPYDS Question	Cohort 4 NFE Girls				Cohort 5 NFE Girls			
	ML1	EL	Diff.	P-Value	ML2	EL	Diff.	P-Value
Lack of control of life	45.4	60.6	15.2*	0.00	71.3	63.8	-7.5	0.08
Lack of solutions to problems	56.0	51.1	-4.9	0.31	68.3	48.1	-20.2*	0.00
Inability to change life	42.3	56.3	14.1*	0.00	66.1	50.5	-15.6*	0.00
Helplessness	48.0	50.1	2.1	0.64	59.6	41.9	-17.7*	0.00
Fate not in hands	59.0	61.2	2.3	0.56	70.7	58.9	-11.8*	0.00
Determine own life	55.5	72.7	17.2*	0.00	79.4	75.4	-4.0	0.24
Ability to complete tasks	75.8	86.6	10.8*	0.00	87.3	81.1	-6.2*	0.04
CPYDS Index	45.6	45.7	0.1	0.96	52.8	43.8	-8.9*	0.00

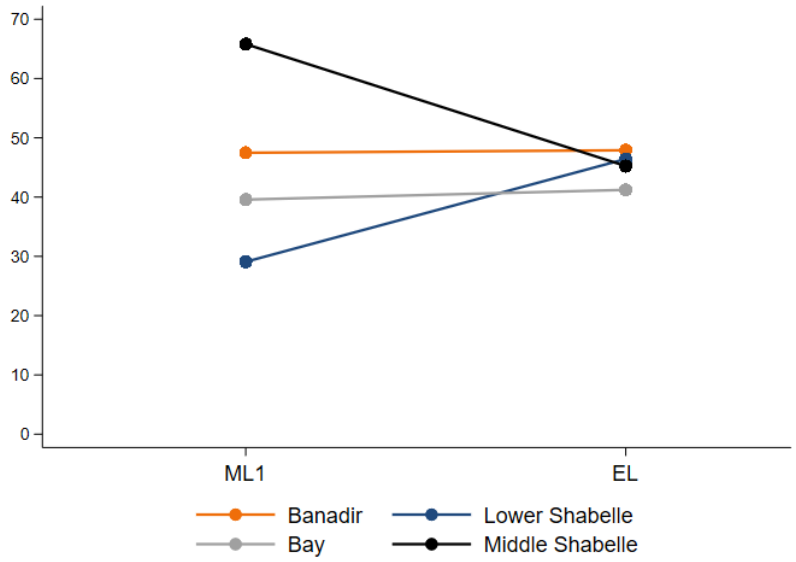
For C4 NFE girls, the table highlights large and significant increases in four out of seven indicators. Two are negative self-efficacy measures, while the other two are positive measures. To measure what this means on balance, we construct an index aggregating the seven items: each of the first five items is given a score of 1 if a respondent agrees or strongly disagrees, and 0 otherwise. On the other hand, the last two items are given a score of 0 if a respondent agrees or strongly disagrees, and a 1 otherwise. We then take the unweighted average of the seven items and multiply it by 100. In this index, the greater the score, the more negative the self-efficacy measure is. Using this index score, we note that the countervailing patterns of change in the aforementioned CPYDS items among C4 NFE girls nullify each other, leading the overall CPYDS index to register a negligible and non-significant increase of 0.1 point.

The stagnant CPYDS index scores between ML1 and EL for the C4 NFE girls belie substantial geographic variation in score changes over time. As seen in Figure 15 below, girls in Middle Shabelle and in Lower Shabelle registered substantial changes in their CPYDS index scores from ML1 to EL: for girls in Lower Shabelle, we observe a significant average increase of 17.3 points ($p < 0.01$), while we observe a significant average increase decrease of 20.6 points for girls in Middle Shabelle ($p < 0.01$). In other words, while girls in Middle Shabelle exhibited more positive self-efficacy measures over time, Lower Shabelle girls exhibited

[139] As a robustness check, we also ran linear regressions models controlling for zone and age. While the exact magnitude of change varied slightly for each CPYDS item, the overarching trends of growing scores for C4 NFE girls and declining scores for C5 NFE girls held true.

more negative measures over the same time interval. Both Banadir and Bay regions saw stagnated CPYDS index scores across rounds.

FIGURE 15: REGIONAL DIFFERENCES IN CPYDS INDEX SCORES, C4 NFE GIRLS



C5 NFE girls, in general, demonstrate the opposite pattern of change in CPYDS item scores: the percentage of girls agreeing or strongly agreeing to each of the seven items has uniformly declined from ML2 to EL. For five of the seven items, this decrease was statistically significant, with four of them significant at the 1 percent level. The aggregate effect of these changes is a decrease in the overall CPYDS index score, which fell by 8.9 points ($p = < 0.01$) between ML2 and EL. In other words, unlike the C4 NFE cohort, we observe improving self-efficacy among C5 NFE girls between their cohort baseline and the EL evaluation.[140] [141] It is worth noting, however, that C5 NFE girls on average had significantly higher CPYDS scores at their cohort baseline than C4 NFE girls had at theirs. As such, while opposite trends were observed among the two cohorts, their scores

[140] It is worth noting once again that this EL evaluation represents different points in time for the two cohort's "post-AGES" trajectory. Because the C4 NFE girls were first recruited at ML1, this EL evaluation represents data for girls who have completed their initial NFE programme more than a year ago. In contrast, C5 NFE girls were first recruited during the ML2 evaluation, which means the EL evaluation had them surveyed when less time had elapse since the end of their initial NFE programming. Nonetheless, data from the ML2 report indicates that the C4 NFE cohort of girls displayed the same patterns of increasing scores in each of the seven CPYDS items (from ML1 to ML2) even when they were at a similar point in time as the C5 NFE during this EL evaluation. In other words, the different overall trends between C4 and C5 NFE girls are unlikely attributable to time elapsed since the end of the initial NFE programming, but due to some other differences between the two cohorts.

[141] To check for similarities between cohorts, we conducted balance tests on variables that may have theoretical links to CPYDS responses. This included age, (mental health) disability status, whether a girl is currently enrolled in an education program, and unemployment status. We find significant differences between the two cohorts on these measures, as measured at their respective cohort baselines. However, including them in the regression models do not alter the overarching trends observed in Table 86.

by EL were comparable, though it is possible that C5 girls' scores will continue to decline (and C4 girls' scores to increase) after this EL evaluation, which would lead to divergent scores once again.[142]

Disaggregating the CPYDS index score changes over time reveals that the pattern is more evenly distributed amongst various subgroups in the C5 cohort of girls. For instance, we observe less variation in score changes across the different regions. Banadir, Lower Shabelle, and Middle Shabelle all experienced declines in the CPYDS index score between ML2 to EL, while Bay region's score remained unchanged. The declines for the first three regions were 9.3, 8.8, and 17.5 points, respectively. All declines were also significant at the 5 percent level, while Banadir's decline was significant at the 1 percent level as well. As with the C4 NFE cohort, girls in the Middle Shabelle region once again saw the steepest decline in CPYDS index scores between rounds.

At face value, two exceptions to the generally uniform pattern of change across subgroups are worth noting. The first is the index score changes among *af-Maay* speakers and non-speakers. While the latter saw a significant 11.8-point decrease between rounds, the former saw a 2.3-point increase over the same time period. A second divergent trend is that girls with any type of disabilities experienced larger declines in their index scores compared to girls without disabilities, though their average score remains similar at EL. For girls with disabilities, this decline was a significant 11.2 points, compared to the 4.2 points observed among girls without disabilities.

For the remainder of this subsection, we briefly highlight how each item on the CPYDS is presented to girls, and highlight the changes in scores since the respective cohort baselines of C4 and C5 girls.

Control of own life

The first of the seven items in the CPYDS measures girls' perceptions of control over their own lives. Girls were asked the extent to which they agreed with the following statement: "I have little control of things that happen in my life." Among C4 NFE girls, the percentage of girls who reported that they were in agreement grew from 45.4 percent at BL to 60.6 percent at EL, with a statistically significant difference of 15.2 points. Among C5 NFE girls, the percentage of girls agreeing declined by 7.5 points, from 71.3 percent at ML2 to 63.8 at EL.

Solutions to problems

Next is the CPYDS measure of girls' ability to solve problems. When faced with the statement, "I do not have any solutions for some of the problems I am facing", 56.0 percent of C4 NFE girls initially agreed at ML1. This was followed by 51.1 percent of girls at EL, which represented a 4.9-point decrease. C5 NFE girls followed a similar trajectory, with 68.3 percent of girls at ML2 and 48.1 percent of girls at EL – a significant decrease of 20.2 points.

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", 42.3 percent of C4 NFE girls agreed with the sentiment in the ML1 round, while 56.3 percent agreed during the EL round. This 14.1 increase was significant at the 1 percent level. For C5 NFE girls, the ML2 figure stood at 66.1 percent, which declined by a significant 20.2 points, reaching 48.1 percent at EL.

[142] Regression results show that the two cohorts had significantly different CPYDS index scores at their respective baselines, but that these was no significant difference at EL.

Helplessness

The fourth CPYDS item measures the extent to which they feel helpless in the face of challenges. At ML1, 48.0 percent of C4 NFE girls agreed or strongly agreed with the statement "When I face life difficulties, I feel helpless". By EL, this figure had increased by 2.1 points, reaching 50.1 percent of C4 girls. At ML2, 59.6 percent of C5 girls agreed or strongly agreed with the statement. By EL, 41.9 percent of C5 girls claimed as much – a significant decline of 17.7 points.

Fate not in hands

The next item measures how much girls agree or strongly agree with the following statement: "I feel my life is determined by others and fate". Overall, 59.0 of C4 cohort girls agreed or strongly agreed with the statement, followed by 61.2 at EL. This increase was not statistically significant. Among C5 girls, 70.7 percent of girls agreed or strongly agreed with the statement, followed by 58.9 at EL, representing a statistically significant decrease of 11.8 points.

Determine own life

The sixth item measures girls' perception of their ability to determine their own lives, as encapsulated in this statement: "I believe things in my life are mostly determined by me." Note that this indicator differs from the preceding four in that (strong) agreement with this indicator represents strong self-efficacy, whereas agreement with the preceding five represents less self-efficacy. At ML1, 75.8 of C4 NFE girls expressed their agreement with this statement. The figure had increased by a significant 10.8 points by EL. For C5 girls, 87.3 percent had expressed their agreement during ML2, followed by 81.1 percent at EL – a significant decrease of 6.2 points.

Task completion

The final item measures girls' agreement to the following statement: "I can finish almost everything that I am determined to do." Like the item on determining one own's life, agreement with this indicator represents positive self-efficacy. At ML1, 75.8 of girls agreed or strongly agreed with the statement, followed by significant increase of 10.8 points, taking the indicator average for C4 girl at EL to 86.6. For C5 girls, the percentage had dropped from 87.3 to 81.1 from ML2 to EL.

Access to Protection Services

This section evaluates the C4 and C5 NFE girls' access to protection services across ML1 and EL (for the C4 girls) and across ML2 and EL (for the C5 NFE girls). We the channels through which they can report abuse, harassment, or exploitation, at school and the at the community.

School

Between ML1 and ML2, the evaluation team observed a sizable drop in the percentage of C4 NFE girls who responded "yes" when asked if they had any channel through which they can report cases of abuse, exploitation, or harassment that took place in a school setting. When comparing results from ML1 and EL among the panel of C4 NFE girls who were surveyed at both rounds and who remained enrolled in a learning programme at EL, we observe a significant decrease in the percentage of girls who believe they have reporting channels: at ML1, 93.5 percent of girls reported having some avenue to report abuses, whereas only 84.3 percent of girls claimed as much at EL. As with the ML2 round, the most commonly sought after channels are, in descending order, head teachers (66.8 percent of those claiming to have a reporting channel), teachers (49.2 percent), and "others" (13.5 percent), which primarily consists of answers referring to family members.

For C5 NFE girls, there is an increase in the percentage of girls reporting that they are able to report a case of abuse in the school to someone, but this increase of 6.5 points – from 76.6 percent at ML2 to 83.1 percent at EL – was not statistically significant. Like C4 NFE girls, the top three parties to whom C5 NFE girls would report a case at EL are: head teachers (60.9), teachers (48.4), and the "other" option (15.9), which primarily consists of family members. However, among C5 NFE girls, there is substantial regional variation in girls reporting to be able to report abuse cases to someone. For C5 girls in Bay region, there is a sharp decline of 14.5 points in the proportion of girls being able to report a case, whilst in Middle Shabelle, there was a 35.7 increase in the same metric.

One important trend to note across cohorts is that there is also a declining share of students who would report to head teachers and teachers, respectively, between ML1 and EL among C4 NFE girls. In the case of head teachers, a decline of 11.9 points was observed, while in the case of teachers, we note a decline of 27.8 points. Both changes were significant at the five percent level, while the latter is also significant at the one percent level.[143] As with C4 NFE girls, we also observe a sharp drop in the proportion of girls selecting "head teacher" as the person to whom they would report an abuse case. Among C5 NFE girls, this figure declined by 27.6 points between ML2 and EL, and this change is significant at the one percent level. It is unclear what contributed to declines in girls turning to teachers and head teachers to report abuse cases, as the qualitative data – specifically, the risk mapping and vignettes – do not indicate distrust of teachers and head teachers by girls. It is possible that the declining tendency to report to head teachers and teachers may be a product of the end of the AGES-specific NFE courses, which may lead girls to be relatively less willing to turn to teachers to report abuse cases in school.

Community

Between the ML1 and EL rounds, we observe a decline in the share of C4 NFE girls who said they are able to report abuse that had taken place in the community. The share of girls declined from 82.4 at ML1 to 76.6 at EL, which represents a significant decrease of 5.8 points. Somewhat more worryingly, the decline was heavily concentrated among girls who had any reported disabilities: while not reported to have disabilities had relatively unchanged scores on this metric (80.0 percent at ML2 compared to 77.3 percent at EL), girls with disabilities registered a much steeper and significant decline of 16.2 points over the same time period (90.2 percent at ML2 to 74.0 percent at EL). While these changes mean that an equal share of C5 girls living with and without disabilities are able to report cases of abuse occurring in the community, the decline among girls with disabilities might warrant further investigation to ensure that schools and other stakeholders continue to raise awareness and improve access to reporting mechanisms for girls with disabilities.

Among C4 girls, the most commonly selected options for whom they can report to are: "others" (41.0 percent), which mainly consists of family members, followed by head teachers (30.1) and teachers (23.0).[144] Another interesting divergence from abuse cases occurring in schools is that girls are much more likely to report cases to the police, whom 14.1 percent of C4 NFE girls cognisant of a reporting channel selected as one of the parties they would turn to.[145]

[143] Both coefficients remained significant even after controlling for region and girls' current enrolment status.
[144] The relative ranking of whom girls who are no longer enrolled in a learning programme would turn to remains the same. This may be because girls who had participated in an NFE programme still maintain a certain level of trust in their former instructors, or because any safeguarding mechanisms that have been shared during an NFE programme feature prominently in girls' thinking, even after they have completed the NFE course.
[145] Compared to 1.5 percent among C4 NFE girls who know of reporting mechanisms for abuse cases occurring in schools.

Among C5 NFE girls, we observe an increase in the share of girls who are able to report an abuse case occurring in the community to someone. At ML2, 66.8 percent of girls claimed as much, whereas at EL this figure had risen to 70.7 percent of girls, though this increase was not statistically significant. The latter girls most frequently cited "other" (35.3 percent), head teacher (35.3 percent), and teacher (29.1) when asked whom they would report abuse cases to. 11.8 percent of these respondents also cited the police, compared to 2.44 percent of girls when asked about abuse cases occurring in a school setting.

For both cohorts, the increased frequency of reporting to family members and to the police is unsurprising, given that girls may be more likely to see cases of abuse happening outside of school to be more appropriately dealt with by adults who do not work in the school. Still, the fact that a large proportion of these girls are still willing to report abuse cases to head teachers and teachers point to the important role that schools play in girls' safeguarding.

11.3. 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.

Income

We first assess changes in monthly income among the C4 and C5 cohorts of NFE girls. It is worth noting from the onset that the sample size of girls who were able to answer questions about their monthly income is small, and even more so when disaggregated by geographic zone. This is driven by two main factors. First, many girls simply do not know or are unable to give an estimate of their monthly income. We exclude such cases from our analysis. Second, we limit our analysis to girls for whom we have monthly income data in both the cohort baseline and in the EL round. For C4 girls, this means girls who were able to tell enumerators how much they earned in both the ML1 and EL round. For C5 girls, we include girls who were able to do so in the ML2 and EL rounds.

We present our results through two groupings. The first is the full panel of girls for each cohort, consisting of all girls who provided information during the cohort-specific baseline and the current EL round. The second take the same panel of girls, but excludes girls whose self-reported income between the cohort baseline and the EL had changed by $500 or greater. This includes both increases and decreases in self-reported income. The full results are highlighted in Table 87 below. In keeping with our evaluation of AGES in previous rounds, we make this methodological decision as a robustness check to prevent results from being driven primarily by outliers.

TABLE 87: MEAN MONTHLY INCOME AMONG C4 AND C5 NFE GIRLS, IN US DOLLARS

Zone	Cohort 4 NFE Girls				Cohort 5 NFE Girls			
	Count	ML1	EL	P-Value	Count	ML2	EL	P-Value
Full Income Panel								
Overall	262	18.32	66.67*	0.00	106	52.79	76.79	0.43

Banadir	113	24.50	66.13	0.07	61	71.66	50.66	0.52
South West State	116	17.20	49.56	0.09	45	27.22	112.21	0.10
Hirshabelle	33	1.09	128.69*	0.02	0	.	.	.
Income Panel, Excluding Outliers								
Overall	253	15.65	40.00*	0.01	100	32.16	41.81	0.50
Banadir	109	17.90	44.70	0.09	58	36.57	36.47	1.00
South West State	114	17.47	30.72	0.18	42	26.07	49.17	0.35
Hirshabelle	30	0.62	58.22	0.08	0	.	.	.

The first segment of the table highlights the results of the full panel, inclusive of outliers. Among C4 girls, we see a substantial growth of over 350.0 percent in the aggregate monthly incomes, and this change is statistically significant at the 5 percent level. Much of these gains are concentrated in Hirshabelle state, where average monthly income seemingly increased by over $125.00. However, the small sample size in Hirshabelle in particular indicates that caution is needed in interpreting these income changes. Among C5 girls, we see a more moderate aggregate increase of $24.00 in monthly income, though neither this figure nor the zone-specific changes in reported incomes are statistically significant. However, it is worth noting that the C5 NFE girls sample did not include any from Hirshabelle. To the extent that Hirshabelle (Middle Shabelle region) may be a zone where girls are improving their economic circumstances relative to other zones, data from our sample of C5 girls may underestimate actual changes in monthly income.

The second segment of the table excludes outliers. Consequently, and as expected, we observe more moderate increases across zones and in both cohorts. The gains among C4 girls in Hirshabelle, for instance, has nearly halved when excluding three outliers from the analysis, though the gains are still substantial. Likewise, the EL average income for C4 girls on aggregate still represents a large and significant increase of $24.35, but this is considerably less than the figure when outliers are included. These trends are mirrored among C5 NFE girls, for whom average monthly income only grew by $9.65 once outliers are excluded. Given the large changes in the data once a few outliers are excluded, we believe that the more conservative estimates are likelier to be representative of monthly income changes of C4 and C5 NFE girls, respectively.

Importantly, we note that there is an increase in the share of both C4 and C5 girls who reported earning no income.[146] Among C4 girls, there is an aggregate increase from 53.7 percent to 61.3 percent. This is largely driven by increases in South West State: where 50.0 of girls claimed to have no income in the past month, this figure has increased to 68.1 percent by EL.[147] Among C5 girls, there is an aggregate increase from 57.0 to 62.0 percent. This increase is largely driven by girls in Banadir, among whom the share of girls reporting

[146] This excludes the outliers.

[147] Much of this is driven by girls in Afgoye, as none had reported having no income during the ML2 evaluation, compared to 63.3 percent at EL. Baidoa, on the other hand, only registered a small increase from 67.4 percent to 69.8 percent. We note, however, that the small sample size of 30 girls in the panel for Afgoye necessitates caution in interpreting this result.

zero income had risen from 50.0 percent at ML2 to 58.6 percent at EL. Taken together, this suggests that incomes earned became more concentrated among some girls, as compared to each cohort's baseline.

At the same time, it is interesting to note that girls who reported no income at all during their respective cohort baselines made more gains in income compared to girls who already reported an income. Indeed, among C4 NFE girls who reported no income during ML1, their average gain in monthly income is $37.89, compared to an average increase of $8.61 among their counterparts. For C5 NFE girls, girls who reported no income during ML2 had an average gain of $31.16 in monthly income, compared to an average decrease of $18.88 among their counterparts. For both C4 and C5 girls, however, the average income among girls who earned no income during the cohort-specific baselines remains lower than the girls who did earn an income prior to EL.

In general, the small sample sizes and the numerous caveats needed when assessing self-reported income data warrant caution in generalising the above results to the AGES population of beneficiaries. Nonetheless, the data presented in this sub-section may be useful starting points for future programme management teams to monitor internally over the course of programme implementation.

Employment Outcomes

Turning now to girls' employment outcomes, we observe a significant decrease between ML1 and EL in the share of C4 NFE girls who are unemployed.[148] The share of C4 NFE girls unemployed at ML1 stood at 52.4 percent, whereas the share of unemployed C4 NFE girls was 31.3 percent at EL, representing a significant decrease of 20.6 percentage points over the course of two years. However, it is worth noting that the ML1 to EL change observed here is similar to the change between ML1 and ML2, where the share of unemployed girls among the C4 NFE cohort also dropped by 19 points. In other words, while substantial progress was observed between ML1 and ML2, it appears that there has been more minimal progress since then. Still, given the NFE programming's focus on improving girls' employment prospects, this remains an encouraging sign.

As is true during the ML2 evaluation round, the biggest increase in job category is seen with domestic work, where we observe a 14.8-point increase in the share of girls claiming to work in such occupations. This increase is significant at the 1 percent level. In addition to domestic work, we also observe statistically significant increases in the share of girls working as a sales or service worker, and those working in farming, fishing, or pastoralism.

TABLE 88: JOB CATEGORIES OF NFE GIRLS, BY COHORT

Job Category	Cohort 4 NFE Girls				Cohort 5 NFE Girls			
	ML1	EL	Diff.	P-Value	ML2	EL	Diff.	P-Value
No occupation	52.4	31.8	-20.6*	0.00	33.6	31.5	-2.1	0.65
Domestic Work	15.2	29.9	14.8*	0.00	28.3	30.6	2.3	0.66

[148] Note that this question was only asked to NFE girls who were 18 and above, or who were under 18 but has at one point been married. This means that not all NFE girls had this information collected from them. However, the proportion of NFE girls asked this question remains high, with 92.3 percent of C4 NFE girls (531/575) in the ML1 to EL panel, and 85.7 percent (317 out of 370) C5 NFE girls asked this question.

Unskilled sales/service worker	9.6	9.0	-0.5	0.83	9.3	6.9	-2.4	0.29
Student	17.0	12.2	-4.8	0.25	15.8	18.3	2.5	0.58
Sales/Service worker	0.8	5.3	4.5*	0.01	2.8	5.0	2.2	0.17
Trades, craft workers, extractive industries	1.9	1.9	0.0	0.97	2.8	0.3	-2.5*	0.04
Farming, fishing, pastoralism	1.6	4.5	2.9*	0.03	4.9	4.1	-0.8	0.68
Professional or managerial positions	0.0	0.6	0.6	0.18	0.4	0.9	0.5	0.50
Other	1.6	4.7	3.1*	0.01	2.0	2.2	0.2	0.87

Among the C5 NFE cohort of girls, the changes in employment outcomes are less clear cut. We observe a sizable but statistically non-significant increase in the share of girls who report to be unemployed. Similarly, other job categories do not register statistically significant changes in the proportion of girls working in said sector, with the exception of traders and craft workers, which saw a decline in the proportion of employed girls selecting it as an answer choice. Given that the C5 NFE cohort had completed their NFE programming relatively recently, this finding is not entirely surprising, as job seekers will likely take some time before finding employment. Additional time may be needed after the conclusion of their AGES NFE coursework before significant changes are observed in employment outcomes.

More encouragingly, however, is that the proportion of girls who claim to have a small businesses showed significant increases for each cohort since their cohort-specific baselines. Among C4 NFE girls, the proportion of business owners grew from 8.0 percent at ML1 to 20.5 at EL. Similarly, the proportion among C5 NFE girls grew from 12.1 percent at ML2 to 18.3 at EL. Both changes are significant at the five percent level, while the change among C4 girls was also significant at the one percent level. That a higher proportion of C5 NFE girls were able to start a business at EL is particularly encouraging given the comparatively short time elapsed since the conclusion of their NFE programme.

Needless to say, reading and numeracy skills are essential to be able to run a business, which explains a clear mechanism by which NFE participation helps with running a business. As one girl readily acknowledged:

> *Education plays a crucial role in enabling her to engage in business effectively. Being able to read, write, and perform calculations makes it easier for her to manage the finances and savings. Thus, she can contribute to her family's welfare through her business endeavours.*
>
> *- FGD (Vignettes Exercise) with Girls, Banadir, Int. 608*

Another possible avenue impacting girls' ability to run businesses is participation in savings groups, which the NFE programme encourages. Doing so may allow girls to have the initial seed capital to start a business. As one CEC member explains:

> *They (girls) are given books which are intended to be given to them by NFE. After that they are told to do savings called 'Ayuuto" which is to motivate them, and awareness is given to them. And then savings associations are made for them and then they learn about the savings, which they can use to make a small business such as sells things on small tables.*
>
> *- FGD with CEC members, Lower Shabelle, Int. 101*

Indeed, some respondents acknowledge that initial capital is often the one resource lacking that prevents girls from starting businesses.[149] For both C4 and C5 NFE girls, there was an increase in the share of respondents who claimed to have participated in savings groups: among C4 girls, the figure increased from 3.7 to 28.2 percent between ML1 and EL, while for C5 girls, the share grew from 10.3 percent to 23.5 percent. It is important to note, however, that the overall proportion of girls running businesses remains relatively small, and that in the qualitative data, few girls directly claimed to have established and run their own business. Nonetheless, the quantitative findings offer suggestive evidence that NFE programming has supported girls in this endeavour.

11.4. 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 EL across the following indicators: participation in youth groups, local political forums, and 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.

To assess covariation among participation in youth groups, local political forums, and service delivery improvement discussions, correlation and covariance matrices were utilised among the three outcomes (this was conducted for C4 NFE and C5 NFE cohorts separately). The data for all pairings of the three outcomes shows an overall positive but weak linear relationship between the three outcomes, indicating that C4 NFE girls are likely to have participated in more than one of these activities each.[150]

While participation more than doubled overall from ML1 to ML2, aside from participation in service delivery improvement discussions, the participation rate remained consistent from ML2 to EL, as shown in Table 89. Additionally, while participation in each and any aforementioned activities increased significantly in every state from ML1 to ML2, only one state observed a significant increase from ML2 to EL. While girls in Hirshabelle had the highest level of participation in any activity at EL, the participation rate was relatively constant between ML2 and EL (increasing only from 71.8 to 75.0 percent). Even then, while engagement increased in youth groups and service delivery improvement discussions, participation in local governance discussions in Hirshabelle decreased substantially between ML2 and EL from 40.8 to 29.8 percent. Additionally, the overall rate of participation decreased in Banadir from 64.4 percent in ML2 to 55.9 in EL.

[149] FGD with mothers, Bay, Int. 202
[150] covar ≈ 0.09 - 0.12; r ≈ 0.50 - 0.63

Only the overall participation rate in South West State increased substantially in the past year, increasing from 53.8 percent in ML2 to 70.8 percent in EL.

TABLE 89: PARTICIPATION IN COMMUNITY ACTIVITIES AND DISCUSSIONS, BY ROUND FOR C4 NFE GIRLS

Community Activity	Participating in ML1 (%)	Participating in ML2 (%)	Participating in EL (%)
Youth groups or networks	14.8	32.2	33.4
Service delivery improvement discussions	14.1	38.4	45.2
Local governance discussions	8.9	28.4	30.6
Received humanitarian assistance	19.8	30.1	30.6
Any of the above	32.5	61.4	64.3

Humanitarian assistance receipt overall held relatively constant between ML2 and EL; surprisingly, only marginal changes (either in rise or decline) in the rate of receiving humanitarian assistance were observed between these rounds. Even with the consistent rates between ML2 and EL in all states, we observe notable differences between states at EL. Though girls in Banadir and South West State maintain a relatively similar rate of humanitarian assistance (29.8 percent and 28.3 percent, respectively), they are both approximately 10 points below the rate of humanitarian assistance in Hirshabelle (39.3 percent) at EL.

The rate of participation – both the doubling of participation rates between ML1 and ML2 and the rate constancy between ML2 and EL – similarly mirrored the participation rates of the C4 NFE girls identifying as IDPs. Out of the three activities, the rate of engagement with service delivery improvement discussions observed the largest increase among IDPs, with the participation rate among IDPs increasing from 36.4 to 43.5 percent from ML2 to EL. Receipt of humanitarian assistance saw minimal change among the IDP subgroup, only decreasing marginally from 30.4 to 29.4 percent.

We also observed nearly unchanged overall participation rates between ML2 and EL among C4 NFE girls with any disabilities (across all alternative codings). Youth group participation had the biggest decline from 37.0 percent in ML2 to 30.1 percent in EL (original coding), followed by service delivery improvement discussion participation (43.5 to 41.9 percent from any disability, original coding); only the local governance discussion engagement rate increased between rounds (31.5 to 35.3 percent any disability, original coding). Looking at specific disability trends, participation rates in community activities maintained a continuous growth from ML1 to EL, going from 20.0 percent in ML1 to 64.7 percent in ML2 and reaching 80.0 percent in EL (original coding); for participation rates among girls with mental health disabilities, from ML1 to ML2, the rate went up from 29.8 to 63.5 percent, but from ML2 to EL, the rate only went up from 63.5 to 64.4 percent (original coding). It is important to note the low sample size count for girls with physical disabilities,

so the trend observed over multiple rounds is unlikely to be representative of the trend of participation rates among the subgroup within the overall CARE AGES population.[151]

Looking at humanitarian assistance received among those with disabilities, the rate of reception among girls with any disabilities stayed relatively constant from ML2 to EL among all coding methods (28.7 to 27.9 percent via original coding; 28.5 to 26.5 via alternative coding 1; 30.6 to 30.0 via alternative coding 2). The rates of receiving humanitarian assistance among girls with cognitive or mental health disabilities also remained relatively constant between these rounds, while the rate among girls with physical disabilities increased from 17.6 to 25.0 percent.

In terms of socioeconomic status (SES), from ML1 to ML2, overall participation rates approximately doubled for girls reporting poor roof quality at their house (28.3 to 60.2 percent), no food availability for at least 10 days (27.3 to 65.9 percent), and their HoH not earning a wage (28.1 to 51.4 percent). However, from ML2 to EL, while the rates of those with poor roof quality and whose HoH is unemployed increased substantially to 71.0 and 68.5 percent, respectively, those who reported no food availability for at least 10 days saw a decline from 65.9 to 61.8 percent by EL. Additionally, the rate of participation in youth groups shrank drastically among those reporting no food availability in the past year, going from 41.2 percent at ML2 to 30.9 percent by EL.

However, all three SES subgroups increased their reception of humanitarian assistance substantially, with the proportion of girls with poor roof quality increasing from 28.3 percent in ML2 to 34.8 percent in EL, those who reported no food availability for at least 10 days from 29.4 to 36.4 percent, and those whose HoH does not earn a wage from 24.3 to 37.1 percent.

Despite this growth in humanitarian assistance reception, the analysis among the C4 NFE cohort highlighted slower growth in other indicators of social support, showcasing an overall stagnating progression in the provision of social support between ML2 (2023) and EL (2024). This may be in part due to the substantial gains from 2022 to 2023 that reached the more accessible girls, whereas more time and effort may be needed to access the remaining, 'hard-to-reach' marginalised girls.

Cohort 5 NFE Girls

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 the comparison between ML2 and EL whether these girls were receiving humanitarian assistance, as well as evaluate the influence and degree of variables such as IDP status and household economic status.

As with the C4 NFE girls, in terms of covariation among the three activity engagement outcomes, statistical tests show an overall positive but weak linear relationship between the three outcomes, indicating the girls are likely to have participated in more than one of these activities each.[152]

Unlike with the C4 NFE cohort, the overall participation rate remains similar across ML2 and EL rounds, with the exception of the participation rate of service delivery improvement discussions, which increased from 36.0 percent at ML2 to 45.4 percent at EL. This trend is reflected among the rates reflected in most of

[151] The sample size for girls with physical disabilities (original coding) was 20 in ML1, 17 in ML2, and 20 in EL, far fewer compared to the sample size of girls with mental health disabilities (original coding), which was 104 in ML1, 85 in ML2, and 104 in EL.
[152] covar ≈ 0.10 – 0.13; r ≈ 0.47 – 0.57

the subgroups analysed for the C5 NFE cohort. Humanitarian assistance reception remains the same however, only going from 29.8 percent in ML2 to 30.5 in EL.

Regionally, while the rate of participation in any of the community activities went up substantially among girls from South West State (from 53.8 to 69.9 percent) and Hirshabelle (58.1 to 67.7 percent) between ML2 and EL, the rate among girls from Banadir actually decreased from 64.8 to 59.4 percent. Notably, the rate of engagement in service delivery improvement discussions among girls in South West State increased the most between rounds, going from 29.4 percent to 51.7 percent, while the rate for this discussion among the other states remained relatively constant. In terms of receiving humanitarian assistance, only the rate among girls in Hirshabelle increased substantially, going from 35.5 percent in ML2 to 45.2 percent in EL; the rates from girls in the other states remained relatively constant.

TABLE 90: PARTICIPATION IN COMMUNITY ACTIVITIES AND DISCUSSIONS, BY ROUND FOR C5 NFE GIRLS

Community Activity	Participating in ML2 (%)	Participating in EL (%)
Youth groups or networks	30.1	33.5
Service delivery improvement discussions	36.0	45.4
Local governance discussions	26.0	30.5
Received humanitarian assistance	29.8	30.5
Any of the above	59.5	64.9

In terms of IDP status, the rates of participation mirror that of the overall cohort including the substantial increase in participation in service delivery discourse from 40.1 percent at ML2 to 50.3 in EL. Humanitarian assistance rates declined from 32.0 percent in ML2 to 29.3 percent in EL.

Rates of participation among those with any disabilities (in the original coding) did deviate from the trend in the overall sample, with the rates of youth group participation and service delivery participation increasing by approximately 7 to 8 percentage points, respectively (25.8 to 33.1 percent for youth group participation; 35.2 to 43.8 percent for service delivery improvement). However, the rate of humanitarian assistance distribution decreased from 32.7 percent to 28.8 percent. These trends are mirrored among the alternative coding schemes for girls with any disabilities in the C5 NFE cohort. Looking at the different disability types, service delivery improvement discourse participation decreased between ML2 and EL among girls with a physical disability (alternative coding) from 42.3 to 38.5 percent, while the rates among those with cognitive (38.5 to 46.2 percent, original coding; 41.2 to 44.1 percent, alt. coding) and mental disabilities (36.7 to 43.6 percent, original coding; 37.6 to 45.2 percent, alt. coding) increased to varying degrees. In terms of humanitarian assistance distributions, rates among girls with physical disabilities (34.6 to 23.1 percent, alt. coding) and cognitive disabilities (34.6 to 19.2 percent, original coding; 35.3 to 17.6 percent, alt. coding) drastically decreased between ML2 and EL. It is important to note and take into consideration that the sample sizes for the physical and cognitive disability subgroups were substantially lower compared to that for the mental disability subgroup (between 20 and 34 for the former, compared to 139 to 165 for the latter), so the magnitude of the differences between rounds for the former two are most likely larger and not as representative of the true difference in participation between ML2 and EL.

In terms of socioeconomic status indicators, notably, the rate of participation in service delivery discussions declined between rounds among girls who reported having poor quality roofing (from 48.2 to 43.4 percent), while the rate among girls who reported having no food availability for at least 10 days and girls whose HoH does not earn any wages increased from 40.9 to 57.6 percent and from 25.0 to 45.5 percent, respectively. Participation rates for local governance discussions among girls who reported having no food availability for at least 10 days increased from 31.8 percent in ML2 to 47.0 percent at EL. It is important to note that the sample size kept into consideration for those whose HoH is unemployed is low,[153] so the generalization of the participation rate changes between rounds to the C5 NFE cohort population should be taken with caution.

Reception rate of humanitarian assistance among those with a poor roof quality in their home reduced from 33.7 to 22.9 percent from ML2 to EL, while the rate among those who had little to no food availability for at least 10 days increased from 21.2 to 37.9 percent.

While community activity participation remains steady in terms of reaching more girls between 2023 and 2024, the analysis – particularly among those with disabilities, IDP status, and indicators of low SES like poor roof quality – reveals that humanitarian assistance reception waned among the more marginalised girls in the cohort in the past year. This might indicate that the effort needed to provide key humanitarian assistance was likely insufficient to reach more marginalised girls, as girls from relatively better backgrounds were more able to access the assistance.

12. Value for Money

Previous sections have evaluated the impact of the AGES program in meeting its original objectives, such as improving girls' learning and life skills, enhancing the practices and attitudes of teachers and caregivers, and fostering better school management. While the data indicate that the program was generally successful in achieving many of its goals, it remains unclear whether it provided value for money by being cost-effective: Did the intervention deliver the expected results while minimizing costs?[154]

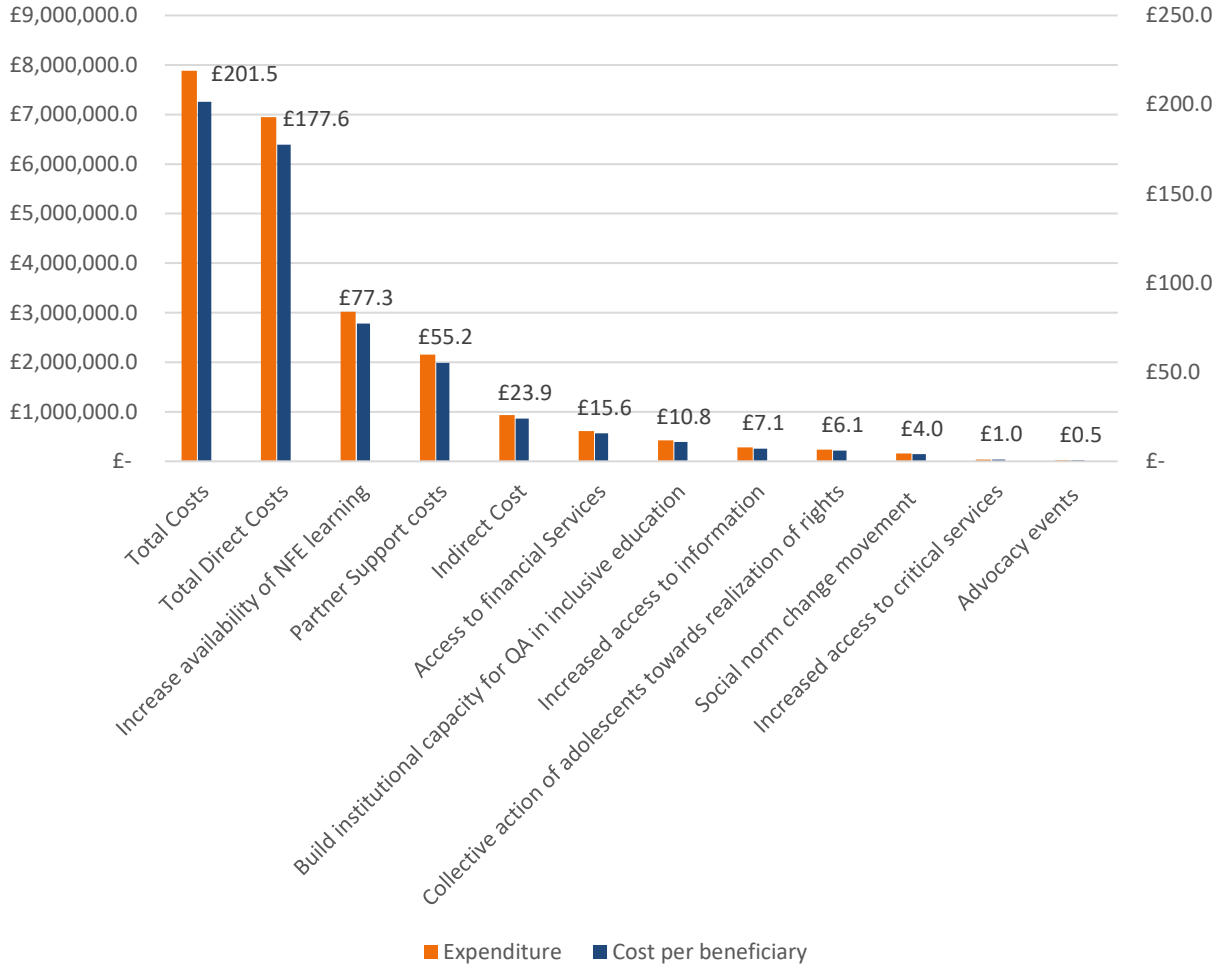
USAID funded CARE AGES activities are all targeting out of school girls and not supporting formal education but only NFE learning programs.

The figure below shows the breakdown of the cost of the program disaggregated by type of activities.

[153] HoH earns no wage, n=44

[154] The VfM analysis provided in this section does not include M&E and CA costs as they are funded by both FCDO and USAID and their source cannot be derived from the budget data available. It is also important to note that the lack of quality data on cost per student in Somalia does not allow for a comparison with national benchmarks.

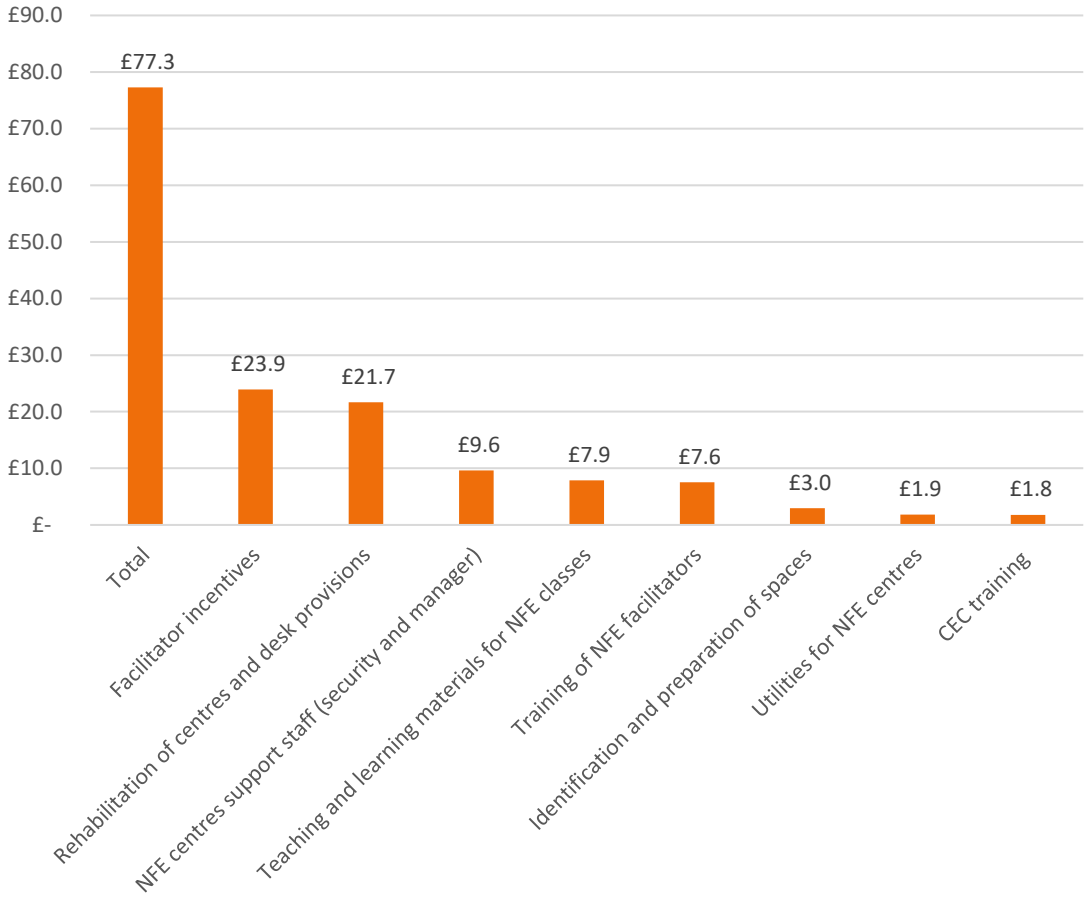
FIGURE 16: COST BREAKDOWN FOR USAID FUNDED ACTIVITIES



The total direct cost per girl of USAID funded activities was £177.6 (£201.5 when including indirect costs). The largest share of the budget was dedicated to activities directly related to the provision of the NFE curriculum (£77.3 per girl, the 43.5% of all direct costs), and a significant share (£55.2 – the 31.1%) was expended for partner support costs (mainly program staff salaries).

The direct costs related to the provision of the NFE curriculum can be further disaggregated to better understand the main cost drivers. The figure below presents the breakdown of costs related to activities aimed at increasing availability of NFE learning.

FIGURE 17: BREAKDOWN OF COSTS RELATED TO THE PROVISION OF NFE LEARNING



As shown in the figure above, the main cost drivers are facilitator incentives (£23.9 per girl, the 30.9%) and the provision of NFE classes (£21.7 per girl, the 28.1%). The latter includes costs for the construction and rehabilitation of classrooms (£12.6 per girl) and the provision of desks (£9.1 per girl).[155] Other significant cost drivers include salaries for support staff at NFE centres (£9.6 per girl), learning materials (£7.9 per girl), and the training of NFE facilitators (£7.6 per girl).

After calculating the costs per beneficiary, we assess how these expenditures translated into improvements in the three key outcomes targeted by the program (learning outcomes, positive transitions, and the sustainability of interventions) by recalling findings reported in previous sections.

In terms of learning outcomes, girls enrolled in the NFE program from both Cohort 4 and Cohort 5 showed significant improvements since the program's inception. Specifically, NFE Cohort 4 girls gained 28.8 points in numeracy and 28.2 points in literacy, while NFE Cohort 5 girls improved by 17.7 points in numeracy and 18.3 points in literacy.

[155] The cost per girl of desks for USAID NFE classes was the approximately half than the cost of desk for ABE spaces funded by the FCDO.

The program also succeeded in supporting positive transitions for girls participating in NFE programs. More than half of the beneficiaries assessed at the endline achieved successful transitions, with 38.1% of Cohort 4 and 31.4% of Cohort 5 NFE girls moving into employment, and 19.7% of Cohort 4 and 21.1% of Cohort 5 NFE girls progressing to formal education.

Finally, the program effectively enhanced social support for girls, as evidenced by significant increases in participation in community activities and discussions among both Cohort 4 and Cohort 5 NFE girls.

13. Recommendations

Although the AGES programme has concluded, in this brief section, the evaluation team offers several recommendations that may be help in the design of future, successor programmes to AGES, or to inform programme closeout actions.

Continuing Education

An intuitive finding, further substantiated in this evaluation, is that girls who continue their enrolment in an education programme are likelier to sustain their learning gains, and further develop their numeracy and literacy skills. This trend is clearly seen across almost all cohorts examined under this study, where girls who remain enrolled in some kind of learning programme outperform their counterparts who had left school. Short of expanding successor programmes to retain more girls for longer periods of time, 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 help sustain learning gains made over the course of a programme. Continuing education could be delivered virtually or remotely; alternatively, in-person sessions could be an effective way to sustain numeracy and literacy skills.

Tailoring Instructional Levels and Class Materials

Analysis of learning assessment data revealed that there is substantial variation across geographic zones/regions in learning gains, with the Banadir region in particular experiencing slower growth than other regions. In part, this may be due to the higher share of Banadir girls who have moved on to employment, compared to other zones, and thus no longer pursue an education (the differential employment rates are most evident in the C4 and C5 NFE cohorts). In comparing C4 and C5 NFE cohorts, we also find that the former, starting from lower base scores, made much more substantial gains in learning compared to C5 girls, whose cohort baseline scores were higher on average. These findings together suggest a need for more tailored approaches for girls based on context. For instance, teaching materials and programme foci may have been ill-suited for those girls who were comparatively high-achieving at the outset of their learning programme. In addition, and in relation to the previous recommendation, programmes that are light touch and can be done with less time spent per day may be needed in areas, such as Banadir, where girls are likelier to seek and find employment, and thus may not have as much time to devote to their studies.

Adapting to Girls' Capacity to Physically Attend Classes

The evaluation highlights trends in girl's attendance and household chores that may warrant further monitoring. When examining the effect of household chores on girls' attendance in school among Cohort 1 girls, we find that there are significant increases in both time spent doing household chores, and, for ABE girls in particular, the likelihood that chores prevent girls from physically attending class. Surveys with caregivers also reveal some backsliding in caregiver's attitudes towards girls' education, as the share of caregivers who believed that it is justifiable for girls to not attend school if they need to help with household chores has increased substantially since ML1, and has reverted to a level similar to BL. The factors behind these trends, including poverty and increasing household work burdens as girls increase in age, are likely difficult for the AGES programme to address fully. As such, future programming may benefit from ensuring that learning curricula are sufficiently adaptive to girls' changing abilities to attend class. For instance, introducing more modules that are conducive to self-study, coupled with policies to allow class materials to be brought home, may help ensure learning continues even when girls are not physically in class. This would be particularly relevant for older FE girls, as well as ABE and NFE girls in general.

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, which is particularly evident in the declining shares of caregivers who claim that schools being too expensive is a valid reason for girls to not attend school. At the same time, girls in ABE and NFE programs (for Cohort 1) are now more likely to have attendance affected by their household and/or work responsibilities. The progress already made can be maintained and deepened through relatively low-cost awareness campaigns in future programmes.

Sharing Teacher Experience, Expertise, and Materials

The report has documented progress in several, but not all, teaching practices. This increased uptake and implementation of positive 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). Doing so may help instructors better understand how to implement these practices, while also building their capacity to implement additional ones that they do not currently employ in the classroom. 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.

Improved Measures on Classroom Gender Equality and Equity

The report indicates that there is a mismatch in girls' responses to gender equality questions, as they were less likely to disagree when asked more generally about whether girls and boys are treated differently, but there is an increasing share of girls who claim that they observed equal treatment of boys and girls when probed about specific practices in the classroom. This suggests that the specific gender equality items in the survey may not fully capture the range of teaching practices, some of which may be unequally implemented for boys and girls. Conceptually, there is also a question of whether the tool questions adequately address gender equity. For instance, is it not immediately clear whether the aforementioned unequal treatment of boys and girls, which girls are more likely to report at EL, entails increased additional support to girls, bearing in mind that girl learners face additional barriers to education that boys are less likely to encounter in the Somali context. To the extent that it does, the finding may not necessarily be a negative one. In future evaluations, this element of teaching practices could be expanded upon to provide more nuanced insights on teaching practices.

Continuous Monitoring and Reducing of Corporal Punishment Use

The analysis of teaching practices in formal schools highlighted that the AGES programme has made strides in reducing the use of corporal punishment on learners, as evident in drastic decreases in both observed use of corporal punishment (which we previously verified through class observations) and girls' reports on. However, since ML1, there is a resurgence in the share of girls who witnessed corporal punishment being used, while girls' reports indicate that, among teachers who punish students for providing wrong answers, corporal punishment is increasingly becoming used as the mode of punishment. These findings suggest that the initial drop in corporal punishment use between BL and ML1 have seen some reversals. Consequently, continued monitoring of corporal punishment use – including after the end of the programme, possibly

through further empowering CECs – could be an important step in ensuring that the initial reductions are sustained.

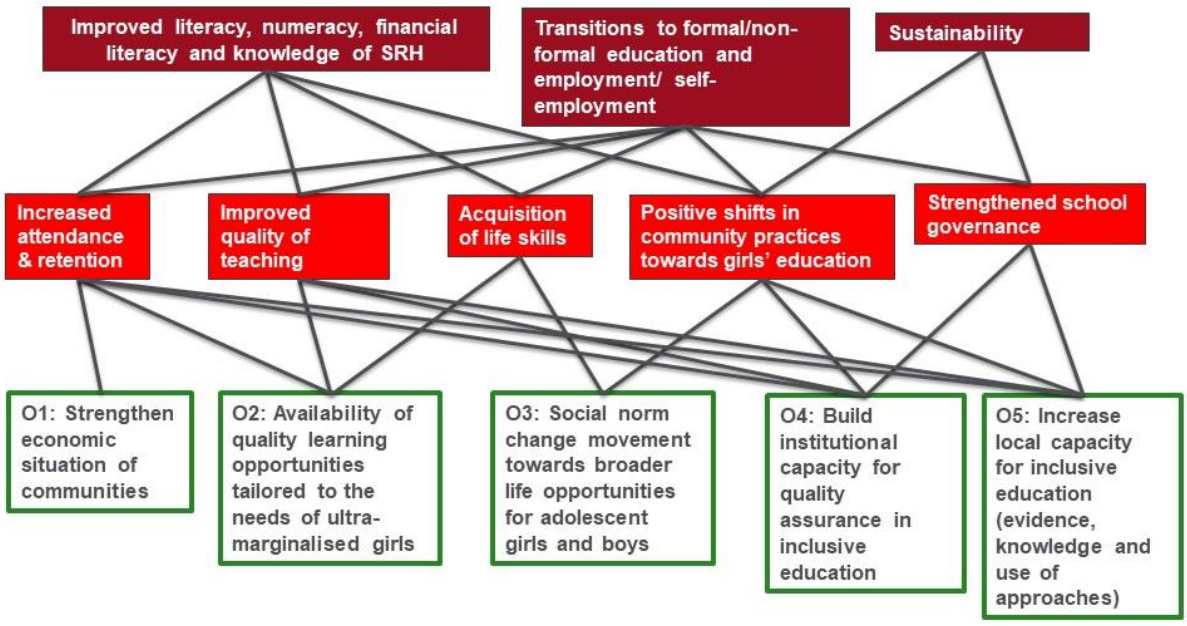
Annexes

Annex 1 - AGES Programme 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 marginalised 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 (ToC) 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 with improved institutional capacity for quality assurance and increased local capacity for inclusive education.

FIGURE 18: AGES PROGRAMME THEORY OF CHANGE



Annex 2 - Evaluation Methodology

This annex outlines the AGES Endline (EL) 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 EL assessment fits within the overarching, multi-year longitudinal evaluation design, which encompasses several rounds of assessment.

The primary objective of the AGES EL 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, ML1, or ML2 rounds. Our choice of temporal comparison – whether to the BL, ML1, or ML2 – is driven, first, by whether the cohort's initial participation was at the baseline (ABE, FE, and C1 NFE girls), at ML1 (C4 NFE girls), or at ML2 (C5 NFE girls). In the case of girls originally recruited at the baseline, we make comparisons from the current round backward to the BL, ML1, and ML2 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 (i.e., mothers, teachers, CECs), 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 considering 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 and evaluation 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 EL 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. 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 programme intervention. 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 endline 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 EL and BL represents the project's naïve impact on learning scores. The difference between 11- and 15-year-old girls (reflecting the 4-year gap between BL and EL) at baseline serves as the benchmark, which is used to assess the naïve change in scores. If the gains between rounds (as a girl transitions from 11 to 15 years old) exceed the differences observed between 11- and 15-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 and since ML1 fieldwork. Aside from these changes, no further schools were removed at EL. However, as there may have been changes in teachers over the past four 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 endline 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 exclude 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, even though it was not participating in the programme. As such, it was excluded from both ML2 and EL evaluation.

A key distinction between the ML2 and EL 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 ML2 round were purposefully brought back into the sample – where they could be located – during EL. 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 2,951 girls. 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 and ML2 rounds. This explains an 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, we successfully recontacted 332 of them and replaced 97. This also applies to the ML1 USAID cohort – C4 NFE girls – with ML1 respondents who were not successfully re-contacted during the ML2 round. 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 and ML2. To be clear, these numbers are indicative of the evaluation team's completion of surveys, not of successful re-contact.

The second and third columns in the table indicate the achieved sample size across cohort groups and the completion rate, respectively. These are the number of interviews completed, during EL, with each cohort; the panel samples available for analysis are typically smaller, for the reasons outlined above.

TABLE 91: 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	566	359	63.42%
ABE Girls	468	263	56.20%
Cohort 1 (C1) NFE Girls	490	263	53.67%
Cohort 4 (C4) NFE Girls	916	575	62.77%
Cohort 5 (C5) NFE Girls	511	374	73.18%

Table 92 expands on the discussion of sample targets by addressing the school-level tools for formal schools, the achieved sample for each, and their respective targets. Because the earlier cohorts of NFE girls (C1, C4, and C5) and ABE girls participated in learning programmes that have now ended, school-level data collection did not take place in their centres. Because one formal school in Howlwadaag (HOW05) was found permanently closed by the evaluation teams, only 36 school surveys were able to be completed, falling short of the sample 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. Without knowing, *a priori,* the number of classrooms in each school, it is not possible to pre-specify a sample target.

TABLE 92: SAMPLE TARGETS AND ACHIEVED SAMPLE FOR SCHOOL-LEVEL TOOLS

Cohort or Girl Type	Sample Target	Achieved Sample	Completion Rate (Share of Target, %)
Formal Schools			
Headcount	N/A	163	N/A
School survey	37	36	97.3%

The EL assessment, as was true of previous rounds, 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 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

Unlike the previous rounds, if the recontact procedure for the girls fails and the enumerator marked the girl as "not found", no replacement procedure would take place since this is the final assessment.

In many instances, girls were unable to be contacted 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

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 EL. 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 EL. 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-EL panel that excludes data from ML1, rather than relying on a smaller set of girls for whom data is available across all four rounds.

Broadly speaking, our preference is to maximize sample size for BL-to-EL 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-EL changes, since ML1-to-EL and ML2-to-EL 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-EL panels, supplementing our analysis by looking at the "full panel" samples (BL-ML1-ML2-EL panel) in some cases, where we want to understand something specific about the trends between rounds, rather than the aggregate trend from BL to EL.

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 considered 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 72 qualitative interviews, with the following breakdown by respondent type:

- 12 FGDs with CEC members
- 12 FGDs with teachers
- 12 FGDs with mothers
- 12 vignette FGDs with girls
- 12 risk mapping FGDs with girls
- 12 KIIs with religious leaders

Fieldwork

In this section, we describe critical aspects of data collection and discuss how the data was analysed. Regarding data collection, this section includes details on the number and reasons for school-level removals (i.e. replacement and 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

17 data collection teams were deployed, each consisting of one team leader and between four and five enumerators. Enumerators and team leaders participated in a 6-day training, which included a 1-day pilot. A total of 99 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 17 team leaders and 65 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. Five of the team leaders were female; as at ML2, 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.

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 eight schools in 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.

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 (QC) tracking tool was developed and used daily 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 QC checks were accompanied by daily data cleaning. Any inconsistencies or mistakes were discussed with the field teams and, if necessary, corrected in the data.

Continuing from ML2, audio-recorded interviews were implemented this round for further quality assurance. We generally record interviews and have a QC 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. 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 specific enumerators or interviews, we were able to review audio transcripts to correct the data, identify the source of data issues, etc.

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.

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 analysing 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. 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

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.

Cooperation Issues with Head Teachers

The evaluation teams encountered cooperation issues with the head teachers and principals for a number of schools. One school in Afgoye refused to interact with the team due to repeated INGO interventions with the school, resulting in contacting the girls for interviews outside of school. Two schools in Baidoa were not cooperative in supporting the mobilization for qualitative interviews with key stakeholders, resulting in delays to qualitative data collection. Another school in Balcad had a head teacher who cooperated at a low level with

the team, resulting in request for support from CARE to ensure better cooperation, with the ordeal delaying data collection.

Examinations and School Closures

Examinations in several schools limited the data collection process, including school-level data for formal schools, resulting in delays with the teams visiting the schools.

Permanent school closures in Kismayo, Howlwadaag, Danyile, and Baidoa increased the difficulty for the teams to contact the girls. For one school in Kismayo, in particular, due to the school's permanent closure via structural collapse, recontacting efforts were hampered as the head teacher noted their lack of knowledge about the location of the girls, stating that they were transferred to other schools, got married, or were in a position where they could not be located.

Flight Delays and Accessibility Issues

The team experienced a series of delays during fieldwork due to flight delays and cancellations, impacting the teams working in Mogadishu, Jowhar, Baardheere, and Dhoobley. Due to a series of unexpected flight cancellations for one team, data collection for this team continued past the expected data collection deadline.

Incidents of Data Manipulation, Fraudulent Interviewees, and Double-Registration

Incidents of data manipulations – including headteachers providing fake girls (i.e., trained to provide information about the target AGES participants), girls whose caregiver double-registered them into CARE AGES for the purpose of perceived additional benefits, and girls sending proxies to complete the interview for them – were identified and reported by field staff to the core team in real-time, which were promptly investigated and reported to CARE Somalia.

Annex 3 - Learning and Transition Beneficiaries Tables

The below tables highlight the number of girls, disaggregated by cohort, who meet the designated learning improvement and transition criteria.

TABLE 93: NUMBER OF LEARNERS WITH IMPROVED LITERACY AND NUMERACY SCORES

Cohort	Evidence line 1: Mean score change from BL	Statistically significant (five percent level) improvement since BL	Total # girls in cohort	Total # of girls with improved scores
Formal Education (FE)	Lit: +34.2 Num: +19.9	Yes	20,452	Lit: 16,873 (82.5%) Num: 15,748 (77.1%)
ABE	Lit: +22 Num: +10.7	Yes	13,276	Lit: 8,882 (66.9%) Num: 8,629 (65.8%)
C1 NFE	Lit: +3.0 Num: +0.7	No	13,439	Lit: 5,792 (43.1%) Num: 6,393 (46.9%)

C4 NFE	Lit: +28.2 Num: +28.8	Yes	13,439	Lit: 9,359 (68.9%) Num: 10,980 (81.7%)
C5 NFE	Lit: +18.3 Num: +17.7	Yes	12,942	Lit: 8,710 (67.3%) Num: 9,616 (74.3%)
Total girls with improved learning[156]		FE + ABE + C4 NFE + C5 NFE		Lit: 49,616 Num: 51,366

TABLE 94: NUMBER OF GIRLS IN VARIOUS TRANSITION OUTCOMES[157]

Transition Pathway	Target # of Girls	Estimated Actual # of Girls[158]
Formal Education		
Retention in FE program	**17, 385**	15,386
Progressed adequately in grade level	*N/A*	*10,867*
Did not progress adequately in grade level	*N/A*	*4,519*
Skills or Vocational Training		0
Employed		740
Self-Employed		103
Underage Employment		103
Out of School, Idle		4,519
ABE		
Transition to FE, higher level ABE, or vocational training	11,285	5,602
In ABE, adequate advancement	*N/A*	*358*
In formal school, not at a more advance level than ABE equivalent	*N/A*	*4,740*
In formal school, at a more advanced grade than ABE equivalent	*N/A*	*451*
Vocational training	*N/A*	*53*
In NFE		451
In ABE, Inadequate Advancement		199
Employed		3,332
Self-Employed		53

[156] This will be reported in the AR submitted to FCDO.
[157] All decimals are rounded up to the closest whole number
[158]Calculated as percentage of EL evaluation sample in each pathway (see also respective cohort transition sections in the main report), multiplied by the number of girls in each cohort (see Table 93: Number of learners with improved literacy and numeracy scores)

Underage Employment		106
Out of School, Idle		3,531
C1 NFE		
In NFE		1,438
In ABE		54
In Formal School		1,949
In Vocational Training		54
Employed or Self-Employed	5,376	6,262
Employed	*4,032*	*5,953*
Self-Employed	*N/A*	*309*
Out of School, Idle		3,749
C4 NFE		
In NFE		2,083
In ABE		67
In Formal School		2,782
In Vocational Training		27
Employed or Self-Employed	5,376	5,066
Employed	*4,032*	*4,717*
Self-Employed	*N/A*	*349*
Out of School, Idle		3,414
C5 NFE		
In NFE		2,834
In ABE		142
In Formal School		2,873
In Vocational Training		181
Employed or Self-Employed	5,177	3,883
Employed	*3,882*	*3,637*
Self-Employed	*N/A*	*246*
Out of School, Idle		3,041

Annex 4 - Characteristics and Barriers

Sample Composition of Characteristics and Barriers

This section's tables present the profiles and educational obstacles encountered by girls in the five AGES cohorts. Each group is detailed through two tables — one for sample traits like the proportion of girls from female-headed home, and another for the proportion of girls facing particular educational barriers. Both baseline (BL) and endline (EL) data is included, showing the prevalence of each trait or obstacle within the cohort.

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 present data on changes that occurred within the same girl over time, such as a girl who was single at BL but is married at EL. 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 EL values we report are the share of girls in the EL sample (panel sample) who fell into a given subgroup at their cohort-specific 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 EL calculation, this outcome is assessed at BL. Among the baseline sample, 10.7 percent of girls had a single living parent. For the EL 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 EL had a single living parent at BL. This share has fallen to 9.4 percent of the sample due to attrition. To be clear, the 21 single-parent girls in the EL 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 95: CHARACTERISTICS OF FE COHORT AT BL AND EL

	Baseline		Endline		
Subgroup	**Pct of Total**	**N**	**Pct of Total**	**N**	**Variable Name & Source**
Household Characteristics					
Girl has only one living parent	11	45	9	21	orphan_single
Girl has no living parents	1	3	0	0	orphan_double

Girl does not live with either parent in her HH	4	15	2	5	no_parents
Girl is currently married	0	1	0	1	married
Girl has ever been married	0	1	0	1	ever_married
Girl is a mother and is under 16 years of age	0	0	0	0	mother_16
Girl is a mother and is under 18 years of age	0	0	0	0	mother_18
Female-headed household	35	147	30	66	female_hoh
Parental Education Background					
HoH has no education of any kind (no Quranic)	22	93	21	46	hoh_noeduc
HoH has no formal education	81	340	79	177	hoh_noformal
Caregiver has no education of any kind (no Quranic)	29	122	30	68	cg_noeduc
Caregiver has no formal education	86	361	87	193	cg_noformal
Household Economic Status					
HH has a poor-quality roof	18	76	15	34	poor_roof
HH went to sleep hungry most nights, last 12 months	7	29	5	11	nofood_most
HH went without clean water most days, last 12 months	5	20	4	8	nowater_most
HH went without medicine most days, last 12 months	14	61	14	32	nomeds_most
HH went without cash income most days, last 12 months	14	57	11	24	nocash_most
HH owns lands	38	160	42	93	owns_land
Caregiver has savings of some form	3	11	3	6	savings
HH owns a phone	85	357	84	188	owns_phone
HH owns a smartphone	24	87	27	51	owns_smartphone
HoH does not earn a regular wage	37	156	32	72	hoh_nowage
Household head is engaged in pastoralism	1	6	0	1	pastoral
Disability Status					
Vision disability	0	0	0	0	disab_vis
Hearing disability	0	0	0	0	disab_hear
Disability the impedes mobility	0	0	0	0	disab_mob
Disability of the arms/hands	0	0	0	0	disab_arms

Disability that impedes self-care	0	0	0	0	disab_selfcare
Disability that impedes communication	0	0	0	0	disab_comm
Cognitive disability	0	0	0	0	disab_cog
Behavioral disability	0	0	0	0	disab_behavior
Mental health disability	0	0	0	0	disab_mh
Mental health disability, alternative coding	12	49	12	26	disab_mh_alt1
Physical disability, any type	0	0	0	0	disab_phys
Physical disability, any type, alternative coding	1	3	1	2	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	0	0	0	0	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	0	2	0	1	disab_ccb_alt1
Displacement and Language					
Household are IDPs	39	143	40	87	idp
Household speaks af-Maay	28	117	26	58	maay
School Facilities					
Girl will not use drinking facilities at school	17	72	17	38	wontuse_drinking
Girl will not use toilet facilities at school	9	37	10	23	wontuse_toilet
No computers available for use at school	94	396	96	213	no_computers
Girl cannot use books/learning materials at school	30	126	33	73	no_materials
Not enough seats for every student in class	9	36	9	20	no_seats
Textbooks are shared between students	40	168	46	103	shared_textbooks
Girls are able to take textbooks/materials home at night	35	146	35	79	materials_night
School has reliable electricity	67	281	63	140	electric
School has water access within 1 km	55	230	56	125	water
School has only cement floors (no dirt)	68	285	68	152	cement_floor
School has separate toilets for girls	71	301	68	151	separate_toilets

School provides at least one meal for students	29	120	30	68	school_feeding
School Environment and Teacher Behaviours					
Teachers do not make girl feel welcome at school	14	57	13	30	unwelcome
Teachers are often absent from class	30	125	30	67	teacher_absent
Teacher rarely or never encourages participation	8	35	10	22	teacher_participation
Teachers punishes students who give wrong answer in class	79	331	81	180	punish_wrong
Teacher used corporal punishment in last week	80	152	76	82	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2	9	1	2	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	2	10	1	2	unsafe_journey_cg
Parental and Girl Attitudes Toward Schooling					
Girl spends a few hours or more per day doing HH chores	58	243	57	127	chores
Caregiver aspires to university education for girl	89	373	89	199	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	88	370	89	198	invest_girls educ
Caregiver believes work or HH chores are acceptable reason to not attend school	43	183	49	110	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	59	247	60	133	school_expensive

TABLE 96: CHARACTERISTICS OF ABE COHORT AT BL AND EL

	Baseline		Endline		
Subgroup	**Pct of Total**	**N**	**Pct of Total**	**N**	**Variable Name & Source**
Household Characteristics					
Girl has only one living parent	10	49	11	28	orphan_single

Girl has no living parents	1	6	2	5	orphan_double
Girl does not live with either parent in her HH	5	24	5	12	no_parents
Girl is currently married	1	4	0	1	married
Girl has ever been married	1	7	1	3	ever_married
Girl is a mother and is under 16 years of age	0	0	0	0	mother_16
Girl is a mother and is under 18 years of age	0	2	0	0	mother_18
Female-headed household	34	163	32	84	female_hoh
Parental Education Background					
HoH has no education of any kind (no Quranic)	22	107	24	64	hoh_noeduc
HoH has no formal education	83	401	82	215	hoh_noformal
Caregiver has no education of any kind (no Quranic)	28	135	26	69	cg_noeduc
Caregiver has no formal education	86	416	82	215	cg_noformal
Household Economic Status					
HH has a poor-quality roof	22	107	22	59	poor_roof
HH went to sleep hungry most nights, last 12 months	8	38	8	21	nofood_most
HH went without clean water most days, last 12 months	6	30	6	15	nowater_most
HH went without medicine most days, last 12 months	17	80	12	32	nomeds_most
HH went without cash income most days, last 12 months	18	87	18	48	nocash_most
HH owns lands	31	152	31	82	owns_land
Caregiver has savings of some form	2	8	1	3	savings
HH owns a phone	86	418	89	234	owns_phone
HH owns a smartphone	15	64	16	37	owns_smartphone
HoH does not earn a regular wage	37	177	29	76	hoh_nowage
Household head is engaged in pastoralism	3	13	1	3	pastoral
Disability Status					
Vision disability	0	0	0	0	disab_vis
Hearing disability	0	1	0	1	disab_hear

Disability the impedes mobility	0	0	0	0	disab_mob
Disability of the arms/hands	0	0	0	0	disab_arms
Disability that impedes self-care	0	0	0	0	disab_selfcare
Disability that impedes communication	0	0	0	0	disab_comm
Cognitive disability	0	0	0	0	disab_cog
Behavioral disability	1	3	1	3	disab_behavior
Mental health disability	0	0	0	0	disab_mh
Mental health disability, alternative coding	17	82	18	47	disab_mh_alt1
Physical disability, any type	0	1	0	1	disab_phys
Physical disability, any type, alternative coding	0	1	0	1	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	1	3	1	3	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	1	5	1	3	disab_ccb_alt1
Displacement and Language					
Household are IDPs	45	190	46	116	idp
Household speaks af-Maay	31	151	33	87	maay
School Facilities					
Girl will not use drinking facilities at school	24	118	22	59	wontuse_drinking
Girl will not use toilet facilities at school	13	61	13	34	wontuse_toilet
No computers available for use at school	96	465	94	248	no_computers
Girl cannot use books/learning materials at school	23	113	24	62	no_materials
Not enough seats for every student in class	10	50	11	29	no_seats
School Environment and Teacher Behaviours					
Teachers do not make girl feel welcome at school	12	58	13	33	unwelcome
Teachers are often absent from class	28	136	32	83	teacher_absent
Teacher rarely or never encourages participation	13	64	10	27	teacher_participation
Teachers punishes students who give wrong answer in class	70	339	70	185	punish_wrong
Teacher used corporal punishment in last week	71	105	76	62	corp_pun

Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	1	4	1	2	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	0	1	0	0	unsafe_journey_cg
Parental and Girl Attitudes Toward Schooling					
Girl spends a few hours or more per day doing HH chores	63	307	65	172	chores
Caregiver aspires to university education for girl	89	432	89	235	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	87	419	86	225	invest_girls educ
Caregiver believes work or HH chores are acceptable reason to not attend school	38	184	35	92	Work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	65	314	65	170	school_expensive

TABLE 97: CHARACTERISTICS OF C1 NFE COHORT AT BL AND EL

	Baseline		Endline		
Subgroup	**Pct of Total**	**N**	**Pct of Total**	**N**	**Variable Name & Source**
Household Characteristics					
Girl has only one living parent	17	90	16	41	orphan_single
Girl has no living parents	1	5	0	1	orphan_double
Girl does not live with either parent in her HH	9	44	7	19	no_parents
Girl is currently married	9	48	10	25	married
Girl has ever been married	18	91	19	49	ever_married
Girl is a mother and is under 16 years of age	0	0	0	0	mother_16
Girl is a mother and is under 18 years of age	2	8	0	0	mother_18
Female-headed household	43	220	45	117	female_hoh
Parental Education Background					
HoH has no education of any kind (no Quranic)	20	102	17	44	hoh_noedu c

HoH has no formal education	82	421	80	210	hoh_noformal
Caregiver has no education of any kind (no Quranic)	24	126	21	55	cg_noeduc
Caregiver has no formal education	84	432	84	220	cg_noformal
Household Economic Status					
HH has a poor-quality roof	20	101	19	49	poor_roof
HH went to sleep hungry most nights, last 12 months	6	32	4	10	nofood_most
HH went without clean water most days, last 12 months	5	24	6	15	nowater_most
HH went without medicine most days, last 12 months	19	97	19	49	nomeds_most
HH went without cash income most days, last 12 months	13	66	9	23	nocash_most
HH owns lands	29	150	32	85	owns_land
Caregiver has savings of some form	2	8	1	3	savings
HH owns a phone	88	454	88	230	owns_phone
HH owns a smartphone	23	104	21	48	owns_smartphone
HoH does not earn a regular wage	44	225	42	109	hoh_nowage
Household head is engaged in pastoralism	1	6	0	1	pastoral
Disability Status					
Vision disability	0	1	0	1	disab_vis
Hearing disability	0	0	0	0	disab_hear
Disability the impedes mobility	0	0	0	0	disab_mob
Disability of the arms/hands	0	0	0	0	disab_arms
Disability that impedes self-care	0	0	0	0	disab_selfcare
Disability that impedes communication	0	0	0	0	disab_comm
Cognitive disability	0	1	0	1	disab_cog
Behavioral disability	0	2	0	0	disab_behavior
Mental health disability	0	0	0	0	disab_mh
Mental health disability, alternative coding	17	87	15	40	disab_mh_alt1
Physical disability, any type	0	1	0	1	disab_phys
Physical disability, any type, alternative coding	1	4	1	3	disab_phys_alt1

Cognitive, communicative, or behavioral disability, any type	1	3	0	1	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	1	6	1	3	disab_ccb_alt1
Displacement and Language					
Household are IDPs	38	163	41	103	idp
Household speaks af-Maay	25	130	27	71	maay
School Facilities					
Girl will not use drinking facilities at school	25	127	28	74	wontuse_drinking
Girl will not use toilet facilities at school	23	119	27	72	wontuse_toilet
No computers available for use at school	97	498	96	251	no_computers
Girl cannot use books/learning materials at school	26	133	29	75	no_materials
Not enough seats for every student in class	20	105	24	63	no_seats
School Environment and Teacher Behaviours					
Teachers do not make girl feel welcome at school	12	61	9	23	unwelcome
Teachers are often absent from class	28	145	26	69	teacher_absent
Teacher rarely or never encourages participation	9	45	7	19	teacher_participation
Teachers punishes students who give wrong answer in class	58	298	57	149	punish_wrong
Teacher used corporal punishment in last week	75	42	77	24	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2	12	1	2	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	2	12	1	3	unsafe_journey_cg
Parental and Girl Attitudes Toward Schooling					
Girl spends a few hours or more per day doing HH chores	72	369	73	190	chores
Caregiver aspires to university education for girl	85	436	86	226	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	89	457	89	234	invest_girls_educ

Caregiver believes work or HH chores are acceptable reason to not attend school	39	202	42	109	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	63	323	61	160	school_expensive

TABLE 98: CHARACTERISTICS OF C4 NFE COHORT AT BL AND EL

	Baseline		Endline		
Subgroup	**Pct of Total**	**N**	**Pct of Total**	**N**	**Variable Name & Source**
Household Characteristics					
Girl has only one living parent	19	53	20	39	orphan_single
Girl has no living parents	1	3	1	1	orphan_double
Girl does not live with either parent in her HH	11	30	9	18	no_parents
Girl is currently married	3	7	2	3	married
Girl has ever been married	4	10	3	5	ever_married
Girl is a mother and is under 16 years of age	0	1	0	0	mother_16
Girl is a mother and is under 18 years of age	1	5	0	0	mother_18
Female-headed household	69	192	69	137	female_hoh
Parental Education Background					
HoH has no education of any kind (no Quranic)	17	47	16	32	hoh_noeduc
HoH has no formal education	85	237	85	169	hoh_noformal
Caregiver has no education of any kind (no Quranic)	15	43	14	28	cg_noeduc
Caregiver has no formal education	84	236	84	166	cg_noformal
Household Economic Status					
HH has a poor-quality roof	22	201	24	138	poor_roof
HH went to sleep hungry most nights, last 12 months	7	61	7	40	nofood_most
HH went without clean water most days, last 12 months	8	74	9	51	nowater_most

HH went without medicine most days, last 12 months	14	129	15	85	nomeds_most
HH went without cash income most days, last 12 months	20	184	22	126	nocash_most
HH owns lands	23	211	22	124	owns_land
Caregiver has savings of some form	3	9	3	6	savings
HH owns a phone	92	258	90	178	owns_phone
HH owns a smartphone	12	30	11	19	owns_smartphone
HoH does not earn a regular wage	46	128	45	89	hoh_nowage
Household head is engaged in pastoralism	1	3	1	2	pastoral
Disability Status					
Vision disability	0	4	1	3	disab_vis
Hearing disability	0	1	0	1	disab_hear
Disability the impedes mobility	2	22	3	16	disab_mob
Disability of the arms/hands	0	1	0	0	disab_arms
Disability that impedes self-care	0	2	0	0	disab_selfcare
Disability that impedes communication	0	0	0	0	disab_comm
Cognitive disability	3	28	3	20	disab_cog
Behavioral disability	2	15	2	10	disab_behavior
Mental health disability	16	142	18	104	disab_mh
Mental health disability, alternative coding	17	156	20	117	disab_mh_alt1
Physical disability, any type	3	29	3	20	disab_phys
Physical disability, any type, alternative coding	3	31	4	22	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	4	40	5	27	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	6	51	6	34	disab_ccb_alt1
Displacement and Language					
Household are IDPs	41	375	46	262	idp
Household speaks af-Maay	33	302	36	207	maay
School Facilities					
Girl will not use drinking facilities at school	17	157	18	105	wontuse_drinking

Girl will not use toilet facilities at school	21	191	22	126	wontuse_toilet
No computers available for use at school	89	812	89	511	no_computers
Girl cannot use books/learning materials at school	6	59	6	36	no_materials
Not enough seats for every student in class	4	35	4	24	no_seats
School Environment and Teacher Behaviours					
Teachers do not make girl feel welcome at school	13	117	14	80	unwelcome
Teachers are often absent from class	17	159	17	100	teacher_absent
Teacher rarely or never encourages participation	5	48	6	34	teacher_participation
Teachers punishes students who give wrong answer in class	28	258	26	151	punish_wrong
Teacher used corporal punishment in last week	8	70	7	38	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2	15	2	9	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	1	4	2	3	unsafe_journey_cg
Parental and Girl Attitudes Toward Schooling					
Girl spends a few hours or more per day doing HH chores	92	844	93	532	chores
Caregiver aspires to university education for girl	63	577	63	359	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	80	736	80	458	invest_girls educ
Caregiver believes work or HH chores are acceptable reason to not attend school	23	65	23	46	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	60	168	60	119	school_expensive

TABLE 99: CHARACTERISTICS OF C5 NFE COHORT AT BL AND EL

	Baseline	Endline	

Subgroup	Pct of Total	N	Pct of Total	N	Variable Name & Source
Household Characteristics					
Girl has only one living parent	20	28	20	22	orphan_single
Girl has no living parents	0	0	0	0	orphan_double
Girl does not live with either parent in her HH	9	12	7	8	no_parents
Girl is currently married	3	4	4	4	married
Girl has ever been married	7	10	9	10	ever_married
Girl is a mother and is under 16 years of age	0	1	0	0	mother_16
Girl is a mother and is under 18 years of age	1	6	0	1	mother_18
Female-headed household	71	99	75	83	female_hoh
Parental Education Background					
HoH has no education of any kind (no Quranic)	19	26	18	20	hoh_noeduc
HoH has no formal education	75	104	75	82	hoh_noformal
Caregiver has no education of any kind (no Quranic)	21	29	19	21	cg_noeduc
Caregiver has no formal education	69	94	67	72	cg_noformal
Household Economic Status					
HH has a poor-quality roof	21	101	23	83	poor_roof
HH went to sleep hungry most nights, last 12 months	12	59	13	46	nofood_most
HH went without clean water most days, last 12 months	18	87	17	61	nowater_most
HH went without medicine most days, last 12 months	18	86	17	62	nomeds_most
HH went without cash income most days, last 12 months	35	169	33	117	nocash_most
HH owns lands	21	109	22	82	owns_land
Caregiver has savings of some form	11	15	10	11	savings
HH owns a phone	93	127	94	101	owns_phone
HH owns a smartphone	27	34	28	28	owns_smartphone

HoH does not earn a regular wage	41	57	40	44	hoh_nowage
Household head is engaged in pastoralism	4	5	4	4	pastoral
Disability Status					
Vision disability	2	12	2	8	disab_vis
Hearing disability	1	3	1	3	disab_hear
Disability the impedes mobility	4	21	4	14	disab_mob
Disability of the arms/hands	1	1	0	0	disab_arms
Disability that impedes self-care	0	2	0	1	disab_selfcare
Disability that impedes communication	1	6	1	3	disab_comm
Cognitive disability	4	18	3	11	disab_cog
Behavioral disability	5	24	4	16	disab_behavior
Mental health disability	36	185	38	140	disab_mh
Mental health disability, alternative coding	43	219	45	166	disab_mh_alt1
Physical disability, any type	6	31	5	20	disab_phys
Physical disability, any type, alternative coding	7	37	7	26	disab_phys_alt1
Cognitive, communicative, or behavioral disability, any type	8	40	7	26	disab_ccb
Cognitive, communicative, or behavioral disability, any type, alternative coding	10	49	9	34	disab_ccb_alt1
Displacement and Language					
Household are IDPs	38	192	40	147	idp
Household speaks af-Maay	30	151	30	112	maay
School Facilities					
Girl will not use drinking facilities at school	16	83	15	55	wontuse_drinking
Girl will not use toilet facilities at school	24	124	24	88	wontuse_toilet
No computers available for use at school	92	472	92	339	no_computers
Girl cannot use books/learning materials at school	7	34	5	18	no_materials
Not enough seats for every student in class	4	21	4	15	no_seats
School Environment and Teacher Behaviours					
Teachers do not make girl feel welcome at school	17	89	16	61	unwelcome

Teachers are often absent from class	23	118	23	85	teacher_absent
Teacher rarely or never encourages participation	7	35	6	24	teacher_participation
Teachers punishes students who give wrong answer in class	51	262	53	197	punish_wrong
Teacher used corporal punishment in last week	18	84	19	65	corp_pun
Safety of School and Surrounding Area					
Girl does not feel safe traveling to school	2	11	2	6	unsafe_journey_girl
Caregiver does not feel it is safe for girls to travel to the school	10	14	8	9	unsafe_journey_cg
Parental and Girl Attitudes Toward Schooling					
Girl spends a few hours or more per day doing HH chores	87	424	86	304	chores
Caregiver aspires to university education for girl	74	364	74	262	aspire_univ
Caregiver believes girls' education worthwhile, even if funds are limited	58	282	57	201	invest_girls educ
Caregiver believes work or HH chores are acceptable reason to not attend school	39	53	44	48	work_over_school
Caregiver believes cost of education is acceptable reason to not attend school	42	58	48	52	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 100: 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		21	66	46	34	11	0	26	0	0	87	30	82	127	110
Girl has only one living parent	21	100.0	8.1	2.2	0.9	1.3	0.0	1.3	0.0	0.0	5.0	3.1	6.5	5.4	2.2
Female HoH	66	8.1	100.0	7.2	2.7	1.8	0.0	2.7	0.0	0.0	14.6	4.5	25.9	17.0	14.8
HoH has no formal education	46	2.2	7.2	100.0	0.9	3.1	0.0	0.0	0.0	0.0	5.0	5.4	16.7	9.4	13.0
HH has a poor-quality roof	34	0.9	2.7	0.9	100.0	0.9	0.0	4.0	0.0	0.0	5.5	1.8	13.9	9.4	4.5
HH went to sleep hungry most nights, last 12 months	11	1.3	1.8	3.1	0.9	100.0	0.0	0.0	0.0	0.0	1.4	2.2	5.6	1.8	2.7
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	26	1.3	2.7	0.0	4.0	0.0	0.0	100.0	0.0	0.0	8.2	1.3	15.7	9.0	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	87	5.0	14.6	5.0	5.5	1.4	0.0	8.2	0.0	0.0	100.0	5.0	38.9	28.8	16.9
Teachers are not welcoming	30	3.1	4.5	5.4	1.8	2.2	0.0	1.3	0.0	0.0	5.0	100.0	13.9	5.8	8.1
Teacher uses corporal punishment	82	6.5	25.9	16.7	13.9	5.6	0.0	15.7	0.0	0.0	38.9	13.9	100.0	51.9	29.6
Girl has heavy chore burden	127	5.4	17.0	9.4	9.4	1.8	0.0	9.0	0.0	0.0	28.8	5.8	51.9	100.0	26.5
Caregiver values work/HH chores over school attendance	110	2.2	14.8	13.0	4.5	2.7	0.0	2.2	0.0	0.0	16.9	8.1	29.6	26.5	100.0

TABLE 101: 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		28	84	64	59	21	0	47	1	3	116	33	62	172	92
Girl has only one living parent	28	100.0	8.4	4.2	1.5	2.3	0.0	1.9	0.0	0.8	5.9	1.1	4.9	8.4	3.4
Female HoH	84	8.4	100.0	9.1	4.9	3.4	0.0	5.3	0.0	0.4	16.9	3.4	19.5	20.5	12.9
HoH has no formal education	64	4.2	9.1	100.0	4.9	6.1	0.0	4.2	0.4	1.1	8.7	6.1	12.2	16.7	9.9
HH has a poor-quality roof	59	1.5	4.9	4.9	100.0	2.7	0.0	6.5	0.0	0.0	12.6	1.1	14.6	13.7	8.7
HH went to sleep hungry most nights, last 12 months	21	2.3	3.4	6.1	2.7	100.0	0.0	1.9	0.0	0.8	3.5	3.0	3.7	6.1	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	47	1.9	5.3	4.2	6.5	1.9	0.0	100.0	0.0	0.0	9.4	4.2	23.2	11.4	6.5

Physical disability, any type	1	0.0	0.0	0.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.4	0.4
Cognitive, communicative, or behavioral disability, any type	3	0.8	0.4	1.1	0.0	0.8	0.0	0.0	0.0	100.0	0.0	1.1	0.0	1.1	0.4
Household are IDPs	116	5.9	16.9	8.7	12.6	3.5	0.0	9.4	0.0	0.0	100.0	5.1	45.7	31.9	11.8
Teachers are not welcoming	33	1.1	3.4	6.1	1.1	3.0	0.0	4.2	0.0	1.1	5.1	100.0	7.3	9.1	5.3
Teacher uses corporal punishment	62	4.9	19.5	12.2	14.6	3.7	0.0	23.2	0.0	0.0	45.7	7.3	100.0	57.3	20.7
Girl has heavy chore burden	172	8.4	20.5	16.7	13.7	6.1	0.0	11.4	0.4	1.1	31.9	9.1	57.3	100.0	20.9
Caregiver values work/HH chores over school attendance	92	3.4	12.9	9.9	8.7	1.9	0.0	6.5	0.4	0.4	11.8	5.3	20.7	20.9	100.0

TABLE 102: 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		41	117	44	49	10	0	40	1	1	103	23	24	190	109
Girl has only one living parent	41	100.0	9.2	2.3	3.1	0.0	0.0	3.8	0.0	0.0	9.6	1.1	29.0	13.0	3.1
Female HoH	117	9.2	100.0	9.2	9.2	1.1	0.0	6.5	0.4	0.0	18.7	3.1	32.3	34.4	19.8
HoH has no formal education	44	2.3	9.2	100.0	2.7	1.1	0.0	1.9	0.4	0.0	4.8	2.7	6.5	11.5	10.3
HH has a poor-quality roof	49	3.1	9.2	2.7	100.0	1.5	0.0	4.2	0.0	0.0	8.8	0.8	19.4	13.7	9.2
HH went to sleep hungry most nights, last 12 months	10	0.0	1.1	1.1	1.5	100.0	0.0	0.4	0.0	0.0	0.8	1.1	0.0	2.7	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	40	3.8	6.5	1.9	4.2	0.4	0.0	100.0	0.4	0.0	8.4	1.9	29.0	11.5	5.3

Physical disability, any type	1	0.0	0.4	0.4	0.0	0.0	0.0	0.4	100.0	0.0	0.0	0.4	0.0	0.4	0.0
Cognitive, communicative, or behavioral disability, any type	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.4	0.0	0.4	0.4
Household are IDPs	103	9.6	18.7	4.8	8.8	0.8	0.0	8.4	0.0	0.0	100.0	2.4	40.0	32.7	12.0
Teachers are not welcoming	23	1.1	3.1	2.7	0.8	1.1	0.0	1.9	0.4	0.4	2.4	100.0	6.5	8.0	3.1
Teacher uses corporal punishment	24	29.0	32.3	6.5	19.4	0.0	0.0	29.0	0.0	0.0	40.0	6.5	100.0	74.2	35.5
Girl has heavy chore burden	190	13.0	34.4	11.5	13.7	2.7	0.0	11.5	0.4	0.4	32.7	8.0	74.2	100.0	29.4
Caregiver values work/HH chores over school attendance	109	3.1	19.8	10.3	9.2	2.3	0.0	5.3	0.0	0.4	12.0	3.1	35.5	29.4	100.0

TABLE 103: 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		39	137	32	138	40	104	117	20	27	262	80	38	532	46
Girl has only one living parent	39	100.0	18.2	3.0	5.6	1.0	2.0	4.0	0.5	0.0	10.1	2.0	1.0	17.2	5.1
Female HoH	137	18.2	100.0	14.1	14.6	6.6	9.1	15.7	1.5	2.0	31.3	7.6	5.6	61.6	18.2
HoH has no formal education	32	3.0	14.1	100.0	2.0	0.5	2.0	3.0	0.5	0.5	6.6	4.0	0.5	13.1	2.5
HH has a poor-quality roof	138	5.6	14.6	2.0	100.0	1.4	3.5	3.7	0.9	2.3	18.1	4.4	2.3	22.1	7.6
HH went to sleep hungry most nights, last 12 months	40	1.0	6.6	0.5	1.4	100.0	3.1	3.3	0.5	0.3	3.3	0.3	0.5	6.4	1.5
Mental health disability, standard coding	104	2.0	9.1	2.0	3.5	3.1	100.0	18.1	0.9	1.6	8.7	3.1	2.3	16.9	2.0
Mental health disability, alternative coding	117	4.0	15.7	3.0	3.7	3.3	18.1	100.0	1.0	1.6	9.4	3.3	2.3	19.2	4.5

Physical disability, any type	20	0.5	1.5	0.5	0.9	0.5	0.9	1.0	100.0	0.2	1.7	0.9	0.2	2.4	2.0
Cognitive, communicative, or behavioral disability, any type	27	0.0	2.0	0.5	2.3	0.3	1.6	1.6	0.2	100.0	3.5	0.5	0.7	4.5	2.0
Household are IDPs	262	10.1	31.3	6.6	18.1	3.3	8.7	9.4	1.7	3.5	100.0	7.0	3.9	43.0	11.1
Teachers are not welcoming	80	2.0	7.6	4.0	4.4	0.3	3.1	3.3	0.9	0.5	7.0	100.0	1.4	11.3	3.0
Teacher uses corporal punishment	38	1.0	5.6	0.5	2.3	0.5	2.3	2.3	0.2	0.7	3.9	1.4	100.0	6.2	3.0
Girl has heavy chore burden	532	17.2	61.6	13.1	22.1	6.4	16.9	19.2	2.4	4.5	43.0	11.3	6.2	100.0	22.2
Caregiver values work/HH chores over school attendance	46	5.1	18.2	2.5	7.6	1.5	2.0	4.5	2.0	2.0	11.1	3.0	3.0	22.2	100.0

TABLE 104: 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		22	83	20	83	46	140	166	20	26	147	61	65	304	48
Girl has only one living parent	22	100.0	18.2	4.5	5.6	3.7	9.1	15.5	0.0	0.9	10.9	4.5	2.9	18.5	9.3
Female HoH	83	18.2	100.0	16.4	16.7	10.2	20.9	40.0	3.6	8.2	30.9	15.5	21.4	67.6	35.2
HoH has no formal education	20	4.5	16.4	100.0	5.6	0.9	6.4	10.9	0.0	2.7	8.2	4.5	1.9	14.8	8.3
HH has a poor-quality roof	83	5.6	16.7	5.6	100.0	6.2	9.6	11.3	2.8	2.8	19.2	5.9	4.6	21.4	14.8
HH went to sleep hungry most nights, last 12 months	46	3.7	10.2	0.9	6.2	100.0	7.6	7.6	2.3	2.5	9.0	3.1	3.7	12.4	3.7
Mental health disability, standard coding	140	9.1	20.9	6.4	9.6	7.6	100.0	37.8	3.0	3.8	18.1	6.8	7.9	33.8	11.1
Mental health disability, alternative coding	166	15.5	40.0	10.9	11.3	7.6	37.8	100.0	3.8	4.3	20.5	8.9	9.4	40.0	22.2

Physical disability, any type	20	0.0	3.6	0.0	2.8	2.3	3.0	3.8	100.0	1.4	3.5	1.9	0.9	5.4	2.8
Cognitive, communicative, or behavioral disability, any type	26	0.9	8.2	2.7	2.8	2.5	3.8	4.3	1.4	100.0	4.9	2.4	2.9	6.2	2.8
Household are IDPs	147	10.9	30.9	8.2	19.2	9.0	18.1	20.5	3.5	4.9	100.0	7.6	8.5	38.3	20.4
Teachers are not welcoming	61	4.5	15.5	4.5	5.9	3.1	6.8	8.9	1.9	2.4	7.6	100.0	4.1	13.8	11.1
Teacher uses corporal punishment	65	2.9	21.4	1.9	4.6	3.7	7.9	9.4	0.9	2.9	8.5	4.1	100.0	16.2	11.9
Girl has heavy chore burden	304	18.5	67.6	14.8	21.4	12.4	33.8	40.0	5.4	6.2	38.3	13.8	16.2	100.0	40.7
Caregiver values work/HH chores over school attendance	48	9.3	35.2	8.3	14.8	3.7	11.1	22.2	2.8	2.8	20.4	11.1	11.9	40.7	100.0

Annex 5 - Learning Tables

In this annex, we provide additional tables documenting learning outcomes for all cohorts and girl types. The purpose of this annex is to document aspects of learning outcomes that are less relevant, analytically, throughout the report, but which may be necessary for internal reporting or other purposes. In particular, we report on learning outcomes for each girl type as of endline, whereas the main report focuses on changes for each cohort and within subgroups. We also report foundational skill gaps for all girl types; 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. 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, though we note that the "overall" results at the top include the subgroup members. 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 cohorts – because the cohorts are considerably different.

Subgroup Learning Outcomes

The tables below report subgroup-specific learning scores for the FE, ABE, C1 NFE, C4 NFE, and C5 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-EL panel sample or the ML1 or ML2 to EL panels, 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 surveyed at EL. 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 EL. For ABE and C1 NFE girls, the sample analysed here includes girls who fell out of the sample at ML1 or ML2 but were re-contacted at EL. Likewise, for C4 NFE girls, the sample includes girls who dropped out of the sample at ML2, but were recontacted at EL. For C5 NFE girls, the sample is exactly the same as the ML2 to EL sample, as no replacements were drawn for attrition.

TABLE 105: SUBGROUP LEARNING SCORES AMONG FE COHORT, AT EL

Subgroup	Sample Size	Somali Literacy	Numeracy (8 Subtask)	Numeracy (11 Subtask)
Overall	359	70.0	77.4	72.6
Geography				
Banadir	149	66.2	72.2	67.1
Jubaland	119	78.4	85.7	82.7
South West State	91	65.2	74.8	68.3
Disability Status				
Physical disability, any type	10	66.4	76.5	73.5

Physical disability, any type, alternative coding	13	67.2	76.1	74.8
Cognitive disability, any type, alt. coding	2	85.6	73.4	67.0
Mental health disability	23	70.4	74.0	69.7
Mental health disability, alt. coding	71	75.0	80.0	76.4
Disability, non-mental health	17	60.9	76.0	70.0
Disability, non-mental health, alt. coding	24	66.4	76.0	71.3
Disability, any	38	65.6	74.1	68.7
Disabiltiy, any, alt. coding	87	72.1	79.0	74.8
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	66	66.4	75.5	68.9
HoH has no formal education	249	69.0	77.3	72.0
Caregiver has no education of any kind (no Quranic)	92	64.7	74.7	69.0
Caregiver has no formal education	272	69.5	77.7	72.6
Household Economic Characteristics				
HH has a poor-quality roof	50	67.6	75.8	71.3
HH went to sleep hungry most nights, last 12 months	21	59.9	71.4	64.5
HH went without clean water most days, last 12 months	34	67.7	77.6	69.2
HH went without medicine most days, last 12 months	49	60.9	73.0	66.4
HH went without cash income most days, last 12 months	62	65.4	75.7	69.8
HH owns lands	118	70.9	78.5	74.7
Caregiver has savings of some form	18	65.6	79.5	75.8
HH owns a phone	304	70.1	77.7	72.9
HH owns a smartphone	81	71.3	79.7	75.4
HoH does not earn a regular wage	126	73.5	77.7	73.3
Parental Circumstances				
Girl has only one living parent	34	68.4	76.0	68.9
Girl does not live with either parent in her HH	12	67.8	70.0	65.3
Female-headed household	139	72.7	77.7	72.6
Displacement and Language				
Household are IDPs	126	73.6	78.8	74.0

Household speaks af-Maay	101	64.0	74.5	68.0

TABLE 106: SUBGROUP LEARNING SCORES AMONG ABE COHORT, AT EL

Subgroup	Sample Size	Somali Literacy	Numeracy (8 Subtask)	Numeracy (11 Subtask)
Overall	263	58.2	73.1	66.4
Geography				
Banadir	101	49.4	66.6	59.6
Jubaland	60	62.0	82.5	77.6
South West State	102	64.8	74.0	66.6
Disability Status				
Physical disability, any type	1	23.1	79.8	60.3
Physical disability, any type, alternative coding	1	23.1	79.8	60.3
Mental health disability, alt. coding	47	52.6	69.3	62.9
Disability, non-mental health	4	23.3	70.7	53.7
Disability, non-mental health, alt. coding	4	23.3	70.7	53.7
Disability, any	4	23.3	70.7	53.7
Disabiltiy, any, alt. coding	51	50.3	69.4	62.2
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	64	52.4	69.3	62.9
HoH has no formal education	215	56.6	73.4	66.6
Caregiver has no education of any kind (no Quranic)	69	53.7	71.3	64.5
Caregiver has no formal education	215	57.1	73.8	67.1
Household Economic Characteristics				
HH has a poor-quality roof	59	51.5	68.3	60.7
HH went to sleep hungry most nights, last 12 months	21	46.5	62.1	53.7
HH went without clean water most days, last 12 months	15	47.7	72.3	62.7
HH went without medicine most days, last 12 months	32	44.6	66.5	56.9
HH went without cash income most days, last 12 months	48	52.9	66.9	59.0
HH owns lands	82	64.8	77.4	71.0

Caregiver has savings of some form	3	24.7	56.3	46.1
HH owns a phone	234	56.6	71.8	64.7
HH owns a smartphone	37	58.0	71.7	63.9
HoH does not earn a regular wage	76	58.1	76.5	70.3
Parental Circumstances				
Girl has only one living parent	28	58.0	71.8	64.8
Girl has no living parents	5	60.5	77.8	69.5
Girl does not live with either parent in her HH	12	56.9	73.8	66.8
Female-headed household	84	60.5	75.0	69.2
Displacement and Language				
Household are IDPs	116	52.0	72.3	64.9
Household speaks af-Maay	87	58.6	71.1	64.1

TABLE 107: SUBGROUP LEARNING SCORES AMONG C1 NFE COHORT, AT EL

Subgroup	Sample Size	Somali Literacy	Numeracy (8 Subtask)	Numeracy (11 Subtask)
Overall	263	54.5	70.3	63.4
Geography				
Banadir	115	54.8	65.5	57.7
Jubaland	72	54.6	79.5	74.2
South West State	76	53.7	68.7	61.7
Disability Status				
Physical disability, any type	1	100.0	88.5	91.7
Physical disability, any type, alternative coding	3	72.4	81.4	78.8
Cognitive disability	1	74.8	69.2	57.1
Cognitive disability, any type, alt. coding	2	61.5	74.0	61.1
Mental health disability	1	99.2	88.3	82.0
Mental health disability, alt. coding	41	50.8	70.7	62.6
Disability, non-mental health	2	87.4	78.9	74.4
Disability, non-mental health, alt. coding	6	69.6	78.4	72.5
Disability, any	3	91.3	82.0	76.9
Disabiltiy, any, alt. coding	46	52.2	71.3	63.3
Parental Educational Attainment				

HoH has no education of any kind (no Quranic)	44	56.5	73.2	67.5
HoH has no formal education	210	51.8	69.3	62.5
Caregiver has no education of any kind (no Quranic)	55	58.2	72.6	66.6
Caregiver has no formal education	220	52.9	70.3	63.6
Household Economic Characteristics				
HH has a poor-quality roof	49	45.7	65.7	58.7
HH went to sleep hungry most nights, last 12 months	11	61.4	69.9	62.2
HH went without clean water most days, last 12 months	16	51.9	64.6	58.8
HH went without medicine most days, last 12 months	50	52.9	60.8	54.1
HH went without cash income most days, last 12 months	23	56.1	68.8	62.4
HH owns lands	86	57.3	73.5	68.3
Caregiver has savings of some form	3	42.3	53.4	47.3
HH owns a phone	230	54.6	70.9	63.8
HH owns a smartphone	48	60.7	72.7	66.7
HoH does not earn a regular wage	109	57.5	70.6	63.5
Parental Circumstances				
Girl has only one living parent	41	45.7	65.9	56.9
Girl has no living parents	1	90.6	83.5	81.7
Girl does not live with either parent in her HH	19	51.3	77.4	71.5
Female-headed household	118	52.8	71.9	64.2
Displacement and Language				
Household are IDPs	104	46.4	65.1	58.2
Household speaks af-Maay	72	44.1	68.0	61.8

TABLE 108: SUBGROUP LEARNING SCORES AMONG C4 NFE COHORT, AT EL

Subgroup	Sample Size	Somali Literacy	Numeracy (8 Subtask)	Numeracy (11 Subtask)
Overall	575	45.1	63.6	56.2
Geography				
Banadir	272	43.6	62.4	54.2
South West State	219	47.1	65.6	59.5

Hirshabelle	84	45.1	62.5	54.3
Disability Status				
Physical disability, any type	20	50.7	63.8	58.1
Physical disability, any type, alternative coding	22	52.4	65.2	59.0
Cognitive disability	20	39.6	64.0	55.9
Cognitive disability, any type, alt. coding	25	42.2	64.1	55.4
Mental health disability	104	34.2	58.0	50.5
Mental health disability, alt. coding	117	35.5	58.8	51.3
Disability, non-mental health	46	43.1	62.0	55.1
Disability, non-mental health, alt. coding	53	45.6	63.8	56.3
Disability, any	136	38.0	60.7	53.3
Disabiltiy, any, alt. coding	155	39.5	61.5	53.9
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	32	55.8	68.3	60.1
HoH has no formal education	169	53.0	66.0	59.0
Caregiver has no education of any kind (no Quranic)	28	54.7	67.4	59.7
Caregiver has no formal education	166	53.7	66.2	59.5
Household Economic Characteristics				
HH has a poor-quality roof	138	32.0	56.9	48.3
HH went to sleep hungry most nights, last 12 months	40	37.2	56.1	49.2
HH went without clean water most days, last 12 months	51	36.3	58.2	50.7
HH went without medicine most days, last 12 months	85	42.4	61.7	54.5
HH went without cash income most days, last 12 months	126	37.7	60.6	53.3
HH owns lands	124	50.8	66.8	60.6
Caregiver has savings of some form	6	70.9	80.7	74.9
HH owns a phone	178	53.3	66.8	59.9
HH owns a smartphone	19	57.1	72.1	67.8
HoH does not earn a regular wage	89	54.4	65.7	58.4
Parental Circumstances				
Girl has only one living parent	39	52.0	64.2	57.4
Girl has no living parents	1	72.6	65.0	54.1

Girl does not live with either parent in her HH	18	65.9	73.2	64.8
Female-headed household	137	51.7	67.1	59.8
Displacement and Language				
Household are IDPs	262	34.8	58.2	50.2
Household speaks af-Maay	207	39.1	61.9	54.9

TABLE 109: SUBGROUP LEARNING SCORES AMONG C5 NFE COHORT, AT EL

Subgroup	Sample Size	Somali Literacy	Numeracy (8 Subtask)	Numeracy (11 Subtask)
Overall	374	46.9	65.2	57.5
Geography				
Banadir	169	42.1	63.0	54.4
South West State	143	46.8	66.5	60.0
Hirshabelle	62	60.3	68.0	60.3
Disability Status				
Physical disability, any type	20	35.0	51.2	44.7
Physical disability, any type, alternative coding	26	36.5	51.2	44.7
Cognitive disability	11	39.1	67.6	61.7
Cognitive disability, any type, alt. coding	16	43.1	63.4	58.1
Mental health disability	142	44.1	62.7	54.6
Mental health disability, alt. coding	168	46.5	64.0	56.3
Disability, non-mental health	41	40.0	58.3	51.3
Disability, non-mental health, alt. coding	49	42.3	59.8	52.9
Disability, any	162	43.3	62.5	54.5
Disabiltiy, any, alt. coding	187	46.1	64.5	56.9
Parental Educational Attainment				
HoH has no education of any kind (no Quranic)	20	44.0	68.0	59.0
HoH has no formal education	82	48.5	66.8	59.2
Caregiver has no education of any kind (no Quranic)	21	39.1	62.1	51.5
Caregiver has no formal education	73	47.2	65.2	57.5
Household Economic Characteristics				
HH has a poor-quality roof	83	27.5	55.2	46.9

HH went to sleep hungry most nights, last 12 months	47	39.6	63.9	57.5
HH went without clean water most days, last 12 months	62	33.7	61.9	55.5
HH went without medicine most days, last 12 months	63	46.6	68.8	61.4
HH went without cash income most days, last 12 months	118	39.8	61.0	54.6
HH owns lands	82	54.3	71.7	64.1
Caregiver has savings of some form	11	57.8	70.3	62.6
HH owns a phone	102	51.2	67.6	60.1
HH owns a smartphone	29	58.5	74.7	69.0
HoH does not earn a regular wage	44	51.1	63.6	55.0
Parental Circumstances				
Girl has only one living parent	22	52.0	67.8	62.4
Girl does not live with either parent in her HH	8	42.9	49.2	41.7
Female-headed household	83	50.3	69.6	61.7
Displacement and Language				
Household are IDPs	150	35.4	58.7	50.8
Household speaks af-Maay	112	38.2	61.9	55.8

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 EL for a given cohort, without reference to whether she is part of a particular panel sample.

TABLE 110: FOUNDATIONAL SKILL GAPS IN NUMERACY, FE COHORT AT EL

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	1.4	1.4	9.5	3.1	12.8	8.1	17.3	4.2	25.1	22.3	16.4
Emergent Learner	0.0	0.3	60.7	1.1	8.1	3.3	18.9	0.6	9.5	7.0	20.6

Established Learner	0.3	5.6	19.2	12.8	30.1	8.9	32.6	18.1	28.4	25.9	36.8
Proficient Learner	98.3	92.8	10.6	83.0	49.0	79.7	31.2	77.2	37.0	44.8	26.2

TABLE 111: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, FE COHORT AT EL

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	2.8	6.4	13.1	14.2	13.6	20.1
Emergent Learner	1.7	6.7	6.4	2.2	8.6	35.4
Established Learner	10.9	18.9	18.1	45.4	41.2	32.9
Proficient Learner	84.7	68.0	62.4	38.2	36.5	11.7

TABLE 112: FOUNDATIONAL SKILL GAPS IN NUMERACY, ABE COHORT AT EL

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	0.8	1.4	12.5	3.3	13.1	10.0	18.4	3.3	28.1	27.3	15.9
Emergent Learner	0.3	0.3	48.7	2.5	5.8	2.8	12.8	0.6	7.0	4.7	20.6
Established Learner	1.1	6.4	6.4	6.1	22.6	10.0	22.8	19.8	17.0	17.3	22.8
Proficient Learner	71.0	65.2	5.6	61.3	31.8	50.4	19.2	49.6	21.2	24.0	13.9

TABLE 113: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, ABE COHORT AT EL

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	4.5	10.0	18.1	20.3	20.3	27.0
Emergent Learner	3.9	7.5	6.4	2.2	7.2	24.5
Established Learner	8.9	20.1	13.9	32.0	26.2	16.7

Proficient Learner	56.0	35.7	34.8	18.7	19.5	5.0

TABLE 114: FOUNDATIONAL SKILL GAPS IN NUMERACY, C1 NFE COHORT AT EL

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	3.3	4.2	19.8	7.5	14.8	12.8	20.6	8.9	33.4	28.7	20.9
Emergent Learner	1.1	1.7	40.4	1.1	6.1	1.4	14.2	0.3	6.4	4.5	18.4
Established Learner	0.6	4.7	7.5	5.6	20.3	5.3	19.5	13.1	13.6	14.5	22.0
Proficient Learner	68.2	62.7	5.6	59.1	32.0	53.8	18.9	51.0	19.8	25.6	12.0

TABLE 115: FOUNDATIONAL SKILL GAPS IN SOMALI LITERACY, C1 NFE COHORT AT EL

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-Learner	11.1	17.5	22.3	24.0	23.4	29.0
Emergent Learner	3.3	3.6	3.9	1.9	7.5	20.3
Established Learner	10.6	20.3	13.6	28.4	23.4	16.4
Proficient Learner	48.2	31.8	33.4	18.9	18.9	7.5

Annex 6 - Logframe and Indicator Performance Tracking Table

To be sent separately.

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 - External Evaluator Declaration

Name of Project: Adolescent Girls' Education in Somalia (AGES)

Name of External Evaluator: Nicolas G. Wicaksono, Consillient Research

Contact Information for External Evaluator: nicolas.wicaksono@consilientresearch.org

Names of all members of the evaluation team: Nicolas Wicaksono, Giorgio Monti, Shreyas Kumar, Emilio Lopez de Romana, Salwa Yusuf, Ahmed Hersi, Nimco Suleiman Hussien, Nasir Saeed Abdi

I, Nicolas G. Wicaksono, 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: NW
- The data analysis was conducted independently by the EE and provides a fair and consistent representation of findings. Initials: NW
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed. Initials: NW
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by Consilient Research. Initials: NW
- All child protection protocols and guidance have been followed. Initials: NW
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols. Initials: NW

Nicolas G. Wicaksono

Consilient Research, July 28th, 2024

Annex 12 - Project Beneficiary Tables

Direct beneficiaries[159]

[159] 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).

	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[160] REOs/ DEOs 8[161] Gender Focal points	9,502 mothers		9,502 [162]mothers	431 CEC members 36 community health workers 700 female mentors 193 VSLA facilitators	724 CEC members 244 male mentor 43 VSLA facilitators	1,155[163] CEC members 36 community health workers 944[164] GEF/BEF mentors 277[165] 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									
[Cohort 6, NFE]	12,731		12,731	36	54	90									
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									

[160] MoE staff trained as ToTs and coaches for quarterly coaching of ABE/NFE facilitators
[161] MoE Gender Focal Points staff trained on Girls and Boys Empowerment concept.
[162] Parents of the enrolled girls participating in the VSLA groups
[163] Trained Community Education Committee Members (452 cohort 1, 23 cohort 2, 350 cohort 4, 145 cohort 5 and 185 cohort 6
[164] Trained GEF/BEF mentors (cohort 1; 60 F, 60 M; cohort 4; 283 F, 87 M; cohort 5; 80 F and 39 M; cohort 6; 277 F, 58 M)
[165] Trained VSLA facilitators (79 cohort 1, 91 cohort 4, 66 cohort 5 and 41 cohort 6

Indirect beneficiaries[166]

	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												

[166] 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.

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[167]	600						7,241
[Cohort 2, ABE]							6,035
[Cohort 1, NFE][168]		4250					6,604
[Cohort 2, NFE]		2475					4,319
[Cohort 3, NFE][169]							4,198
[Cohort 4, NFE]	2,830	2,059	880				13,439
[Cohort 5, NFE]	2,830	1,942	711				12,942
[Cohort 6, NFE]	2,830	2,115					12,731

[167] 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.

[168] 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.

[169] 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.

Annex 13 - Programme Management Response

The programme team agrees with the recommendations, while noting how those are either already under implementation or may face practical limitations. Specific responses are provided below.

#1 Continuing education. As expected, the transition into further levels of education is associated with higher learning outcomes. The programme agrees that there is a need to support transitions into formal, accelerated, and technical-vocational education and training/ TVET, particularly through linkages with other initiatives. Examples of transition support includes AGES' referrals to vocational training opportunities; linkages between AGES and system-level programmes such as the Global Partnership for Education/GPE-funded ESPIG, which provided capitation grants to primary schools to absorb new enrolment; and facilitating access to VSLAs and cash transfers to enable households to support the costs of education. The recommendation does include a suggestion to provide opportunities for continued education virtually or remotely. It is possible to structure follow-up courses for remote or virtual delivery, such as those provided by a2i in Bangladesh or Noon in Afghanistan; however, these opportunities should not be ad-hoc and project-based, but rather integrated into the national qualifications framework to enable participants to earn certificates or credits towards a certification.

#2 Tailoring Instructional Levels and Class Materials. The evaluator recommends tailoring the programme and associated materials to learning levels, specifically aiming at supporting high-achieving students. AGES agrees that this is indeed critical. However, AGES also notes that in some cases, students with higher baseline scores represent a case of incorrect identification of the target population and/or misrepresentation during enrolment and placement tests. In these cases, students with higher scores are often individuals with a history of prior exposure to education, whose learning levels would be more suitable for enrolment in primary grades or advanced ABE levels. Due to the cost of primary education and the limited offer of ABE levels 2-4, out-of-school adolescents may not disclose their education history to obtain an opportunity for re-enrolment. AGES' placement tests are designed to reduce this practice and to re-route candidates to opportunities better suited for their learning levels. AGES has already catered to a limited range of learning levels in NFE, using formative assessments and remedial materials; however, NFE is not designed to respond to the learning needs of students with levels equivalent to a grade 4-5 student. It is designed to respond to the needs of those who have never attended school or dropped out in grade 1.

#3 Adapting to Girls' Capacity to Physically Attend Classes. As noted by the evaluator, the domestic chore burden increases with age and caregivers' attitudes towards the allocation of girls' time shift towards prioritizing domestic work vs education as girls become older. This reflects both traditional gender norms as well as the opportunity cost of education in female-headed households (30% of the formal education students, 32% of ABE students, and 45% of NFE cohort 1 students are living in female-headed households). In female-headed households, the entire family depends on the meagre income of the mother, who depends on the children to share household chores and caregiving responsibilities in order to allow her to work outside the home. The evaluator's recommendation of providing self-study materials **has already been in place since 2020** – AGES provides home-based learning materials to support students with high absenteeism to study remotely.

#4 Promoting Community Prioritisation of Girls' Education. While AGES agrees that awareness raising activities are important to strengthen parental support for girls' education, it should be noted that (a) those have already been conducted by the programme since its inception, including through community leaders, media, religious leaders, and Girls' Empowerment Forums; and (b) additional campaigns are being conducted by the MoECHE through the GPE-funded Girls' Education Accelerator.

#5 Sharing Teacher Experience, Expertise, and Materials. The programme agrees with the recommendation to reinforce knowledge-sharing on the use of improved teaching practices. This is being done through FMS MoEs' coaches and through WhatsApp groups bringing teachers together to share experiences. It can be further advanced by working with FMS MoEs to set up district-level experience-sharing platforms for teachers.

#6 Improved Measures on Classroom Gender Equality and Equity. The programme agrees that there is scope to expand the measurement of gender equality practices in class through scales used with students. This will be piloted under the upcoming Accelerated Education Activity.

#7 Continuous Monitoring and Reducing of Corporal Punishment Use. The partial resurgence in the use of corporal punishment was observed in the evaluation and during monitoring rounds in 2024 and is associated with the shocks observed in November 2023. The new Safeguarding Policy provides the legal background for the MoECHE / FMS MoEs to address the use of corporal punishment, and the new Girls' Education Accelerator will support its operationalization through training and coaching at district level. However, there is also a need for further supporting teachers to address mental health issues. There is a clear pattern of surges in the use of verbal violence and corporal punishment during and after shocks; while this does not justify violence against students, it is important to acknowledge that the experience of violence/extreme conditions often translates into violent reactions, and to invest in psychosocial support for its prevention.

consilient