consilient

Somali Girls Education Promotion Project - Transition

Endline Evaluation Report

25 March 2022
Version 3

USAID FROM THE AMERICAN PEOPLE

Girls' Education Challenge

UKaid from the British people

Table of Contents

List of Tables

List of Figures

Abbreviations

ABE	Accelerated Basic Education
ALP	Alternative Learning Program
CEC	Community Education Committee
EAC	Educate A Child
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
ESPDG	Education Sector Plan Development Grant
ESPIG	Education Sector Program Implementation Grant
FCDO	Foreign, Commonwealth and Development Office
FGD	Focus Groups Discussion/s
FGS	Federal Government of Somalia
GEC	Girls Education Challenge
GEC-T	Girls Education Challenge - Transition
GEF	Girls' Empowerment Forum
GPE	Global Partnership for Education
GwD	Girls with Disabilities
MOE	Ministry of Education
REO	Regional Education Officer
SeGMA	Secondary Grade Math Assessment
SeGRA	Secondary Grade Reading Assessment
SOMGEP-T	Somali Girls' Education Promotion Project -Transition
DFID	UK's Department for International Development
YLI	Youth Leadership Index

Executive Summary

The Somali Girls' Education Promotion Project – Transition (SOMGEP-T), funded by UK's Foreign and Commonwealth Development Office (FCDO) and USAID, was implemented from 2017-2022 in rural and remote areas of Somaliland, Puntland, and Galmudug, reaching an estimated 20,000 girls and 10,000 boys directly and another 20,000 students through indirect benefits. The implementation of SOMGEP-T followed on the successes of SOMGEP (2013-17, funded by FCDO), with a particular focus on enhancing learning outcomes and transition rates for marginalised adolescent girls. SOMGEP-T was implemented by a consortium formed by CARE International, ADRA, local women's rights network NAGAAD, and local non-governmental organisations HAVOYOCO (a youth-led committee) and TASS. The project's activities were conducted in close collaboration with state- and national-level Ministries of Education, responding to priority areas identified in state and Federal-level sector development plans.

SOMGEP-T used a mixed-methods, quasi-experimental design for impact measurement. The endline evaluation sample included 69 primary schools, split between 37 intervention and 32 comparison schools. Additionally, a pre-post evaluation design was used to assess progress on accelerated education programming. Data collection took place in an additional 32 Alternative Learning Programme (ALP) centres and 35 Accelerated Basic Education (ABE) centres, which are located in the same communities as SOMGEP-T intervention schools. In total, the endline sample included 1,802 girls and their households, 965 of whom were re-contacted from the baseline and interviewed successfully. The endline data collection took place in December 2021.

Learning

A few key findings emerged from the learning analysis. Firstly, although improvements were observed in numeracy, Somali literacy, English literacy, and financial literacy amongst girls in intervention schools, these improvements were also simultaneously observed in girls in comparison schools. Secondly, learning improvements occurred largely within the first two years of the programme, prior to the ML2 evaluation – a finding which can largely be attributed to the COVID-19 pandemic and the resulting long-running school closures that occurred between the ML2 and endline evaluations.

At the endline, it became evident that SOMGEP-T had a much larger impact on learning among a few specific groups of ultra-marginalised girls, especially those marginalised along multiple overlapping axes, such as girls from relatively poor or pastoralist households who were out-of-school at the baseline, girls with physical disabilities, and the lowest-performing students at the baseline. For instance, girls from pastoralist households who were out-of-school at baseline gained an average of 10.8 percentage points in Somali literacy, over and above the comparison group. A similar, but less stark, pattern was observed in average numeracy scores (3.9 percentage points over and above the comparison group). In both cases gains among this subgroup were larger than among other out-of-school girls or pastoralist girls who were already in school when the programme started.

Gains were much more limited or non-existent in more advanced literacy and numeracy tasks, and among students at more advanced grades. This finding is consistent with the fact that the largest learning gains were among girls starting from the lowest level of numeracy and literacy in the baseline. These results both reflect the limited capacity of teachers, particularly those without formal training, to deliver more advanced subject content in an effective manner. While the proportion of non-learners has decreased over time, this was more pronounced in lower-level literacy and numeracy tasks. The result is illustrative of the structural barriers to

improving learning outcomes in upper grades among rural Somali girls. While in-service training may be sufficient to ensure major gains in the acquisition of foundational skills, this is not the case for more advanced skills, and there is an urgent need to increase the number of qualified teachers and/or provide alternative means of increasing access to quality delivery – including, but not limited to, remote teaching.

Participation in Girls' Empowerment Forums (GEFs) was associated with significantly higher gains in learning – a difference of 6.6 percentage points over and above the comparison group in terms of aggregated learning scores. GEF participation was also associated with a 4.4 percentage point increase in English literacy, over and above the comparison group. The result highlights how the development of girls' social-emotional skills contributes to learning processes, increasing their ability to participate actively in class, envision non-traditional gender roles for themselves, and build self-confidence and social capital.

Transition

The transition results indicate that the programme has had a substantial impact on transition rates in SOMGEP-T communities, with higher re-enrolment of out-of-school girls, and more in-school girls remaining enrolled and advancing in grade level over time. In the aggregate, from baseline to endline, the programme is associated with a 5.1-point improvement in transition rates among in-school girls. From baseline to endline, the programme produced an 18.0-point improvement in transition rates among the cohort of out-of-school girls. Interestingly, the timing of the impact on in-school girls differed from that on OOS girls. Whereas the timing of the programme's impact on in-school girls was concentrated more heavily in the latter years of implementation, the timing of the impact on OOS girls was concentrated in the first two years of implementation. This is likely because both the ALP and ABE centres had been established and were active by the time of the ML2 evaluation.

Sustainability

Lack of continued government funding and exogenous shocks, such as drought and COVID-19, appear to have had a negative impact on the sustainability of programme impacts. Although the results suggest that Community Education Committees (CECs) are more active and engaged than at baseline, these shocks have affected their ability to raise funds, leaving them to focus more heavily on awareness-raising and school monitoring activities, which do not necessarily require funding. All interviewed Regional Education Officers (REOs) reported gender departments were formed as a part of the programme, but in most cases, these departments have already lost significant funding, been completely defunded, or been absorbed into another part of their respective Ministries.

However, there were also some notable positive impacts captured through the sustainability indicators. As noted above, CECs are reportedly more engaged in communicating with parents, forming school management plans, monitoring schools, enrolling and re-enrolling girls, handling conflicts, and raising funds to cover salaries and school improvement projects when they can. The presence of a critical mass of community champions for marginalised girls' rights among CEC members is a key legacy of SOMGEP-T, particularly in the Somali context, where education is largely financed and managed by communities. Among REOs, awareness of ALP and ABE increased dramatically from midline, and REOs report these programs are largely effective. Additionally, teacher trainings and efforts to focus on female teachers have reportedly led to dramatic improvements in teaching quality and increases in the numbers of female teachers.

Attendance

In the aggregate, SOMGEP-T programming does not appear to have had any impact on attendance rates among girls since the baseline. Classroom headcounts showed no improvement in intervention schools, vis-

à-vis comparison schools, since either the baseline or ML2. School records show a positive impact since baseline but only among the extremely small sample of students who were enrolled and successfully contacted at both baseline and endline; a larger sample of students tracked from ML2 to endline show a negative – though not statistically significant – programme impact. Among caregiver reports of attendance, the programme seems to have had a null or negative impact over both the full programme lifecycle and since ML2.

The COVID-19 pandemic and other exogenous shocks likely affected school attendance and undermined the small gains that were previously documented between baseline and ML2.In a context where girls' workload, malnutrition and health issues have been exacerbated due to the ongoing drought, economic crisis, and COVID-19, gains in attendance are unlikely to be observed. Still, in spite of high absenteeism rates, the participation in the program contributed to higher learning gains among the most marginalised in relation to the comparison group, potentially due to the use of remedial learning strategies.

Teaching Quality

Overall, the results suggest the programme has had a positive impact on teaching quality. Two main indicators were used to measure teaching quality. The first was the use of formative assessments to improve numeracy and literacy outcomes in the classroom. The use of formative assessments has increased from baseline to endline in intervention schools, over and above comparison schools. Additionally, the program exceeded its target, a 60 percent increase in the proportion of teachers using formative assessments, with a final result of 82.4 percent. Gains in the use of active teaching methods in relation to comparison schools were particularly high among teachers exposed to coaching, confirming the validity of the project's Theory of Change.

The second target was to shift teacher perceptions of teaching quality, which was measured in four ways: perception of corporal punishment, equity in expectations of girls and boys, remedial teaching, and gender-sensitive lessons. Improvements were observed for two of the four (corporal punishment, adjusting lessons to help struggling students) topics. Even for the two topics where perceptions worsened over time, the decline for intervention schools (2.6 percentage points) was less than the decline for comparison schools (9 percentage points).

School Management and Governance

School management and governance results differed by group. The results in this section are reflective of the perceptions of two main groups: head teachers and caregivers. Based on the results from the head teacher survey, the number of CECs functioning has dropped, as has the number of functions being carried out by CECs. The opposite result was found for the caregiver survey. Based on the results from the caregiver surveys, both intervention and comparison schools saw increases in all metrics of CEC activities between baseline and endline, including monitoring of teacher attendance, improving school infrastructure, and supporting students financially. Particularly during the last two years of the programme, similar interventions on CEC training were rolled out by system strengthening projects across all three states, including interventions inspired by SOMGEP-T. Therefore, the results observed in this endline differ considerably from those in previous evaluation rounds, showing improvements across both groups (intervention and comparison) but also showing that SOMGEP-T schools had larger improvements in specific areas, such as teacher monitoring.

There were some positive results from the head teacher survey. For example, in the head teacher survey, both intervention and comparison schools saw increases in the percentage of schools that have a management plan, have CEC members visit, and have the following four factors included in school management plans:

monitoring, child protection, enrolment encouragement, and follow-up with students who had dropped out. However, for most indicators in this section it is important to note that comparison schools had a much lower starting point, which led to negative difference-in-differences results.

Life Skills

The programme used two main metrics to measure progress in life skills: the Youth Leadership Index (YLI) and the Life Skills Index (LSI). Based on the results for these two indices, the impact of the programme on life skills is inconclusive. The majority of programme impact results were not statistically significant, but the vast majority did indicate a positive association between the programme and life skills. For example, the mean YLI score for intervention girls was 50.6 points at baseline and increased to 59.6 points at endline, but comparison girls experienced a similar increase over the same period, resulting in an estimate of programme impact of just 0.35 points. Similar results occurred for the LSI: the programme's impact from baseline to endline was positive, though not statistically significant, and this held true when we considered shorter analysis periods, such as baseline to ML2 and ML2 to endline, reinforcing the idea that the programme had a small but consistent impact.

Given the consistency of small positive effect, these results should not necessarily be interpreted to mean the intervention was not effective – rather, the lack of significance may simply be the result of difficulties in isolating programme effects. As noted above, the impact of participation in GEFs – a platform for life skills development – had an outsized impact on learning outcomes. In addition, qualitative data collected through a new 'Girl Networking Tool' provided insights into some of the less tangible gains in terms of social capital; peer support networks fostering increased motivation and agency; and girls' ability to apply newly learned knowledge skills at home and at the community. Last, but not least, intervention schools had a near-significant, 10.6 percentage point larger gain of girls who stated they felt confident answering questions in class, over and above comparison schools, likely reflecting a combination of both girls' increased agency and shifts in teachers' behaviour.

Community Attitudes

In the context of the SOMGEP-T evaluation, community attitudes are expected to have effects on transition rates by encouraging re-enrolment of OOS girls and continued enrolment for girls already enrolled. The endline analysis shows that perceptions by caretakers of the value of a girl's education, the likelihood that she uses it, and whether her opinions are an input to the decision to re-enrol her in school have improved from baseline to endline and have improved to a greater degree in intervention communities. These positive attitudinal shifts occurred largely between the baseline and the first midline, which is not necessarily surprising given several of the indicators rose to such an extent that there was not much room for further improvement.

However, despite the positive attitudinal shifts, there were also some results that were more worrisome. Parents assert less support for girls who perform poorly in school, and just 59 percent of caregivers state that secondary school would help a girl in her role as a homemaker. In the qualitative interviews, respondents mentioned that religious leaders only promote girls' education insofar as it relates to Qur'anic learning and only until the girl is able to marry. The results reflect a transitional moment in Somali society, where conflicting views about girls' and women's roles in the economy and community lives coexist. It is interesting to note, for instance, that the project has had a major impact on learning outcomes among some of the hardest to reach groups – pastoralists, ABE students - but had less impact among ALP students, many of whom are married and have children.

School-Related Gender-Based Violence

For in-school girls, improvements were observed from baseline to midline, but these positive results largely either reversed or slowed down from midline to endline. Whereas in the baseline, the proportion of girls who said they feel unsafe *traveling to school* was much higher in comparison schools, by the endline, the proportion of girls who said they feel unsafe was actually higher in intervention schools. The share of girls who feel unsafe *at school* sharply decreased in both comparison and intervention schools from baseline to midline, but in both areas the proportion had begun to reverse from midline to endline. By the endline, the proportion of girls from comparison schools who reported feeling unsafe at school was slightly lower than the proportion from intervention schools but was roughly the same.

For out-of-school (OOS) girls, the share of caregivers from intervention schools who cited it is unsafe for their girls to be in school and that girls will be mistreated by teachers decreased slightly from baseline to endline, and the share who cited it is unsafe to travel to school and girls will be mistreated by other children increased very slightly. Despite large differences in results between comparison and intervention schools at either the baseline, midline, or both, by the endline, the share of caregivers citing a given reason was roughly the same in intervention and comparison schools. The increased occurrence of violent incidents in Somalia in 2021, as well as the additional vulnerabilities resulting from drought and the COVID-19 crisis, are likely to have contributed to such results.

The qualitative interviews suggest girls still face varied and multiple risks in public spaces in their communities, on the way to school, and in school. However, there was also evidence that programme activities with CECs and GEF girls have been effective in reducing conflict in schools and creating a safer and happier learning environment.

Conclusions

The project's focus on marginalised girls is largely reflected in the greater learning gains achieved among particularly at-risk subgroups and in its positive impact on transition. The findings also confirm the Theory of Change's assumptions regarding the close association between social-emotional skills development and improved learning outcomes for marginalised girls. SOMGEP-T's interventions have resulted in gains in the use of active teaching practices; shifts in social norms at community level; and improvements in the participatory management of schools. The findings provide valuable lessons on what works to reach the most marginalised in a context where access to education remains a major challenge, particularly among nomadic populations and girls. The results also demonstrate the impact of sustainable and affordable approaches, such as investments in participatory management and Girls' Empowerment Forums.

Conversely, the intervention did not have an impact on attendance, reflecting context-specific challenges related to pastoralism, migration and ongoing crises. SOMGEP-T also did not have an impact on higher level numeracy tasks, showing the limitations of the in-service teacher training approach when applied with a population of teachers with limited or no formal qualifications. The latter provides valuable lessons for implementers and government partners, highlighting the extent of the gaps and the need for dramatically different, long-term approaches to transform learning outcomes in upper grades at a moment when the education system is expanding at a fast pace.

Introduction

The Somali Girls' Education Promotion Project – Transition (SOMGEP-T) is a large-scale education programme that began implementation in May 2017 in Galmudug, Puntland, and Somaliland. The programme is funded by the UK's Foreign and Commonwealth Development Office (FCDO) as part of the global Girls' Education Challenge (GEC), which supports programmes specifically designed to improve educational outcomes among the most marginalised girls in developing countries. SOMGEP-T builds on an earlier iteration of GEC programming – the Somali Girls' Education Promotion Project (SOMGEP) – by CARE in an overlapping set of communities. SOMGEP, which was part of GEC phase I, was implemented from 2013 to 2017 and targeted many of the same outcomes as SOMGEP-T, including learning, transition, attendance, school management, and community attitudes. SOMGEP-T worked in the same communities, seeking to learn from CARE's experience in the same communities and extend its impact. Implementation of SOMGEP-T ended after just over 4.5 years, in late 2021, a period in which the programme directly benefitted approximately 20,000 girls and 12,000 boys directly and another 20,000 students, and over 10,000 community members indirectly.

SOMGEP-T targets a range of barriers to girls' educational inclusion and achievement. The programme's core outcomes are: learning, broader defined as literacy, numeracy, and financial literacy; transition, the extent to which girls remain in school, move up in grade levels, or transition into appropriate alternative educational options or employment; and sustainability. A multidimensional programme, it includes a wide variety of interventions, including teacher training, support for better community-based school management, improving community attitudes and norms around girls' education, improving access to alternative education options for post-primary girls, facilitating the development of leadership skills and girls' self-esteem, and supporting relevant government ministries to enhance the inclusion of girls and marginalised students more generally.

Project Context

SOMGEP-T is being implemented in a complex environment, with a number of contextual factors that are important for framing the current educational circumstances facing girls in programme communities. At a broad level, education in Somalia is defined – as with all government services in the country – by the civil war that erupted in 1987 and the eventual collapse of the Siad Barre regime and the central government in Mogadishu in 1991. The varied historical trajectories and experiences during the civil war directly impact current conflict dynamics, government administration, and economic marginalisation in the areas where SOMGEP-T works. Below, we discuss how these factors vary across space and influence educational outcomes; we also discuss contextual factors which link all three regions, especially their recent experience of drought and the impact of COVID-19.

Ongoing Conflicts

While SOMGEP-T communities largely fall outside areas directly impacted by ongoing, widespread conflict, they can be and are buffeted by internecine conflicts of three main types. The first is centred on a territorial dispute between states, which overlap with clan-based disputes in the area. The second type of conflict centres on violent extremist organisations that operate in the region. Several SOMGEP-T communities are impacted, a fact which is reflected by the continued exclusion in the endline of at least one sample school.

The third type is inter-clan conflict that can arise in many areas where the program is being implemented. Clan disputes impact every region, though outright violence tends to be short-lived. Most inter-clan disputes do not come to wide international attention because they occur in rural areas and villages and involve a relatively small number of deaths; however, they disrupt the provision of public services and drive displacement. In addition, the continual exposure to violence – or the latent threat of violence – can have significant mental health consequences even outside of direct, physical harm caused by conflict.

Educational Marginalisation

In light of its history of recent and ongoing conflict, it is not surprising that educational outcomes in Somalia and Somaliland are particularly poor. Across all of Somalia and Somaliland, just 48.4 percent of girls aged 15-19 have *ever* attended school, a finding that represents a marked improvement – among the 20-24 and 25-29 age groups, the share of women who have ever attended school is 34.1 and 21.0 percent, respectively.[1] The modal woman in every age group has completed less than a single year of formal schooling. Literacy rates are extremely low: just 36.1 percent of girls 15-24 years old in Puntland were capable of reading a short, common sentence, in 2011.[2] And, in Somaliland, only 18.3 percent of girls were entering primary school at the appropriate age level.[3]

As discussed in previous evaluation reports, early marriage and higher dropout rates for girls are prime drivers of the gap between female and male educational attainment. In the regions where SOMGEP-T works, 13.2 percent of girls currently under 18 years of age have been or are married. And, consistent with assumptions underlying the SOMGEP-T theory of change, the share of girls attending school is *nearly* equal to that of boys in the youngest ages, with the gender gap beginning to widen at age 13 and accelerating around age 16, with fewer and fewer teenage girls enrolled in school.

Importantly, educational outcomes tend to be worse in rural areas and in the specific regions targeted by SOMGEP-T. For instance, Sool and Sanaag have the lowest share of girls 15-24 years who have ever attended school – as low as 31.7 percent in Sool.[4] Moreover, school attendance in rural areas – and, especially, among nomadic populations – is far worse on average. In the four main regions where SOMGEP-T is implemented, the share of girls 15-24 who had attended any schooling beyond primary school was 31.2 percent in urban areas, 11.9 percent in rural areas, and 0.3 percent among nomadic populations.[5] These disparities – between urban, rural, and nomadic communities, and across regions – are also reflected in gross attendance rates. The female gross attendance ratio in Sool and Sanaag is just 32.8 and 38.8, respectively.[6] Among nomadic

[1] Somali National Bureau of Statistics. 2020. "Somali Health and Demographic Survey 2020." Data provided by SNBS: http://microdata.nbs.gov.so/
[2] UNICEF and Ministry of Planning and International Cooperation. 2011. "Multiple Indicator Cluster Survey – Northeast Zone, Somalia, 2011."
[3] UNICEF and Somaliland Ministry of National Planning, and Development. 2011. "Multiple Indicator Cluster Survey: Somaliland, 2011."
[4] Somali National Bureau of Statistics. 2020. "Somali Health and Demographic Survey 2020." Data provided by SNBS: http://microdata.nbs.gov.so/
[5] Somali National Bureau of Statistics. 2020. "Somali Health and Demographic Survey 2020." Data provided by SNBS: http://microdata.nbs.gov.so/
[6] Puntland Statistics Department, Puntland State of Somalia, *The Puntland Health and Demographic Survey,* p.32.

populations, it is a mere 2.2.[7] Because SOMGEP-T operates primarily in rural communities, the children targeted by the programme face intersecting axes of marginalisation – they live in peripheral regions, neglected by the relevant governing authorities, and they live in rural parts of those regions, where the state's reach and service provision are weakest.

Drought, Population Mobility, and Economic Marginalization

Somalia and Somaliland are currently in the grip of a long-term drought, which has spanned the duration of SOMGEP-T implementation. The most severe impacts of the drought were felt in 2017, during the early stages of SOMGEP-T programming, but poor and inconsistent rainy seasons in the ensuing years has caused continuing humanitarian emergency in much of the region, which has worsened in recent months. According to the Food Security and Nutrition Analysis Unit (FSNAU), the 2021 *deyr* rains (September – December) failed, marking "the third consecutive below-average rainfall season since late 2020."[8] The current drought conditions are most pronounced in central Somalia, but are also severe in areas with large concentrations of SOMGEP-T communities.[9]

The ongoing and worsening drought conditions interact with the broader economic marginalisation of these regions and their lack of diversified livelihoods. The regions in question are deeply dependent on livestock production, with 61.4 percent of households owning livestock of some kind. To illustrate the impact of the current drought, 40.3 percent of the rural or nomadic households in SOMGEP-T regions lost at least one camel due to drought in 2018-2019, a period when the drought was less severe than at present.

Programme regions are disproportionately populated by nomadic households. In Sool, the Population Estimation Survey of Somalia estimated that 57.3 percent of the population was nomadic, and this same rate was 64.8 percent in Sanaag.[10] Even less nomadic regions represented in the sample – such as Mudug and Toghdeer– have nomadic populations comprising more than 20 percent of their populace. Moreover, the rural areas of these regions have much *higher* rates of nomadism, and these are precisely the areas where SOMGEP-T schools are located.[11] The result of a highly mobile population that is highly dependent on rainfall for their livelihoods is that consistent school enrolment is extremely difficult to maintain, especially in the face of drought or conflict.

In essence, SOMGEP-T schools operate in an environment of extreme resource limitations, population mobility, and economic and environmental fragility, even by Somali standards. While geographically-disaggregated data is extremely limited, the Somali context as a whole presents a challenging environment for girls' education and the regional figures presented here suggest that SOMGEP-T communities face many additional challenges with regard to livelihoods, climate resilience, and educational attainment.

The COVID-19 Pandemic

Finally, the newest contextual factor to consider since the completion of the second midline evaluation (ML2) is the outbreak of COVID-19 and its spread through Somalia and Somaliland. The ML2 evaluation took place

[7] Puntland Statistics Department, Puntland State of Somalia, *The Puntland Health and Demographic Survey,* p.32.
[8] FSNAU. 2021. "Somalia Food Security Alert: December 20, 2021." Available at: https://fsnau.org/downloads/Somalia-Food-Security-Outlook-December-2021.pdf
[9] Ibid.
[10] United Nations Population Fund. 2014. *Population Estimation Survey 2014: For the 18 Pre-War Regions of Somalia*.
[11] For example, SOMGEP-T schools in Mudug are concentrated in the dry pasturelands along the Ethiopian border.

immediately prior to the start of the global pandemic, with data collection occurring in late 2019. In mid-March 2020, Somalia had its first documented case of COVID-19, and public health authorities quickly closed down international and domestic air travel, closed schools, and put in place limits on internal movement and group gatherings. Somaliland experienced its first documented case around the same time, implementing quarantine measures for new arrivals and closing schools, but did not aggressively limit group gatherings or restrict internal movements.

Across both Somalia and Somaliland, there have been 23,532 confirmed cases thus far, with 1,333 recorded deaths.[12] These figures hint at a relatively low direct impact of COVID-19 in Somalia, but relatively low mortality as a share of total population is misleading, given the limited health surveillance systems even in major urban areas. In practice, the mortality burden is much higher than that reported by national health authorities, with a marked increase in burials in, for instance, Mogadishu, over the period January to September 2020.[13] Statistics on positivity rates also point to a significant impact of the pandemic, as do findings from other education evaluations in Somalia – for instance, the Educate Your Children II programme's baseline study reported that 57 percent of households in Puntland self-reported a death or deaths due to COVID-19.

The most direct impact on the education sector arose from the mandated school closures in both Somalia and Somaliland. Schools in Somalia were closed for approximately five months, from March to early September 2020 and closed again in early 2021.[14] In Somalia, some variation in the duration of closures should be expected, because most education providers are privately funded; while the closure mandate would keep schools closed, there is little guarantee that schools would reopen immediately upon easing of the mandate. In Somaliland, schools were closed for a similar period, approximately five months.[15] The tenuous enrolment status and attendance rates of many students in rural schools, when coupled with this long-term disruption in regular learning, should be expected to have major impacts on enrolment and transition rates.

A less direct, but still important, impact on the education sector stems from the economic effects of the pandemic. The livestock sector, in particular, was badly hit because it is driven by livestock exports to Saudi Arabia, particularly during the Hajj season. The suspension of pilgrimage visits in 2020 and continued restrictions in 2021 reduced demand for livestock exports that dominate the economy of Somaliland, Puntland, and Galmudug.[16] Remittance inflows also declined in response to the pandemic, with some sources stating that remittances fell by more than 50 percent.[17] Given the centrality of remittances to household resilience and households' ability to pay for educational expenses, this impact is particularly acute. Combined

---

[12] As reported by the Somalia Ministry of Health (MoH), via Twitter, on December 23, 2021. Note that cases in Somaliland are tracked by the Somaliland Ministry of Health Development (MoHD) but the Somalia MoH includes Somaliland's cases in its own count as well. For simplicity, we use the aggregate numbers reported by Somalia's MoH.

[13] Warsame, Abdihamid, Farah Bashiir, Terri Freemantle, Chris Williams, Yolanda Vazquez, Chris Reeve, Ahmed Awels, Mohamed Ahmed, Fransesco Checchi, and Abdirisak Dalmar. 2021. "Excess Mortality During the COVID-19 Pandemic: a Geospatial and Statistical Analysis in Mogadishu, Somalia." *International Journal of Infectious Disease* 113: 190-199.

[14] UNICEF. 2021. "Puntland – Learning Continuity during COVID-19 and Beyond." August 27. Available at: https://www.learningpassport.org/stories/puntland-learning-continuity-during-covid-19-and-beyond.

[15] Heering, Eric, et al. 2020. "COVID-19 Responses and Education in Somalia/Somaliland." *Sociology Mind* 10: 200-221.

[16] Mtimet, Nadhem, Francis Wanyoike, Karl M. Rich, and Isabelle Baltenweck. 2021. "Zoonotic Diseases and the COVID-19 Pandemic: Economic Impacts on Somaliland's Livestock Exports to Saudi Arabia." *Global Food Security* 28.

[17] International Organisation for Migration. 2020. "COVID-19 and the State of Remittance Flows to Somalia." Available at: https://www.iom.int/sites/g/files/tmzbdl486/files/press_release/file/covid-19_and_the_state_of_remittance_flow_to_somalia_-_iom_somalia_august_2020.pdf.

with the damage wrought by locusts during 2020 and the effect on staple food prices, the economic situation has worsened significantly since the ML2 evaluation, prior to the start of the global pandemic.

Methodology

Evaluation Questions

SOMGEP-T is being implemented as part of the broader GEC-T programme, an FCDO-funded initiative that continues support to individual country-level programmes from an earlier round of GEC interventions. In the case of SOMGEP-T, it builds on the Somali Girls' Education Promotion Project (SOMGEP), which was implemented between 2013 and 2016. Being part of the umbrella GEC-T programme shapes the outcomes of interest to the evaluation and the metrics which are used for assessing impact. The research questions addressed in this evaluation are tightly focused on two goals. The first set of questions focus on whether SOMGEP-T has had a discernible impact on intervention girls, schools, and communities, across many different core and intermediate outcomes. The second set of questions are focused on understanding the nature of impact, how it is mediated, through which activities and mechanisms it operates, whether some girls are being left behind or could be better targeted, and whether gains made are likely to be sustained past the end of active SOMGEP-T implementation.

The first set of questions are broader and generally employ similar methodologies for obtaining answers. These questions include:

- What impact did the GEC Funding have on the transition of marginalised girls through education stages and their learning?
- To what extent did the intervention result in additional gains in learning (literacy and numeracy) among the intervention group, in relation to the comparison group?
- What impact has the program had on intermediate outcomes in the intervention group, relative to the comparison group.

The second set of questions are slightly more varied:

- What works to facilitate transition of marginalised girls through education stages and increase their learning?
- How sustainable were the activities funded by the GEC and was the program successful in leveraging additional interest and investment?
- Is there a significant difference between the acquisition of literacy/ numeracy/ English skills among ASLP participants and marginalised girls enrolled in formal secondary school?
- Are the intermediate outcomes identified by the project contributing to the accelerated acquisition of literacy/ numeracy skills and improved transition rates? Are there intermediate outcomes that do not seem to be influencing outcomes at all? Do the findings support the ToC or challenge its assumptions?
- What are the key factors influencing the acquisition of literacy, numeracy and English language skills?
- What are the specific literacy/ numeracy/ English competencies that marginalised girls are lagging behind on?

- Is there a difference in the rate of acquisition of literacy / numeracy subtasks that girls are able to practice in their daily lives, vis-à-vis the acquisition rate of subtasks that are not used on a regular basis by the girls targeted by the project?
- To what extent are extremely marginalised sub-groups, such as pastoralist girls and disabled girls, able to attain basic competencies in literacy, numeracy and English? Are there other sub-groups who are demonstrating a pattern of lagging behind in performance? What are the potential reasons for this pattern?
- Is there a relationship between the acquisition of leadership skills and learning outcomes? If yes, how?
- To what extent the participation of mothers in VSL may influence the acquisition of numeracy and financial literacy skills by girls?
- What are the key factors influencing transitions to more advanced levels of education?
- To what extent are extremely marginalised sub-groups, such as pastoralist girls and disabled girls, able to transition into upper primary/ post-primary education? Are there other groups who are lagging behind in transition rates? What are the potential reasons for this pattern?
- To what extent is the acquisition of leadership skills influencing transition outcomes?
- Did the intervention contribute to a shift in traditional gender norms and power relationships at the household and community levels? If yes, what types of changes have occurred? How are these changes affecting adolescent girls and boys?
- How did the intervention affect boys' learning and retention?
- What are the key changes identified by the girls themselves in terms of their capacity to engage in non-traditional roles at the household, school and community? To what extent are those claims supported by quantitative evidence?
- Is there a difference in the learning outcomes for students targeted in areas heavily affected by drought, compared to those that were less affected? Likewise, is there any difference for transition?

The overall evaluation design, described in the next section, was designed with these questions in mind.

Evaluation Design

In this section we provide a brief review of the methodological design of the SOMGEP-T evaluation from baseline through this endline evaluation. Few aspects of the design have changed appreciably over the course of the evaluation. In previous reports, we have described the evaluation design in significant detail, and we refer readers to the baseline and ML1 reports, in particular, for greater detail on several aspects of the design, including the joint sampling approach and the matching methods used to identify matched-pairs of schools between the intervention and comparison groups. In this section we provide background on the overarching design, its advantages, and limitations but do so without extensive technical detail.[18]

SOMGEP-T employs a quasi-experimental difference-in-differences design, with pre-selected comparison and intervention groups of schools. Prior to the baseline, CARE identified a set of schools where their interventions would be implemented. These schools became the intervention population. At the same time,

[18] Previous evaluation reports are available on the Girls' Education Challenge website, using the following links:
Baseline: (https://girlseducationchallenge.org/media/xy0b1ufb/somgep-t-gect-baseline-evaluation.pdf);
Midline #1: (https://girlseducationchallenge.org/media/0acewdvh/somgep-t-gect-midline-evaluation-ii.pdf);
Midline #2: (https://girlseducationchallenge.org/media/kcujhfvw/somgep-t-gect-midline-evaluation.pdf).

CARE identified a pool of schools where the programme would not be implemented, and which would be suitable as possible comparison schools. As we note below, the distinction between intervention and comparison schools is not precise, from the perspective of experimental ideas of compliance and spillover, because comparison schools have potentially benefitted from *some* SOMGEP-T interventions, especially those implemented at a system level, rather than within communities or schools. Nonetheless, the design allows us to assess the impact of the overall suite of SOMGEP-T interventions, while recognising that spillover of this kind may dampen our estimates of programme impact.

From the pool of intervention and comparison schools, the evaluation team constructed matched pairs – one intervention school matched to the comparison school to which it was most similar. Intervention and comparison schools were matched on the basis of geographic zone, school size, the existence of a Community Education Committee, and whether the school was receiving assistance from another NGO. Only rural schools were considered for sampling. In total, 38 matched pairs of schools were identified, which were balanced exactly in terms of zone, and nearly balanced in terms of school size and the receipt of assistance from other NGOs.[19] The matched-pair design has lost *some* of its value over time, because – as we discuss in more detail in the Sampling Methodology section, below – there has been school-level attrition from the sample over time. In total, one intervention school was lost from the sample between baseline and endline, compared to six comparison schools. This has resulted in less balance between the groups in terms of the schools' pre-existing characteristics, because when a school fell out of the sample, we did not drop its pair-matched peer, in an effort to preserve the available sample size.

Importantly, any imbalance between intervention and comparison groups in terms of pre-existing characteristics – either due to the limitations of our ability to match schools or the differential attrition that has occurred since baseline – is controlled explicitly through the difference-in-differences design. This design looks for differences between intervention and comparison groups in their trends from baseline to endline. The analysis does not rely on simple differences between the two groups at endline or any other point in time. Rather, we investigate whether – for instance – girls in intervention schools improved more than girls in comparison schools, regardless of their relative starting point at baseline. Indeed, the design is robust to a number of sources of bias that plague pre-post analyses without a comparison group, specifically:

- **Pre-intervention differences between intervention and comparison groups –** if the intervention and comparison group have different baseline performance in terms of learning or transition, this gap is controlled for explicitly in the design. If project schools already had higher learning achievement prior to the program's start, a simple comparison between intervention and comparison schools at endline would not be valid, because intervention schools were *already* higher-achieving than their comparison counterparts. By accounting for baseline differences in learning outcomes, this source of bias is eliminated.
- **Systematic changes in outcomes over time that are not attributable to the project –** if the areas where SOMGEP-T is being implemented experience a broad-based change, such as multiple productive rainy seasons in a row, this would bias simple pre-post analyses of program impact. In the

[19] The matching method employed – Coarsened Exact Matching (CEM) – matches units within a set of bounds, rather than enforcing exact matches. This means the method can often produce less-than-perfect balance; its strength is that it generates a greater number of matches and that it can be used to match units on continuous variables or variables with many discrete values, because exact matches are not required. In our use case, the initially stricter matching requirements – matching across a wider range of variables – produced too few matches, and the set of matched variables had to be reduced. Thus, CEM was well-suited to our use, because a more exacting method would have produced even fewer matched-pairs.

example given, we might expect enrolment and attendance rates to rise broadly in all areas where the rains have been reliably good. If the evaluation used a pre-post design without a comparison group, and the analysis revealed improving transition rates from baseline to endline, this increase would be incorrectly attributed to the program, when it may have arisen from the improved rains. By incorporating a comparison group, the SOMGEP-T design allows us to identify general trends in the area and isolate those from program-specific impacts.

These two advantages motivate most uses of difference-in-differences as a method for programme evaluation, and it provides a comparatively rigorous base on which to draw inferences about the impact of SOMGEP-T on learning, transition, and other outcomes.

The overall evaluation design outlined here has not changed appreciably since the baseline. Three substantive changes are worth describing in more detail, to understand their impact on the design's rigour. The first change made to the primary design has been in the sample itself, which was adjusted at the school level in the manner noted above. This shift has minor methodological implications, insofar as the intervention and comparison groups are slightly less well-matched than they were at baseline; as noted above, however, the difference-in-differences approach does not *require* matched-pairs and is, by design, robust to baseline differences between the two groups.

The second change concerns spillover of SOMGEP-T interventions into comparison communities. SOMGEP-T is an integrated programme, consisting of interventions at the system, community, school, and individual levels. System-level interventions target MOEs and their staff, who also interact with and oversee comparison schools, resulting in exposure of comparison schools to, for instance, Regional Education Officers (REOs) with enhanced capacity. In addition, the success of CEC-focused programming has prompted other programmes to engage in CEC capacity-building, which may impact comparison schools. Spillovers of this kind produce bias *against* finding programme impacts and we note this, at times, when discussing specific null results in the report.

The third change is much broader and was an adjustment in the overall programme, in response to learning from SOMGEP and the baseline SOMGEP-T evaluation. The programme was expanded in important ways in the form of ABE and ALP programming. This shift affected the evaluation design to the extent that it necessitated bringing in additional cohorts of girls. However, the recruitment of ALP and ABE girls into the evaluation does not influence the primary design used for studying learning, transition, or intermediate outcomes, as these additional samples are treated separately from the primary cohort recruited at baseline.

Data Collection Tools

The design of this endline evaluation closely matches that of the three previous evaluation rounds – baseline, ML1, and ML2. The consistency maintained across rounds extends to the data collection tools employed; by maintaining the same tools over time and making changes to the tools sparingly and carefully, the comparability of data across rounds has been maximized to the extent possible. A somewhat fuller description of each tool has been provided in the ML2 report; the full survey scripts and qualitative interview guides are annexed to this report. Briefly, the quantitative tools included:

- Household Survey – modules with girls tracked as part of our learning and transition cohorts, their head of household, and their primary caregiver.

- Learning assessments – covering numeracy, Somali literacy, English literacy, financial literacy, and working memory. The numeracy assessment is based on the Secondary Grade Maths Assessment (SeGMA) while both literacy assessments were based on the Secondary Grade Reading Assessment (SeGRA).
- Classroom Observation – capturing classroom environment, teaching quality, and pedagogy.
- Attendance Headcounts
- Head Teacher Survey – capturing school characteristics, record-keeping and policies, and attendance of cohort girls.

Importantly, the same girls who complete the household survey also complete the learning assessment, as part of the joint sampling methodology for tracking learning and transition among the same set of girls.

Small adjustments have been made to the quantitative tools – primarily the household survey – over time. At endline, the main adjustments involved incorporating additional questions on community attitudes toward girls' education. Specifically, during the ML2 evaluation we noted that the tendency for community members and parents to report high support for girls' education, in the abstract, might be because they were imagining a high-achieving girl. If the cognitive frame used for the question is a high-achieving girl, this would make it easy to report a strong belief in the value of girls' education.[20] In response, the household survey and head teacher survey were revised to address this specific question, using both direct questions and survey experiments designed to test hypotheses about how the frame – a high- or low-achieving girl – can influence opinions.

Additional revisions were made to the household survey to shed light on transition rates, especially among girls who had previously fallen out of the panel sample, e.g., a girl who was in the baseline, was not located during the ML1 or ML2 rounds, but was successfully brought back into the sample at endline. For such a girl, her transition data will be extremely incomplete. In this round, we added retrospective questions regarding a girls' enrolment status and grade level for each year from 2017 to 2021 to fill in gaps of this kind and provide an additional data source for assessing transition rates.

As in previous rounds, tool revisions were made with attention to the goal of making comparisons across time. In past rounds, revisions were possible because we could keep those revisions in place in later rounds, making comparisons across those rounds where the revisions were present. For instance, tool revisions made in ML1 were kept in ML2 and endline, facilitating ML1-to-ML2-to-EL analysis, but did not allow analysis using that question from baseline to ML1. Because this is an endline evaluation, any additions to the quantitative tools made in this round do not allow comparison backward to earlier rounds. In general, this means that revisions were motivated by a desire to understand a particular dynamic that could be studied exclusively using cross-sectional data at the endline (e.g., differential support for girls' education as a function of the girls' perceived performance) or by the goal of "looking backward" with retrospective questions (e.g., capturing past enrolment status to facilitate additional analysis of transition outcomes).

The table below documents the achieved sample size for each quantitative tool at baseline and endline. The main numbers reflect the overall total of completed observations for each tool; the numbers in parentheses

[20] The ML2 evaluation provided impressionistic evidence that respondents were interpreting questions in this way or basing their answers on "best-case assumptions" about the girl in question, based on qualitative interviews in which interviewees often focused on a single high-achieving girl in their village and argued that girls' education had now caught up completely to boys' education.

reflect the number of observations contained in the "panel sample" from baseline to endline, which is the set of comparable observations actually utilised for baseline-to-endline analysis.[21] As the table shows, the panel sample of household surveys is significantly smaller than the overall sample; we unpack the evolution of the cohort girl samples in more detail in our discussion of re-contact rates in a later section.

TABLE 1: ACHIEVED QUANTITATIVE SAMPLE AT BASELINE AND ENDLINE

Tool	Baseline Observations (Bl-EL Panel Observations)	Endline Observations (BL-EL Panel Observations)
Household Survey and Learning Assessments	1741 (965)	1802 (965)
Head Teacher Survey	76 (68)	68 (68)
Classroom Observation	152 (131)	127 (127)
Headcounts	506 (422)	499 (452)

The qualitative tools underwent more significant revisions than the quantitative tools. First, KIIs with Girls with Disabilities (GWDs) were removed from the evaluation. In their place, the number of participatory exercises with girls – especially members of the Girls Empowerment Forums – were expanded.[22] At ML2, we conducted four risk mapping and four vignette story-telling exercises with these girls; at endline, this was expanded to ten of each type of exercise. A new participatory exercise, designed to understand girls' social and support networks, was added. In addition, the tools were heavily revised by CARE and the evaluation team, in line with methodological recommendations made during the ML2 round. These changes were designed to shift their focus toward assessing impact and studying differential programme impact across subgroups, including those that are marginalised. In total, 70 qualitative interviews were completed – ten each with the following respondent types:

- FGD with Community Education Committee (CEC) members
- FGD with teachers
- FGD with mothers
- KII with Ministry of Education (MOE) officials

[21] The "panel sample" is most clearcut in the context of the household survey, where the panel is defined as the set of girls who appeared in both the baseline and endline data. For headcounts and classroom observations, the panel is defined by inclusion of the *school*, not the specific classroom. For instance, if headcounts were conducted in a given school at both baseline and endline, the school is considered part of the "baseline-to-endline panel" for analysis of the headcount data.

[22] This decision was made for both substantive and methodological reasons. Substantively, during earlier evaluations the programme and evaluation teams noted that there would be value in capturing additional qualitative data on girls' own perceptions of GEF participation, the benefits of peer networks, participating in school-based projects and girl-led activities, and the importance of social capital. Participatory exercises were both expanded in number from previous rounds and a new participatory exercise was designed to help capture these outcomes, but this necessitated reducing the number of other qualitative interviews. From a methodological perspective, the evaluation team found the KIIs conducted with GWDs in earlier rounds difficult to implement and interpret. Direct, targeted questioning regarding the challenges they face at school was not possible, because direct questions imply challenges where the girl may not perceive any and could produce harm as a result. Less direct questions were necessary, but produced fewer concrete findings, reducing the overall value of the KIIs.

- Participatory risk mapping with girls who are members of the Girls Empowerment Forum
- Participatory story-telling exercises (vignettes) with girls who are members of the Girls Empowerment Forum
- Participatory girls' network mapping exercises with girls who are members of the Girls Empowerment Forum

Sampling Methodology

School Sample

The discussion of the overall evaluation design, above, described the approach to selecting schools at the baseline, which utilised pair-matching to select the intervention and comparison groups. In total, 76 schools were sampled at baseline. We refer readers to the baseline evaluation report for more details on the initial sample selection.

Since baseline, the school-level sample has shrunk somewhat, as schools were dropped from the sample for a variety of reasons. No replacement schools were selected (replacement schools are distinct from the expansion of the sample to include ALP and ABE centres). In total, 7 schools fell out of the sample in a manner that made their attrition, effectively, permanent. First, five schools were dropped from the sample because they were outliers in terms of learning scores at baseline.[23] All five schools were in the comparison group, and it was decided to exclude them going forward. Two additional schools were removed from the sample for accessibility and security reasons, including intra-clan conflict. Finally, a number of schools were not visited during the ML1 round, but this reflects the smaller scope of the ML1 round; those schools excluded were not removed permanently and were brought back into the sample in ML2. The school-level sample from ML2 was replicated precisely during the endline, with no further attrition or re-incorporation of schools from earlier rounds.

With regard to ALP and ABE centres, no changes to the sample occurred from ML2 to endline. In fact, the centre-level samples were maintained across all applicable rounds: the same 32 ALP centres were visited at ML1, ML2, and endline; the same 35 ABE centres were visited at both ML2 and endline.

Sample Characteristics

During the analysis conducted in later sections of this report we will utilise varied samples or subsets of girls to facilitate our difference-in-differences regressions and assist in determining programmatic impact. In this section, we will investigate the differences between these samples and subsets with the aim of determining how these differences will impact our analysis. We will make three separate comparisons as detailed below:

Baseline to Baseline-Endline Panel. As we note during individual analyses, we utilise a variety of subsamples to make the most rigorous possible comparisons to answer a given research question. In the context of the household survey and learning assessments, this consists of different subsamples of girls,

[23] We discuss this issue in greater detail in the baseline and ML1 reports. In short, these schools exhibited especially high baseline scores, particularly in terms of English literacy. Together, the evaluation team and CARE's monitoring and evaluation staff followed up in detail and found that English literacy scores in these schools were likely driven by the impact of specific teachers and the fact that instruction was largely in English, both of which set them apart from their matched schools.

typically comprising alternative "panel samples" of girls who were tracked across at least two rounds. Our primary panel sample in this endline is the baseline-to-endline panel, which consists of girls from the baseline sample who were successfully recontacted during the endline survey. Note that, in order to appear in this sample, girls do not need to have been interviewed successfully during the midline #1 and midline #2 evaluation rounds.

The panel samples in question include many fewer girls than the original baseline sample drawn in 2017, due to school- and individual-level attrition. Although replacements were recruited into the sample for many of the girls who could not be re-contacted, replacements still reduce the available sample size for panel data analysis, because the girl herself has changed. For instance, a girl who fell out of the sample between ML2 and the endline would not be eligible for inclusion in the BL-to-EL panel analysis, because she did not appear in the endline data.

In the discussion that follows, we focus on the baseline-to-endline panel sample, because it is the focus of the majority of our analysis in this report. The comparison we make here is between the full baseline sample – which included 1,741 respondents – and baseline samples that have been affected by attrition in later rounds. We first compare outcomes and characteristics within the full baseline sample to outcomes in the same sample, after removing the seven schools that fell out of the sample over time. In both cases, we report baseline outcomes; this comparison allows us to see how school-level attrition affected the composition of the tracked sample.

The second comparison we make is between the full baseline sample and the baseline-to-endline panel sample. Again, we report baseline outcomes and characteristics for both sets of girls. This comparison allows us to see the impact the combination of school- and individual-level attrition has had on sample composition at a fixed point in time, prior to any programme impacts. The overall baseline sample included 1,741 girls; school-level attrition after the fact – i.e. in later rounds – reduced the available sample to 1,581; individual-level attrition between baseline and endline reduced the sample further to 965 girls who appear in both the baseline and endline rounds of data collection.

The goal of these comparisons is twofold. First, we hope to demonstrate the importance of utilising a panel data set in our comparisons. We also will show how attrition across survey rounds has affected the composition of our sample. This will allow us to understand the risks of utilising the various panels.

Intervention / Comparison. When investigating programmatic impact, we will often compare girls in intervention schools to girls in comparison schools. In this sub-section we will investigate systematic differences between the two groups that would impact our difference-in-differences estimations. It is important to note that the difference-in-differences methodology accounts design controls for systematic changes in outcomes over time that are not attributable to the programme itself. This means that by incorporating an explicit comparison group, the difference-in-differences approach is able to control for systematic shocks that affect girls in both intervention and comparison schools. Therefore, this comparison is not concerned with pre-existing differences between the two groups but instead differences that may influence how a girls' learning or enrolment changes in response to shocks.

Baseline to Endline Panel / ABE Girls / ALP Girls. The cohorts of girls in this comparison will comprise the main panel datasets utilised in the analysis of this report. While the groups themselves are defined in other section of this report, here we will analyse how they differ from each other among the selected variables.

The comparisons discussed above will be made using a wide selection of covariates. Each comparison will be made using a slightly different subset of variables based on the goals of the comparison being made. However, the three main groupings of variables that will be utilised are discussed below. The exact variables used in each of the subsections and the selection criteria used will be discussed in greater depth later in this section.

Demographic Characteristics. These variables include household characteristics such as marital status, education, and being an orphan as well as information on food security, disability, income, and proxies for wealth.

Household Barriers to Education. Variables in this group are household characteristics that may impact a girl's access to education. These include a high chore burden, a far distance to the nearest school, or a lack of familial support for education.

School-Oriented Barriers to Education. This group of variables includes factors at a girl's school that may impact her access to education. These include if a girl feels unsafe travelling to school, if a teacher uses physical punishment, or if there are no learning materials available at the school.

Full, School-Level Attrition, and Individual-Level Attrition Samples

As discussed above, the main purpose of this sub-section is to show how attrition across survey rounds has impacted the composition of our primary sample, which is a panel of girls from baseline to endline. This will help us determine the risks associated with using the various panels in our later analysis. This analysis exclusively uses data collected in 2017 at the baseline; thus, the comparisons between samples are all based on data collected at the same time. The first column reports outcomes among the full baseline sample, as it stood at baseline. The second and third column reports outcomes among smaller and smaller versions of the baseline sample, removing the girls who would eventually fall out of the panel sample due to school- and individual-level attrition.

TABLE 2: BASELINE OUTCOMES AMONG THE FULL BASELINE SAMPLE AND SUBSAMPLES AFFECTED BY ATTRITION

	Full Baseline Sample (N=1741)	Baseline Sample with School-Level Attrition (N=1581)	Baseline Sample with Individual-Level Attrition (N=965)
Living without both parents	11.2%	11.4%	9.7%
Orphan	0.3%	0.4%	0.5%
Female HOH	47.6%	46.0%	42.9%
HOH – No formal education	72.6%	71.8%	69.8%
Caregiver – No formal education	83.0%	82.9%	82.7%
Currently married	3.6%	3.7%	3.2%
Ever married	4.2%	4.3%	3.6%

Is a mother	2.4%	2.5%	2.0%
Became a mother before age 16	0.3%	0.4%	0.1%
Household Characteristics			
Pastoralist HH	10.7%	11.0%	10.7%
HOH – no wage earning	42.7%	41.7%	38.5%
Moved in past 12 months	3.2%	3.2%	2.3%
Seasonal migration	6.6%	6.8%	5.6%
Owns camels	10.3%	10.9%	9.3%
Owns medium-sized livestock	57.4%	59.9%	59.2%
Owns land	69.3%	69.9%	70.8%
Informal house / temporary	4.6%	4.6%	4.4%
Poor quality roof	32.0%	34.1%	32.2%
Food Security			
Sleeps hungry many / most days	11.9%	12.0%	10.2%
No water many / most days	31.1%	31.4%	28.7%
No medicines many / most days	43.3%	42.9%	40.1%
Disabilities and Impairments			
Girls with any disability	6.4%	6.3%	6.7%
Girls with any disability except mental health	1.8%	1.8%	1.9%
Vision impairment	0.3%	0.4%	0.3%
Hearing impairment	0.3%	0.4%	0.3%
Mobility impairment	0.6%	0.6%	0.6%
Cognitive impairment	0.5%	0.5%	0.5%
Self-care impairment	0.4%	0.4%	0.2%
Communication impairment	0.7%	0.8%	0.5%
Mental health impairment	17.6%	17.9%	16.8%
Anxiety	14.6%	14.9%	13.8%
Depression	14.2%	14.0%	11.5%
Other Characteristics			

Language of instruction not Somali	8.0%	5.9%	6.2%
Affected by drought	47.4%	49.3%	48.9%

The table documents two broad themes that should be borne in mind during our analysis. The first is that school-level attrition accounts for comparatively less attrition – just 160 girls fell out of the sample due to school-level attrition – than does the loss of individuals from the sample. Indeed, the samples before and after seven schools were removed are remarkably similar to one another in most ways. Disability rates, food security, and measures of household asset ownership are all similar between the two subsamples, in part because school-level attrition accounts for such a small share of attrition, in the aggregate.

The second theme is that individual-level attrition, which reduces the sample size another 616 respondents, leaving our baseline-to-endline panel data consisting of 965 total girls. As the table shows, girls who remain in the panel sample from baseline to endline are slightly less likely to have a disability and have slightly better household finances and asset ownership. With that said, the gaps documented between the full baseline sample and the baseline sample that remains after attrition that has cumulated over multiple years are relatively small and not statistically significant.

Intervention and Comparison Girls

As discussed above this comparison will investigate systematic differences between the girls in intervention and comparison schools that would impact our difference-in-differences estimations. We will therefore be utilising variables for this comparison that may influence how a girls' learning or enrolment changes in response to shocks. The comparison presented in the table below uses the school-oriented barriers to education as well as the household barriers to education and data on shocks.

TABLE 3: COMPARISON OF INTERVENTION AND COMPARISON GIRLS

	Intervention (N=518)	Comparison (N=447)
Caregiver believes travel to school is unsafe for girls	3.0%	3.0%
Girl feels unsafe travelling to school	2.4%	0.8%
Doesn't feel safe at school	1.5%	1.2%
Difficult to move around school	17.5%	18.1%
Doesn't use drinking water facilities	10.3%	6.7%
Doesn't use toilet at school	13.0%	15.7%

Doesn't use areas where children play/socialize	53.6%	49.8%
Disagrees teachers make them feel welcome	2.9%	2.4%
Agrees teachers treat boys and girls differently in the classroom	28.7%	25.0%
Agrees teachers often absent from class	21.2%	20.4%
Afraid of teacher	67.6%	72.4%
Does not feel comfortable asking teacher questions	2.9%	1.2%
Teacher punishes students who get things wrong in a lesson	54.0%	52.0%
Teacher uses physical punishment	53.9%	52.2%
No computers at school	84.0%	83.5%
Cannot use books or other learning materials at school	10.6%	13.9%
Not enough seats for all students	5.9%	5.5%
Caregiver says principal is of poor quality	2.3%	2.6%
Caregiver says teaching is of poor quality	0.6%	0.8%
Girls says teacher does not ask girls and boys an equal number of questions	8.0%	6.3%
Girl says teacher does not ask girls and boys questions of equal difficulty	10.9%	11.5%
Household Barriers to Education		
High chore burden	23.8%	27.1%
Doesn't get support from family to stay in school	3.7%	5.9%
Girl feels that she has no choice whether to attend or stay in school	81.4%	77.1%
Girl feels family makes schooling decisions for her	24.3%	31.3%

Distance to school is greater than 30 minutes	1.8%	2.7%
Caregiver and family members are not involved in CEC	17.7%	17.4%
Shocks		
School affected by conflict	0.0%	7.8%
Affected by drought	49.2%	48.5%

Overall, the characteristics of girls in intervention schools and girls in comparison schools are remarkably well balanced. However, there are several places where there are small differences between the groups. The first is if a girl's school has been affected by conflict, 7.8 percent of girls in comparison schools reported that their school had been impacted by conflict while 0 percent of girls in intervention schools reported the same. Similarly, 72.4 percent of girls in comparison schools reported being afraid of their teacher while only 67.6 percent of girls in intervention schools reported the same. Finally, 2.4 percent of girls in intervention schools reported feeling unsafe traveling to school while only 0.8 percent of girls in comparison schools reported feeling unsafe during the same trip.[24] Importantly these differences do not all trend the same direction, or for example, girls in one group are not systematically more afraid of their living environment than girls in the other group. This again points to balance between the groups among the selected covariates and is favourable for our ability to compare the two groups using a differences-in-differences methodology.

Baseline Panel, ABE Girls, and ALP Girls

The comparison presented in the table below represent our main cohorts for analysis in this report. Therefore, we will utilise all three data types in our comparison, demographic characteristics, household barriers to education, and school-oriented barriers to education. Based on the definitions of these groups we expect them to differ in significant ways, however understanding the nature of these differences is important to our analysis.

TABLE 4: COMPARISON OF BASELINE PANEL, ALP GIRLS, AND ABE GIRLS

	Baseline / Endline Panel (N=965)	ALP Girls (N=211)	ABE Girls (N=347)
Living without both parents	25.2%	20.4%	8.4%

[24] This gap is interesting, in part, because relatively few caregivers in the endline reported that conflict had occurred in their communities in the past year. Despite the lack of reported conflict, girls do not always feel safe on their way to school. One explanation – consistent with other research we have conducted in Somalia and Somaliland – is that respondents interpret questions around conflict to be asking about outright, widespread conflict, as opposed to internecine, small-scale tensions and occasional killings across clan lines. A second possible explanation is that the communities in question may be relatively safe from conflict, but that girls understand well the risks they face outside of direct conflict, including harassment and sexual assault. Girls may feel unsafe for reasons unrelated to outright conflict.

Orphan	0.9%	0.0%	0.0%
Female HOH	46.1%	38.9%	42.5%
HOH – No formal education	66.0%	60.9%	79.4%
Caregiver – No formal education	73.2%	69.7%	87.3%
Currently married	13.3%	24.2%	0.9%
Ever married	16.7%	28.4%	1.2%
Is a Mother	2.0%	23.7%	0.6%
Became a mother before age 16	0.1%	1.0%	0.0%
Household Characteristics			
Pastoralist HH	9.6%	9.5%	5.8%
HOH – no wage earning	43.9%	40.8%	54.9%
Moved in past 12 months	2.3%	0.9%	1.2%
Seasonal migration	7.9%	2.4%	1.4%
Owns camels	14.0%	13.7%	9.8%
Owns medium-sized livestock	63.0%	67.3%	62.1%
Owns land	68.0%	68.6%	70.4%
Informal house / temporary	9.0%	6.2%	5.1%
Poor quality roof	21.6%	36.0%	46.8%
Food Security			
Sleeps hungry many / most days	7.4%	7.6%	11.0%
No water many / most days	24.3%	31.3%	34.1%
No medicines many / most days	31.6%	52.4%	40.1%
Disabilities and Impairments			
Girls with any disability	10.7%	9.5%	7.2%
Girls with any disability except mental health	5.4%	2.4%	2.6%
Vision impairment	1.9%	0.5%	1.2%
Hearing impairment	0.4%	0.0%	0.0%
Mobility impairment	0.6%	0.5%	0.3%
Cognitive impairment	0.2%	0.0%	0.0%
Self-care impairment	0.2%	0.0%	0.0%

Communication impairment	0.2%	0.0%	0.0%
Mental health impairment	5.3%	7.6%	5.2%
Disability of arms/hands	2.3%	1.9%	1.4%
Anxiety	3.6%	5.7%	4.6%
Depression	3.7%	5.2%	3.5%
Shocks			
Affected by drought	48.9%	24.6%	46.1%
School-Oriented Barriers to Education			
Caregiver believes travel to school is unsafe for girls	3.0%	8.9%	2.1%
Girl feels unsafe travelling to school	1.7%	2.7%	2.0%
Doesn't feel safe at school	1.3%	0.5%	1.5%
Difficult to move around school	17.7%	17.8%	16.0%
Doesn't use drinking water facilities	8.8%	21.7%	19.2%
Doesn't use toilet at school	14.2%	20.7%	23.3%
Doesn't use areas where children play/socialize	52.0%	52.2%	35.5%
Disagrees teachers make them feel welcome	2.7%	2.7%	5.8%
Agrees teachers treat boys and girls differently in the classroom	27.1%	36.0%	40.7%
Agrees teachers often absent from class	20.9%	21.9%	32.0%
Afraid of teacher	69.7%	60.5%	65.1%
Does not feel comfortable asking teacher questions	2.2%	2.7%	5.8%
Teacher punishes students who get things wrong in a lesson	53.1%	54.3%	56.9%
Teacher uses physical punishment	53.1%	25.8%	51.7%
No computers at school	83.8%	96.2%	93.8%
Cannot use books or other learning materials at school	12.0%	22.2%	29.1%

Not enough seats for all students	5.7%	13.0%	3.8%
Caregiver says principal is of poor quality	2.4%	0.9%	1.7%
Caregiver says teaching is of poor quality	0.7%	0.0%	0.3%
Girls says teacher does not ask girls and boys an equal number of questions	7.3%	53.4%	17.5%
Girl says teacher does not ask girls and boys questions of equal difficulty	11.1%	54.3%	18.7%
Household Barriers to Education			
High chore burden	25.3%	28.4%	7.2%
Doesn't get support from family to stay in school	4.7%	4.3%	6.6%
Girl feels that she has no choice whether to attend or stay in school	79.4%	66.5%	80.1%
Girl feels family makes schooling decisions for her	27.6%	17.1%	35.2%
Distance to school is greater than 30 minutes	2.2%	1.9%	4.6%
Caregiver and family members are not involved in CEC	17.6%	15.1%	16.6%

As expected, the three cohorts compared in this subsection differ in significant ways. First, we will compare ALP girls to the baseline and endline panel. ALP girls are much more likely to be married, be a mother, and to have a mother who was under sixteen than girls in the baseline sample. Among ALP girls 28.4 percent reported being married compared to 16.7 percent in the baseline sample. Similarly, 3.8 percent of ALP girls reported being a mother compared to 0 percent of the baseline sample. However, girls in the baseline sample are more likely to migrate seasonally, with 7.9 percent in the baseline sample reporting this compared to 2.4 percent of ALP girls. Another point of difference between the groups is the proxy for wealth of having a poor-quality roof, here 36.0 percent of ALP girls report having a poor-quality roof while 21.6 percent of girls in the baseline survey report the same. Girls in the ALP sample also report much higher rates of not having access to medicines many or most days, with 52.4 percent compared to 31.6 percent of girls in the baseline sample.

Finally, the sample of ALP girls has a much lower exposure to the current drought, with just 24.6 percent of ALP girls in drought-exposed communities, compared to 48.9 percent of girls from the in-school and out-of-school cohorts tracked since baseline. This difference does not reflect structural differences in girls'

exposure to climate shocks; rather, it highlights the slightly different geographic dispersion of the ALP cohort, as ALP centres in the sample are more likely to be situated in Sanaag and other areas where drought has been less severe to date.[25] It is also important to note that our coding of drought is not based on household-level exposure to the impacts of drought, but a broader assessments of which regions and districts have been most affected, based on remote sensing data.

Next, we will compare ABE girls to the baseline and endline panel. ABE girls are much more likely to be living with both parents. With 8.4 percent of ABE girls reporting living without both parents compared to 25.2 percent of girls in the baseline survey. However, ABE girls are much more likely to have a head of household with no formal education with 79.4 percent compared to 66.0 percent in the baseline survey. Similarly, ABE girls are more likely to have a head of household that does not earn a wage with 54.9 percent compared to 43.9 percent. ABE girls also have much higher rates of having a poor-quality roof on their home with 46.8 of ABE girls reporting this compared to 21.6 of girls in the baseline sample. Girls in the ABE cohort also have less access to water with 34.1 percent reporting they have no access to water most or many days compared to 24.3 percent among the baseline cohort. In a related finding 19.2 percent of ABE girls don't use the drinking water facilities at school compared to 8.8 percent of girls in the baseline survey. Finally, 40.7 percent of ABE girls report that their teachers threat boys and girls differently in the classroom while only 27.1 percent of girls in the baseline survey report the same.

Re-Contact Procedures

The evaluation's design leans heavily – even more so than in previous rounds – on the fact that a panel of girls has been tracked over time across rounds.[26] Each cohort of girls has been tracked since the girls were first recruited: in-school and out-of-schools at baseline, ALP girls at ML1, and ABE girls at ML2. A key goal of the evaluation team, throughout the rounds, has been to minimise sample attrition to the extent possible, to preserve the available panel sample and statistical power of the sample to draw inferences regarding programme impact.

Re-contact procedures were designed with this goal in mind, and the procedures were emphasised repeatedly during training. A highly specific set of steps were required to be completed before a girl could be replaced – in earlier rounds – or dropped from the sample, in the present round. The steps for attempting re-contact were:

- Call every phone number listed for the household a minimum of three times, allowing at least six hours between phone call attempts. Calls must span two days in total; in other words, not all calls can be completed on the same day. In each round of data collection, households were asked to provide two contact phone numbers; in practice, this means that each household had between one and five numbers listed.
- Visit the girl's household a minimum of two times, allowing at least six hours between visit attempts.

[25] Note that ALP centres are all located in communities which also have formal schools supported by SOMGEP-T. However, the ALP and formal school samples were drawn separately, resulting in a sample that has incomplete overlap in locations.

[26] We say the endline relies more heavily on the panel sample because no replacements were selected at endline. In previous evaluation rounds, we reported findings from a "repeated cross-sectional" sample, which included replacement girls and the girls they replaced in the analysis. The panel sample – those girls who were successfully re-contacted, excluding replacements entirely – was used, as well, but we were able to leverage both datasets at times. In the endline, no replacements were selected, meaning that we can only report panel-based findings. Use of the panel sample is more rigorous, ceteris paribus.

- Ask the head teacher and other teachers in the school for contact information for the girl or her family, and how they can be reached.
- Ask other girls in the same grade or age group if they know the girl and how she can be reached.

These procedures generally mirror standard practice in panel surveys in underdeveloped areas. If a girl could not be reached after completing these steps, enumerators informed their team leader, who verified the process had been completed and approved the removal of the girl from the endline sample.

The importance of maintaining the integrity of the sample across rounds is reflected by two additional steps taken by the evaluation team, in consultation with CARE's monitoring and evaluation staff. The first is that a significant number of girls were tracked to nearby cities to which they had migrated; when tracked in this way, we were able to complete a household survey and learning assessment with the girl. As a rule, we imposed common-sense limits on the length of time team members should spend traveling to find a girl in her original village. For instance, if a girl was in another nearby village that would take 20 minutes to reach by car, a member of the team was assigned to visit the girl's household. If a girl was in another village 3 hours away by car, they did not travel to reach her. This is the same fuzzy standard that we have applied in earlier rounds of SOMGEP-T.

However, this round found many more girls who had migrated away from their home village. At ML2, 35.9 percent of the in-school and out-of-school girl cohort was in grade 7 or grade 8. Given that two years passed between ML2 and endline, this set of girls would, if they progressed two grades, moved into secondary school by the time of the endline. This is important because many of the evaluation villages do not have secondary schools, and it is common for girls to migrate away from home to attend secondary school, often living with extended family members in urban areas. In addition, as the girls are now two years older than at ML2, out-migration and marriage rates among the cohort have naturally increased. And the disruption of COVID-19 has had an unknown impact on out-migration. This situation prompted our concern and concern from the CARE team regarding re-contact rates.

During fieldwork, when enumerators encountered girls who could not be located in their villages, they systematically collected information on her current whereabouts, including phone numbers for contacting her and her family members. Using information on their locations, the evaluation team identified cities to which a high number of girls had migrated and organised time for a fieldwork team to visit those cities to contact as many girls as possible. Fieldwork teams visited six cities and successfully interviewed approximately 150 girls in these locations. This approach increased the aggregate re-contact rate at endline by 5.4 points.

The second step targeted girls who had moved away or could not be contacted in-person, but who had not moved to an urban area our teams could reasonably visit. This included girls who had moved to Mogadishu or areas of south-central Somalia, because this would have necessitated a significant investment in transport and logistics. It also included girls who had moved to remote villages, into Ethiopia, or abroad. While attempting to locate them in their original villages, enumerators collected as much information as possible about these girls and their families. A small team, operating out of Consilient's home office in Hargeisa, phoned girls to attempt to locate and complete a short survey with them over the phone.

Unfortunately, the length of the household survey prevented us from completing the full household survey over the phone; learning assessments also cannot be completed by phone. The survey focused, instead, on collecting basic transition-related information from the girls, namely: her enrolment status, grade level, and a few demographic details, such as marital status. In total 286 girls were contacted in this way. Most (82.5

percent) lived outside their original village, while the vast majority of the remainder were away from their village at the time our teams visited. Interviews with these girls do not shed light on any of the programme's intermediate outcomes or learning outcomes. However, they do provide useful – if incomplete – information regarding transition status. The data is insufficient for use in our main transition analysis, but allows analysis of the retrospective transition questions, increasing our sample size for this analysis by 9.7 percent.

Re-Contact Rates

The endline evaluation sought to re-contact a total of 2,811 girls who had been interviewed in a previous round of SOMGEP-T. As was the case during the ML2 evaluation, we took an all-inclusive approach to defining who should be re-contacted: put simply, every girl who had participated in at least one previous SOMGEP-T evaluation was sought out, with the exception of girls in communities that had been entirely removed from the sample. We included, for example:

- Out-of-school girls recruited at baseline but who could not be located during ML1, ML2, or both. These girls were not replaced during the midline rounds. They were re-contacted, despite having fallen out of the sample earlier, to generate the largest possible baseline-to-endline (or ML1-endline, etc.) sample for analysis.
- In-school girls recruited at baseline who fell out of the sample during ML1 or ML2 and were replaced. In these cases, the endline attempted to follow up with both the replacement girl selected in an earlier round and the "original girl" she replaced.

The sample universe of 2,811 girls includes every girl – in-school girls, out-of-school girls, ALP girls, and ABE girls – who had been contacted at least once in prior rounds. In some cases, an original cohort girl may have fallen out of the sample at ML2; by bringing her back into the sample at endline, we can include her in analyses of changes from baseline to endline and from ML1 to endline. This logic motivated the inclusive approach to re-contacting girls.

The table below documents the number of girls, by cohort, who were last successfully contacted in each round. It is important to emphasise that the achieved re-contact rates at endline are much lower than would have been the case if the evaluation only re-contacted girls who had appeared in the ML2 round. By also attempting to re-contact girls who were last interviewed during the baseline round in 2017 – who we have already attempted and failed to find 1-2 times during midline rounds – our aggregate re-contact rate is suppressed. In total, we successfully re-contacted 64.2 percent of the 2,811 girls. For reference, typical re-contact rates on previous SOMGEP-T rounds and other GEC programmes in Somaliland, Puntland, and Galmudug have been around 80 percent. However, these re-contact rates were typically based on re-contacting only girls who had remained in the sample up until the immediately prior round. The table shows that our targeted sample includes 449 girls who we have previously failed to contact in both the ML1 and ML2 rounds and a further 254 girls who were contacted at ML1 but could not be located at ML2.

TABLE 5: NUMBER OF ENDLINE RESPONDENTS, BY COHORT AND ROUND OF LAST SUCCESSFUL CONTACT

Cohort	Last Contact at BL	Last Contact at ML1	Last Contact at ML2	Total
In-school girls	124	146	922	1,192

Out-of-school girls	325	0	368	693
ALP Girls	N/A	108	336	444
ABE Girls	N/A	N/A	482	482
Total	449	254	2,108	2,811

The table below documents re-contact rates as a function of cohort type and the round in which the girl was last contacted. As we would expect, girls who had previously fallen out of the sample – last re-contact at baseline or ML1 – were much less likely to be successfully included in the endline round. Of the 449 girls who were initially contacted at the baseline but never successfully interviewed since, we were able to locate and interview just 32.1 percent. Re-contact rates are much higher among the set of girls who were successfully interviewed at ML2 – 73.1 percent in the aggregate. Even this re-contact rate is slightly lower than during past rounds; however, it is important to remember that two years have passed since ML2, whereas previous rounds were conducted with just a one-year gap. Because most failures to re-contact are driven by girls who have migrated away from their villages, a two-year gap provides more time for girls to leave the village, and increases the attrition rate.

TABLE 6: RE-CONTACT RATES AT ENDLINE, BY COHORT AND ROUND OF LAST SUCCESSFUL CONTACT

Cohort	Last Contact at BL	Last Contact at ML1	Last Contact at ML2	Total
In-school girls	35.5%	50.7%	78.2%	70.4%
Out-of-school girls	30.8%	N/A	69.0%	51.1%
ALP Girls	N/A	42.6%	64.0%	58.8%
ABE Girls	N/A	N/A	72.7%	72.7%
Total	32.1%	47.2%	73.1%	64.2%

Beyond the last successful re-contact, it is clear that re-contact rates are systematically higher among girls who were originally recruited as in-school girls. Their enrolment rates in each round were relatively higher, as well, and their links to the local school likely enhanced our ability to locate and interview them. ALP girls also have lower re-contact rates; among girls contacted at ML2, ALP girls have the lowest re-contact rate at endline. This can be explained by their higher average age: the mean age of ALP girls at ML2 was 18.1 years, compared to 13.7 years for ABE girls, 15.4 years for out-of-school girls, and 14.5 years for in-school girls. Older girls are more likely to get married and leave their village or migrate for school or work, driving up attrition rates among ALP girls.

Re-contact rates are also considerably higher in intervention communities. Among in-school girls, re-contact was successful 73.9 percent of the time in intervention communities, and just 66.7 percent of the time in comparison communities. Among out-of-school girls, however, comparison communities actually had higher re-contact rates – 56.7 percent versus 51.5 percent. Overall, transition rates were 4.4 points higher in

intervention communities, owing to the fact that in-school girls make up a higher proportion of the overall sample. It is worth noting that the re-contact advantage seen in intervention communities is actually exacerbated in a regression framework which controls for age, geographic zone, and the round in which a girl was last contacted. While re-contact rates in comparison communities are still higher for out-of-school girls in a regression framework, the gap is much smaller; and the positive gap between intervention and comparison communities in terms of in-school girls widens when controlling for other predictors of successful re-contact. In the section that follows, we unpack the predictors of successful re-contact further.

TABLE 7: AGGREGATE RE-CONTACT RATE

	Re-Contact Rate Among Girls Last Contacted at ML2	Aggregate Re-Contact Rate
Total	73.1%	64.2%

Predictors of Successful Re-Contact

Attrition from the panel sample poses three dilemmas when we wish to draw inferences regarding the impact of SOMGEP-T. First, attrition reduces the sample size available for analysis, reducing the power of our statistical tests – their ability to distinguish real programme impacts from null or zero effects. The original evaluation design factored attrition into the calculation of sample sizes needed to achieve conventional levels of statistical power. However, all else equal, larger samples are unambiguously better at detecting true effects and distinguishing between null and alternative hypotheses. And our sample sizes are sufficiently small that high levels of attrition present very real threats to drawing inferences.

Second, attrition is unlikely to be distributed randomly within the population. That is, not all girls are equally likely to fall out of the sample. As shown in the previous section, there are significant differences across zones in attrition and individual-level factors also likely cause differential attrition. If girls exit the sample as a function of pre-existing characteristics, it can produce bias in our estimates of programme impact. Even more problematic would be if attrition is *correlated* with the outcomes of interest to the programme. In other words, if girls who drop out of school are less likely to be located and interviewed, this may cause us to overstate the rate of successful transition, because girls with unsuccessful transition outcomes are less likely to remain in the sample. In the context of SOMGEP-T's difference-in-differences design, this is only problematic if there are differences in this pattern between the intervention and comparison groups. For instance, if programme schools are better at facilitating re-contact, even for girls who have dropped out of school, it may drive down their estimated transition rates.

Third, even in the absence of differential attrition across intervention and comparison groups, attrition alters the underlying sample employed for analysis. For instance, as older girls drop out of the sample, the sample becomes younger. At baseline, the average age of girls was 13.3 years. At the same time, the mean age of baseline girls who were successfully tracked through to the endline was 13.0 years, a difference in age that is statistically significant at the 1 percent level. If marginalised girls are more likely to fall out of the sample, the sample may become less representative of the targeted communities over time, and our inferences may not reflect the true impact of the programme.

In this section, we analyse differences in re-contact rates as a function of geography, a girl's characteristics, and the characteristics of her household and the school in her community. Our analysis employs linear (OLS) regression models predicting the binary outcome of successful or unsuccessful re-contact at endline.[27]

Each panel is derived from a linear regression predicting successful re-contact. Each age group's relationship to re-contact is calculated as a binary or dummy variable for that age group, where girls aged 14 years are the omitted category. Therefore, the impact of being 15 years old on re-contact rates is measured as the difference in re-contact rates between 15-year old girls and 14-year old girls, controlling for geography and the round in which the girl was last contacted. In general, age is positively related with attrition – older girls are much less likely to stay in the sample and be re-contacted at endline. This is illustrated by the fact that our estimates are consistently negative for higher age groups. Recall that the comparison group are girls aged 12-14 years. Therefore, a negative coefficient (a dot to the left of the vertical line at zero) for the group of girls aged 17 years means that girls in this age group were less likely to be re-contacted than girls in the 14-year age group.

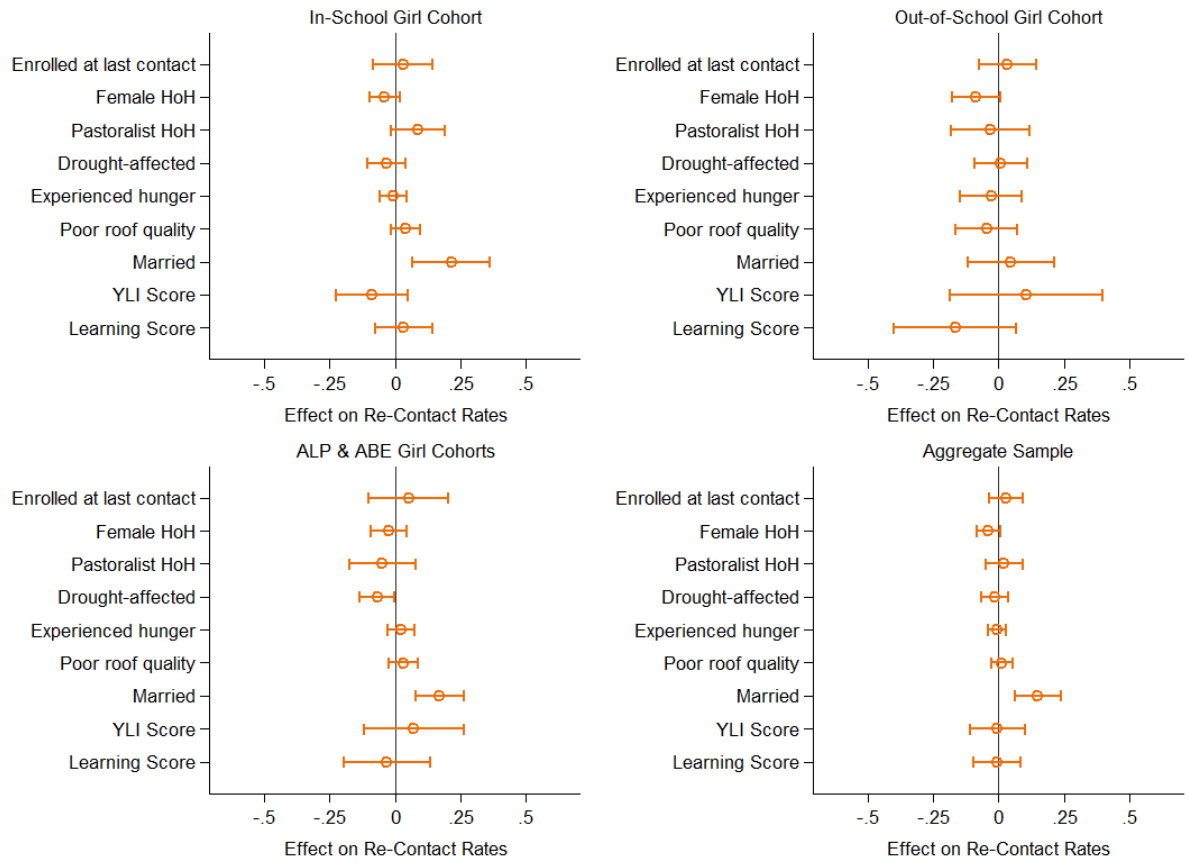
The findings are slightly obscured by the fact that we divide age groups up into individual year buckets. To illustrate the nature of the relationship, we can say that in-school girls aged 14-15 years old at present had a re-contact rate of 73.1 percent in the endline evaluation. In comparison, in-school girls aged 19-22 years had a re-contact rate of just 55.3 percent. The systematically lower re-contact rates among older girls are clearest among ABE and ALP girls, but this is a relationship that holds among all of the cohorts tracked over time for this evaluation.

Beyond age, we also analysed the relationship between re-contact and a girl's household, demographic, and learning characteristics. The figure below reports the results of a series of linear regressions that predict successful re-contact. The regressions control for geographic zone, age, and the round of last successful contact with the girl, in light of the very different re-contact rates that obtain between girls last contacted at baseline and those last contacted at ML2. For instance, the finding related to female-headed households (female HoH) is that in-school girls from female-headed households were less likely to be successfully re-contacted (though this finding is not statistically significant), holding zone, age, round of last contact, and all the other factors in the graph constant.

The two most interesting findings concern the relationship between marriage and gender of the head of household, on one hand, and re-contact rates, on the other. Among all cohorts – but especially in-school girls, ALP girls, and ABE girls – marriage is associated with higher re-contact rates, a finding that is surprising at first glance. However, it is important to note that all predictor variables are coded on the basis of a girls' information *at the time of her last contact*. We cannot use a girl's *current* characteristics for this analysis, because current characteristics are unobservable for girls who were not successfully re-contacted. Therefore, we use a girl's characteristics in the most recent previous round in which she was interviewed to predict successful re-contact at endline. In light of this, the marriage finding makes more sense, because girls who are *already* married are less likely to migrate away. Girls who were married when we last interviewed them likely have deeper ties to the villages where they were located, because they live there with their husbands. In addition, because the act of getting married is a major driver of out-migration, girls who are already married are less likely to migrate away from their villages, and more likely to remain in the sample.

[27] We use linear models despite the fact that the outcome is binary, primarily thanks to the benefits of the linear probability model in terms of directness of interpretation.

FIGURE 1: DEMOGRAPHIC AND OTHER PREDICTORS OF SUCCESSFUL RE-CONTACT AT ENDLINE, BY COHORT



The second finding that emerges from the figure concerns the female-headed households. While girls in female-headed households have – in past evaluation rounds – achieved similar learning scores and performance on other indicators as girls from male-headed households, this does not hold true for re-contact rates at endline. Girls in female-headed households have lower, by 3.8 percentage points, re-contact rates after controlling for other factors, a finding that is significant at the 10 percent level. Among out-of-school girls, this gap is even larger: girls from female-headed households have 8.5 point lower re-contact rates.

The figure above also reports a number of null results within individual cohorts and in the aggregate sample. Perhaps the most surprising are those related to YLI and learning scores. The Youth Leadership Index (YLI) is a measure of a girl's self-perceived ability to lead; broadly, it measures their self-confidence, decision-making and problem-solving skills, and their ability to motivate their peers.[28] We would expect self-confidence and performance in school (measured as mean performance in numeracy and Somali literacy) to predict a girl's likelihood of staying in school. We would also expect girls who remain enrolled to be more

[28] CARE. 2014. "CARE's Youth Leadership Index: Toolkit 2014." Available at https://www.care.org/wp-content/uploads/2020/05/CARE-YLI-Toolkit-FINAL-WEB.pdf.

likely to be re-contacted. However, this appears to be too simplistic a view, given the finding that YLI scores and learning scores have, essentially, no impact on re-contact rates. It may actually be that self-confident and higher-performing girls are more likely to remain in school, but that this outcome has mixed effects on re-contact success: girls who remain in school in their home village are more likely to be located and interviewed, but girls who remain in school but migrate to urban areas to pursue secondary education are *less* likely to be located and interviewed.

The important takeaways from this analysis concern the effect of differential attrition on the composition of the panel sample, and the effect of differential attrition on the comparability of the intervention and comparison groups. With regard to sample composition, panel samples tracking girls from ML2 to endline and baseline to endline, respectively, tend to be slightly younger than they would have been if no attrition had occurred. Similarly, they tend to underrepresent girls from female-headed households somewhat, because such girls were less likely to be re-contacted at the endline and less likely to be included in our panel samples as a result. However, in most other respects, attrition was relatively evenly spread across subgroups, without significant impact on sample composition.

With regard to the composition of intervention and comparison groups, there is some limited evidence that differential attrition will affect their comparability. In short, while intervention and comparison groups lost girls from female-headed households at largely similar rates, this is not true for all correlates of attrition. For instance, drought and marital status have meaningfully different impacts on attrition rates between intervention and comparison schools: among comparison communities, drought drives lower re-contact rates, while re-contact rates are actually higher in drought-affected intervention communities than in unaffected communities. The result is that the intervention group has become more drought-affected than its comparison group as a result of differential attrition related to drought. In most ways, attrition affects intervention and comparison communities similarly, but the presence of differential attrition across the two groups suggests that caution is warranted when interpreting findings regarding programme impact on outcomes that are highly correlated with, for instance, drought exposure at present.

Learning

Learning represents one of the core outcomes for SOMGEP-T, which sought to improve learning outcomes among each of the four distinct cohorts of girls the programme targeted. The programme's Theory of Change regarding learning focuses on changing teaching practices, improving attendance, and encouraging retention and enrolment in school – or alternative education – for girls who were not enrolled at the programme's outset. Across the different cohorts, girls took as many as five distinct learning assessments, testing their abilities in terms of numeracy, Somali literacy, English literacy, financial literacy, and their "working memory." In general, our analysis focuses most heavily on the first three assessments, though we report results for all five assessments.

The numeracy and literacy assessments are based on the Secondary Grade Mathematics Assessment (SeGMA) and Secondary Grade Reading Assessment (SeGRA), respectively. The primary skills tested within each assessment are:

- Numeracy – tests students' ability to identify patterns in sets of numbers; perform the most basic forms of addition, subtraction, multiplication, and division; perform higher-level arithmetic

involving, for instance, carrying numbers and multi-digit multiplication; and complete word problems based on relatively simple arithmetic.

- Somali literacy – tests students' ability to read individual words; reading comprehension for both short and multi-paragraph passages; reading speed; and facility with negative and future tenses.
- English literacy – tests students' ability to identify English letters and individual English words; reading comprehension for both short and longer (2-3 paragraph) passages; and facility with negative and future tenses.

The financial literacy assessment consists of 11 subtasks that measure the respondent's ability to apply maths in word problems that mimic everyday situations, including the calculation of profit and interest. Finally, the working memory assessment – which is analysed in a dedicated subsection below – consisted of showing each child a set of 19 images, removing the images, and then checking how many of the images the child could recall. Later in this section, we describe each assessment in more detail, especially when analysing performance on particular subtasks or mastery of specific skills.

To the extent possible, the difficulty and structure of the assessments remained the same across rounds, to facilitate the most rigorous comparisons possible. The one exception occurred in Somali literacy, where the first two evaluation rounds raised concerns about potential "ceiling effects" regarding Somali literacy scores. During the ML1 evaluation, 16.7 percent of cohort girls achieved scores of 90 percent or above; given that two full years of programme implementation remained, this suggested that many girls might achieve the maximum possible score in Somali literacy, dampening the sensitivity of our measure and impeding us from drawing conclusions about the programme's impact. To help guard against ceiling effects, a 9th subtask was added to the Somali literacy assessment; this subtask asked girls to read a story with missing words or phrases and fill in the words or phrases that would logically connect the preceding and subsequent sentences. In other respects, the Somali literacy assessment remained identical to previous rounds, with very minor adjustments that did not affect the difficulty or structure of individual subtasks.

For numeracy, Somali literacy, English literacy, and financial literacy, the exact questions were adjusted slightly from round to round to ensure that students did not recall the questions. For instance, a simple addition problem would have been changed from 3+5 to 2+6, with the difficulty remaining identical. Similarly, reading passages were altered, but with similar difficulty, length, and structure.[29] Moreover, thanks to the difference-in-differences design used for analysis of learning improvements among the main cohort, any minor changes in assessment difficulty across rounds are explicitly controlled for, because they are applied to both the intervention and comparison groups equally. Changes in assessment difficulty are more problematic in the context of simple pre-post comparisons – such as that employed when studying learning among the ALP and ABE cohorts; even here, however, changes in the assessment contents were extremely minor and our investigation of assessment difficulty between the ML2 and endline rounds shows that these rounds of assessments are statistically indistinguishable from one another in terms of difficulty.[30]

[29] The working memory remained exactly the same across both rounds in which it was implemented, but the order of the images shown to the children was changed.

[30] See discussion in Annex FILLIN. In each previous round, the evaluation team collected a set of comparison assessments, with the same girls completing, e.g., the ML1 and baseline numeracy assessments as part of the ML1 evaluation. This allowed us to compare the two assessments while eliminating sampling variation, because the same girls completed both assessments at the same time. In previous rounds, we invariably found that assessments had not changed in difficulty across rounds. The same remains true for the ML2-to-endline comparisons in Annex FILLIN.

Overall Learning

This section focuses on aggregate learning outcomes among the original cohort of SOMGEP-T girls, recruited originally at baseline. Importantly, not all assessments were completed in each evaluation round. The table below documents the rounds during which each assessment was completed by members of the original baseline cohort of in-school girls (top panel) and out-of-school girls (bottom panel). Note that other cohorts – ALP and ABE girls – are discussed in separate subsections, below.[31] The most important omission is the exclusion of the "full" Somali literacy assessment from the baseline and ML1 rounds – as noted above, the "full" Somali literacy assessment, including all nine subtasks, was not developed and implemented until ML2. Therefore, at baseline and ML1, Somali literacy was assessed using the first eight subtasks only.

The ML1 round included more significant omissions, in terms of both the girls interviewed and the assessments completed. As we discussed in the methodology section previously, out-of-school girls – girls who were not enrolled in school at the time of recruitment into the sample at baseline – were not contacted at all during ML1, and did not complete any of the learning assessments. In addition, in-school girls did not complete the English literacy or financial literacy assessments during the ML1 round. In practice, these omissions are inconsequential for our analysis, because our interest is in testing for impact over the life of the programme, from baseline to endline. We are also interested in programme impact over the two years that have elapsed since the ML2 evaluation, because this period has not been studied in previous rounds. Changes in learning or other outcomes since ML1, specifically, are of less interest, because they are encapsulated by the broader and more important question regarding impact since baseline.

TABLE 8: LEARNING ASSESSMENTS IMPLEMENTED ACROSS ROUNDS

Assessment Type	Baseline	Midline #1	Midline #2	Endline
In-School Girls				
Numeracy	Yes	Yes	Yes	Yes
Somali Literacy – 8 Subtasks	Yes	Yes	Yes	Yes
Somali Literacy – Full	No	No	Yes	Yes
English Literacy	Yes	No	Yes	Yes
Financial Literacy	Yes	No	Yes	Yes
Out-of-School Girls				
Numeracy	Yes	No	Yes	Yes
Somali Literacy – 8 Subtasks	Yes	No	Yes	Yes

[31] The nature of the analysis of learning outcomes among ABE and ALP girls, which does not include a comparison group or difference-in-differences models to assess programme impact, makes it simpler to report on improvements in learning for these cohorts separately.

Somali Literacy – Full	No	No	Yes	Yes
English Literacy	Yes	No	Yes	Yes
Financial Literacy	Yes	No	Yes	Yes

Before beginning analysis of aggregate learning outcomes, it is important to reiterate the nature of the samples of girls studied in this section. As we discussed in the context of the recontact analysis in the previous section, each set of girls studied in this evaluation constitute a panel sample, in which the same girl has appeared in each round of the evaluation. Two versions of the panel sample comprise the vast majority of our analysis. The baseline-to-endline panel includes all cohort girls who were successfully interviewed in both the baseline and endline evaluation rounds; importantly, their inclusion or exclusion from the ML1 and ML2 samples does not affect their inclusion in this panel, which focuses on changes over the length of the programme. This sample of 965 girls (n = 1,930 across the two rounds) provides the best sample for assessing change over the length of the programme. In this sample, 63.3 percent of the girls are in the in-school girls cohort, who were enrolled in school at the time of their initial recruitment at baseline.

The second panel runs from ML2 to endline, and includes all cohort girls who appear in these two rounds, regardless of whether they appeared in the baseline or ML1 rounds. Replacement girls can appear in this sample, but only if they were selected as replacements during the ML1 or ML2 rounds. This sample treats the ML2 as a sort of baseline – whatever replacement or attrition that took place before the ML2 is irrelevant, but girls must have been successfully interviewed at ML2 and then recontacted during the endline. This sample provides the most rigorous assessment of change since the ML2 evaluation was conducted in late 2019, exactly two years prior to the endline. For the learning analysis, the ML2-to-EL panel sample includes 975 girls (n = 1,950 across the two rounds), 73.9 percent of whom are part of the in-school girl cohort.

Numeracy

We start our analysis with numeracy, the assessment of which included 11 subtasks; scoring was adjusted to a 100-point, with each subtask equally weighted in the final scoring. In practice, this means that individual test items – the constituent parts of a subtask – are given unequal weight in the final score, because some subtasks consist of ten test items, while others include as few as two items.

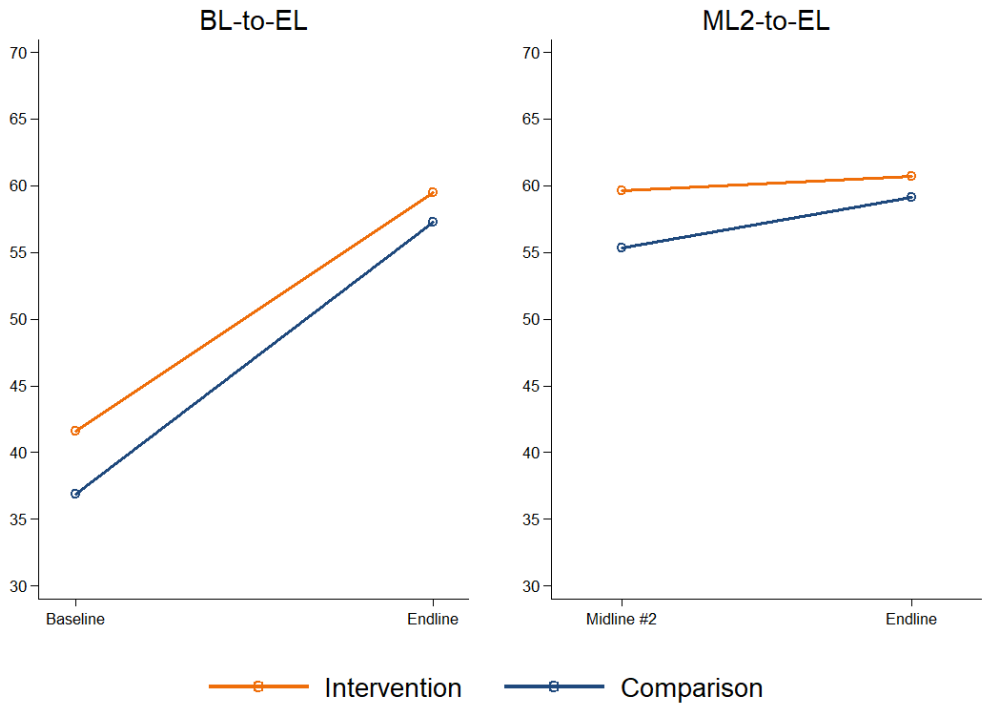
As in previous rounds, achievement in numeracy ran the length of the scoring scale, with 3.4 percent of girls scoring 0 and 0.5 percent of girls achieving a perfect score.[32] Both shares are much lower than in previous rounds – for instance, at baseline, nearly a quarter (24.4 percent) of the cohort girl sample were unable to answer a single numeracy question correctly; even so, performance at the top-end of the scoring range was also more common, with 3.1 percent of girls in the baseline achieving a perfect score. The mean score at endline was 51.5 percent, and scores closely approximated a normal distribution, in contrast to the right-skewed distribution of the baseline.

[32] Note that, throughout this section, we refer exclusively to cohort girls – those recruited originally at baseline – and exclude ALP and ABE girls from our discussion.

The figure below reports changes in mean numeracy scores from baseline to endline (left panel) and ML2 to endline (right panel) among the in-school girls cohort. Before turning to whether the programme had an impact on learning outcomes, it is noteworthy that, while numeracy scores rose broadly and significantly from baseline to endline, almost all of the improvements in numeracy occurred in the first two years of the program, prior to the ML2 evaluation. The right panel of the figure shows that numeracy scores have been flat since ML2. This finding will emerge as a theme across much of the learning analysis that follows, a theme which can be attributed, in large part, to the COVID-19 pandemic and the school closures that were in place for much of 2020. In the transition section of this report, we discuss a second exogenous shock – a drought that worsened as the endline evaluation approached – which affected SOMGEP-T communities. We attribute flat learning outcomes since ML2 to a combination of COVID-19 and – potentially – drought conditions that have worsened recently. Long-running school closures due to COVID-19 have certainly had a direct impact on learning outcomes, but the recent drought is very likely driving higher dropout rates, worse attendance, and lower learning scores through its nutritional impacts, impacts on teachers and teacher absenteeism, and generalised stress within households and communities.

As the figure shows, there have been no discernible improvements in numeracy as a direct result of SOMGEP-T programming. In other words, while numeracy scores in SOMGEP-T schools have increased by 17.9 points – from 41.6 to 59.5 percent – numeracy scores in comparison schools improved by 20.4 points over the same period. Our best estimate of programme impact, using a simple difference-in-differences model, is that the programme resulted in a -2.5 point change in numeracy scores.

FIGURE 2: NUMERACY SCORES AMONG IN-SCHOOL GIRL COHORT



The right panel of the figure shows a similarly bleak outcome over the period since ML2. Over this period, girls in intervention communities achieved improvements of 1.1 points, while girls in comparison communities improved by 3.8 points.

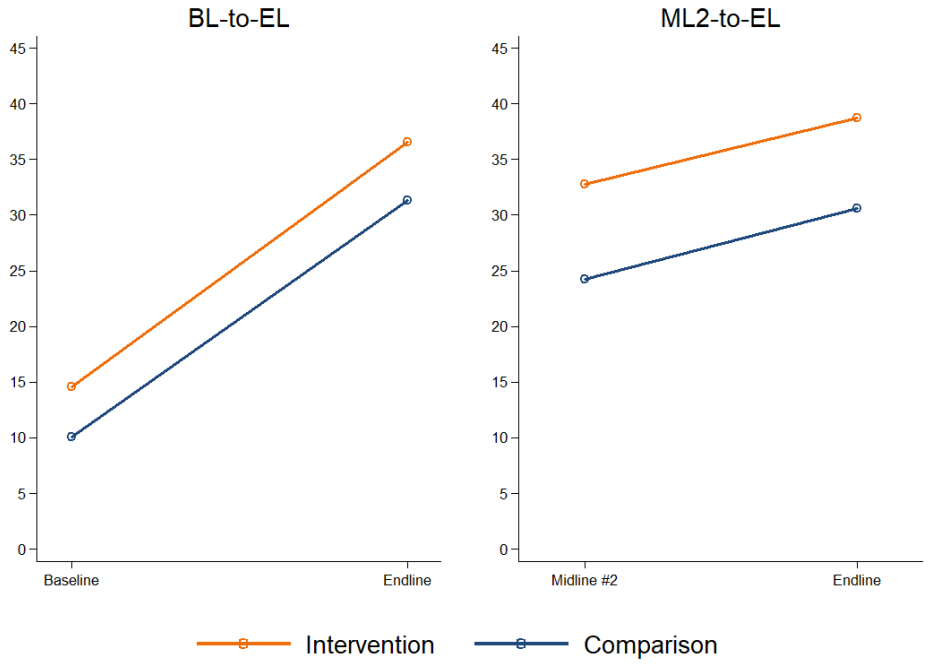
The results thus far are focused on girls who were enrolled at the time of their initial recruitment into the sample. However, this does not mean the girls had uniform or extensive exposure to schooling in their respective communities because this sample includes a number of girls who dropped out of school during the programme. To illustrate, just 65.6 percent of in-school girls in the panel sample were enrolled in school all five years from 2017 to present; some girls (n = 31) appear to have dropped out shortly after the baseline was completed and were not enrolled in any subsequent year. It is important to keep this fact in mind, because the impact of the programme on learning is not strictly a function of improved learning outcomes among girls who were enrolled continuously and deeply exposed to the programme. Instead, programme impact is influenced by the extent of exposure – high dropout rates reduce exposure, and can reduce the impact of the programme.[33] In this sense, impact on learning is mixed with the programme's ability to promote retention and attendance among cohort girls, alongside improved teaching and other in-school factors expected to promote stronger learning. Later in this section, we assess the programme's impact as a function of the number of years a girl spent enrolled in school since 2017.

Our second analysis focuses on out-of-school girls – those who were not enrolled at the time of their recruitment into the sample. The figure below reports changes in numeracy scores for our two panel samples, running from baseline to endline (left panel) and ML2 to endline (right panel). The samples used in this analysis are significantly smaller, because the out-of-school girl cohort was smaller at baseline and because out-of-school girls that fell out of the sample in the intervening years were not replaced during ML1 or ML2. The result is a sample of just 354 girls for the baseline-to-endline analysis and 254 girls for the ML2-to-endline analysis.

As the figure shows, there has been no appreciable improvement in numeracy scores among girls in intervention communities, relative to the comparison group. The two groups had very low starting points – a mean score of 14.6 and 10.1 points, at baseline, among the intervention and comparison groups, respectively. Between baseline and endline, the groups improved almost identically – 22.0 points for the intervention group and 21.3 points for the comparison group. The findings shown in the right panel are similar, with improvements from ML2 to endline nearly indistinguishable from one another across the two groups. From ML2 to endline, we estimate that the programme's impact on numeracy scores among the out-of-school girl cohort was -0.4 points.

[33] Accounting for variation in exposure, we performed supplemental analysis of girls who were enrolled at each evaluation point or self-reported being enrolled in each year since 2017. The programme's impact on learning scores was not systematically different among these subgroups of girls, though our estimates of impact on numeracy and English literacy were slightly more positive. Even in this analysis, we cannot fully account for the extent of programme exposure – a common problem in almost all experimental and quasi-experimental designs – because girls' attendance varies so much and the strength of exposure is also a function of the school and classrooms in which they are educated.

FIGURE 3: NUMERACY SCORES AMONG OUT-OF-SCHOOL GIRL COHORT



The analysis to this point employed a simple difference-in-differences model, which assesses the difference in trends across rounds for the intervention and comparison groups. As a function of the difference-in-differences design, pre-existing differences between the intervention and comparison groups are accounted for controlled for directly. However, this analysis does not account for other factors that might shape numeracy outcomes, including the ongoing drought in the sampled communities. In addition, incorporating control variables – even if they are implicitly accounted for in the difference-in-differences design – can improve the precision of our estimates. For these reasons, we estimated a series of linear regression models that account for additional factors and for the clustered nature of our sample.

The table below reports the results of these models, presenting our estimates of programme impact (the difference-in-differences coefficient) and the p-value associated with that estimate for a series of models. The top panel of the table focuses on the baseline-to-endline panel sample, estimating increasingly saturated models in each row; we also report results for models with only the in-school girl and out-of-school girl cohorts, respectively. The bottom panel repeats this analysis for the ML2-to-endline panel sample.

TABLE 9: DIFFERENCE-IN-DIFFERENCES ESTIMATES OF PROGRAMME IMPACT ON NUMERACY

Regression Specification	Diff-in-Diff (Program Impact)	P-Value
Baseline to Endline Analysis		

Basic model	-1.65	0.5
Controls for region and age	-2.07	0.37
Controls for region, age, drought, and enrolment status	-1.49	0.56
In-school girls cohort only	-3.81	0.22
Out-of-school girls cohort only	0.7	0.82
Midline #2 to Endline Analysis		
Basic model	-2.41	0.09
Controls for region and age	-2.37	0.1
Controls for region, age, drought, and enrolment status	-2.34	0.11
In-school girls cohort only	-3.02	0.04
Out-of-school girls cohort only	0.84	0.78

The top row of the table reports the aggregate programme impact from baseline to endline, averaging across the in-school and out-of-school girl cohorts that were previously reported separately. The results suggest that the programme had a negative impact on numeracy scores since baseline, though this result is far from statistically significant ($p = 0.50$). In other words, we are unable to distinguish the impact of the programme on numeracy scores over this period from a null effect of no impact. The incorporation of additional control variables does not alter this finding appreciably – while the impact estimate changes slightly, none of the results are distinguishable from a null effect. As the bottom two rows show, positive impact in the baseline-to-endline analysis occurs exclusively among the out-of-school girl cohort; among in-school girls, the programme had a more negative impact (-3.81 points), though even this result is not statistically significant.

The findings from ML2 to endline are more compelling, from a statistical perspective, because the effect sizes approach conventional standards of statistical significance. Aggregating across the two groups of girls, our best estimate is that the programme reduced numeracy scores by 2.41 points from ML2 to endline, relative to the counterfactual comparison group of students. This finding is statistically significant at the 10 percent level; as with the baseline-to-endline analysis, the negative impact is concentrated among in-school girls, among whom we estimate that the programme reduced numeracy scores by 3.0 points, relative to girls in comparison communities.

Somali Literacy

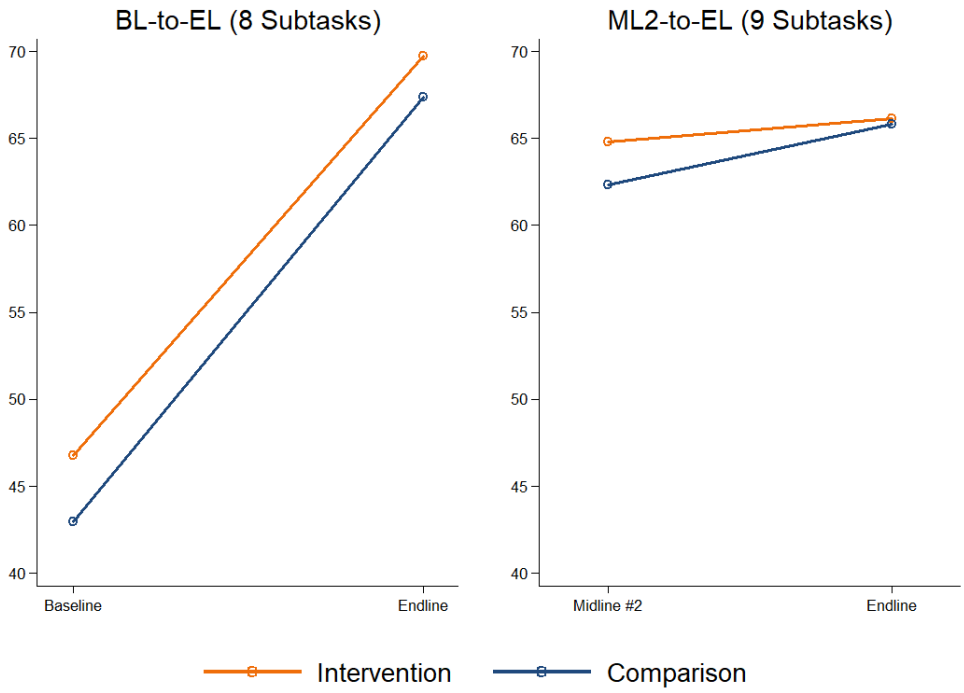
The next outcome we analyse is Somali literacy, which was measured using a SeGRA-based assessment of eight subtask beginning at baseline. Our analysis of baseline-to-endline changes utilises the slightly shorter version of the assessment, with the eight subtasks that were present in both baseline and endline rounds. Our analysis of changes from ML2 to endline utilises the full version of the assessment, with nine subtasks. In both cases, scores are calculated over a 100-point scale, with equal weighting of each subtask.

At endline, the mean score on the full version of the assessment was 54.5 percent, while the truncated version produced slightly higher average scores (58.1 percent), a fact that lines up with the intention – to make the assessment more difficult and avoiding ceiling effects – of adding the ninth subtask. In practice, few girls

achieved a perfect score on either version of the assessment; just 3.1 percent of girls achieved a perfect score on the shorter and easier version, though this does not mean that ceiling effects are irrelevant, as high scores – even if they are below a perfect 100 percent score – produces a situation in which girls are not able to demonstrate substantial improvement, because the potential for improvement is limited to, e.g., only one or two subtasks.

The figure below reports our primary results for Somali literacy among the in-school girl cohort. Both intervention and comparison groups have shown considerable improvement in literacy since the baseline, with mean scores rising by 23.5 points among in-school girls. However, these improvements were shared approximately equally between the intervention and comparison groups – in fact, girls in comparison schools improved at a slightly faster rate (1.4 points) than girls in SOMGEP-T schools.

FIGURE 4: SOMALI LITERACY SCORES AMONG IN-SCHOOL GIRL COHORT

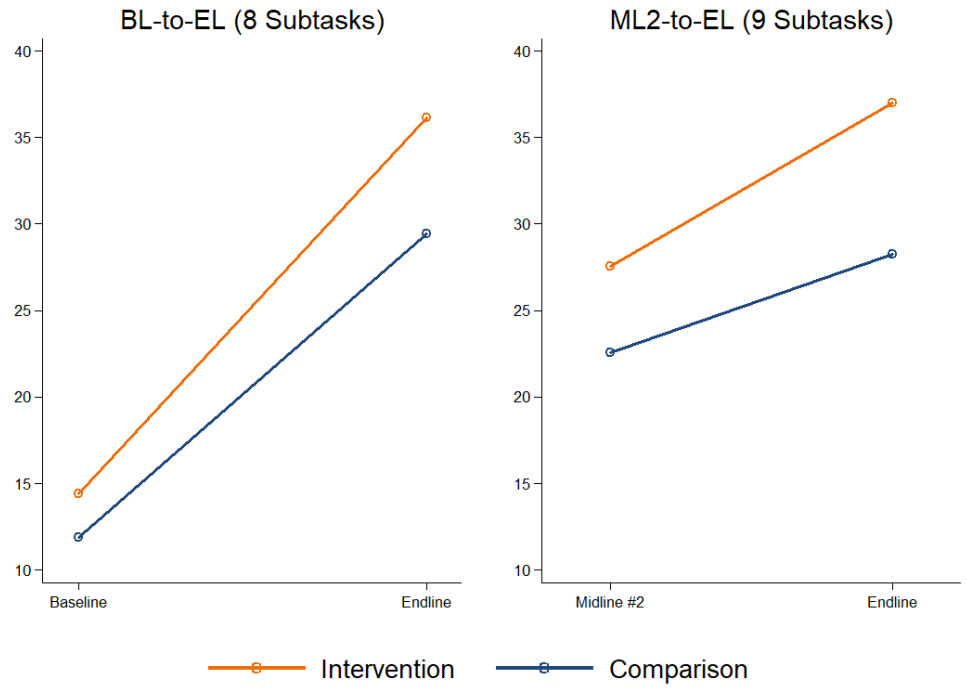


The right panel reports the same analysis over a shorter time period, from ML2 to endline. As shown in the figure, this analysis utilises the slightly longer Somali literacy assessment. Similar to our findings for numeracy, the most negative finding regarding programme impact comes from the in-school girl cohort over the ML2-to-endline period. For Somali literacy, we estimate that the programme caused a 2.1 point decline in Somali literacy scores.

In contrast to our findings for in-school girls, out-of-school girls in intervention schools showed improvements in Somali literacy vis-à-vis their counterparts in the comparison group. As the figure below shows, out-of-school girls in intervention schools improved by 21.7 points over four years, while comparison girls improved by just 17.6 points. The ML2-to-endline analysis reveals similar gains – based on a simple

difference-in-differences model, we estimate that the programme produced a 3.7-point improvement in Somali literacy among out-of-school girls since the ML2 evaluation round.

FIGURE 5: SOMALI LITERACY SCORES AMONG OUT-OF-SCHOOL GIRL COHORT



Throughout this section, the results presented in graphical form are based on the simplest possible difference-in-differences calculation. As a general rule, regression models that incorporate additional control variables rarely alter our findings regarding learning outcomes in a meaningful way. However, this is not true of Somali literacy. In the regression models listed in the top panel, our findings remain unchanged for the aggregate model, which includes both in-school and out-of-school girls. However, our estimates of programme impact among in-school girls changes: a basic model (reported in graphical form only) suggested that the programme reduced Somali literacy among this group by 1.4 points; a more saturated model that accounts for enrolment status, age, and exposure to the current drought suggests the programme reduced Somali literacy by a much larger 2.7 points, though this finding does not approach conventional thresholds for statistical significance.

The same is true of the ML2-to-endline analysis, regression models for which are reported in the bottom panel of the table. In the aggregate, our results do not change appreciably across different versions of the difference-in-differences models. However, when we restrict our attention to out-of-school girls, a basic model – reported only graphically in the figure above – we estimate the programme improved Somali literacy by 3.7 points. However, in a fuller model, accounting for region, age, exposure to drought, and enrolment status, programme impact increases to 5.2 points. This finding of considerable positive programme is based on a comparatively small sample of just 254 girls, which explains why such a large impact estimate is not statistically significant ($p = 0.20$). Nonetheless, the magnitude of the estimated impact is notable.

TABLE 10: DIFFERENCE-IN-DIFFERENCES ESTIMATES OF PROGRAMME IMPACT ON SOMALI LITERACY

Regression Specification	Diff-in-Diff (Program Impact)	P-Value
Baseline to Endline Analysis		
Basic model	1.07	0.69
Controls for region and age	0.19	0.94
Controls for region, age, drought, and enrolment status	0.88	0.74
In-school girls cohort only	-2.69	0.45
Out-of-school girls cohort only	3.8	0.18
Midline #2 to Endline Analysis		
Basic model	-1.07	0.54
Controls for region and age	-1.05	0.54
Controls for region, age, drought, and enrolment status	-1.02	0.58
In-school girls cohort only	-2.42	0.27
Out-of-school girls cohort only	5.15	0.20

One criticism of the more saturated models reported in the table is that factoring in enrolment status unfairly reduces estimates of programme impact. This would be true if the programme increased enrolment and enrolment is associated with higher learning scores, in which case the most saturated models would underestimate programme impact. In practice, this potential pathway does not seem to produce bias in the results, given that the results change very little when we include enrolment status in the models.

Leaving aside the specific model choice, the most clear-cut finding, with regard to Somali literacy, is the sharp divergence between the programme's apparent positive impact among the out-of-school girl cohort and the negative effect within the in-school girl cohort. As shown in the table above, the programme appears to have reduced Somali literacy by 2.69 points since the baseline among in-school girls. At the same time, the programme improved Somali literacy among out-of-school girls (those who were out-of-school at the baseline) by 3.8 points. A similar trend is evident over a shorter period, since ML2, with gains of 5.15 points among the out-of-school cohort, but backsliding among the in-school girls cohort.[34] While neither finding is

[34] One interpretation of these very different trends between in-school and out-of-school girls is that changes *within* schools – changes in teaching practices, improved attendance, etc. – were not particularly effective at or were too minimal themselves to improve learning scores. Thus, girls who were already in school did not see major gains in learning relative to girls in comparison schools, because they were already reaping benefits from being in school, and changes in teaching quality or other in-school changes were not effective. In contrast, as we discuss in the context of the transition outcome, the programme was quite effective at bringing out-of-school girls into either formal school or alternative learning programmes. If the main driver of gains in Somali literacy stem

statistically significant, the estimates are suggestive and – in the case of out-of-school girls – substantively meaningful – a 5.2 point difference over just two years represents a considerable change in a girl's ability to read.

English literacy

In many ways, the findings for English literacy mirror those of Somali literacy, with no identifiable programme impact among the in-school girl cohort and a small positive impact among the out-of-school girl cohort. Unsurprisingly, performance on the English literacy assessment was significantly worse than for Somali literacy – the average score at the endline across all cohort girls was just 21.7 percent. A significant set of girls, 31.0 percent, were unable to answer a single question correctly at endline.

The number of girls who were unable to complete a single test item correctly raises questions about administration of the assessment. Poor performance may be overstated if enumerators held girls to unreasonable standards for the first two subtasks, in which girls read letters and words aloud. In order for a girl to receive a score of zero for the overall assessment, she must fail to read any of the 100 letters and 60 words presented to her in subtasks 1 and 2, respectively. It seems unlikely that girls would be unable to read any English letters, given that English and Somali share the Latin alphabet.

However, several factors suggest that the frequency of girls scoring zero is not a result of poor administration. First, the share of girls failing to read a single letter decreased dramatically from baseline, where 55.5 percent of girls difference-in-differences not read any letters correctly, to 31.2 percent at endline, a decline which fits our theoretical expectations and functions as a check of whether administration has profoundly changed. Second, a significant share of the girls who were unable to read any English letters were also unable to read any Somali words, which also fits the relationship we would expect to see if both tests were administered well. Third, the pronunciation of letters can be confusing, as girls may imitate the sound the letter makes in Somali and read the letters with a Somali "accent," rather than identifying the letters by their English names and sounds. To ensure that all enumerators approached this task similarly, we dedicated time during training to play audio recordings of a staff member reading letters in English with various accents and pronunciations, so that all enumerators would apply a similar standard.

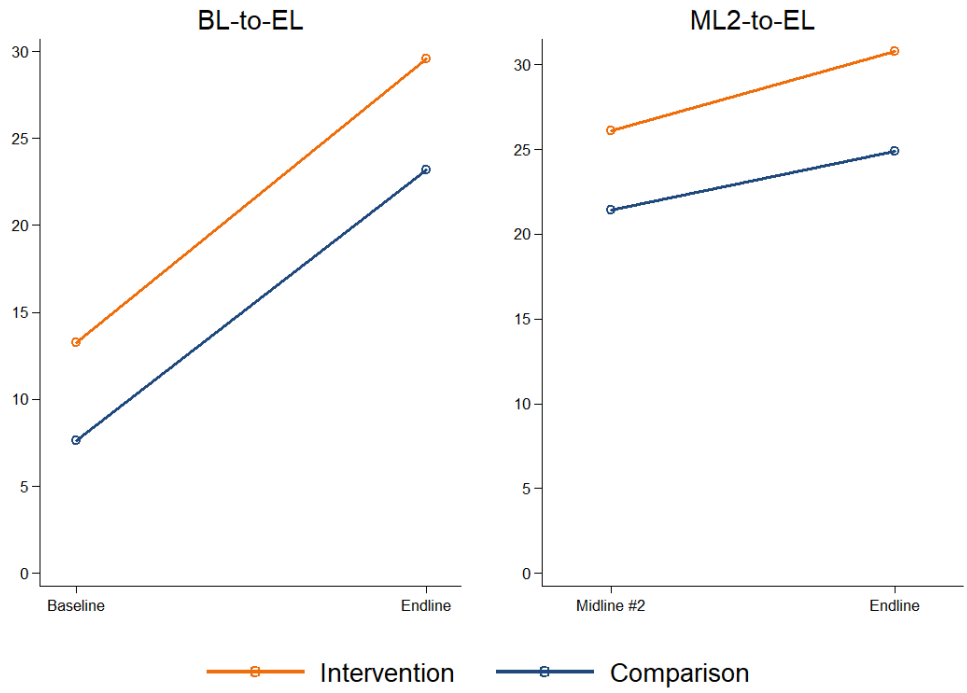
Finally, and most importantly, if grading standards shifted somewhat across rounds, this shift will not bias our results unless it is applied differentially between intervention and comparison schools. Based on analysis of enumerator-specific scores, there is no evidence that either intervention or comparison schools were put at a disadvantage in the grading of English subtask 1 during the endline. Given this, we can be confident in our measure of English literacy and our ability to assess the programme's impact over time.

The figure below reports English literacy scores for the in-school girl cohort. As in previous analyses, the left panel represents the baseline-to-endline trends, while the right panel reports on shorter-term trends, since ML2, with a slightly different sample of girls. While the left panel shows a dramatic increase in English literacy since baseline, this trend is shared by girls in comparison communities. Factoring in the similar trend in

from simply being enrolled and attending a school or ABE/ALP programing, this would explain why the programme was able to "move the needle" among out-of-school girls but had less impact on girls who were already enrolled at the time the programme started.

comparison communities, intervention girls appear to have gained just 0.8 points in English literacy as a result of the programme.

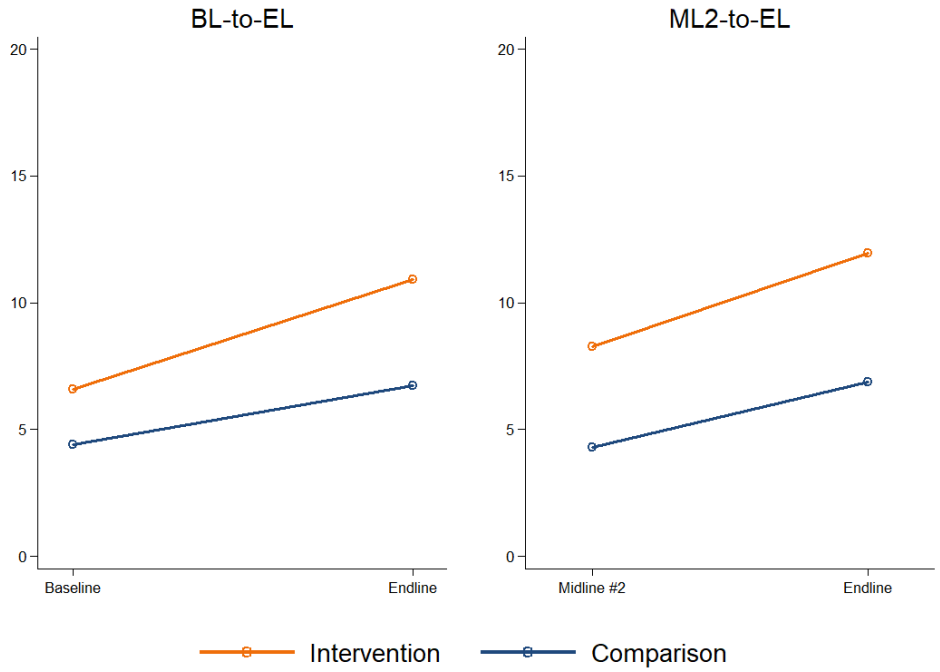
FIGURE 6: ENGLISH LITERACY SCORES AMONG IN-SCHOOL GIRL COHORT



The trends from ML2 to endline are slightly more positive for in-school girls from intervention communities. Here, we estimate that the programme produced a 1.3 point improvement in scores. Girls in intervention communities gained 4.7 points over the two years between 2019 and 2021, while girls in comparison communities improved by 3.4 points over the same period. Relative to other learning outcomes studied thus far, the gains in English literacy experienced by girls – both intervention and comparison – since ML2 are interesting. While we observed improvements of 2.3 points in both numeracy and Somali literacy, aggregating across the full sample of in-school girls, from ML2 to endline, the same set of girls improved by 4.1 points in English literacy. Although intervention girls did not make significant gains above and beyond their peers in the comparison group, the upward trend in the overall sample represents progress in spite of school closures that disrupted learning.

In the context of numeracy and Somali literacy, programme impacts were most positive among out-of-school girls. This remains true for English literacy. As the figure below shows, out-of-school girls in comparison communities experienced little improvement in English literacy, improving from a very low baseline of 4.4 points at baseline, up to 6.7 points at endline. Over the same period, English literacy among intervention girls increased from 6.6 to 10.9 points. After factoring in gains made among comparison girls, we estimate the programme improved English literacy by 2.0 points since baseline among out-of-school girls.

FIGURE 7: ENGLISH LITERACY SCORE AMONG OUT-OF-SCHOOL GIRL COHORT



The right panel of the figure highlights the fact that programme impact was more muted over the ML2-to-endline period. Over these two years, the programme improved English literacy by 1.1 points among out-of-school girls.

Interestingly, our findings are quite different when we combine the in-school and out-of-school girl cohorts into a single estimate of programme impact on English literacy. From baseline to endline, the programme had positive impacts among both in-school and out-of-school girls, at 0.8 and 2.0 points, respectively. When we combine the two cohorts into a single group, we find the programme improved English literacy by 2.6 points, as shown in the top row of the table below.

This finding illustrates the importance of distinguishing the in-school and out-of-school girl cohorts, because this higher estimate of impact is driven by the relative mix of in-school and out-of-school girls across intervention and comparison communities. In this panel, 31.9 percent of the intervention group is composed of out-of-school girls, while out-of-school girls make up 42.3 percent of the comparison group. Under many circumstances, the composition of the sample would not affect our results, because the sample composition remains stable across rounds, and any underlying differences between intervention and comparison groups are accounted for by the difference-in-differences design. However, because the outcome evolved in dramatically different ways between the two cohorts, a sample in which cohort is not balanced across intervention and comparison groups produces bias if cohort is not factored into the analysis. This is one of the key reasons why we generally report findings separate for in-school and out-of-school girls.

The regression models in the table tend to confirm the findings from the earlier graphs. The programme had relatively limited impact among in-school girls; none of the results among this cohort are distinguishable from a null effect. Changes among the out-of-school girl cohort are larger and positive, but the effect sizes are also too small to be distinguishable from zero.

TABLE 11: DIFFERENCE-IN-DIFFERENCES ESTIMATES OF PROGRAMME IMPACT ON ENGLISH LITERACY

Regression Specification	Diff-in-Diff (Program Impact)	P-Value
Baseline to Endline Analysis		
Basic model	2.55	0.35
Controls for region and age	1.87	0.47
Controls for region, age, drought, and enrolment status	2.14	0.4
In-school girls cohort only	-0.43	0.89
Out-of-school girls cohort only	1.71	0.52
Midline #2 to Endline Analysis		
Basic model	1.29	0.49
Controls for region and age	1	0.57
Controls for region, age, drought, and enrolment status	1.02	0.58
In-school girls cohort only	0.49	0.84
Out-of-school girls cohort only	1.46	0.56

Financial Literacy

The final learning outcome we consider in this section is financial literacy. In line with the discussion earlier in this section, financial literacy is – in many ways – an assessment of a girls' ability to apply arithmetic in finance-related word problems. However, it also requires knowledge of financial concepts that might be unfamiliar to many adolescents, such as gross versus net revenue, calculation of profit, and so forth.

Interestingly, the correlation between performance on financial literacy and numeracy, while high (0.54 at the endline), numeracy scores are actually more closely correlated with those of English and Somali literacy. In the endline, numeracy and Somali literacy scores are correlated at 0.79, a very high degree of correlation. The lower correlation between financial literacy and numeracy suggests that the primary difficulty in the former assessment is not simply the arithmetic involved in the word problems, but knowledge of financial concepts and the ability to think through multi-step word problems. This is because the arithmetic used in the assessment is not more difficult than that in the numeracy assessment; but the questions require multiple arithmetic steps, as shown in the example below.

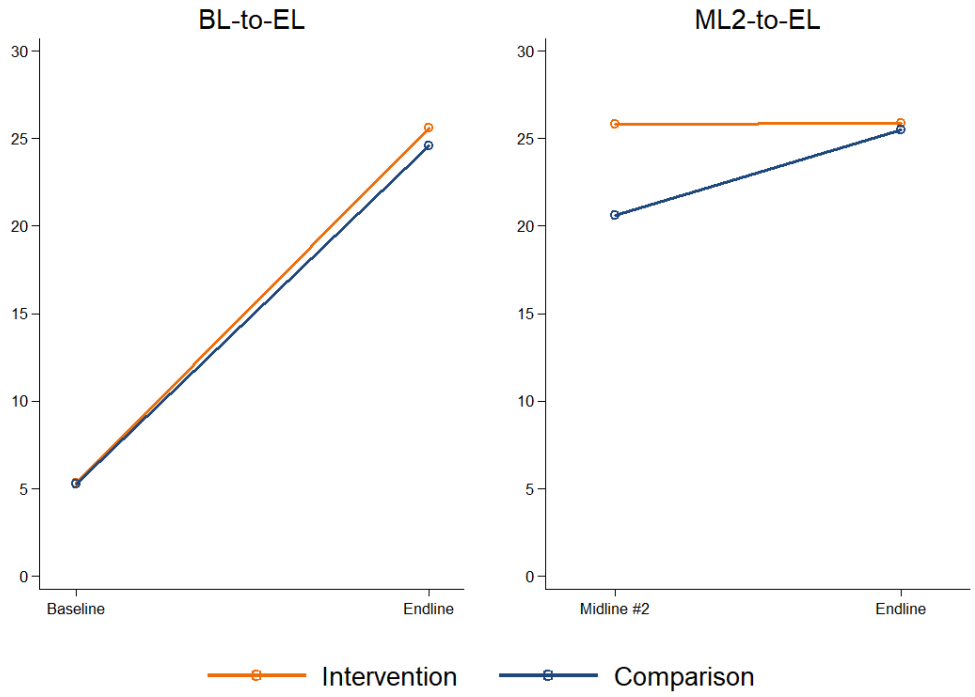
- Fadumo had saved $120. She used $65 to buy a goat. After a year, she had sold the two goat's kids for $50 each. She added the $100 to her savings. How much money does she have now?

In this example, the girl must subtract costs from existing savings, then add the revenue earned from the sale of the two goats to the result of the earlier subtraction problem. A follow-up question asks the girl how much profit Fadumo earned, which necessitates an understanding of what "profit" means.

At baseline, financial literacy was remarkably low, overall; the mean score was just 4.6 percent, and even lower (3.1 percent) among out-of-school girls. As both figures below illustrate, there has been significant improvement since baseline among both in-school and out-of-school girls; the first figure below shows an average gain of 19.9 points over this period, averaging across intervention and comparison groups. The programme's impact over this period was marginal, though, producing a 1.0-point greater improvement among intervention girls in the in-school cohort.

Before discussing the findings from ML2 to endline – which differ considerably – we note that improvements over time among out-of-school girls are less dramatic, overall, and improvements that can be attributed to the programme itself are also smaller (0.2 points). This finding is provided in the left panel of the second figure, below.

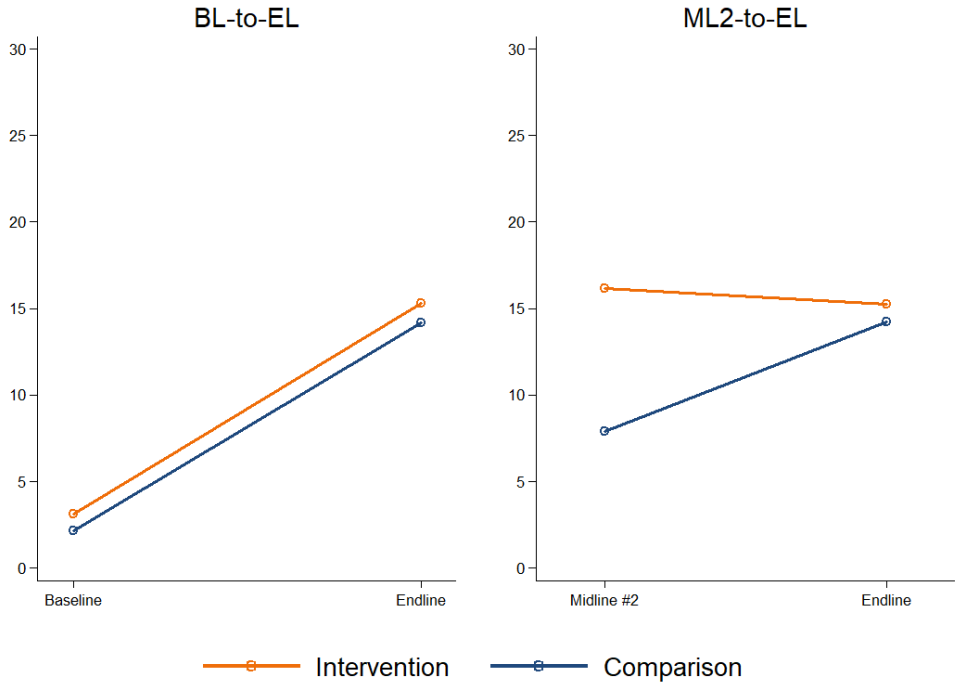
FIGURE 8: FINANCIAL LITERACY SCORES AMONG IN-SCHOOL GIRL COHORT



The change in financial literacy from ML2 to endline – among in-school girls – is shown in the right panel of the figure above. In contrast to impacts over the full programme lifecycle, changes from ML2 to endline have been unambiguously negative. Among in-school girls, the programme produced a -4.8-point change between 2019 and 2021, and girls' scores actually declined in real terms. In the figure below, we document a similar, but even sharper, decline among out-of-school girls. Within this sample, we estimate the programme caused a 7.9-point decrease in financial literacy.

The core finding – very small or no improvements from baseline to endline and sharp declines from ML2 to endline – are robust to the inclusion of additional control variables and use of a combined sample of all cohort girls in the ML2-to-endline panel sample. Indeed, when we investigate the overall impact of the programme – combining in-school and out-of-school girls – we find that the programme reduced scores by 5.5 points, a finding that is statistically significant at the 5 percent level ($p = 0.02$). Regardless of the control variables employed, this result remains stable and statistically significant.

FIGURE 9: FINANCIAL LITERACY AMONG OUT-OF-SCHOOL GIRL COHORT



How can we explain such a large and negative effect of the programme on financial literacy? The most straightforward explanation is that changes from ML2 to endline represent a type of "reversion to the mean" among intervention girls. In the ML2 evaluation, we documented a large 8.4-point impact of the programme since baseline. In that round, girls in intervention communities experienced a sharp improvement in financial literacy; in the current round, this improvement has disappeared almost entirely. This also explains why financial literacy can decline significantly from ML2 to endline, but remain stable over the longer period, from baseline to endline – the programme produced large gains from 2017 to 2019, gains which have eroded since 2019.[35]

Reversion to the mean or "catch-up" within comparison communities is best demonstrated through the analysis of a panel of girls who were successfully contacted in the baseline, ML2, and endline evaluations. The

[35] As with many of the other outcomes where gains made since ML2 have eroded over time, we can likely attribute much of this backsliding to the varied crises facing girls and families in rural areas where SOMGEP-T was implemented – COVID-19 and its economic impacts, especially on the livestock trade; the closure of schools in response to COVID-19; the desert locust swarms that took place in 2020; and the drought that has worsened over time through the start of the endline.

sample is slightly different from others analysed here, but shares many of the same respondents; it is useful specifically because it allows us to track changes across all evaluation rounds.[36] At the baseline, intervention and comparison girls in this sample achieved nearly identical scores, on average – 4.1 percent. At ML2, a gap of 8.4 points had opened up, with intervention girls improving to an average score of 23.3 points, versus 14.9 among comparison girls. In the endline evaluation, however, the same intervention girls showed almost no additional progress, gaining just 0.3 points from ML2 to endline. Over the same period, comparison girls improved by 5.8 points, closing most of the gap with intervention girls that had opened between baseline and ML2.

While the exact scores – and impact estimates – reported for this alternative sample differ slightly from those reported in the figures above, the trends are very similar. And tracking the same panel of girls across three rounds makes clear that the programme's impact on financial literacy was concentrated in the first two years of programme implementation, while comparison girls largely caught up to their intervention community counterparts in the last two years.

Subtask-specific analysis

In our last quantitative analysis of aggregate programme impact on learning, we disaggregate learning scores by skill required (in the case of numeracy) and by subtask (in the case of Somali and English literacy). The motivation for this analysis is the possibility that the programme had particularly strong impact on specific skills but these gains are obscured by null or negative impact on other subtasks. Alternatively, if the programme had impact across most subtasks but large negative impacts on one or two subtasks, this may suggest that the latter subtasks were poorly designed, poorly administered, or otherwise problematic.

For this analysis, we utilised the pre-defined subtasks for Somali and English literacy. For numeracy, we attempted two versions of the analysis: one in which we used individual subtasks and one in which we condensed the subtasks that tested very similar skills (e.g., simple subtraction and addition were aggregated into a single skillset).

The findings of this analysis do not alter our more general conclusions regarding the programme's impact on learning. For instance, across eight Somali literacy subtasks, the programme had a positive impact on scores among intervention girls on four, and negative impact on the remaining four. There is not a compelling pattern to these results – the programme made positive improvements in some easier and some more difficult subtasks. The same is true of numeracy, where negative estimates of programme impact were shared across most subtasks. Two conclusions emerge from this discussion: first, negative or null estimates of programme impact are not driven by poor performance on just one or two subtasks; thus, we cannot attribute negative or null programme impact to poor test item design or administration. Second, there is no evidence that the programme had outsized impact on particular types or difficulties of skills; rather, the trends reported in our aggregate analysis are shared across most subtasks.

Qualitative Analysis

[36] The sample includes 355 comparison girls and 404 intervention girls, for a total panel of 759 girls who were successfully contacted in all three rounds. We exclude the ML1 round from consideration because any panel sample including ML1 is much smaller, due to the exclusion of out-of-school girls and several schools from the ML1 round.

We interviewed teachers, CEC members, mothers, and girls to better understand their perceptions of how learning has changed throughout the programme. In these interviews, we attempted to answer three questions: 1) did SOMGEP-T improve learning outcomes?, 2) Did SOMGEP-T improve learning outcomes among certain subgroups?, and 3) What explains the changes in learning?.

As detailed in the quantitative analysis above, the programme did not have a significant impact on learning compared to the comparison cohort. However, the qualitative perceptions of whether the programme impacted learning demonstrated a more mixed, and largely, positive reflection from teachers, CECs, mothers, and girls on how the programme impacted learning. In this section, we provide an analysis of the qualitative data and what it says about how the programme affected girls' learning, including sub-groups. Throughout the analysis, we have woven reasons provided by interviewees as to why or how the programme did or did not influence girls' learning.

Overall Learning

Interviewees had mixed, but mostly positive, opinions about the impact of the programme on learning on the cohort as a whole. Teachers pointed to improvement in teaching quality, as a result of SOMGEP-T's teacher training programme, as a reason for improved learning: "We have received training from CARE about educating and improving girls' education, such as advising girls on how to reach the same level as boys and helping us with challenges".[37] Mothers concurred that the quality of education had increased: "The quality of the education at the school has improved and it helps the girls to improve their education. My daughter used to study subjects in Somali but now she is studying in English and Arabic and her understanding has increased".[38] A focus group with mothers also pointed to new books and educational materials as reasons for the increase in learning. Contrary to others, several teachers and mothers said that there had been no changes in girls' education during the programme period.

Additionally, many compared the improvement of girls' learning to boys, with one CEC member saying: "The girls were the ones that achieved the best grades in school. The top two girls got a scholarship to university".[39] Others also said that parents had started to focus on the education of their daughters, instead of their sons: "When it comes to education, the parents and community in this area have noticed that the girls are better than the boys. As the boys have adapted to using kat, they have begun to encourage the girls to attend school".[40]

Specific subjects were occasionally mentioned in interviews. As above, several teachers pointed to an increase in teaching quality as a cause of increased learning in specific subjects. Most notably, teachers frequently noted that math teachers had started to use cartoons, drawings, or tangible items like rocks to enhance the lessons: "Math teachers are being encouraged by the trainings they receive. I see them constantly using cartoons which helps the rural students they are teaching a lot".[41] In one instance, a mother described improvements in English literacy by saying: "In the past, students were learning all subjects through Somali language till 8th grade, and this makes it difficult for them to catch up on the secondary level, but now grade 5th is started in the English language and they could reach class 8 with having a good understanding of language".[42] It is unclear whether

[37] FGD with Teachers, Int. 209.
[38] FGD with Mothers, Int. 301.
[39] FGD with CEC Members, Int. 103.
[40] FGD with CEC Members, Int. 105.
[41] FGD with Teachers, Int. 202.
[42] FGD with Mothers, Int. 301.

this change in curriculum was prompted by SOMGEP-T but it may be the driving force behind several comments from teachers and CEC members regarding the improvement in literacy during the programme period.

The impact of COVID-19 on learning cannot be overstated; in fact, it was mentioned in almost every interview as a major barrier to learning. Teachers emphasized the school lockdowns as detrimental to student learning. Some teachers said that they visited homes or used the internet to provide lessons but did not state whether they felt their efforts had a positive impact on learning. A girl said that she tried to keep up with her lessons during the lockdown but, as a result of being home, was also now responsible for more housework than before, impacting her ability to learn. Interestingly, drought was not mentioned as a driver of poor learning outcomes; although, it was frequently discussed as a driver of poor enrolment.

Sub-Group Learning

In the quantitative analysis, we also examined the effects of the programme on specific sub-groups. In that analysis, we did not find significantly improved learning among girls in treatment schools compared to control schools. Discussion of learning among sub-groups was relatively uncommon in qualitative interviews. More frequently, sub-group discussion was related to retention and transition, particularly for GwD and girls from pastoralist households. Below we have provided an analysis of the limited qualitative remarks on how the programme impacted the learning of GwD and girls from either economically marginalized or pastoralist households.

While most interviewees said that GwD were likely to drop out due to physical barriers or bullying, several interviewees commented on improvements in learning among GwD. Most notably, a teacher said that SOMGEP-T had provided glasses to visually-impaired students, in addition to teacher training on disabilities, that had improved GwD ability to learn in the classroom. Likewise, girls in focus groups were aware of the medical treatment that SOMGEP-T had provided to GwD and said, in a response to a vignette, that the programme would supply a GwD with needed equipment.[43] However, this sentiment was not universal and two different CECs stated that children with disabilities had not received any programme support in their region.[44]

The programme also supported pastoralist households, particularly through implementation of ABE: "We have benefited from the CARE project in two major things. Firstly, they had trained our teachers, and secondly they had invested in an afternoon school project, which has benefited many students, almost 60 who are nomadic and in the morning they herd livestock".[45] ABE allowed students from pastoralist households more flexibility than the traditional school system allowed. This was critically important as many interviewees noted that parents would be reluctant to let their children live in a city to received education. By utilizing a more unorthodox approach, pastoralist children were able to complete their home duties and learn at school. However, it should be noted that most comments throughout the interviews described immense challenges to enrolment and so, while, from the teachers' perspective, ABE was effective, girls from pastoralist households still faced many obstacles to accessing school.

[43] Vignette FGD with Girls, Int. 603.
[44] FGD with CEC Members, Int. 106.
[45] FGD with Teachers, Int. 202.

As described above, COVID-19 was a major barrier to learning for all students. However, several interviewees specified that economically marginalised households were more heavily impacted than others, as a teacher described:

> *During COVID-19, there was a decline in the knowledge of students; although students are not the same. Some had access to WhatsApp and TV at home, and some had nothing and they are poor.*
>
> -FGD with Teachers, Int. 208

In focus groups, mothers also picked up on this theme, saying that some households were able to pay for an instructor or internet access during the pandemic but that they were not able to. Others said that they sent their children to either Islamic school or a private school to learn during COVID-19. Additionally, some mothers attempted home learning but complained that they did not have the knowledge, supplies, or even a desk and chair to facilitate learning. At least one mother also said that she was illiterate and could not help her child to learn. Girls in economically marginalised households were much more likely to suffer gaps in learning due to COVID-19 than other girls, according to the qualitative interviews.

Interestingly, another sub-group was also mentioned: girls involved in GEFs. Dissimilar to the above sub-groups, who suffer from stigma and marginalization, interviewees said that girls in GEFs are much more confident and, critically, have higher grades:

> *Some of the changes that we have seen of girls [in the GEF] include being able to speak publicly in front of students, and their grades have risen.*
>
> -FGD with Teachers, Int. 208

In addition, several teachers described girls in GEFs encouraging other girls to learn, both those already enrolled and OOS girls. Mothers also reported that girls in GEFs would work on lessons together outside of the classroom. While not a traditionally labelled sub-group, teachers, mothers, and students alike felt that girls in GEFs learned more and were improved students because of their participation in GEF.

The results from the qualitative interviews are mixed and, at many points, contradict the quantitative findings. Most felt that there were improvements in girls, mostly resulting from improved teaching quality, particularly compared to boys in their same communities. Interestingly, the quantitative results show that girls still lag behind boys in learning, although progress has been made. The improvements in girls' learning may be great enough that community members see them as having reached the same level or greater as boys even though girls still test lower on standardized testing.

GwD learning benefitted from SOMGEP-T's medical treatment and equipment, including hearing aids, eyeglasses, and wheelchairs, which made school more accessible for them. In addition, teachers received training on disabilities and made more appropriate accommodations to GwD to facilitate learning. However, this support was not universal and some interviewees said that the programme had not supported GwD with appropriate equipment. There was limited information on girls from pastoralist households but it seems that more innovative approaches, such as ABE, were thought to be successful in increasing girls' learning.

For all groups, COVID-19 presented an often insurmountable barrier to girls' learning. Girls in relatively wealthier families, that could provide private schooling, tutoring, or internet, fared better than girls in more

economically marginalized households and it is unclear whether the gaps that increased during the school closures will be overcome.

Subgroup Programme Impact

The findings reported thus far have focused on aggregate impact in the full set of cohort girls. In general, our analysis suggests that the programme has not had an appreciable impact on girls' learning, when considering all four outcomes and the full cohort, including both in-school and out-of-school girls.

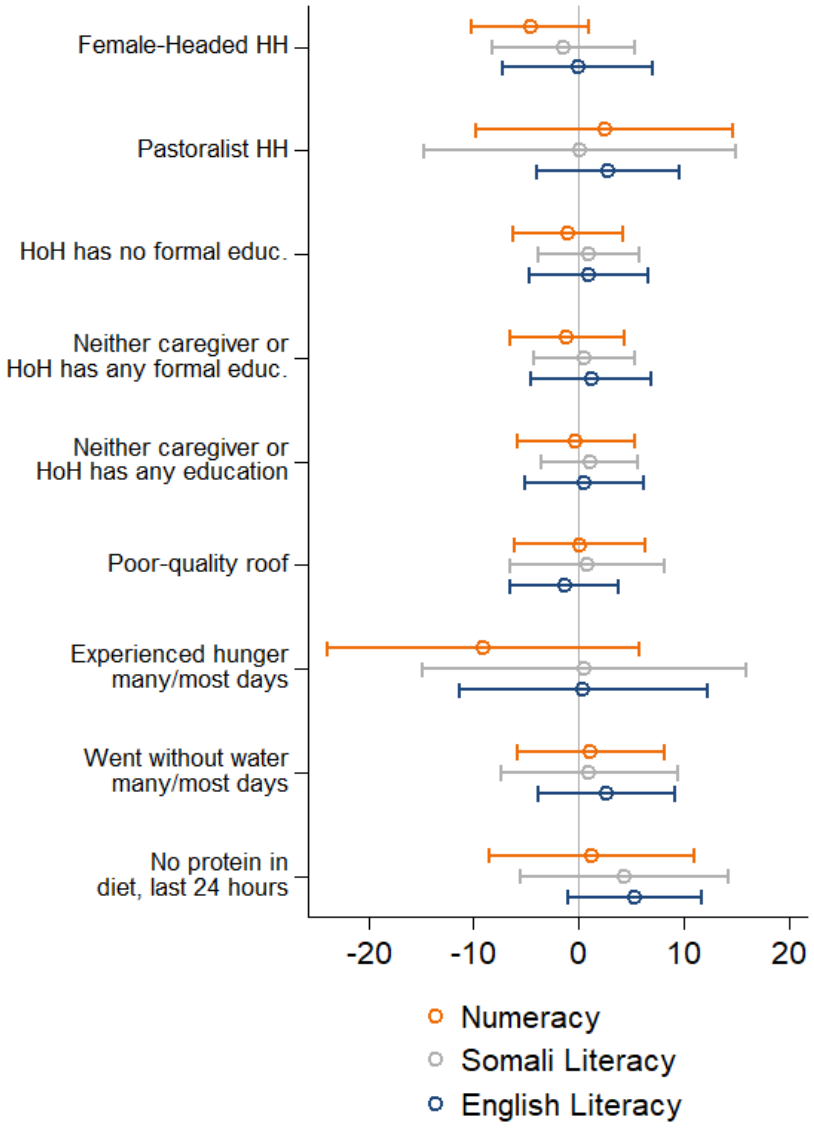
At the same time, there are good reasons to expect that the programme may improve learning among particular subgroups. The first reason is tied to the programme itself, which targeted marginalised girls specifically and provided them with additional support, such as bursaries, among other interventions. If some subgroups receive more intense exposure to programme activities, we would expect the programme to have outsized impact among these groups. The second reason is that students respond at different rates to different interventions. For instance, girls who – at baseline – were responsible for significant household or other non-school duties (i.e. childcare, agricultural work, etc.) might disproportionately benefit from the programme's efforts to encourage caregivers to reduce their girls' domestic workloads. In that case, we might expect particularly large impacts for these girls, if household labour dynamics changed and their attendance in school increased.

Note that our discussion in this section is focused on the estimation of programme impacts within subgroups, not the analysis of differences in learning scores between subgroups. In previous rounds, we dedicated considerable space to analysing subgroup differences, with the goal of understanding the barriers to learning. In this round, we move this analysis to Annex 3, because understanding current differences in learning outcomes across groups is of lesser interest, as programme implementation has ended. In contrast, we remain interested in subgroup-specific programme impact, because it can help us understand why aggregate impact was muted and whether the programme was successful among any particular subgroups.

Our approach to this question is to estimate a series of linear regressions – identical to those employed for the aggregate analysis, utilising difference-in-differences – but restricting the sample to particular subgroups. For instance, we estimate the difference-in-differences among the sample of girls whose head of household has not completed any formal education. The key comparison remains the same within this subsample, which includes both intervention and comparison girls: we are interested in differential trends from baseline to endline between intervention and comparison girls. Note that we have selected a small set of subgroups to assess in this section; a much fuller set of results are provided in Annex 3.[46]

[46] See the subheading in Annex 3 "Subgroup Programme Impact on Learning." The tables provided therein cover a more comprehensive set of subgroups.

FIGURE 10: IMPACT OF THE PROGRAMME ON LEARNING OUTCOMES, AMONG HOUSEHOLD-DEFINED SUBGROUPS



In the figure above, we report the difference-in-differences estimate, a regression coefficient, from a series of regressions. The subgroup among which the regressions are run are listed along the y-axis. For each subgroup, we analyse impact on numeracy, Somali literacy, and English literacy scores. Each dot in the figure is our estimate of the programme's impact (a regression coefficient), while the horizontal bars represent the 95 percent confidence interval around the estimate. To illustrate, the top dot-and-bar in the figure represents the impact of the programme on numeracy scores, comparing only girls from female-headed households to

one another. The sample for this regression model is 414 girls tracked from baseline to endline.[47] The vertical line bisecting the figure at zero provides a visual guide to whether results are statistically significant; where the 95 percent confidence interval (horizontal bars) cross the line at zero, the result cannot be distinguished from a null or zero effect.[48] The first figure below reports results for a series of subgroups defined by a girl's household characteristics, including whether the head of the household is engaged in pastoralism, the educational attainment of adult household members (head of household and caregiver), and indicators of poverty and household economic status.

The overarching finding from the figure above is that any improvements in learning are not especially concentrated among particular subgroups. Our earlier analysis did not reveal any significant impact from baseline to endline among the full cohort in any of the three main learning outcomes. At best, the programme seems to have improved learning outcomes among out-of-school girls, while either reducing performance or having no impact on performance among in-school girls. In the aggregate, the programme had no meaningful impact on learning outcomes from baseline to endline.

When we employ subsamples of girls meeting specific demographic and household criteria, the same is broadly true. Among girls from female-headed households, the programme had a more negative impact on numeracy than it did among the full sample. The estimated -4.6-point impact within this subgroup is not statistically significant at conventional levels ($p = 0.11$), but the finding is relatively compelling, in light of the reduced sample employed. For numeracy, the trends are different. In general, among girls from comparatively disadvantaged backgrounds – e.g., living in a female-headed household, living in a household that experienced hunger more often than others, and feeling like they have little influence over decisions about their schooling – programme impact on numeracy scores were more negative. In the overall sample, the programme produced a 1.5 to 1.7-point drop in numeracy scores; among these subgroups, the programme produced a decline of 4.6 and 5.1 points, respectively.

One interesting partial exception concerns girls in households that have poor diets. We defined one subgroup as the set of households that had not consumed any source of protein (dairy, meat, eggs, legumes, etc.) in the 24 hours prior to data collection. Protein deficiency can inhibit memory and learning, and it is fairly common for households to lack protein sources, especially during climate-related shocks – such as drought – that affect the milk and meat productivity of livestock. In the endline sample, 21.6 percent of cohort girls lived in a household that had not consumed a protein source in the previous 24 hours.[49] Among this group, the programme had more positive impacts on learning scores, especially English and Somali literacy. Among this subgroup, the programme produced a 4.4 point and 5.3 point increase in Somali literacy and English literacy, respectively. Neither result is statistically significant, though the finding for English literacy approaches

[47] Total observations for the regression are 828, with 414 girls in each round. The sample includes 185 comparison and 229 intervention girls, respectively.

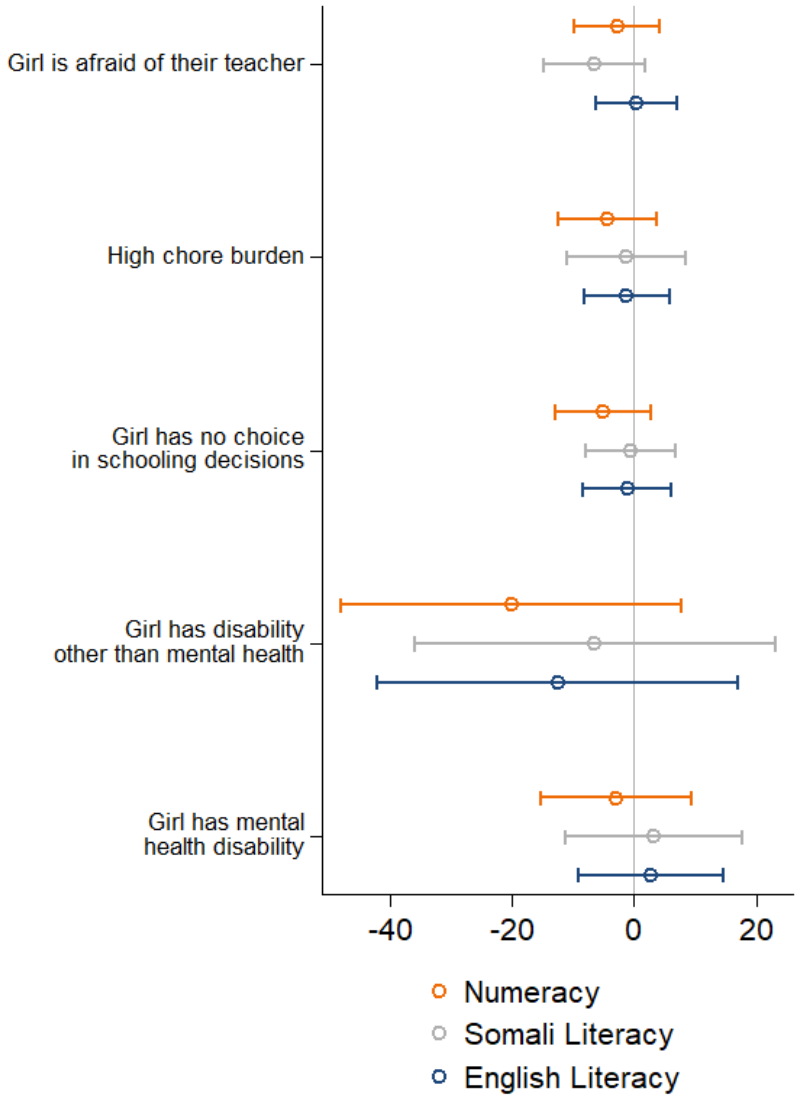
[48] It is worth noting that the use of a 5 percent threshold for statistical significance – implied by the use of a 95 percent confidence interval in the figure – imposes a relatively high bar for significance, especially because subgroup analysis reduces the sample size and the resulting precision of our estimates.

[49] We classified girls into this subgroup on the basis of their responses at endline, which was necessary because dietary diversity information was not collected at baseline. However, we feel it is reasonable to expect that households with limited protein intake at endline probably also experienced insufficient protein intake at baseline – or were more likely to face those deficiencies over the course of the programme – because aggregate dietary diversity over years does not change rapidly. That is, households that face chronic deficiencies in one period are at higher risk of deficiency in a later period, all else equal, though the actual experience of deficiency depends heavily on local environmental conditions.

conventional levels of significance (p = 0.11) despite the small number of girls – just 201 in each round – available for analysis.

In addition to household-level characteristics and household-level poverty, we also studied characteristics and outcomes specific to individual girls. These subgroups are more varied, encompassing girls who report being afraid of their teacher, girls who are assigned a heavy burden of household chores at home, and girls with disabilities. In the figure below, we report subgroup-specific programme impacts among these groups of girls. The analysis mirrors that above: we define subgroups based on survey responses given by either the girl or her caregiver at baseline, and then study whether the programme had impact – using difference-in-differences – among this subsample of girls.

FIGURE 11: IMPACT OF THE PROGRAMME ON LEARNING OUTCOMES, AMONG INDIVIDUAL-LEVEL SUBGROUPS



As with household demographics and household poverty, there are no systematic differences in programme impact between subgroups and the aggregate results reported in earlier sections. Note that girls with disabilities are grouped into two separate analyses, depending on the nature of her disability. For girls experiencing significant levels of anxiety and depression, we group them together in the last category reported in the figure; for girls with other forms of disability – including physical, cognitive, and communicative disabilities – we group them together in the second-from-bottom result. The most compelling finding in this figure concerns this latter group of girls, whose learning outcomes appear to have been harmed somewhat by the programme. However, this finding is based on a total sample of just 18 girls – 11 in intervention communities and 7 in comparison communities. These findings are far too noisy to draw any conclusions regarding programme impact among this specific subgroup.

For nearly all the household-level and demographic subgroups studied, there do not appear to be outsized programme impacts in Somali literacy or English literacy. For each subgroup, our results on these outcomes are null, and our best estimates of programme impact within subgroups are relatively close to zero; in short, even ignoring the criteria of statistical significance, the programme's impact within subgroups for Somali literacy and English literacy are not substantively significant.

Most of the subgroups we studied are defined by some form of disadvantage, typically based on their household characteristics. There are, of course, other groupings within which the programme maybe have had stronger impact. For instance, girls who had greater exposure to schooling may have gained greater benefits, because most of the SOMGEP-T activities intended to improve learning occur within schools.

During the endline evaluation, we collected retrospective data on girls' enrolment status over the life of the programme. While girls were asked their enrolment status during each round's data collection, there were gaps in this data collection: in 2020, when no evaluation was conducted; in 2018 when girls in the out-of-school cohort were not recontacted; and in every evaluation round, when we failed to recontact specific girls.

Before turning to the subgroup analysis, there are several interesting patterns in this data. First, girls in the in-school girl cohort actually did have significantly more exposure to school than girls in the out-of-school cohort. While this is implied by the names of the cohorts, it was not guaranteed, as we defined cohorts by the status of the girls when they were first recruited, regardless of whether they enrolled or dropped out after the baseline. Among girls in the endline sample, in-school girls were enrolled during 4.3 of the 5 years of programming, on average, compared to 1.8 years among out-of-school girls. Second, girls in intervention schools also had greater exposure – they were enrolled in 3.7 years, while girls in comparison schools were enrolled for 3.4 years, on average.[50]

Third, there is a strong correlation between a girls' exposure to schooling and her performance on learning assessments. For instance, for every year of additional schooling (out of five) the girl claimed, her numeracy score improved by 7.1 points, even after controlling for age, region, drought exposure and current enrolment status. A single year of additional enrolment was also associated with a 10.1 point gain in Somali literacy, a

[50] Of course, enrolment and exposure to schooling at slightly different, as girls can be enrolled without actually attending school. Indeed, enrolment can occur without a girl attending school *at all*. We did not attempt to quantify rates of attendance retrospectively, because we do not expect girls to have a strong recollection of their attendance rates from several years in the past. We did ask girls whether they attended school at all during any years in which they reported being enrolled and nearly all girls who were enrolled reported attending at least *some* days during those years. For instance, among 855 girls who were enrolled in 2017, 99.4 percent reported that they attended school on at least one day during the 2017 school year. This measure has severe limitations but suggests that enrolment is a reasonable proxy for at least some exposure to schooling.

4.8 point gain in English literacy, and a 3.2 point gain in financial literacy. Each correlation is statistically significant at the 1 percent level.

In the table below, we report the improvement in learning scores associated with each level of school enrolment over the period from 2017 to 2021. Reading across the table, one year of enrolment during that period is associated with a 6.1 point gain in numeracy, whereas five years of enrolment is associated with a 38.3 point gain, compared to a child who was not enrolled at all during that period. The results show that learning improvements are generally monotonic, improving with each additional year of enrolment. The most surprising result is in *where* the biggest gains occur. In theory, we would expect the biggest improvements in learning – at least in numeracy and Somali literacy – to occur with relatively minimal exposure to schooling, perhaps 1-3 years. During this period, students might learn the basics of numeracy and Somali literacy, the "low-hanging fruit" that requires less schooling to achieve. However, this does not appear to be the case. The biggest year-on-year gains in numeracy and Somali literacy both occur between 4 and 5 years of exposure, suggesting that the biggest improvements in learning arise due to *consistency* of enrolment.

TABLE 12: IMPROVEMENT IN LEARNING AS A FUNCTION OF YEARS OF ENROLMENT, 2017-2021

Assessment Type	1 Year	2 Years	3 Years	4 Years	5 Years
Numeracy	+6.1	+10.4	+17.5	+21.6	+38.3
Somali literacy	+3.9	+11.5	+19.4	+32.5	+51.0
English literacy	+2.8	+5.1	+5.4	+10.9	+27.7
Financial literacy	+3.4	+5.0	+4.8	+11.5	+16.8

Building on this analysis, we investigated whether the programme's impacts were different depending on a girls' exposure to schooling. To be clear, in the analysis that follows we are not comparing learning scores between girls with varied levels of schooling, as the discussion above – and basic theory – already makes clear that greater exposure would produce higher learning scores. Instead, we are interested in whether the differences between intervention and comparison girls are larger among girls with more exposure to schooling.

Unfortunately, there is little evidence for heterogeneous programme impacts of this kind. While the programme's impact on English literacy and financial literacy among the subset of girls who were enrolled all 5 years between 2017 and 2021 – again, comparing intervention and comparison girls, *both* of whom were enrolled for 5 years – the programme had large negative effects on numeracy and Somali literacy among this same group. We interpret these opposing results to be a function of sampling variation and statistical noise, rather than reliable findings, especially given that we generally expect learning scores across disciplines to be correlated with one another.

Our final subgroup-level analysis concerns girls who began from different starting points during the baseline. It is possible that the programme achieved meaningful impact among girls who already had a base level of skills upon which teachers could build; alternatively, the programme may have had particular impact among

girls at the lowest starting points, if basic numeracy and literacy skills are easiest to gain.[51] We divided the panel of girls – those who appear in both the baseline and endline samples – according to their baseline scores on each assessment. We formed three groups of girls for each assessment. Using numeracy to illustrate, we formed a "zero" group for girls who received a score of zero on the baseline assessment; a low-scoring group for those who scored above 0 percent but below 15 percent; and a mid- or high-scoring group for those who scored above 15 percent.[52] We experimented with alternative cutpoints and numbers of groups, but opted for this coding because more groups tended to obscure the key findings; alternative cutpoints generally resulted in substantively similar results – we selected the cutpoints based on empirical examination of the distribution of baseline scores.

The table below reports programme impact among subgroups of girls, defined by their performance on a given learning assessment at baseline. For numeracy, we can see that – among the set of girls who scored 0 percent at baseline – the programme produced a 3.3-point improvement in numeracy. That is, the programme produced gains of 3.3 points above for intervention girls, above and beyond the improvements comparison girls achieved over the baseline-to-endline period.

TABLE 13: PROGRAMME IMPACTS AS A FUNCTION OF GIRLS' BASELINE LEARNING SCORES

	Zero-Group	Low-Achieving Group	High-Achieving Group	Aggregate Impact
Numeracy	3.3 (199)	3.2 (185)	-3.7 (581)	-2.1 (965)
Somali literacy	6.0 (269)	9.2 (136)	-2.23 (560)	0.2 (965)
English literacy	2.7 (372)	9.2 (141)	1.5 (452)	1.9 (965)

The table demonstrates an important caveat to our more general analysis of programme impacts on learning. Programme impacts are relatively high among students who were "low-achievers" at baseline, but these impacts disappear almost entirely among girls who already performed comparatively well at baseline. Consider Somali literacy: if we aggregate the zero and low-achieving group for simplicity, there are 405 girls who scored below 15 percent at baseline. Among this subsample, comparison girls improved from an average score of 1.8 percent at baseline, to a score of 28.7 percent at endline. Over the same period, intervention girls improved from 1.9 percent to 35.1 percent.

[51] Note that we do not claim low-level numeracy and literacy skills are easier to gain. This is an area of educational research that is still open. Our view is that there are reasonable theoretical arguments to suggest that lower-level skills could be easier for schools to inculcate in students *and* reasonable theoretical arguments that those skills are very difficult to teach in a school setting among adolescents who have not acquired the skills previously.

[52] For Somali literacy, the cutpoint between low- and high-achieving was also 15 points. English literacy was difficult to code, because baseline scores were heavily concentrated in the lowest end of the range. We chose a cutpoint of 5 percent, which produced a situation in which 46.8 percent of girls fell into the "mid- or high-achieving" group at baseline. Despite the low threshold used, this is actually fewer girls than were classified as mid- or high-achieving in numeracy or Somali literacy. The choice of cutpoints for analysis of this kind tends to be arbitrary and based on the best judgment of the analyst. We sought to achieve a degree of balance between the sizes of our groups while also respecting what we saw as an important difference between girls who scored 0 percent on an assessment and those who were able to answer at least a few questions correctly.

The findings in the table are extremely suggestive. It appears that the programme was successful in improving learning outcomes among the girls who had been "left behind" previously. Among this group, the programme was able to help them gain core skills in numeracy and literacy, at a faster rate than would have occurred naturally in comparison communities. This finding fits with the programme's interventions to improve remedial education, which we would expect to benefit the lowest-achieving students. It also fits with the focus, in the last two years, on improving fidelity of implementation in schools where poor fidelity of implementation was observed, a fact that may explain outsized benefits among girls who under-performed at baseline.

At the same time, this finding implies that the programme was not effective at improving learning among girls who had already achieved a degree of competency in basic numeracy and literacy. This is an important point because the barriers to higher-level learning – i.e. improvements in learning among high-performing students – may differ from those that prevent learning among students beginning from a lower level. To the extent that the programme was unable to make significant progress against those barriers, it suggests that different strategies may be needed for future programmes to benefit higher-achieving students.

Testing the Theory of Change

The SOMGEP-T Theory of Change identifies a number of intermediate outcomes which are expected to contribute to improvements in the programme's primary outcomes, learning and transition. These intermediate outcomes include improved school attendance, teaching quality, school management, community attitudes, girls' self-esteem and life skills, and a reduction in school-related gender-based violence. In this section, we test aspects of the programme's Theory of Change by investigating whether the links posited by the programme to exist between, e.g., attendance and learning actually operate in practice. In other words, we ask whether improved attendance is actually correlated with and causes improvements in learning outcomes.

We divide this analysis into three sections. First, we consider individual-level intermediate outcomes, which are those that vary across individual girls, even within the same school. Second, we investigate school-level intermediate outcomes, such as teaching quality and school management, and how these intermediate outcomes affect learning scores. Third, we move beyond the intermediate outcomes specified as part of the Theory of Change and study whether exposure to specific programme interventions are associated with higher learning scores. Because SOMGEP-T is an integrated programme with a wide range of interventions, we are generally not able to identify which girls or households received each individual intervention; however, in the case of Village Savings and Loan Association (VSLA) membership and participation in Girls' Empowerment Forums (GEFs), we can identify households and girls, respectively, with exposure and test whether that exposure is correlated with improvements in learning.

Individual-Level Factors

To study the relationship between individual-level intermediate outcomes and learning, we study changes in girls' learning scores over time using a series of regression models. We use two different models: the first is a pooled linear regression, in which we treat repeated observations of the same girls equally. That is, girls can appear in the model multiple times, because they were contacted in multiple rounds. Within this model, we add control variables to account for factors that can influence girls' learning scores *and* simultaneously impact the intermediate outcome being studied. This version of the analysis is essentially a cross-sectional model that

exploits cross-sectional variation – differences in attendance and learning scores between girls are included in the estimation of how attendance and learning are related.

The second model is much better and analyses changes *within the same girl* over time.[53] This model exploits only differences in attendance and learning with the same girl, using changes over time to understand how attendance and learning scores "move together." This model employs girl fixed effects, which accounts for underlying differences between girls, to ensure that our results are driven *entirely* by variation within the same girl over time. This is the most rigorous possible approach to studying the relationship between intermediate outcomes and learning, in the absence of a quasi-experimental design that was established explicitly for this purpose.[54] It is important to note that the fixed effects model explicitly controls for all factors at the girl-, household-, or school-level that might explain learning outcomes and which are fixed across time. This is a fairly conservative model, estimating correlation between the independent and dependent variables using only variation that occurs within individual cases (girls) over time.

To begin, we define a single learning outcome that averages a girl's score on numeracy, Somali literacy, and English literacy, with their score on a 0-100 scale. We aggregate scores in this way for the sake of brevity; throughout the analysis, we also checked whether the results are specific to learning on just one or two subjects, or whether the relationship between the intermediate outcome and learning varied appreciably across subjects.

In the figure below, we report regression coefficients – and 95 percent confidence intervals – for the relationship between a series of intermediate outcomes and learning. The first focuses on caregiver attitudes toward girls' education, which are scored from three questions. In each, respondents were presented with a statement – listed below – and asked their level of agreement:

- Girls' education is worth investing in even when funds are limited
- Girls are just as likely as boys to use their education
- I listen to [girl] when making decisions about her education

For the first two statements, a respondent who agreed strongly was given a score of 1; for the last statement, a respondent who reported listening to their daughter in the context of educational decisions was given a score of 1. Respondents' scores ranged from 0 to 3, with higher scores representing more positive attitudes toward girls' education.

As the results show, there is a very strong cross-sectional relationship between caregiver attitudes and learning outcomes for their daughters – a change in response to just one question (from negative/ambivalent toward girls' education to positive) is associated with a 3.8 point improvement in average learning scores. Importantly, though, the model reported – in orange in the figure – exploits variation *across* girls. There are almost certainly factors which contribute to both more supportive attitudes and improved learning scores;

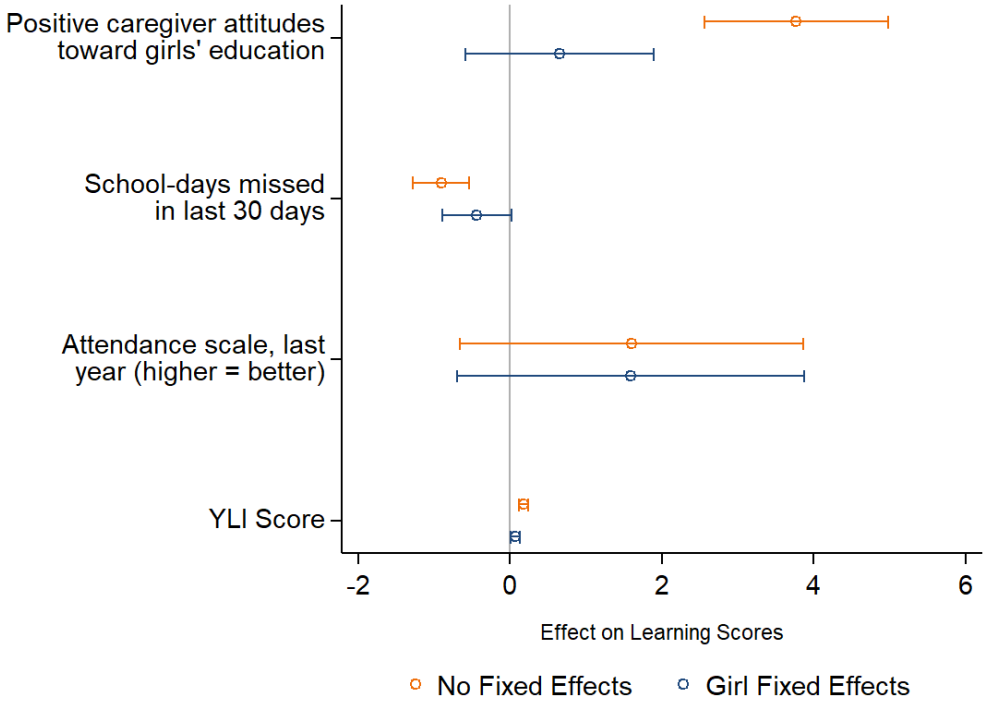
[53] This regression is often referred to as the "within estimator" for this reason; the most typical name, in our experience is a linear panel regression with fixed effects.

[54] While the overall evaluation employs a quasi-experimental design, such a design only applies to the intervention for which it was intended. The difference-in-differences design allows us to study the impact of the overarching intervention. But because this analysis treats attendance rates – and other intermediate outcomes – as "treatment" or independent variables whose effect we wish to study, we would need a quasi-experimental design which artificially applied this treatment (artificially varying attendance, for instance) to a select group of schools or girls. In the absence of such a design, a fixed-effects panel regression is the best approach to understanding the relationship between, for instance, attendance and learning.

for instance, better-educated parents may support education for their daughters, but may also be wealthier and better able to invest in schooling. A number of potential confounding variables exist, which motivates our focus on the second model, employing fixed effects. In this model, a 1-point improvement in attitudes (on a 0-3 scale) is associated with a 0.7-point change in learning scores. This relationship, while showing a positive link between community attitudes and learning, is much weaker and is not statistically significant, helping to illustrate the importance of the fixed effects model for controlling for cross-sectional variation.

The next two results focus on the relationship between school attendance and learning. In many ways, this is the most straightforward linkage between an intermediate outcome and learning. However, we cannot simply assume that attendance produces better learning outcomes, as there are any number of barriers that can prevent students from learning, despite being in school. Attendance alone is not necessarily sufficient to improve learning.

FIGURE 12: INDIVIDUAL-LEVEL INTERMEDIATE OUTCOMES AS PREDICTORS OF LEARNING SCORES



The results in the graph confirm this point – attendance is certainly correlated with learning scores, but the effect sizes are not extremely large. Our first measure of attendance is the number of days of school a girl missed in the previous month, as reported by their caregiver. Absences are capped at 10 to prevent outliers from influencing the results too heavily; a higher number on this measure represents lower attendance rates. As we would expect, a greater number of absences is associated with lower learning scores – for every additional day missed, learning scores drop by 0.4 points, even in the much more conservative model. In a second measure, we also asked caregivers to simply describe their girls' attendance rate since the start of the school year, identifying girls who attended "most", "more than half", "about half," and "less than half" the time. The results below use a scaled measure that runs from 4 (attended most of the time) to 1 (less than half

the time). Both regression models using this measure suggest that higher attendance increases learning scores, though the results are not statistically significant ($p = 0.17$ and $p = 0.18$, respectively).

The final individual-level predictor of learning that we study are YLI scores. YLI scores are measured on a 0-100-point scale, on the basis of 21 survey questions answered by each girl. Higher scores indicate greater self-esteem, self-efficacy, and life skills. In practice, the scale ranges from 0 to 100 and is approximately normally distributed.

The relationship between YLI and learning scores is difficult to see, due to the scale used. However, the relationship is fairly strong and quite robust to the inclusion of girl fixed effects. In the more conservative model, a one-point change in YLI is associated with a 0.07-point improvement in learning scores. While this may not seem like a sizable effect, one must remember two things: first, the improvement in learning is averaged across all three subjects; therefore, a girl who improves by 1 point in terms of their YLI score would be expected to improve by 0.07 points on each of the three learning assessments. Second, a one-point change in YLI scores is exceedingly small, given the 100-point scale. A one-standard deviation change in YLI produces a 1.2-point improvement in learning. While this effect is relatively small, we are confident in the result ($p = 0.04$), especially given that this result employs the more robust fixed effects model.

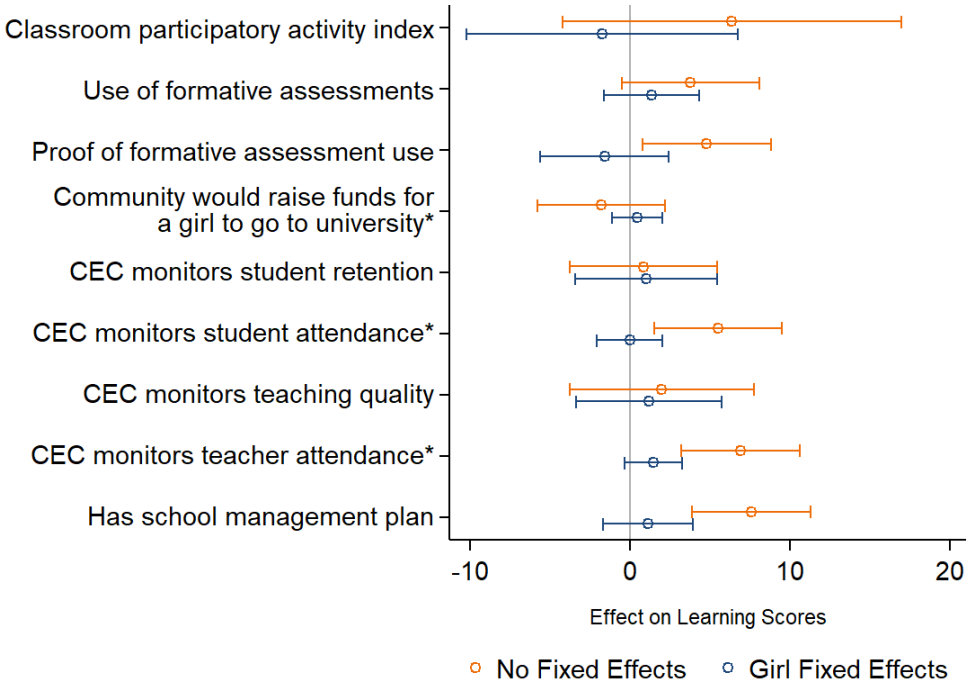
School-Level Factors

In the next set of results, we shift focus from the individual to the school level. While we are still interested in understanding how these factors shape girls' learning scores, the "treatment" in question – e.g., better teaching quality – is applied to all students in the school, to some degree, as opposed to a specific individual. This does not change the structure of our approach in any meaningful way, but it does make our models slightly less efficient, because the "treatment" is clustered at the school level, rather than varying from girl to girl. The result is less precision and a larger confidence interval around our estimates. We continue to use two models, with a strong preference for the panel fixed effects models reported in blue in the figure below.

Based on the results below, teaching quality has a weak and contradictory relationship with learning outcomes, at least once we control for underlying differences between girls. The cross-sectional models, in orange, make it clear that – across schools – there is a positive relationship between teaching quality and girls' learning. However, when we control for differences across schools and girls and focus on how *changes* in teaching quality over time are related to *changes* in learning outcomes over time. In these models, the greater use of participatory methods – calling on students, encouraging participation by unengaged students, using group work, etc. – is associated with – essentially – no impact on learning.[55]

[55] This finding should not be read as an indictment of participatory teaching methods, in general; the more likely explanation for why participatory teaching practices are not linked to more positive learning outcomes concerns their implementation, in practice, within classrooms. It is also possible that teachers who were observed using participatory methods did not use them consistently when their classrooms were not being observed as part of this evaluation, reducing the association between their use and learning outcomes.

FIGURE 13: SCHOOL-LEVEL INTERMEDIATE OUTCOMES AS PREDICTORS OF LEARNING SCORES



By comparison, the self-reported use of formative assessments exerts a more positive impact on girls' learning scores. In schools where both teachers (two teachers were assessed through in-class observations in most schools) self-reported their use, formative assessments are associated with an increase in learning scores – averaging across all girls in the cohort – of 1.36 points. We find this result fairly compelling: while we are often skeptical of self-reports by survey respondents, we have greater faith in this measure than the "proven use" of formative assessments, which required teachers to show documentation of formative assessment use to our field team leaders.[56] Although the relationship between self-reported formative assessment use and learning scores is not statistically significant, the direction of the effect is consistent with our expectations; a stronger study, linking use of formative assessments to students in specific classrooms, may be able to more reliably identify the relationship in question.

Beyond teaching quality, the figure above also documents the relationship between school management – especially CEC activity levels – and learning scores. We asked head teachers, who are themselves members of the CEC in almost all cases, about CEC activities. A core task of the CEC is to monitor the schools, including the quality of infrastructure and teaching, and to check on student retention and attendance. More active CECs also follow up on student absences and dropouts. We employed several measures of CEC activity, including whether they monitored student attendance, teacher attendance, and teaching quality; we also asked head teachers whether the school had a school management plan, a document typically assembled by CECs, and serving as a proxy measure of an effective and organised CEC.

[56] The latter measure imposes a high bar, which teachers may not be able to reasonably meet immediately before or after teaching a class. Moreover, this measure requires that the individual collecting the data know how to identify formative assessments and evidence of their use, which may introduce measurement error.

As the results show, there is a very strong relationship between an active CEC and learning scores, when we look across schools. However, when we consider only changes in these two variables over time, using fixed effects, the results are weaker and more tenuous. Certainly, there still appears to be a relationship between more active CECs and learning scores, but none of the individual measures produce statistically significant results in isolation. A joint test, aggregating across these measures, was also not statistically significant, though the results of this additional analysis are strongly suggestive of a weak, but consistent, relationship between school management and learning.[57]

Testing the effect of specific interventions

In this section, we expand on the analysis above, moving beyond an assessment of the Theory of Change at the level of intermediate outcomes. Here we study the effect of individual interventions, specifically a household's participation in a VSLA and a girl's participation in a GEF. While we are generally unable to identify a girl's exposure to individual interventions – as opposed to exposure to the entire suite of interventions that constitute SOMGEP-T – we are able to identify these interventions because survey respondents were asked explicitly about their participation in each type of group.

The analysis largely follows that of the previous section, employing the more rigorous panel fixed effect regression models to study the relationship between VSLA and GEF membership, on one hand, and learning scores, on the other hand. All of our models have the same basic structure, but are estimated on different subsamples of the data. In all cases, we study a panel of girls tracked over at least two rounds, and use a fixed effect for each girl.

To start, we classify girls as being part of a VSLA household if their caregiver states that they participate in a VSLA, whether the VSLA is explicitly supported by SOMGEP-T or not. In total, 17.3 percent of respondents at the endline participated in a VSLA. Highlighting the fact that some VSLAs exist in these communities outside the scope of the SOMGEP-T programme, 13.4 percent of respondents in comparison communities were members of VSLAs. At the same time, 20.6 percent of respondents in intervention communities were members.

In our preferred model, VSLA membership is associated with 1.8-point increase in aggregate learning scores, which are an average of numeracy, Somali literacy, and English literacy scores. This result is marginally significant ($p = 0.06$), and substantively large, especially when we consider that this average score obscures more substantial effects in the context of English literacy (3.1 points). Across multiple models, this pattern is replicated: VSLAs tended to have the most positive and consistent effect on English literacy scores, while effects on other subjects were marginal and less consistently positive. This fact may be a function of the role that savings and income play in allowing girls to attend secondary school, where English language facility may be particularly enhanced. However, the positive effect remains even when we explicitly factor in grade level and age, suggesting that there are other mechanisms at work as well.

GEFs also appear to have a positive impact on learning, but this relationship is much stronger. We define girls as participating in a GEF if they report having participated in at least one GEF activity in the last year. This is a fairly narrow definition of membership, and we might expect to find that such a low threshold (instead of requiring frequent GEF participation) weakens the relationship with learning scores. That does not appear to

[57] The joint test is simply an F-Test assessing whether the set of coefficients was jointly different from the null hypothesis of zero effect.

be the case, however. In fact, participation is associated with a 6.6-point improvement in aggregate learning scores. This effect holds even when we control for enrolment status, as out-of-school girls are not participants in GEFs and enrolment should be predictive of learning scores.

The improvements in learning scores as a function of GEF participation are also broad-based. The smallest effects were in the context of English literacy, where GEF participation was, nonetheless, associated with an increase of 4.4 points. Again, it is important to emphasise that these effects are estimated *within* girls, rather than *across* girls, a much stricter approach to drawing inferences. In addition, the findings are robust to models in which we limit our classification of GEF exposure to GEFs that exist within SOMGEP-T intervention schools. A majority of girls who participated in GEFs did so within an intervention school, but this is not universally true. However, when we limit the definition of exposure in this way, GEF participation is still associated with a substantively and statistically significant increase in learning scores over time.

Conclusions

The findings in this section provide fairly strong evidence in favour of the SOMGEP-T interventions and Theory of Change, though we are not able to confirm all aspects of the programme design. Some intermediate outcomes appear to be well-connected to girls' learning scores, especially attendance and YLI scores. In both cases, improvements yield improvements in learning scores, even after controlling for differences across individual girls and estimating the models in as rigorous a manner as the research design will allow.

At the school level, the results are less positive and far less conclusive with no real evidence that teaching quality and school management improvements yield significant gains in learning. However, this analysis is less compelling than that of individual-level factors, due to the clustered nature of our measures of both intermediate outcomes. Moreover, we cannot identify the girls who are exposed to particularly low- or high-performing teachers within each school, dampening the precision of our estimates.

Finally, the findings regarding VSLAs and GEFs give us additional confidence in the SOMGEP-T Theory of Change, because these two specific interventions appear to be strongly related to learning scores. While we cannot test all SOMGEP-T interventions individually, this finding gives us a degree of confidence in at least two of the more prominent interventions sponsored by SOMGEP-T.

Of course, these results can easily be overstated. The difference-in-differences research design employed for the overall evaluation does not allow us to use the same quasi-experimental methods for analysis of how intermediate outcomes or individual interventions impact learning scores. We have used the best possible approach to studying these relationships, but even fixed effect models can be biased if there are confounding variables that impact both the dependent and independent variables and which change over time. A more compelling study would involve setting up GEFs and VSLAs in a handful of locations – without being part of a package of interventions – and studying learning outcomes vis-à-vis a set of comparison, non-intervention communities. Nonetheless, the findings above do provide relatively strong evidence, given limitations in the research design, and should be taken as support for the overarching Theory of Change.

ALP Girls Learning

We now analyse the learning performance of girls involved in the Alternative Learning Programme (ALP) who were recruited during ML1. At EL, 444 ALP girls were contacted.[58] This includes 211 girls from the ML1 sample (the ML1-EL panel) and 215 girls from the ML2 sample (the ML2-EL panel) who were successfully re-contacted at EL.[59] Only 165 girls sampled at ML1 were successfully re-contacted at both ML2 and EL (the ML1-ML2-EL panel). The remaining 46 girls contacted at ML1 and EL—but not ML2—may have dropped out of programming in ML2 or been otherwise unavailable at the time of ML2 data collection.

In this section, we analyse changes among the ML1-EL, ML2-EL, and ML1-ML2-EL panels for ALP girls. Our below analysis presents the learning outcomes for these groups of girls; we focus primarily on ML1 to EL results because ML1 represents girls' first exposure to ALP programming, while at ML2, some gains in learning outcomes may already have been made among girls who had been enrolled in ALP in the previous year. Supplementary analysis, however, focuses on changes from ML2 to EL.[60]

TABLE 14: LEARNING OUTCOME SCORES, BY PANEL, ROUND, AND TASK

Panel	ML1	ML2	EL
Numeracy			
ML1-EL	47.3%	N/A	47.3%
ML2-EL	N/A	49.7%	47.7%
ML1-ML2-EL	49.1%	52.9%	49.1%
Somali literacy[61]			
ML1-EL	45.9%	N/A	46.0%
ML2-EL	N/A	49.4%	44.2%
ML1-ML2-EL	48.1%	57.5%	49.9%
English literacy[62]			
ML2-EL	N/A	11.2%	11.1%
Financial literacy[63]			

[58] We attempted to contact all girls contacted at ML1 (including those who were not successfully re-contacted at ML2) as well as all girls contacted at ML2.
[59] At ML1, 365 girls were recruited for the sample. Of these girls, 336 were successfully re-contacted at ML2. However, of these girls, only 189 were in ALP at ML1 and remained in ALP programming at ML2.
[60] In general, we do not present results for the full panel of girls contacted in all three rounds (ML1, ML2, and EL) due to limited sample size.
[61] Only eight Somali literacy sub-tasks were included at baseline and ML1. An additional ninth task was included at ML2 and EL. As such, this table presents results for the eight-task Somali literacy score for ML1-EL and ML1-ML2-EL panels, and for the nine-task score for the ML2-EL panel.
[62] English literacy was not assessed at ML1.
[63] Financial literacy was not assessed at ML1.

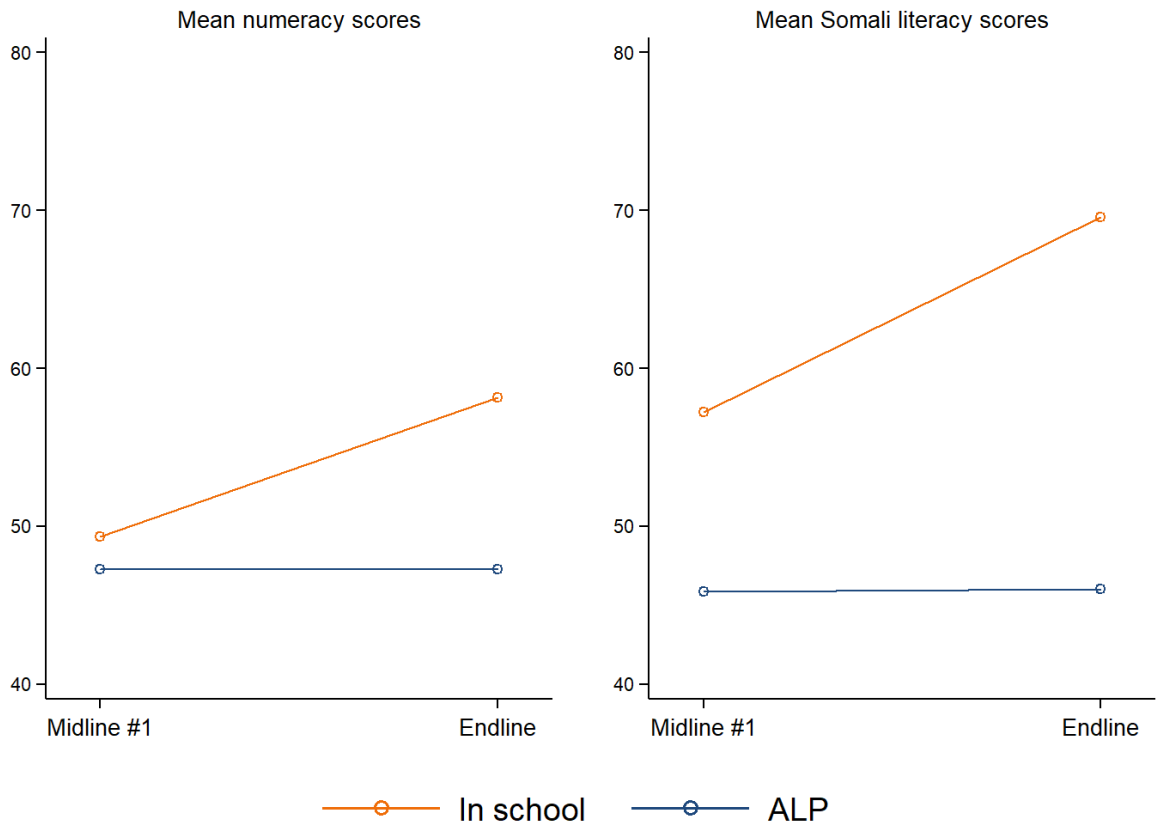
ML2-EL	N/A	22.6%	20.9%

The above table shows scores for numeracy, Somali literacy, English literacy, and financial literacy for each panel of girls. Scores were substantially higher for numeracy and Somali literacy than for English literacy and financial literacy; this was expected, as girls are more likely to have had previous exposure to numeracy and Somali literacy education than education in English or financial literacy. Notably, for each learning outcome and each panel, scores had generally increased marginally or not at all at EL. In general—and as was found at ML2—we would expect scores to improve over time as girls grow older and have greater exposure to education. However, external shocks, such as the COVID-19 pandemic and its effect on schools and learning, may have depressed learning scores. The scores for numeracy and Somali literacy in the above table suggest that this may indeed have been the case; among the ML1-ML2-EL panel, scores increased at ML2, but returned to approximately ML1 values at EL, suggesting that gains in learning through ALP programming were undone in the last year of the programme.

To better understand the potential impact of external shocks, in the below figures, we first compare ALP girls to in-school girls. ALP activities were designed to allow enrolled girls to "catch up" to in-school girls; as such, we would expect ALP girls to have more rapid gains in learning outcomes compared to in-school girls. However, in the context of the negative impact of COVID-19 on learning, instead of more rapid gains, we might expect the ALP programme to mitigate learning losses compared to in-school girls.

However, the figures below suggest that this was not the case. While ALP girls' mean numeracy and Somali literacy scores remained stagnant between ML1 and EL, in-school girls' mean scores increased by an average of almost 9 percentage points for numeracy and more than 12 percentage points for Somali literacy in this same time period. Relative to ALP girls, in-school girls had positive and significant gains in learning outcomes for numeracy and Somali literacy between ML1 and EL. Analysis of ML2-EL results shows, similarly, positive and significant improvements in mean numeracy, Somali literacy, and English literacy scores for in-school girls relative to ALP girls; in-school girls also had relatively greater gains in financial literacy scores than ALP girls from ML2-EL, but this relative difference was not significant. Overall, these findings suggest that ALP activities may not have effectively worked to "catch up" enrolled girls to in-school girls.

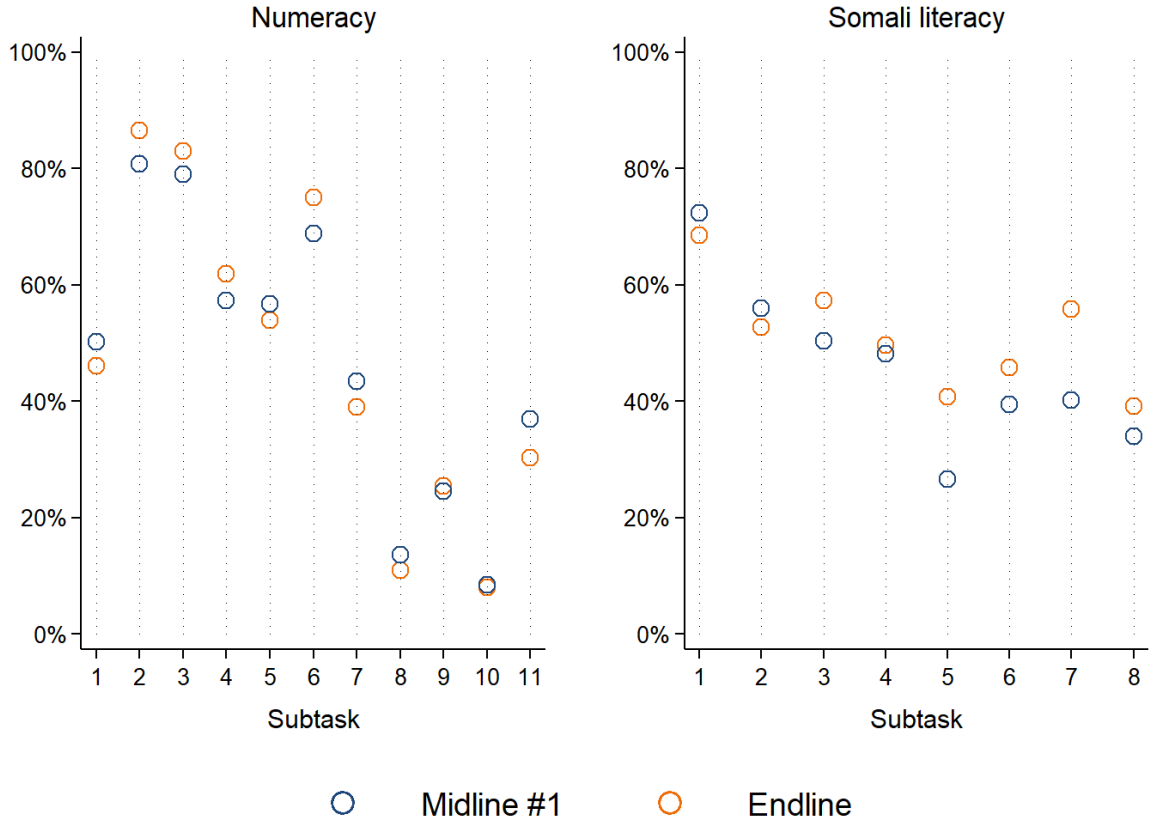
FIGURE 14: TRENDS IN NUMERACY AND SOMALI LITERACY SCORES FOR ALP GIRLS, ML1-EL



Out-of-school girls offer a further benchmark; we would expect participation in ALP activities to result in positive changes in learning outcomes over time relative to these girls, who are not involved in any form of education. As out-of-school girls were not contacted at ML1, we cannot compare results for the ML1-EL panel, but rather use the ML2-EL panel. Using this panel, as with in-school girls, we find that improvement in numeracy, Somali literacy, and financial literacy learning outcomes was significantly greater for out-of-school girls relative to ALP girls from ML2 to EL. This finding is difficult to explain, as we would generally expect ALP girls who have had some exposure to programme activities to at least match the learning gains of out-of-school girls, simply given the effects of maturation. It is possible that, because the ALP programme targeted highly-marginalized girls, these girls had more barriers to learning than out-of-school girls. However, including controls for demographic variables and measures of marginalization—including age, geographic zone, whether the head of household is female, and whether the head of household has an education—we still find significantly less learning gains for ALP girls relative to out-of-school girls.

We now analyse subtask-specific changes in order to better understand patterns in learning for ALP girls. We utilize the ML1-EL panel for numeracy and Somali literacy subtasks, and the ML2-EL panel for English and financial literacy subtasks. The below figures show mean scores on each subtask across rounds.

FIGURE 15: NUMERACY AND SOMALI LITERACY SUB-TASK SCORES FOR ALP GIRLS, ML1-EL



The above figure shows that for numeracy, ALP girls' mean scores improved on sub-tasks 2, 3, 4, 6, and 9 at endline compared to midline, and worsened on the remaining tasks. Positive changes in scores, however, were only significant for sub-tasks 2 and 6, while there was a significant decrease in mean scores for sub-task 7. Sub-task 2 involved basic addition of single-digit numbers, while sub-task 6 involved addition and subtraction word problems. Sub-task 7, meanwhile, involved multiplication of two-digit numbers. In general, the results shown above suggest some improvement in girls' ability to do addition and subtraction (with the exception of sub-task 5, subtraction of two-digit numbers), but stagnation or worsening of abilities to do multiplication and division.

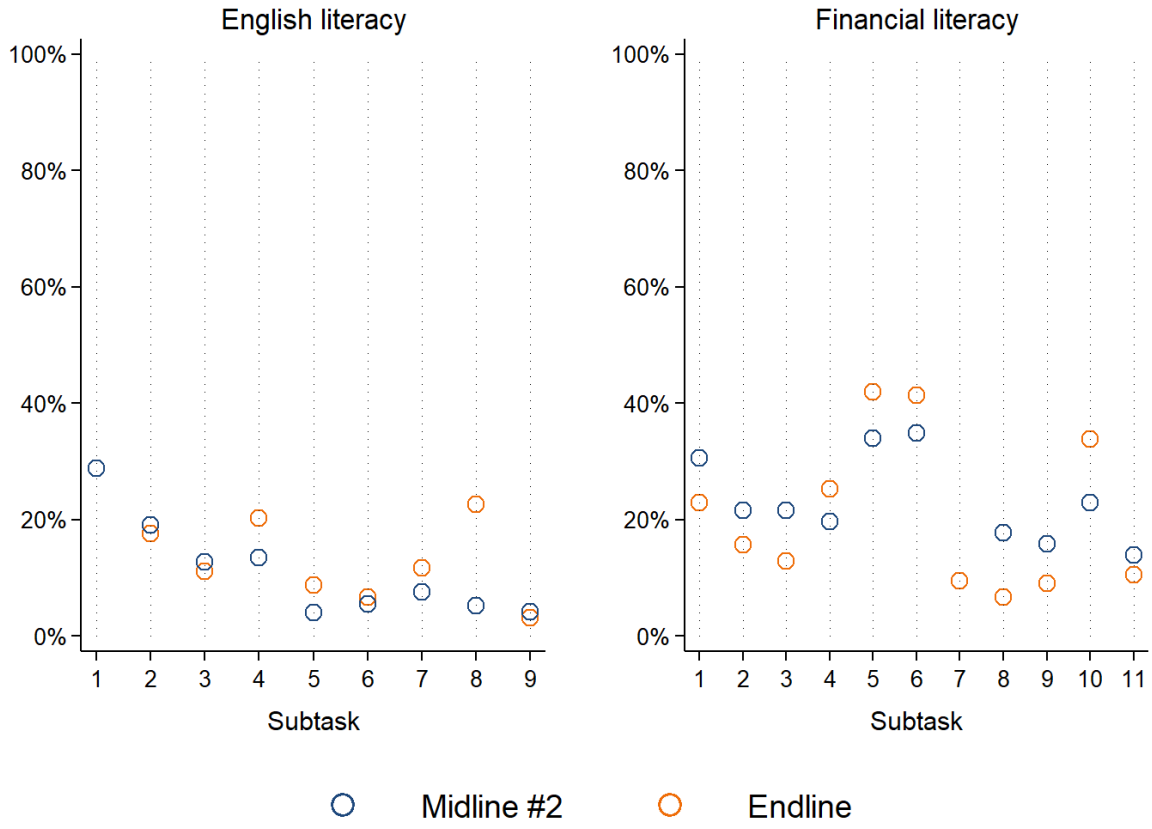
Additionally, it is worth noting that ALP girls scored highest on sub-tasks 2, 3, and 6 at endline, again showing a relatively higher level of ability to complete tasks related to addition and subtraction. There was a substantial decline in scores after sub-task 6 (and, in particular, for sub-tasks 8 through 10) for tasks involving multiplication and division. This suggests, as at midline, that while many ALP girls have some level of understanding of addition and subtraction—especially of one-digit numbers—there remains a substantial skills gap in multiplication and division.

For Somali literacy, ALP girls' mean scores improved on all sub-tasks except for 1 and 2, and significantly increased for sub-tasks 3, 5, and 7. There was no significant decrease in scores for any sub-task. Sub-tasks 3 and 5 involved reading comprehension, while sub-task 7 involved converting a sentence to the negative form. It is difficult to draw a clear conclusion from these results, though they generally accord with the programme's

emphasis – in the context of ALP education – on reading comprehension over reading fluency or the speed of reading. If girls' reading comprehension were improving but fluency remained unchanged, we would expect improvements on subtasks 2, 3, and 5, without appreciable gains on subtask 4. This pattern holds in the figure above, with the exception of subtask 2, where ALP girls experienced a very small degree of backsliding. Nonetheless, the pattern suggests the programme's focus on reading comprehension is being reflected in the outcome of the Somali literacy assessments.

Beyond reading comprehension, there were occasionally significant – but inconsistent – improvements on higher-level skills related to literacy. For instance, sub-task 8 – which, similarly to sub-task 7, involved converting a sentence to the future tense – saw no significant improvement from ML1 to EL. In addition to these findings, it is important to note that while scores were highest at both ML1 and EL for recognition of basic words, only 60.4 percent of ALP girls were proficient in this task, which is a fundamental skill for performance on subsequent tasks.

FIGURE 16: ENGLISH AND FINANCIAL LITERACY SUB-TASK SCORES FOR ALP GIRLS, ML2-EL



Looking now at results for English literacy in the above figure, we find significant increases in mean scores for sub-tasks 4 and 8 from ML2 to EL, and no significant decreases in scores for any sub-task. Similar to Somali literacy, sub-task 4 involved reading fluency, while sub-task 8 involved converting a sentence to the negative form. These findings, in conjunction with those above for Somali literacy, suggest that teachers may have more focused or more effective in teaching conversion of sentences to the negative form than other

literacy-related skills, such as basic letter and word identification. However, it is important to acknowledge that mean scores for all sub-tasks in English literacy were fairly low; no more than one-third of ALP girls answered questions correctly for any English literacy sub-task. Furthermore, at EL, only 19.0 percent of ALP girls were proficient in reading basic English words. This suggests a need to continue strengthening the foundations of basic literacy in order to improve girls' performance in subsequent literacy tasks.

For financial literacy, we find a significant increase in mean score for sub-task 10, and significant decreases in mean scores for sub-tasks 3, 8, and 9.[64] As above, it is difficult to draw clear conclusions from these results; the financial literacy sub-tasks with significant decreases in mean scores involved calculating net and gross profit, but on other sub-tasks related to profit calculation, mean scores improved at EL.

Lastly, we analyse predictors of change in ALP girls' learning outcomes, including household characteristics and community attitudes.[65] For this analysis, we focus on change in numeracy and Somali literacy outcomes from ML1-EL. The below table shows changes in mean numeracy and Somali literacy scores between ML1 and EL by characteristic. We restrict the analysis to girls who belonged to the characteristic group at both ML1 and EL. For example, only girls whose head of household had no occupation at both ML1 and EL are included in the relevant row of the table; girls whose head of household became employed between ML1 and EL, or who lost his or her job between ML1 and EL, are not included, as these girls did not belong to the characteristic group at both ML1 and EL. This restriction, while limiting sample size, allows for more rigorous comparisons over time.

TABLE 15: CHANGE IN MEAN NUMERACY AND SOMALI LITERACY SCORES FOR ALP GIRLS BY CHARACTERISTIC

	Numeracy			Somali literacy		
Characteristic	**ML1**	**EL**	**Change**	**ML1**	**EL**	**Change**
All ALP girls	47.3%	47.3%	0.0	45.9%	46.0%	0.1
HH demographics						
Female-headed household (n = 49)	45.7%	47.2%	1.5	43.4%	45.8%	2.4
HoH has no education (n = 40)	50.5%	51.3%	0.8	53.0%	54.7%	1.7
HoH has no formal education (n = 83)	47.4%	49.3%	1.9	46.9%	51.1%	4.2
CG has no education (n = 50)	50.0%	46.5%	-3.5	46.0%	46.2%	0.2
Pastoralist household (n = 20)	37.3%	40.0%	2.6%	34.2%	40.3%	6.2
HH assets						
HoH no occupation (n = 46)	47.5%	48.2%	0.7	43.9%	46.5%	2.6

[64] There are no responses to financial literacy sub-task 7 for ALP girls at ML2.

[65] We do not analyse by disability due to low sample size (for both non-mental health and mental health disabilities). Similarly, we exclude from the analysis any characteristics for which sample size is low (including those for which few girls belonged to the characteristic group, as well as those for which few girls did *not* belong to the characteristic group).

Owns medium livestock (n = 96)	42.1%	45.8%	3.7	39.2%	41.9%	2.7
Owns land alone (n = 85)	48.5%	48.9%	0.4	46.1%	48.7%	2.6
Household participates in a VSLA (n = 37)	49.9%	48.0%	-1.9	54.6%	51.2%	3.3
Girl Characteristics						
Girls with any disability (n = 20)	42.0%	46.0%	4.0	49.8%	43.9%	-5.9
Girls with any disability other than mental health (n = 5)	40.4%	46.1%	5.7	45.5%	60.2%	14.7
Mental health impairment (anxiety or depression) (n = 16)	40.3%	46.1%	5.8	48.2%	37.7%	-10.6
Community attitudes						
High chore burden (n = 36)	55.0%	56.0%	1.0	53.1%	50.8%	-2.3
Miscellaneous Subgroups						
Participation in a Girls' Empowerment Forum (n = 53)	44.9%	49.2%	4.3	45.7%	52.1%	6.4
No protein intake, last 24 hours (n = 30)	44.6%	47.6%	3.0	44.0%	40.8%	-3.2

This table shows that several characteristics are associated with negative changes in mean numeracy scores from ML1 to EL. In particular, girls whose caregiver had no education had substantially greater declines in numeracy scores from ML1 to EL than the mean of all ALP girls, and girls who had a high chore burden had substantially greater declines in Somali literacy scores. In the context of the above findings, one plausible reason for the former finding may be that caregivers with educations were more able to assist their girls to learn at home, thus mitigating the negative impact of COVID-19. Likewise, girls with higher chore burdens may have had less ability to study at home during school interruptions (or at any time), which may thus have reduced learning outcomes.

In addition to this summary analysis, we run a regression analysis of the change in numeracy and Somali learning outcomes by demographic variable included above, as well as by age. We find no significant relationship between the above characteristics and change in numeracy scores at the 5 percent level, although at the 10 percent level, we find that ownership of medium livestock is associated with greater improvements in numeracy scores. Ownership of medium livestock may serve as a proxy for household wealth or ownership of assets, thus suggesting that ALP girls living in higher-income households were more likely to improve their numeracy abilities. For Somali literacy, we find that only age is a significant predictor of changes in literacy scores; however, surprisingly, we find that older ages are associated with *smaller* increases (or greater decreases) in Somali literacy scores. It is difficult to explain this finding, as in general, age would be expected to be positively correlated with learning outcomes due to the effects of maturation and greater exposure to education (formal or informal). It is possible that this correlation between age and declining Somali literacy

scores is capturing another, unmeasured dimension, such as girls' motivation during the learning assessment or their increased levels of responsibility at home or in the community.

ABE Girls Learning

This section analyses the learning outcomes for ABE girls in numeracy, Somali literacy, and financial literacy. ABE girls were first contacted at ML2; 482 ABE girls were interviewed during this round, of which 347 were re-contacted during EL.[66] As such, this section will focus on changes in learning outcomes for the ML2-EL panel. We will also compare changes in learning outcomes among ABE girls to changes among in school and out-of-school girls. In general, we expect ABE girls to have lower scores for each learning outcome than in-school girls, as the ABE programme was designed to reach girls who have never attended school or who dropped out early in primary school. ABE girls also may—or may not—have lower scores for each learning outcome than out-of-school girls, depending on when these girls dropped out of school (for example, if many out-of-school girls completed primary school before dropping out, they may have higher base levels of education than ABE girls). However, we would expect ABE girls' change in learning outcomes from ML2 to EL to be significantly greater than out-of-school girls, given their exposure to ABE programming.

TABLE 16: COMPARISON OF CHANGE IN LEARNING OUTCOMES FOR ABE GIRLS, ML2-EL

	ABE		ISG			OOS		
Learning outcome	ML2	EL	ML2	EL	DiD	ML2	EL	DiD
Numeracy	25.4%	39.9%	57.7%	60.0%	12.2*	28.0%	34.2%	8.3*
Somali literacy	15.1%	32.5%	63.7%	66.0%	15.1*	24.8%	32.1%	10.1*
Financial literacy	6.5%	12.1%	23.5%	25.8%	3.3	11.5%	16.4%	0.7

The above table shows ABE girls' numeracy, Somali literacy, and financial literacy scores at ML2 and EL. Mean numeracy scores were highest at ML2 compared to those for other learning outcomes and remained highest at EL, with a substantial improvement of 14.5 percentage points. While Somali literacy scores started at a lower point—with a mean score of only 15.1% at ML2—ABE girls saw the greatest increase in these scores at EL compared to other learning outcomes. Mean scores for financial literacy, meanwhile, both started at the lowest level and increased by the least between ML2 and EL. Financial literacy questions involved asking girls to solve a series of word problems related to their ability to manage money in personal and business contexts; these questions thus require both numeracy and Somali literacy skills, as well as some knowledge of financial management. As a result, it is unsurprising that ABE girls performed relatively worse on this learning outcome.

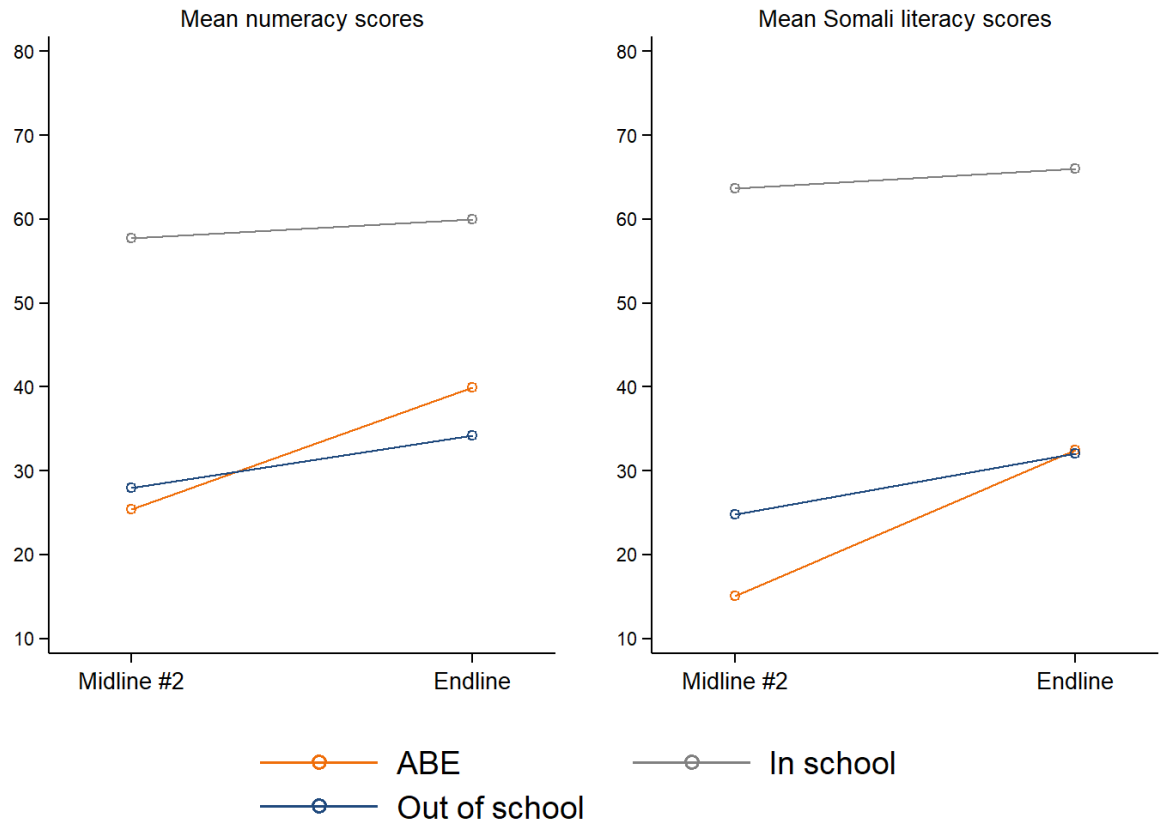
The table also provides comparisons to in school and out-of-school girls, represented further in figures below. This comparison allows us to better understand whether these improvements in learning outcomes were the result of ABE programming or, rather, simply due to maturation or other external effects driving improvement over time. While neither of these groups provides a perfect benchmark by which to judge ABE

[66] However, 481 total ABE girls were interviewed at EL.

girls given the different characteristics of girls in these groups and their differing exposure (or lack thereof) to education, the analysis still allows for a better understanding of the extent to which results were driven by ABE programming, as opposed to external factors. The below figures show the change in mean scores for numeracy and Somali literacy for each of these three groups of girls between ML2 and EL; subsequent analysis also controls for age, a key demographic variable found to influence learning outcomes at ML2.

The figure below shows two interesting and positive trends. First, while ABE girls began from a much lower point in terms of learning outcomes compared to in-school girls—scoring around 30 percentage points lower, on average, on numeracy at ML2 and nearly 50 percentage points lower, on average, on Somali literacy at ML2—and remained at a lower point at EL, their mean numeracy and Somali literacy scores improved at a much higher rate than in-school girls. As shown in Table 16 above, ABE girls' mean numeracy and Somali literacy scores improved, respectively, by 14.5 percentage points and 17.4 percentage points from ML2 to EL. In comparison, from ML2 to EL, in-school girls' mean numeracy and Somali literacy scores improved only by a mean of 2.3 percentage points for both learning outcomes. This relative improvement in the scores of ABE girls compared to in-school girls is significant, including when controlling for age.

FIGURE 17: TRENDS IN NUMERACY AND SOMALI LITERACY SCORES FOR ABE GIRLS, ML2-EL



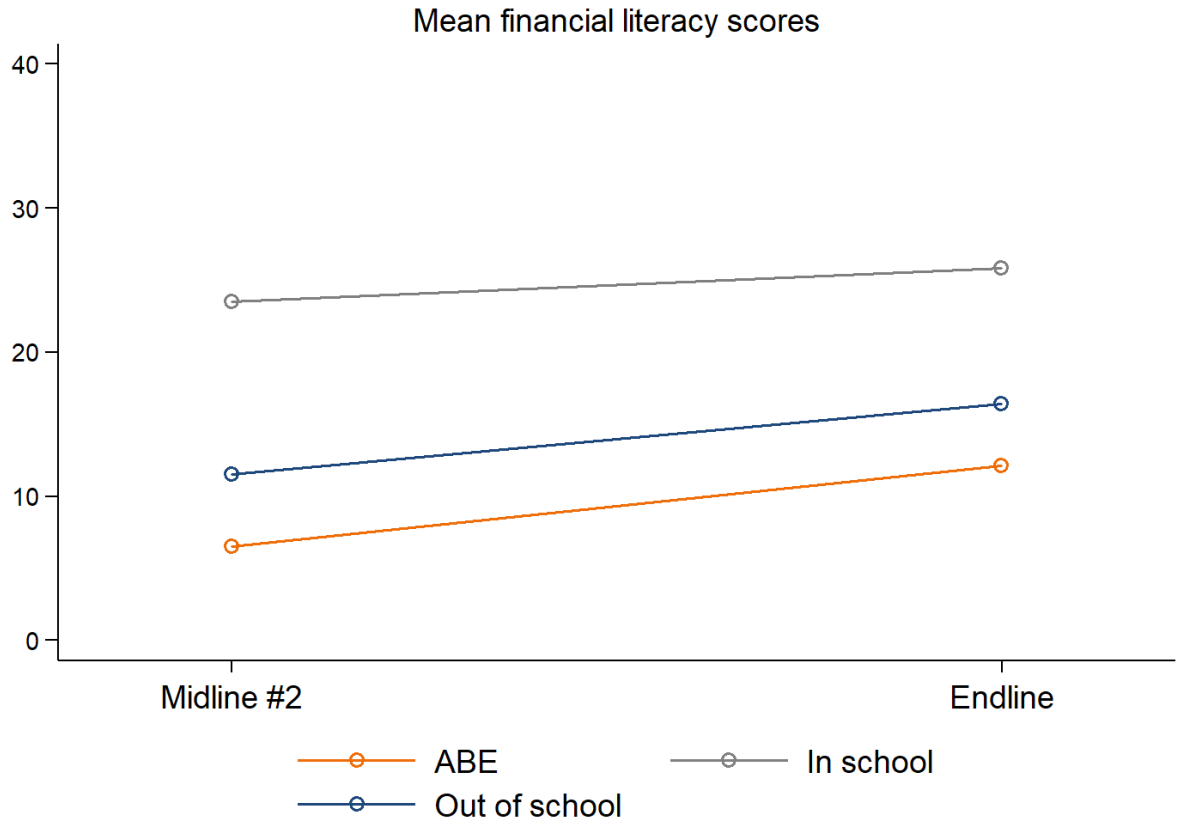
It is worth noting that it is somewhat unsurprising that ABE girls would improve at a faster rate than in-school girls, as it is often easier to learn basic skills than more difficult skills—although, given the high levels of marginalization and low levels of basic education of ABE girls, this effect should not be overstated. In-school

girls, in contrast, may be learning more difficult numeracy and Somali literacy skills that take longer to master. Regardless, these findings still suggest a positive effect of ABE programming to help ABE girls begin to catch up to in school peers.

The second trend of note is that while ABE girls had, at ML2, lower average learning outcomes than out-of-school girls—suggesting a high level of marginalization of ABE girls in line with programme goals—by EL, ABE girls had either surpassed (numeracy) or caught up to (Somali literacy) out-of-school girls. As above, we find that this higher relative rate of improvement in mean scores is significant for both numeracy and Somali literacy, including when controlling for differences in age profiles. These findings first suggest that ABE programming was generally successful in targeting highly marginalized girls with very little educational background, given that their initial learning outcomes were below those of out-of-school girls. Secondly, as with the comparisons to in-school girls, the findings show a positive impact of ABE programming on helping ABE girls "catch up" to their peers.

In comparison to these positive findings for numeracy and Somali literacy, the below figure shows that ABE girls' financial literacy scores, while improving between ML2 to EL, did not improve relative to in school or out-of-school girls. ABE girls' mean financial literacy scores were substantially lower than those for out of school and (especially) in-school girls at ML2, and remained roughly the same number of percentage points below these two groups of girls' scores at EL. Difference-in-differences analysis indeed finds no significant improvement for ABE girls relative to either in school or out-of-school girls, including when controlling for age. Overall, these findings suggest that while ABE programming may have effectively worked to rapidly improve ABE girls' numeracy and Somali literacy scores, it was not particularly effective in improving financial literacy knowledge.

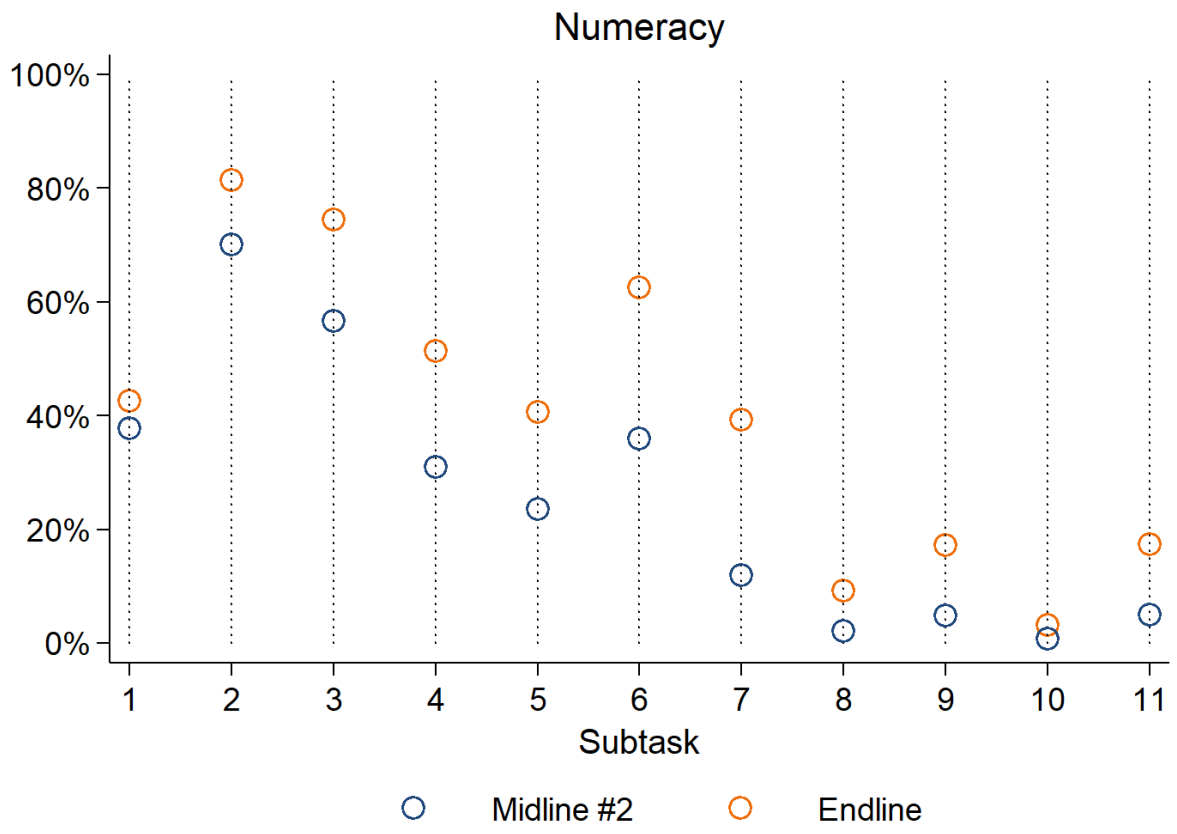
FIGURE 18: TRENDS IN FINANCIAL LITERACY SCORES FOR ABE GIRLS, ML2-EL



Given these findings, we now analyse sub-task specific scores for numeracy and Somali literacy to better understand the specific areas in which ABE girls improved (both in general and relative to in school and out-of-school girls). At ML2, for numeracy, this analysis showed that the majority of girls had achieved proficiency in one digit addition and subtraction, although girls tended to struggle more with subtraction than addition. However, there was a steep decline in abilities to perform two-digit addition and subtraction, and a further decline in abilities to multiply or divide (both one- and two digit numbers). For Somali literacy, we found that ABE girls struggled to identify and read a set of common Somali words, a foundational skill for subsequent literacy tasks.

The below figure shows numeracy sub-task scores across ML2 and EL. We analyse scores at EL as well as girls' improvement on sub-tasks from ML2 to EL. It is first worth noting that, as at ML2, girls performed relatively poorly on sub-task 1, showing relatively low ability to recognise basic number patterns. Scores improved substantially on sub-tasks 2 and 3, one digit addition and subtraction, and then fell for sub-tasks 4 and 5, which assess two-digit addition and subtraction. The sharpest decline, however, came after sub-task 7; scores for sub-tasks 8 through 11, which assess two-digit multiplication and division (one and two digit), were substantially lower than scores for all other sub-tasks.

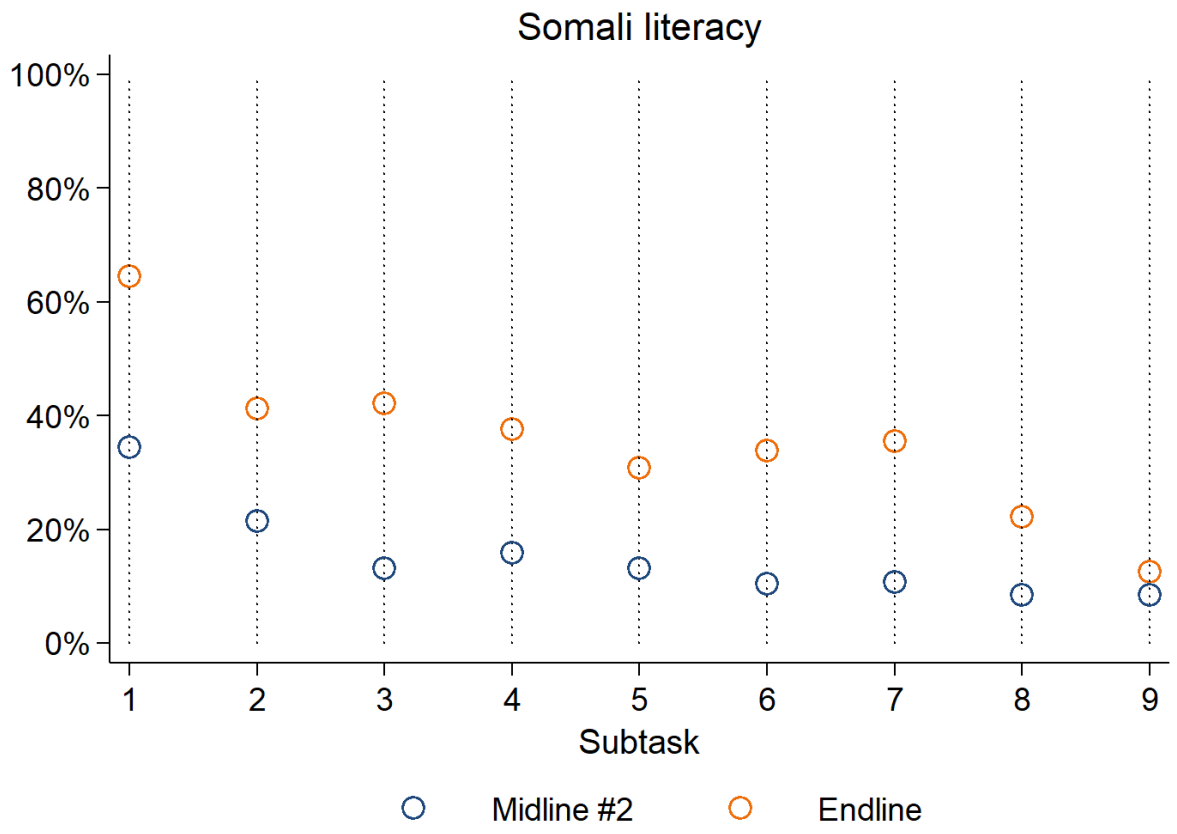
FIGURE 19: NUMERACY SUB-TASK SCORES FOR ABE GIRLS, ML2-EL



Turning now to comparisons of results across ML2 and EL, we first note that ABE girls' scores improved significantly on every sub-task. Given the results above, it is also worth noting the substantial and significant improvement in ABE girls' mean scores on sub-task 7, involving one digit multiplication. At midline, the steepest decline in scores occurred between sub-task 6 and 7, and scores remained very low for all subsequent tasks; in contrast, at EL, girls demonstrated a relatively higher level of mastery of one digit multiplication, although scores for subsequent tasks remained very low. This suggests that ABE activities led to targeted improvement in a foundational skill for later tasks—one digit multiplication—thereby improving girls' likelihood of being able to later master more difficult multiplication and (potentially) division skills.

The below figure shows, as with numeracy, significant improvement in scores on every Somali literacy sub-task from ML2 to EL. The largest improvement, of around 30 percentage points on average, occurred for sub-task 1, identification and reading of basic Somali words. At ML2, only 17.3 percent of ABE girls re-contacted at EL achieved proficiency in identifying and reading common words; at EL, this increased to 50.4 percent of girls. While this is still a low level of reading proficiency, particularly for a fundamental task (for example, at endline, 85.4 percent of all interviewed in-school girls achieved proficiency in this sub-task), these results show marked improvement in just one year. It is also a positive finding that mean scores increased the most for this sub-task, as this suggests that ABE activities successfully targeted the fundamental skills needed for girls to progress to higher levels of literacy.

FIGURE 20: SOMALI LITERACY SUB-TASK SCORES FOR ABE GIRLS, ML2-EL



Finally, we now analyse predictors of change in ABE girls' learning outcomes, including household characteristics and community attitudes.[67] For this analysis, we focus on change in numeracy and Somali literacy outcomes from ML2-EL. The below table shows changes in mean numeracy scores between ML2 and EL by characteristic. As with ALP girls, we restrict the analysis to girls who belonged to each characteristic group at both ML2 and EL. This restriction, while limiting sample size, allows for more rigorous comparisons across rounds.

TABLE 17: CHANGE IN MEAN NUMERACY AND SOMALI LITERACY SCORES FOR ABE GIRLS BY CHARACTERISTIC

	Numeracy			Somali literacy		
Characteristic	**ML2**	**EL**	**Change**	**ML2**	**EL**	**Change**
All ABE girls	25.4%	39.9%	14.5	15.1%	32.5%	17.4

[67] We do not analyse by disability due to low sample size (for both non-mental health and mental health disabilities). Similarly, we exclude from the analysis any characteristics for which sample size is low (including those for which few girls belonged to the characteristic group, as well as those for which few girls did *not* belong to the characteristic group).

HH demographics						
Female-headed household (n = 95)	29.1%	42.8%	13.7	17.6%	36.3%	18.7
HoH has no education (n = 89)	19.5%	34.5%	15.0	7.8%	25.1%	17.3
CG has no education (n = 110)	21.0%	35.1%	14.1	11.0%	28.8%	17.8
Pastoralist household (n = 20)	18.8%	37.5%	18.6	10.9%	37.3%	26.4
HH assets						
HoH no occupation (n = 112)	25.7%	38.5%	12.8	15.8%	31.4%	15.6
Owns medium livestock (n = 172)	22.2%	36.4%	14.2	11.5%	29.1%	17.6
Has regular access to water (n = 97)	26.7%	39.0%	12.3	14.4%	30.5%	16.1
Owns land alone (n = 158)	26.8%	38.7%	11.9	15.5%	32.3%	16.8
Owns land (n = 171)	26.2%	38.8%	12.6	15.1%	32.5%	17.4
Poor roof quality (n = 77)	20.5%	32.9%	12.4	8.3%	25.6%	17.3
Household participates in a VSLA (n = 31)	24.6%	42.0%	17.4	16.6%	33.5%	16.9
Girl Characteristics						
Girls with any disability (n = 25)	25.5%	34.0%	8.5	10.5%	26.1%	15.7
Girls with any disability other than mental health (n = 9)	19.8%	26.3%	6.5	10.0%	17.4%	7.4
Mental health impairment (anxiety or depression) (n = 18)	31.6%	40.0%	8.4	12.0%	31.4%	19.4
Community attitudes						
Girl feels family makes schooling decisions for her (n = 46)	15.5%	31.7%	16.2	9.0%	22.4%	13.4
Miscellaneous subgroups						
Participation in a Girls' Empowerment Forum (n = 20)	23.3%	43.3%	20.0	21.2%	41.0%	19.8
No protein intake, last 24 hours (n = 103)	26.0%	41.1%	15.1	15.2%	37.0%	21.8

This summary table shows little substantial variation in change in numeracy or Somali literacy scores by characteristic compared to the mean change in scores for all ABE girls. However, a further regression analysis of the change in numeracy and Somali learning outcomes by the characteristics included above, as well as by

age, shows some surprising results. First, age is significantly but negatively correlated with changes in numeracy outcomes; in other words, as the age of girls increased, their numeracy scores were significantly less likely to have improved in comparison to other girls, all other demographic characteristics held constant. One possible mechanism for this finding is that older girls may have greater responsibilities in the home, thus giving them less time to study and improve their numeracy scores. Second, we find that girls whose head of household had no education had significantly greater improvement in numeracy and Somali literacy scores from ML2 to EL. Again, this is a surprising finding; we would generally expect girls with access to educated household members to have greater improvements in learning scores. Finally, we find that girls whose head of household had no occupation had significantly less improvement in numeracy and Somali literacy scores; this variable likely serves as a proxy for household poverty, suggesting that girls living in greater levels of poverty may have benefitted somewhat less from ABE programming.

Boys Learning

As part of data collection in each round, the evaluation team completed learning assessments with boys in the same households where girls' learning assessments and household surveys were completed. These learning assessments were intended to provide insight into: the performance of boys in general; their improvements over time, as a function of programme interventions; and the differential in performance between girls and boys and its evolution over time.

Boys were sampled as part of the household survey. At the time of the household survey – in each evaluation round – enumerators completed a learning assessment with a randomly chosen boy from the household, if there was a boy between the ages of 10 and 19 (at baseline). The target age range for boys increased in each round to mirror the increasing age of the girls' cohort. It is important to emphasise that boys were not part of a tracked cohort, and the set of boys assessed in each round changed, depending on their availability, the re-contact status of the girl whose household the boy lived in, and the random selection process. The table below lists the sample sizes of boys' learning assessments in each round. The sample was reduced significantly during the ML1 round, due to the much smaller sample of households that were included.

TABLE 18: BOYS' AND GIRLS' LEARNING ASSESSMENTS COMPLETED, BY ROUND

Gender	Baseline	Midline #1	Midline #2	Endline
Girls	1741	807	1290	1194
Boys	466	189	268	328

The table below reports the grade and age distribution of boys in the baseline and endline samples. The shift toward higher grades is a byproduct of the intentional sampling of older boys as each evaluation round progressed. For instance, the mean age of boys increased from 12.9 years at baseline to 16.4 years at endline, because the target age ranges were adjusted, as noted above. Therefore, it is not surprising that the sample of boys includes a larger share in higher grades than at baseline.

It is interesting to note, however, that the share of boys who are out-of-school decreased. We might expect a higher rate, given that the boys aged and some may have stopped attending after, for instance, finishing

grade 8. One possible explanation is that the composition of the sample of cohort girls shifted away from out-of-school girls over time, as out-of-school girls were more likely to fall out of the sample and were not replaced in the sample. To the extent that enrolment status of boys and girls is correlated within households, this might explain why the number of out-of-school boys fell over time.

TABLE 19: GRADE DISTRIBUTION OF BOYS, ACROSS ROUNDS

Grade Level	Baseline	Endline
Out-of-School	33.7	26.2
Grade 3 or below	19.5	5.8
Grade 4	14.2	4.6
Grade 5	18.2	11.0
Grade 6	14.4	7.6
Grade 7	0.0	7.6
Grade 8	0.0	11.0
Grade 9	0.0	9.2
Grade 10	0.0	6.4
Grade 11	0.0	7.9
Grade 12	0.0	2.7

Our first analysis consists of a straightforward comparison of boys' learning scores to girls' scores across rounds. Again, it is important to note the structure of the sample and how it was selected, because boys were not tracked over time, leaving open the possibility that the sample changed systematically across rounds. In addition, the sample of girls also shifted over time, with major changes occurring after the baseline, when five schools were dropped from the sample because they were positive outliers in terms of learning performance and two schools were dropped for security and accessibility reasons. Usefully, school-level sample changes also excluded boys from the same communities; while it is imperfect, this does suggest that the boys' and girls' samples changed over time in similar ways, due to the exclusion of the same sample points. This argument does not apply to individual-level attrition of girls, however, of which there is a significant amount since baseline. In those cases, boys in households that fell out of the sample would also be excluded, and there is no guarantee that boys in households with girls who fell out of the sample – due to migration, marriage, or for other reasons – are equivalent to said girls. For this reason, caution is warranted in interpreting results comparing boys' and girls' scores over time, because changes over time could be a function of shifting sample composition, rather than true changes in the gender gap.

The table below reports boys' and girls' learning scores for numeracy, Somali literacy, and English literacy, in three separate panels. We employ a "naïve" t-test to check for differences between boys' and girls' scores in each round; we use the term "naïve" to denote the fact that this approach does not account for omitted variables that may produce bias in our findings. For instance, if the average age of boys and girls in a particular

round is different, it may bias our results because age is associated with higher learning scores, even in the absence of enrolment.

TABLE 20: BOYS' AND GIRLS' LEARNING SCORES, BY ROUND

Round	Boys	Girls	Significant Difference?
Numeracy			
Baseline	41.7%	29.1%	Yes
Midline #1	54.4%	50.4%	Yes
Midline #2	60.1%	49.9%	Yes
Endline	62.4%	51.5%	Yes
Somali Literacy			
Baseline	42.6%	31.7%	Yes
Midline #1	56.6%	57.9%	No
Midline #2	55.7%	55.5%	No
Endline	61.7%	58.1%	Yes (p = 0.08)
English Literacy			
Baseline	8.9%	10.3%	Yes (p = 0.07)
Midline #2	25.8%	19.1%	Yes
Endline	29.3%	21.7%	Yes

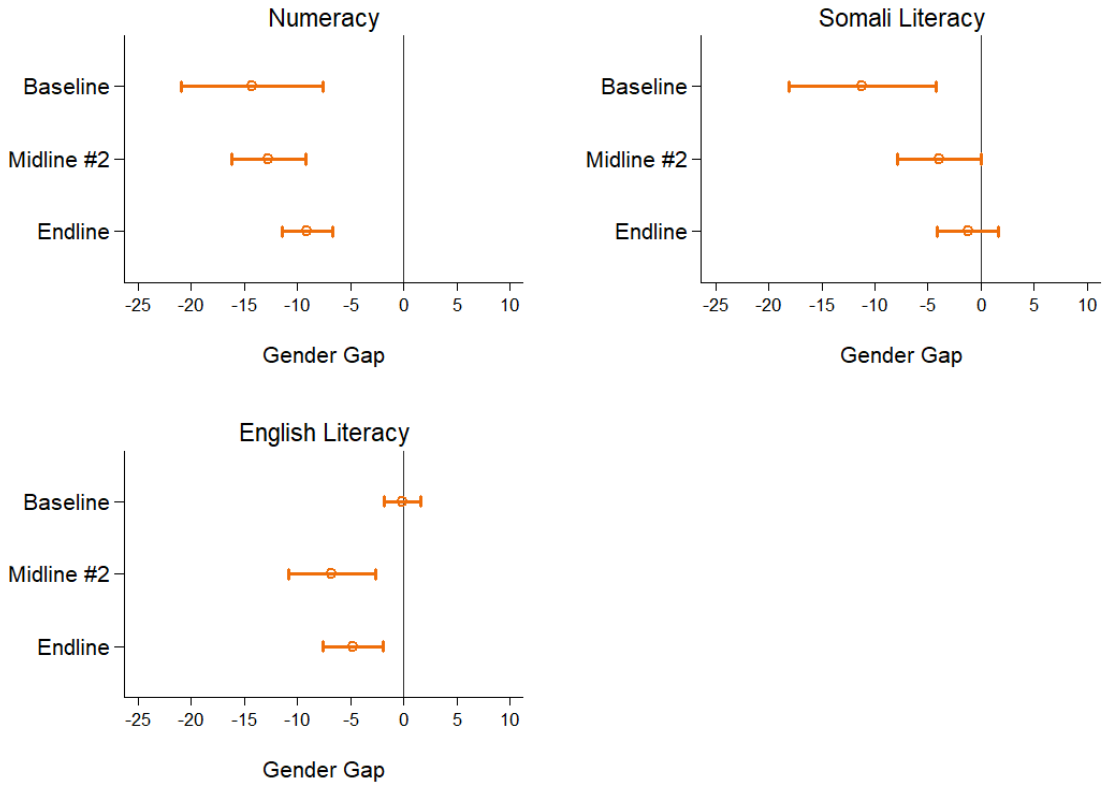
The most compelling finding in the table concerns the persistent gender gap in numeracy scores across rounds. While there is some variation round-to-round, boys' numeracy scores are consistently higher than those of girls – about 12.6 points higher at baseline and 10.9 points higher at endline. At the same time, girls appear to have "caught up" with boys with respect to Somali literacy – at ML2, girls and boys are indistinguishable from one another statistically, and at endline boys score just 3.6 points higher than girls, compared to the 10.9-point gap that was present at baseline. Interestingly, the "catch-up" in Somali literacy is driven almost entirely by girls and boys in intervention schools, where the gender gap disappeared entirely (0.8 points) by the endline, compared to a gender gap of 6.9 points that remained in comparison schools. This strongly suggests the role of the programme in improving girls' learning outcomes, though it is important not to overstate this result, as "catch-up" by girls in intervention schools was only slightly greater in numeracy than their comparison-school peers, and both sets of girls continue to languish equally behind boys in terms of English literacy.

Due to concerns about the comparability of samples across time, we took a further step in the analysis that would allow us to control for some of the major differences in sample composition that might drive learning results. For each evaluation round, we used a linear regression to predict learning scores on the basis of gender, zone, age, and grade level. The goal in this analysis was to control for age, grade level, and geographic

zone, to isolate the relationship between gender and learning scores. While we cannot attribute the gender gap identified through this method as *caused* by gender, it is sufficient for our purposes to understand whether the gender gap changed over time, after accounting for other factors.

The figure below reports the regression coefficients that represent the difference between male and female students' performance, on average. The horizontal bars show the 95 percent confidence interval around the gender gap; where the bars do not overlap with the vertical line at zero, the difference between boys' and girls' performance is statistically significant. In other words, the gender gap is statistically distinguishable from a null effort of zero.

FIGURE 21: GENDER GAP IN LEARNING SCORES, BY ROUND



As the top-left panel in the figure shows, the gender gap in numeracy scores declined slightly over time, from 14.3 at baseline to 9.1 points at endline.[68] The sharpest decrease in the gender gap took place in Somali literacy, where the gap between boys' and girls' performance effectively disappeared by the time of the endline. Unfortunately, English literacy moved in the opposite direction over time, with a widening gap

[68] Note that these gaps are different from those reported in the previous table. As previously indicated, the table reports a simple difference-in-means without control variables; in contrast, the results in the figure account for other factors that are known to affect learning scores.

between girls and boys. At baseline, boys scored just 0.1 points higher than girls, but this had increased to 4.7 points at endline.

It is interesting that differences between boys and girls moved in opposite directions over time. We would normally expect that, if a group caught up in terms of achievement in one school subject, they would also improve, relatively, in other subjects. One explanation for the increased gender gap in English literacy – noted by the project's Monitoring & Evaluation team during the ML2 round – is that boys are exposed to greater levels of English speaking and writing because they have greater autonomy and may travel outside of the village more frequently. They may also have greater incentives to learn English – either in school or on their own – if they plan to move to an urban area or abroad to try to find work. These incentives are less compelling for girls, who are less likely to migrate for employment and much less likely to do so abroad.

In light of the fact that the gender gap in learning scores fell over time, and the programme had limited impact on girls' learning scores in our difference-in-differences models, it is not surprising that the programme has had no discernible effect on boys' learning in intervention communities, relative to comparison communities. To assess the programme's impact in the same rigorous, quasi-experimental framework used for girls' learning, we estimated a difference-in-differences model testing for changes in learning scores among boys in intervention communities, relative to those in comparison communities.

Although the sample of boys is not a panel, difference-in-differences can be employed in repeated cross-sectional samples, as long as the samples are collected using equivalent means from an equivalent sample frame. This assumption is violated to some degree by the changing nature of the household sample, but it is not clear to what extent this would produce bias. Nonetheless, the results do not suggest that the programme has improved boys' learning scores relative to the comparison group of boys – the programme's estimated impact from this method is -0.1 points, -1.4 points, and 0.2 points for numeracy, Somali literacy, and English literacy, respectively. None of the impact estimates noted are statistically significant. As was the case with girls' learning outcomes, boys' performance has improved dramatically over time, but we cannot attribute these changes to the programme itself, as very similar trends occurred within comparison communities.

Grade Level Achievement

In this section, we begin to examine the specific skills that students have achieved in literacy and numeracy, in order to understand where cohort girls fall with regard to broader state and national standards. Our first approach is to assess students relative to the expected skill acquisition for a particular grade level, as laid out in national and state curricula. In the section that follows, we classify students according to their achievement of learning skills that are specifically measured in our learning assessments.

Grade level achievement provides a relatively simple way to describe whether a student has fallen behind their expected skill acquisition and how far behind they have fallen. Educational systems each have their own standards for what constitutes achievement of, for instance, a "grade 8 reading level." This situation is especially complicated in the context of SOMGEP-T, which operates under the jurisdiction of distinct Ministries of Education. The overlapping jurisdictions noted in the introduction to this report means that there is not a single standard curriculum to which all SOMGEP-T schools should be, or are, held. Importantly, the Federal Government of Somalia (FGS) completed development of its curriculum in 2017; the textbooks for this curriculum were distributed in Galmudug schools in 2019-20.

We classified students only in terms of mathematics and English literacy, because the curricula do not include detailed skill acquisition guidelines for Somali literacy. Because Somali (af-Mahatiri dialect) is the mother tongue of nearly all SOMGEP-T students, while English is their second or third language, applying the skills guidelines for English to Somali literacy is inappropriate.

The table below outlines the grade-level standards for mathematics and English literacy and how specific subtasks from the learning assessments map onto these standards. It is important to emphasise the occasionally subjective nature of our mapping of skills to grades – in particular, with regard to English literacy, our assessment of what constitutes "simple sentences" may differ from those who designed the curriculum. Note that some grade levels do not have a standard specified. This occurs in cases in which the evaluation learning assessments did not include a skill specific to that grade level. In some cases, a single subtask was split across multiple grade levels, as when a single subtask includes both simpler and slightly more difficult addition; in these cases, we classify test items on an item-by-item basis.

TABLE 21: GRADE LEVEL STANDARDS FOR MATHEMATICS AND ENGLISH LITERACY

Grade Level Achieved	Mathematics Skills	English Literacy Skills
1	• Addition without carrying numbers (portion of subtask 2) • Subtraction without borrowing (subtask 3)	N/A
2	• Addition carrying one number (portion of subtask 2) • Addition with 3 digits, carrying up to 1 number (subtask 4) • Subtraction carrying one number (portion of subtask 5) • Addition and subtraction word problems with simple underlying arithmetic (subtask 6) • Multiplication of 1-digit numbers (subtask 7) • Division of 2-digit number by 1-digit number (subtask 9)	• Letter identification (subtask 1)
3	• Subtraction carrying two numbers (portion of subtask 5) • Multiplication of 2-digit numbers (subtask 8) • Word problems with simple multiplication and division (subtask 11)	• Identification of basic words, e.g., classroom objects, foods, animals (subtask 2)
4	• Identifying number patterns (subtask 1) • Division of 3-digit number by 2-digit number (subtask 10)	• Reading simple sentences (subtask 3 and portion of subtask 4)

5	N/A	• Reading low-medium difficulty sentences (subtask 5; portion of subtask 4)
6	N/A	• Reading medium-difficulty sentences (subtask 6) • Filling in missing words with medium-difficulty words (subtask 7) • Converting to negative form (subtask 8) • Converting to future tense (subtask 9)

Assignment of students to grade levels is completed in line with FM guidance for grade level achievement. In order to achieve a given grade level, a student must achieve a score of approximately 80 per cent on subtasks (or relevant, grade-specific portions of a subtask) for that grade, and those for the preceding grades.[69] To elaborate, consider a student being assessed for grade 1-level numeracy. They would need to achieve scores of approximately 80 per cent or higher on subtask 3. They would also need to achieve a score of approximately 80 per cent on the grade 1-level portion of Subtask 2, which tests addition that does not require "carrying" numbers. They would *not* need to achieve a passing score on the grade 2-level portions of subtask 2, which tests addition that requires carrying numbers. A student being assessed for grade 2-level competency would need to complete each of the subtasks specified for grade 1 *and* those specified for grade 2.[70]

It is also important to note the relatively small number of test items that are often used to determine whether a student meets a particular standard. For instance, achieving Grade 1 numeracy requires students to answer five addition and ten subtraction questions. Students must answer four of five addition questions and eight of ten subtractions questions correctly to achieve Grade 1 numeracy. While this is a reasonable standard, the relatively low number of test items means that students who make just a few small mistakes may not achieve the standard.

The table below reports grade level achieved in mathematics. This table can be difficult to interpret: the grade level achieved is specified by row, while the grade in which the girl is enrolled is specified in columns across the top of the table. Each column sums to 100 percent, and the interpretation of a single cell is that, e.g., 33.3 percent of cohort girls enrolled in Grade 4 failed to achieve a Grade 1 level in mathematics, and 63.0 percent of cohort girls enrolled in Grade 4 achieved a Grade 1 level in mathematics. To reiterate a point made earlier: each student is classified into exactly one grade level of achievement; therefore, a student who achieved grade-4 competency also met the standards for the lower grades. Within the table, we report achievement for cohort girls – in-school girls and out-of-school girls only, excluding ALP and ABE girls – in intervention communities, with achievement for cohort girls in comparison areas reported in parentheses.

[69] Where there are many items in a subtask, we follow the 80 per cent rule. Where there are four items, we allow one wrong answer (75 per cent correct) without disqualifying a student from achievement of a grade level.
[70] Note that the grade-level analysis includes only the in-school girls cohort – girls who were enrolled at baseline.

TABLE 22: NUMERACY GRADE LEVEL ACHIEVED BY INTERVENTION (COMPARISON) GIRLS, BY CURRENT GRADE

Grade Level Achieved	OOS	Under Grade 4	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10+
Below Grade 1	28.4% (38.6%)	45.8% (50%)	33.3% (26.3%)	25% (7.1%)	27.5% (12.2%)	12.2% (14.1%)	4% (7.5%)	2.7% (5.8%)	4.5% (2.6%)
Grade 1	66.5% (54.6%)	54.2% (50%)	63% (68.4%)	75% (78.6%)	65% (73.2%)	60% (70.3%)	64% (62.7%)	52% (59.6%)	51.5% (48.7%)
Grade 2	5.2% (3.9%)	0% (0%)	3.7% (5.3%)	0% (14.3%)	5% (7.3%)	13.3% (4.7%)	13% (17.9%)	30.7% (11.5%)	22.7% (30.8%)
Grade 3	0% (1.9%)	0% (0%)	0% (0%)	0% (0%)	2.5% (7.3%)	11.1% (10.9%)	14% (7.5%)	10.7% (17.3%)	13.6% (10.3%)
Grade 4	0% (1%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	3.3% (0%)	5% (4.5%)	4% (5.8%)	7.6% (7.7%)

Note: Grade level achievement is specified in the first column; the grade level in which the girl is enrolled is specified in the first row. To determine the typical grade level achievement of a girl in grade 4, one should identify the Grade 4 column and read down the column to obtain Grade 4 students' achievement levels.

The clearest finding from the table mirrors that from previous rounds – very few girls achieve competency at the level of the grade appropriate to the grade level in which they are enrolled. For instance, no girls enrolled in Grade 4 have achieved Grade 3 competency in mathematics, which is the standard they should have achieved when being promoted into Grade 4. Out of 132 girls enrolled in grades 2-5, only 6.1 percent of them achieved mathematics competency at the level of the grade prior. More disconcerting, however, is the number of students enrolled in higher grades who have not reached Grade 4 achievement – just 5.0 percent of Grade 8 girls in intervention schools can perform maths at a Grade 4 level, and relatively few – 13.0 and 14.0 percent, respectively – have achieved Grade 2 or Grade 3 competencies either.

Similar arguments pertain to English literacy, shown in the table below. Just 2.3 percent of girls – intervention or comparison – in grades 3-7 have achieved competence at the grade level they just left (i.e. grade 5 reading for a student currently enrolled in grade 6). Also similar to mathematics, achievement of lower-level reading skills – e.g., grade 3 reading – is relatively uncommon even in upper primary grades. Among intervention girls enrolled in grade 8, only 31.8 percent can read at a 3rd-grade level.

TABLE 23: ENGLISH LITERACY GRADE LEVEL ACHIEVED BY INTERVENTION (COMPARISON) GIRLS, BY CURRENT GRADE

Grade Level	OOS	Under Grade 4	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10+
Below Grade 2	83.5% (83.6%)	87.5% (81.8%)	96.3% (84.2%)	75% (75%)	70% (63.4%)	53.3% (62.5%)	36% (40.3%)	20% (42.3%)	16.7% (30.8%)
Grade 2	11.3% (10.1%)	8.3% (18.2%)	3.7% (15.8%)	20% (17.9%)	27.5% (17.1%)	26.7% (20.3%)	33% (38.8%)	25.3% (28.8%)	18.2% (20.5%)

Grade 3	2.1% (2.9%)	0% (0%)	0% (0%)	5% (3.6%)	0% (12.2%)	10% (10.9%)	7% (7.5%)	14.7% (13.5%)	15.2% (10.3%)
Grade 4	2.1% (2.9%)	4.2% (0%)	0% (0%)	0% (3.6%)	2.5% (4.9%)	6.7% (6.3%)	17% (11.9%)	25.3% (13.5%)	33.3% (28.2%)
Grade 5	0.5% (0.5%)	0% (0%)	0% (0%)	0% (0%)	0% (2.4%)	3.3% (0%)	6% (1.5%)	12% (0%)	12.1% (7.7%)
Grade 6	0.5% (0%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)	1% (0%)	2.7% (1.9%)	4.5% (2.6%)

Note: Grade level achievement is specified in the first column; the grade level in which the girl is enrolled is specified in the first row. To determine the typical grade level achievement of a girl in grade 4, one should identify the Grade 4 column and read down the column to obtain Grade 4 students' achievement levels.

It is not clear what conclusions should be drawn regarding changes since the baseline in terms of grade-level achievement. In practice, fewer girls achieve grade-appropriate competency at endline than baseline (results not shown). However, this may be due – in part – to the fact that the learning assessments do not assess skills appropriate to higher grades; skills included in our assessments map to a maximum of grade 4 mathematics and grade 6 reading. It is also the case that girls that grade advancement has driven a larger wedge between actual skills acquisition and expectations based on official curricula. At the same time, the fact that so few upper-primary girls demonstrate skills considered central to lower-primary learning suggests that, if we were able to assess higher-level skills, the vast majority of girls would not perform at the standard suggested by their grade level.

The nature of cohort girls' pathways through school naturally complicates the story presented in the tables above. Many of the cohort girls were not enrolled when initially recruited and either remained out-of-school or enrolled in school late. Other girls have moved in and out of schooling in ways that inevitably reduce their total learning. While girls should not be enrolled in or promoted to grades significantly higher than their skills acquisition warrants, the reality of how Somali students proceed through their education necessitates understanding that this model does not apply to many students. This argument suggests excusing or rationalising findings that few girls achieve grade-appropriate competency.

Unfortunately, this argument contradicts the reality even for students who have remained enrolled relatively consistently over time. To assess whether grade-level competency is achievable for students who had less complicated transition and enrolment pathways, we identified a subset (n = 176) of cohort girls in intervention schools who reported being enrolled in school in each of the four evaluation rounds. These girls have a greater degree of verified exposure to schooling than most cohort girls, though we did not verify whether they *attended* school frequently. We further restricted this subset to girls who are currently enrolled in grades 7-9 (n = 120).

Among this set of girls – who have been enrolled more consistently than most and who have advanced to upper-primary or secondary grades – just 38.6 percent achieved grade-2 competency in mathematics and just 60.0% achieved the equivalent competency in English. This finding, more than any other, hints at the remarkably poor learning outcomes that even significant exposure to formal schooling produces and suggests that even greater efforts are needed to improve teaching quality and incentivise learning.

Foundational Skill Gaps

The discussion of grade-level achievement in the previous section established that an extremely small share of cohort girls – whether enrolled in school at endline or not – have acquired grade-appropriate skills in mathematics or English literacy. However, the analysis presented in the previous section focuses on a dichotomous assessment of whether a student met curricular standards; It tends to obscure the specific skills actually held by girls because multiple skills are grouped together into a single grade-level standard and because we employ a dichotomous metric of whether a girl met the standard. In this section, we unpack skill acquisition and skill gaps in more detail by analysing scores on individual subtasks and allowing for greater variation in performance than a simple binary "achieved" versus "not-achieved" standard.[71]

As in previous rounds, we place girls into one of four categories on each subtask, depending on their performance *on that subtask alone*. These categories are:

- Non-learners (0%) – these girls are entirely unable to perform the skills captured in a subtask
- Emergent learners (1-40%) – these girls may have a basic grasp of the skills required, but are unable to apply those skills widely
- Established learners (41-80%) – these girls understand the skills required but cannot apply those skills in all cases or to the more complex questions in a given subtask
- Proficient learners (81-100%) – these girls have achieved relative mastery of a particular skill

The idea underlying this analysis is to identify patterns in girls' performance that reveal something systematic about the ways in which they do or do not learn. For instance, if girls tend to underperform on word-based mathematics problems but can perform the arithmetic operations that underlie those word problems, it suggests that they understand the mathematical principles, but lack the ability to practically apply mathematical skills to real-world problems. The analysis can also reveal the levels at which most girls begin to struggle, splits in the sample between girls who perform very well and those who cannot complete a subtask at all, and so forth.

In the tables below, we report the subtask-specific performance of two sets of girls: all cohort girls in intervention schools, in the top panel of each table, and cohort girls in intervention schools who were enrolled at the time of the endline evaluation, in the bottom panel of each table. Additional tables, which include the full set – across both intervention and comparison schools – of cohort girls, are provided in Annex FILL-IN.[72] We report on ALP and ABE girls' foundational skills and skill gaps in a separate section below. The tables are best understood by considering a single subtask, such as numeracy subtask 3, in which 76.6 percent of students are classified as "proficient" because they scored 81 percent or higher on the ten individual test items contained

[71] In previous rounds, this section was framed as an opportunity to adjust programming, because it allowed readers to see where skills "dropped off" – potentially allowing teachers to target their lessons more accurately at students' skill gaps. While SOMGEP-T interventions have now ended, this analysis is still useful because it helps us understand where gaps continue to exist and the exact extent of functional skills girls have gained. The findings help contextualise the analysis of aggregate learning scores by clarifying the level of learning cohort girls have achieved in practical terms that all readers can understand.

[72] In past years, we have reported data for all enrolled girls, from both intervention and comparison schools but in this round we focus our main analysis only on intervention schools. The primary reason we relegate comparison schools to an annex is because this analysis is not intended to draw distinctions between intervention and comparison schools or to assess programme impact. We analysed subtask-specific programme impacts – using difference-in-differences – earlier in this report. Instead of programme impact on subtask-specific skills, our interest here is in understanding patterns of achievement among girls to help demonstrate what learning gaps continue to exist even among girls in SOMGEP-T schools.

within the subtask. In contrast, 11.2 percent of students achieved a mean score of 41-80 percent on the same set of ten test items, and are classified as "established learners," in line with the FM's guidance.

The table below documents skill achievement and skill gaps in numeracy across 11 subtasks. The first pattern that emerges from this table is that the identification of patterns in numbers is more difficult than we would anticipate for the assessment's first subtask. Just 11.1 percent of cohort girls in intervention communities achieved a score of 80 percent or higher across ten pattern-identification items. Children are generally able to complete these items, but not reliably, even though the underlying arithmetic – requiring very basic addition or subtraction – is no more difficult than the addition and subtraction problems presented in subtasks 2 and 3. One possible explanation for comparatively poor performance on subtask 1 is that performance requires understanding the logic of the exercise, pattern recognition, which is less obvious than arithmetic operations that are presented in exactly the same way as in their classes. In other words, asking students to complete "5 + 3" is identical to the *structure* of tasks they have undoubtedly completed in class, while identification of number patterns may be a comparatively foreign concept, even if the underlying arithmetic is no more difficult.

Also notable is the gap in performance that arises when the patterns in subtask 1 become slightly more complicated. Girls tend to perform well on tasks that essentially consist of counting – e.g., subtask 1 item 2, which asks girls to fill in the blank in the list "40 41 __ 43." Among these tasks, girls achieved a mean score of 88.5 percent. Slightly more difficult items move beyond counting but involve a pattern centred on common "round numbers," such as:

- A pattern incrementing by ten, using round numbers, e.g., "60 70 80 __"
- A pattern incrementing by five, also using relatively round numbers, e.g., "260 265 __ 275"

The most difficult pattern identification tasks are those that are not simply counting (increment by an integer above one) and that do not increment by round numbers.[73] Girls' performance decreased monotonically across these types of tasks: they perform best on the simplest "counting" items, somewhat worse on items moving beyond counting but utilising round numbers and increments of 5 or 10, and still worse on items that do not utilise round numbers. On the two most difficult items, girls averaged a score of just 13.9 percent, compared to 88.5 percent on the easiest counting-based items.

Subtasks 2 and 3 demonstrate that girls, in general, are quite proficient in basic addition and subtraction. Those who do not meet the standard of proficiency fall into the "established learner" category; very few girls (7.3 and 12.2 percent) received scores below 40 percent on these two subtasks, respectively. Interestingly, girls show no decline in performance when moving from tasks involving addition below ten to tasks involving addition above ten. But girls *do* show a decline – though one which is small – between subtraction problems involving only single-digit natural numbers and those that use the numbers ten or zero. This suggests that a small number of girls are able to subtract numbers on the 1-9 spectrum but struggle with slightly more complicated operations.

[73] An example is subtask 1, item 10, which asks girls to fill in the blank in "7 12 __ 22." This pattern increments by five, but does not use round numbers as the individual values, obscuring the pattern slightly.

TABLE 24: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION GIRLS, NUMERACY

Subtask	1	2	3	4	5	6	7	8	9	10	11
Skill Assessed	Missing number	Addition (Level 1)	Subtraction (Level 1)	Addition (Level 2)	Subtraction (Level 2)	Word problems (add/sub)	Multipl. (Level 1)	Multipl. (Level 2)	Division (Level 1)	Division (Level 2)	World problems (mult/div)
Intervention Girls											
Non-Learner (0%)	3.6%	5.3%	7.8%	16.8%	25.7%	12%	26.3%	65.9%	41.4%	77.4%	52.8%
Emergent Learner (1-40%)	34.3%	2%	4.4%	13.9%	12%	5.1%	11.1%	12.5%	21.3%	10%	0%
Established Learner (41-80%)	50.8%	15.4%	11.2%	31.6%	32.2%	28.8%	33.2%	10.6%	17.6%	7.8%	19.9%
Proficient Learner (81-100%)	11.4%	77.3%	76.6%	37.7%	30.1%	54%	29.4%	11.1%	19.6%	4.8%	27.3%
Intervention Girls Enrolled in School											
Non-Learner (0%)	2.2%	3.1%	4.7%	9.6%	18.5%	7.8%	15.4%	55.1%	28.3%	70.5%	42.4%
Emergent Learner (1-40%)	28.8%	1.6%	3.3%	13.6%	10.3%	3.8%	10%	15.6%	25.4%	12.9%	0%
Established Learner (41-80%)	54%	13.4%	10%	31.3%	34.2%	27.9%	36.8%	14.1%	21%	10%	23.9%
Proficient Learner (81-100%)	15%	81.9%	81.9%	45.5%	37.1%	60.5%	37.7%	15.2%	25.2%	6.5%	33.7%

As in previous evaluation rounds, there is a notable decline in girls' performance between easier (subtasks 2 and 3) and more difficult (subtasks 4 and 5) arithmetic. The evidence suggests that a set of girls can only complete more difficult addition problems through the application of rules of thumb. This is demonstrated by the gap in performance between items 1 and 2 of subtask 4. Both involve two-digit addition, but the former (10+13) can be answered by relying on a very simple shortcut, remembering that adding ten simply changes the first digit of the other number in the problem; the latter (21+43) does not have a concrete rule of thumb on which students can rely and fewer students (13 percentage points fewer) answered this question correctly. Similarly, a subset of girls are able to complete subtraction involving two 2-digit numbers without any borrowing, but 3-digit by 2-digit subtraction that requires even simple borrowing precipitates a significant drop in scores.

Unsurprisingly, performance on arithmetic is worse for multiplication and division than for easier arithmetic operations. Notably, girls perform almost identically on more complex subtraction (subtask 5) as on the simplest multiplication problems (subtask 7), suggesting that many girls do not understand the basic tenets of multiplication, as applied to 1-digit multiplication problems. Two-digit multiplication (subtask 8) is noticeably more difficult for girls in the sample than 1-digit multiplication (subtask 7). However, this gap obscures the fact that the decline in performance began *within* subtask 7 when girls were asked to multiply higher numbers (6x8) versus lower numbers (5x5 or 7x2). This decline is surprising because the former problem does not require application of a different rule; it simply uses larger numbers for which students may not have memorised the answer.

This surprising finding is mirrored by the pattern of performance in subtasks 9 and 10, focused on division. A subset of students does not seem to understand the rules that underlie division, but are sufficiently familiar with certain values that they can answer them correctly. Two examples illustrate this point:

- Girls are more likely to answer "36/6" correctly than to answer "42/7" correctly (44.9 percent versus 32.3 percent).
- Girls are more likely to answer "132/11" correctly than to answer "299/13" correctly (23.4 percent versus 17.6 percent)

In both cases, the rules required are identical. The first pair of problems cannot be structured in "long division" format and the whole dividend must be divided by the divisor, but students are more likely to be familiar with the "36/6" problem because students often remember "like-by-like" multiplication problems (such as 6x6 or 7x7) more easily. The second pair of problems can both be structured as long division problems where the division of the first two digits of the dividend by the divisor produces a remainder that can be applied to the last digit of the dividend. However, students perform better when asked to solve "132/11," possibly because they have been exposed to rote memorisation of "multiplication tables" for relatively small numbers, e.g., up to 12x12, but not for higher numbers.

The final pattern of interest concerns students' performance on arithmetic word problems. For both sets of word problems (subtask 6 and 11), students appear to perform reasonably well, relative to subtasks that involve equivalently difficult arithmetic. For instance, the subtask 6, item 1 involves simple addition and girls answer correctly at rates nearly identical to their performance on subtask 2, involving simple addition. A

similar pattern can be seen in the context of multiplication.[74] The natural conclusion is that girls who are able to solve arithmetic operations are generally able to conceptualise word problems into their arithmetic forms, a key skill.

At a macro level, there are fairly large differences in performance between girls who were enrolled in school at the time of endline data collection and those who were not enrolled. The table obscures these differences somewhat, because the majority of cohort girls in intervention communities were enrolled during the endline. As a result, comparing the full intervention girl sample to the sample of intervention girls who were enrolled understates the difference between enrolled and non-enrolled girls, as the sample reported in the bottom panel of the table includes only 194 non-enrolled girls, compared to 448 enrolled girls. In practice, the gaps are sizable: for instance, 13.9 percent of non-enrolled girls are proficient at complex subtraction (subtask 5), compared to 37.1 percent of enrolled girls. We do not unpack these results further, because our purpose is in understanding patterns of skill gaps, rather than the correlation between enrolment status and performance and acquisition of specific skills.[75]

Moving to Somali literacy, the table on the next page documents proficiency levels for specific Somali literacy subtasks among cohort girls in intervention communities (top panel) and among the subset enrolled in school at endline (bottom panel). Unsurprisingly, word recognition produces the highest level of proficiency, with 74.6 percent of girls able to recognize 80 percent or more of the 60 words listed. Even among girls who were not enrolled in school at the time of the endline, 54.6 percent met this standard of proficiency.

Performance on reading comprehension is less impressive, but the vast majority of girls fall into the proficient or established learner categories for the first two reading comprehension subtasks (2 and 3). Notably, girls performed better, on average, on subtask 3 than subtask 2, although the former was designed to be a more difficult comprehension exercise.

[74] For example. 41.6 percent of girls correctly solved a word problem requiring calculation of 2x15. On non-word multiplication problems, girls achieved comparatively higher scores on the simplest 1-digit-by-1-digit problems and comparatively lower scores on more difficult 2-digit-by-2-digit problems, suggesting that an equivalent 2-digit-by-1-digit problem might result in scores similar to those seen on the word problem referenced.

[75] Effective analysis of the relationship between school exposure and skill acquisition would require a more complex design and coding scheme, as many non-enrolled girls have been exposed to significant schooling, and many girls enrolled at endline were out-of-school previously. Therefore, using enrolment status at endline exclusively provides only an approximate measure of a girls' actual exposure to schooling.

TABLE 25: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION GIRLS, SOMALI LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Reading Words	Reading Comp (easy)	Reading Comp (medium)	Reading Fluency	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)	Sentence completion
Intervention Girls									
Non-Learner (0%)	8.7%	16.8%	18.4%	13.7%	27.6%	31.3%	37.4%	44.9%	59.3%
Emergent Learner (1-40%)	4.5%	7.5%	3.3%	16%	16.8%	13.4%	3.6%	3.4%	17.4%
Established Learner (41-80%)	12.1%	36.6%	28.5%	25.1%	43%	23.5%	5.9%	6.9%	7.9%
Proficient Learner (81-100%)	74.6%	39.1%	49.8%	45.2%	12.6%	31.8%	53.1%	44.9%	15.3%
Intervention Girls Enrolled in School									
Non-Learner (0%)	4.2%	9.2%	11.2%	7.4%	20.3%	21.7%	26.8%	34.4%	51.8%
Emergent Learner (1-40%)	3.3%	7.6%	3.1%	12.7%	16.7%	13.8%	3.6%	4%	19%
Established Learner (41-80%)	9.2%	35.7%	28.8%	24.3%	48%	26.6%	6%	7.1%	10.5%
Proficient Learner (81-100%)	83.3%	47.5%	56.9%	55.6%	15%	37.9%	63.6%	54.5%	18.8%

Reading comprehension declined significantly in subtask 5, though this decline was driven primarily by the last two test items within subtask 5. The test items for subtask 5 consist of the following comprehension questions:

- Who was washing the dishes?
- Who spoke with Abdi about Hamda's problem?
- How was Hamda's problem solved?
- Why did Nimco's teacher tell Abdi he should be proud of Ibrahim?

On the first two items, girls answered correctly 58.7 and 71.7 percent of the time, respectively. On the next two items, these scores declined to 43.1 and 22.1 percent. One possible explanation is that the latter two questions were based on information provided near the end of the story and girls may not have been given adequate time to finish reading the entire passage. Although subtask 5 is not a time-limited task, poor administration of the assessment could result in giving girls insufficient time to read the passage, especially because it uses the same passage as the timed reading exercise in subtask 4.[76] To explicitly combat this potential problem, enumerators were repeatedly advised during training to provide girls ample time to finish the passage and to verify that they were finished reading before proceeding to the comprehension questions.[77] A more likely explanation is that the two test items that produced lower scores are genuinely more difficult: the first two test items are based on information directly and unambiguously provided in the story. For instance, the story states "One morning, Ibrahim and Hamda were helping their mother to wash the dishes"; the answer to the first question is provided explicitly. In contrast, the answer to the last question must be inferred from multiple pieces of information in the story, information which is scattered across multiple sentences and paragraphs. This suggests that girls' reading comprehension is fairly good, even in the context of longer or more complex stories, but that comprehension tends to be limited to relatively direct inferences.

Reading fluency is captured primarily in subtask 4, in which girls were presented with a passage with 413 words. Median reading speed, among girls who read at least one word, was 83 words per minute, which is equivalent to a Grade 3 reading level in the US system, according to Hasbrouck and Tindal's reading fluency metrics.[78] To achieve proficiency on this subtask, girls needed to read 80 words or more in one minute.

Multiple mid-range subtasks in Somali literacy – especially subtasks 4-6 – show a continuum of achievement, with girls falling into each of the four proficiency levels in relatively large numbers. In short, there is not a sharp and unambiguous distinction between girls who have the skill tested by the subtask and those who do

[76] To be clear on this point, rushing girls through subtask 5 would constitute mis-administration of the assessment. However, we have encountered confusion among enumerators in the past when administering this subtask. The previous subtask asks girls to read a passage with a 60-second time limit. After the 60-second limit, girls are supposed to be given additional time – without a time limit – to read the remainder of the passage – or re-read it – for reading comprehension to answer the questions in subtask 5. However, enumerators have – in the past – failed to provide girls an opportunity to read the passage again, misinterpreting the instructions. While we do not believe this is a systematic problem, because this specific issue was covered repeatedly during training, it is possible that a few enumerators may have incorrectly administered the assessment, dragging down scores for comprehension questions that rely on a girl having finished the entire passage.

[77] Subtask 5 asks girls questions about a passage that was read during subtask 4, which is a timed reading fluency exercise in which girls are given 60 seconds to read as much of the passage as possible. After the timed exercise, girls are asked to read the remainder of the passage, even starting from the beginning if they would like, to prepare for the reading comprehension questions that follow.

[78] Hasbrouck, Jan, and Gerald A. Tindal. 2006. "Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers." *The Reading Teacher* 59 (7): 636-644.

not; rather, some girls have mastered the skill, others have the skill but cannot apply it widely, and still others have not acquired the skill at all. This pattern breaks down in subtasks 7-9, where the table shows polarisation – girls tend to either achieve proficiency or fail the subtask entirely. The most likely explanation is that each item in subtask 7 is based, effectively, on application of the same process, converting a sentence to its negative form. Acquisition of this skill is closer to binary than is reading comprehension.[79] As a result, girls who have acquired the skill are generally proficient at converting sentences, while those who have not are unable to correctly complete any of the associated test items.

Relative to Somali literacy, girls' performance in English literacy is poorer and – as shown in the table below – declines rapidly from the first subtask. In other words, aggregate English literacy is weaker and proficiency beyond basic skills – letter recognition and word recognition – is limited to a very narrow subset of girls. Letter identification is the only foundational skill on which a majority of girls achieved scores above 40 percent and only 41.7 percent of girls achieved proficiency in this skill.

Girls demonstrate a precipitous decline in performance when moving from letter identification to word identification, a fact which is consistent with the significantly higher difficulty of reading complete words versus individual letters. The gap in difficulty is especially pronounced because official Somali and English both use the Latin alphabet, meaning that English letter identification should be comparatively straightforward. This decline continues, naturally, when shifting toward reading complete sentences and paragraphs for comprehension, as in subtasks 3-5. Especially noteworthy is performance on subtask 5, where the majority of girls were unable to correctly answer a single question, out of four possible. The simplest item in subtask 5 asked respondents "How did the monkey wake the lion?" The correct answer is provided directly in the second sentence of the story, which reads "The lion woke up when the monkey stepped on his tail." Just 17.8 percent of girls correctly answered this question, which reinforces the low level of even basic English literacy among the sample.

[79] The skill of converting a sentence to its negative form is not entirely binary, as application of the rules can vary in complexity, but only within a comparatively narrow spectrum.

TABLE 26: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION GIRLS, ENGLISH LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Letter Identification	Word Recognition	Reading Comp (easy)	Reading Fluency (medium)	Reading Comp (medium)	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)
Intervention Schools									
Non-Learner (0%)	28.7%	39.4%	59.5%	53.7%	75.7%	78%	76.9%	86%	91.7%
Emergent Learner (1-40%)	13.7%	14.8%	8.1%	11.7%	7.6%	6.4%	9%	0%	0%
Established Learner (41-80%)	15.9%	21.8%	16.8%	15.4%	14.3%	12.9%	11.1%	5.8%	2.3%
Proficient Learner (81-100%)	41.7%	24%	15.6%	19.2%	2.3%	2.6%	3%	8.3%	5.9%
Intervention Girls Enrolled in School									
Non-Learner (0%)	17.4%	25%	47.5%	40.6%	67.9%	70.5%	69.6%	80.8%	89.1%
Emergent Learner (1-40%)	13.2%	16.1%	9.6%	14.3%	9.6%	8.7%	11.2%	0%	0%
Established Learner (41-80%)	16.5%	27%	21.7%	19.4%	19.4%	17.2%	15.2%	8%	3.3%
Proficient Learner (81-100%)	52.9%	31.9%	21.2%	25.7%	3.1%	3.6%	4%	11.2%	7.6%

Working Memory

Good working memory is a functional and practical tool. Working memory helps with organization, planning, and decision making. It also can help with learning and socialization. While some amount of working memory is an innate trait, there is speculation that we can train our working memory to function more effectively.[80] Other factors like diet, sleep, and age more conclusively influence working memory.[81] Working memory aids literacy, both in speaking and in reading. Working memory can also help with vocabulary, conjugation and tense.

Note from the project:

The working memory test is built off of Nelson Cowan's. His working memory assessments typically involved random word or digit recall. Tests similar to ours are commonly used in clinical psychology nowadays, using computer-based platforms. A useful snapshot of common phonological span working memory tasks can be found in this review by Cabbage et al. (2017):

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608376/[82]

Somali girls in our panel do not have high levels of numeracy or literacy, and because we do not want to test for either of those skills in the working memory assessment, we chose to use images. These images changed from ML2 to EL. The images are also relevant in that the interviewee will be able to tell what they are. While images are often easier to recall than words, the structure of working memory has precedent in old clinical memory tests, which used image cards instead of words or digits.[83]

Classic working memory tests generally include three components – phonological span, visuospatial tasks, and central executive tasks, thus testing the three components of Baddeley's working memory model (phonological loop, visuospatial sketchpad and central executive). The task included in this study assesses phonological span only, as it intends to assess the extent to which working memory affects the acquisition of reading skills.

During training enumerators were trained on grading working memory. Because the test uses images instead of words there is not necessarily one answer for each image – a hill can also be a mountain, a cow can also be a heifer. Enumerators were told to use their best judgment.

[80] Melby-Lervåg, M., Redick, T. S. and Hulme, C. 2016 'Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review', Perspectives on Psychological Science, 11(4), pp. 512–534

[81] "A large amount of epidemiological data indicates that a deficient provision of nutrients during in utero development and/or childhood, results in long-lasting impaired attention and learning, decreased IQ scores and reduced visuospatial working memory". – Pérez-García, G., Guzmán-Quevedo, O., Da Silva Aragão, R., & Bolaños-Jiménez, F. 2016. Early malnutrition results in long-lasting impairments in pattern-separation for overlapping novel object and novel location memories and reduced hippocampal neurogenesis. *Scientific reports*, *6*, 21275.

[82] Cabbage, K., Brinkley, S., Gray, S., Alt, M., Cowan, N., Green, S., Kuo, T., & Hogan, T. P. 2017. Assessing Working Memory in Children: The Comprehensive Assessment Battery for Children - Working Memory (CABC-WM). Journal of visualized experiments: JoVE, (124), 55121.

[83] Paivio, A., Rogers, T.B. & Smythe, P.C. 1968. Why are pictures easier to recall than words?. Psychon Sci 11, 137–138.

During the ML2 our enumerators administered a working memory assessment to both girls and boys. This assessment was simple. Enumerators gave a piece of paper with 19 images of different items. All items are topical to Somalia and to the child. The child was asked to memorize each image. The enumerator than took the paper, preventing the child from referencing it, and asked the child to list the images they had seen. This score, the number of images a child remembered, was then normalized on a 0-100 point scale. 100 points indicates that the girl remembered all 19 items, a difficult task.

TABLE 27: WORKING MEMORY SCORES BY COHORT AND ROUND

Working Memory Scores ML2 to EL of Panel Samples			
Cohort	ML2	EL	Change
ISG	55.3	52.0	-3.3
OOS Treatment	48.5	50.8	2.3
OOS Comparison	45.0	47.8	2.8
ABE	51.5	51.2	-0.3
ALP	54.4	47.0	-7.4
ISG Treatment	56.3	53.2	-3.1
ISG Comparison	54.4	51.5	-2.9
Cross Section of Boys Working Memory Scores			
Boys Treatment	57.5	53.47	-4.03
Boys Comparison	54.1	52.5	-1.6

In general, from ML2 to EL, girls' working memories have declined. Table 27 indicates a broad, but small reduction of working memory for all cohorts except for OOS girls. This is a counterintuitive finding. Access to education seems a likely corollary with working memory. Except for ALP girls, OOS girls start and end with the lowest working memory scores, an expected finding, but their increase in scores require regression analysis to explain.

TABLE 28: WORKING MEMORY DIFFERENCE AND DIFFERENCE REGRESSION COEFFICIENTS

Regression Details	Impact Estimate	95% Confidence Interval	P Value
No controls	-1.2	-6.4 - 4.0	0.645
Control for zone	-1.2	-6.4 - 4.0	0.645
Control for region	-1.2	-6.4 - 4.0	0.645
Control for age	-1.3	-6.5 - 3.9	0.625
Control for diet	-1.3	-6.7 – 4.0	0.611

Control for disability	-1.3	-6.5 – 4.0	0.624
Control for clean water	-1.4	-0.34 - 5.3	0.604
Control for poor roof	-1.3	-0.4 - 4.8	0.612
Control for enrolment	-0.4	-5.9 – 5.1	0.878

Each one of these models uses weights and clusters standard errors by school.

The difference-in-differences regression models did not yield significant difference-in-differences coefficients, regardless of the controls. Predominately the controls behaved as expected, disabilities are negatively associated with working memory as well as lack of clean water. These two stresses are likely to decrease cognitive ability.

Note from the project: No differences in working memory are expected from the project's intervention. Other types of interventions, most notably school feeding and micronutrient supplementation in contexts with high levels of malnutrition, may result in changes to working memory scores (as observed in the present results). The working memory supplemental analysis is intended to provide information on a potential factor affecting learning outcome results, particularly on reading comprehension, which is typically affected by lower working memory spans.[84]

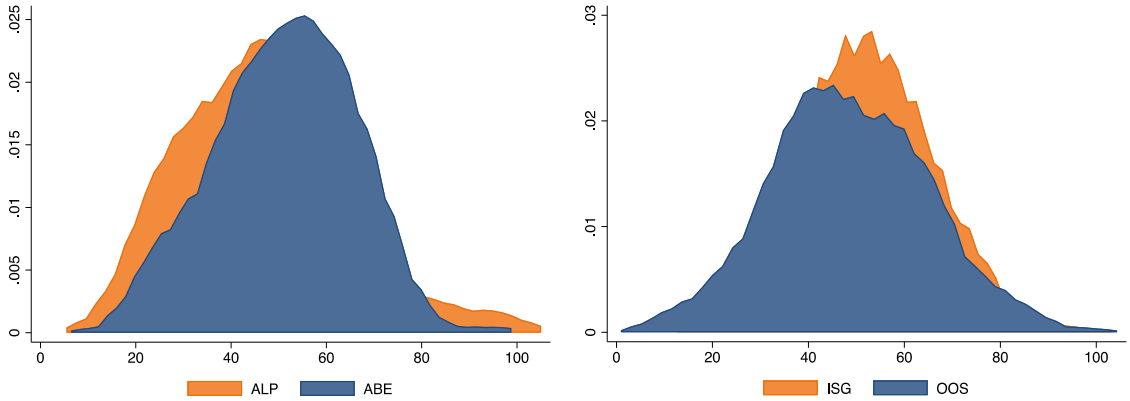
Diverse diets were associated with a 0.6 point increase in working memory score with a p value of 0.805. Diverse diets were binned to extract the extremes of diets. We created a dummy variable to indicate diverse diets, diets with more than nine food groups, and diets lacking in diversity, diets with less than nine food groups.

Disability is strongly negatively associated with working memory. As a control, disability has a coefficient of negative 5.9 points is significant at the 98.3 confidence level. Disability is defined as an aggregate of all disabilities except for mental health. This is an expected outcome – disability likely effects nutrition and sleep.

Working Memory Score Distribution

[84] Abadzi, H. (2008) *Efficient Learning for the Poor: New Insights into Literacy Acquisition for Children.* International Review of Education, 54:581-604.

FIGURE 22: WORKING MEMORY KERNEL DENSITY PLOTS AT THE ENDLINE



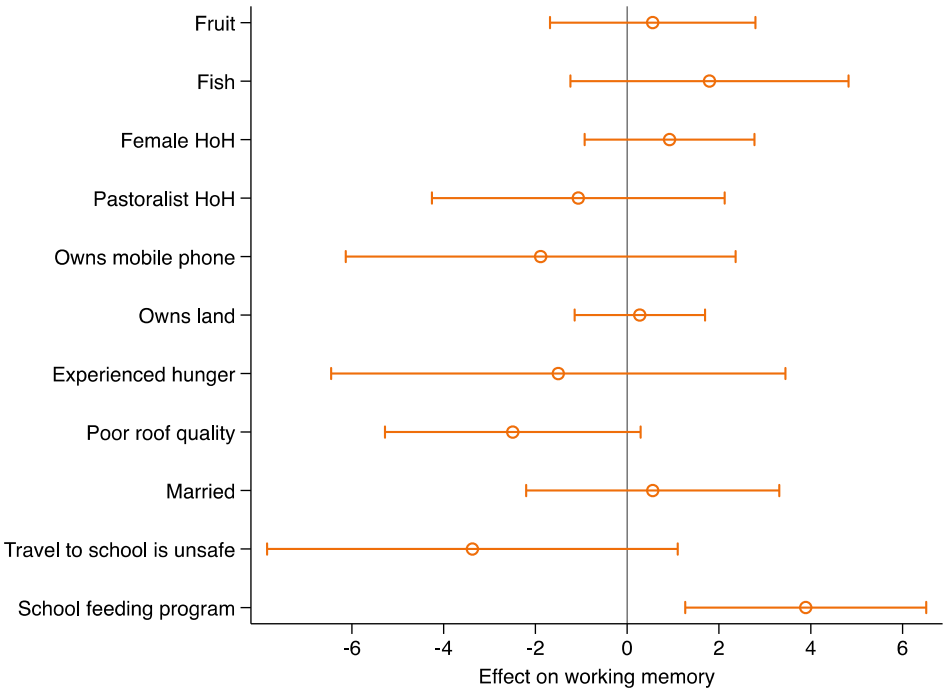
The four kernel densities in Figure 22 estimate the probability density of a working memory scores by Cohort in the EL. ISG girls have a density of higher scores compared to OOS girls, the same is true of ABE girls. The distribution is flatter for ALP and OOS girls than it is for ABE and ISG girls. ALP girls are more likely than ABE girls to score under 45 and over 85. OOS girls are more likely than ISG girls to score over 80 and under 40 on the working memory assessment.

The distribution of working memory score ML2 to EL shows comparison girls are more likely to have a positive increase in working memory from the ML2 to the EL.

Household Factors and Working Memory

Working memory is influence by a number of household factors such as poor roofs, clean water, and presence of hunger. These indirectly effect working memory through their influence on the physical and mental state of girls. A hungry girl with malnutrition is more likely is more likely to suffer working memory difficulties.

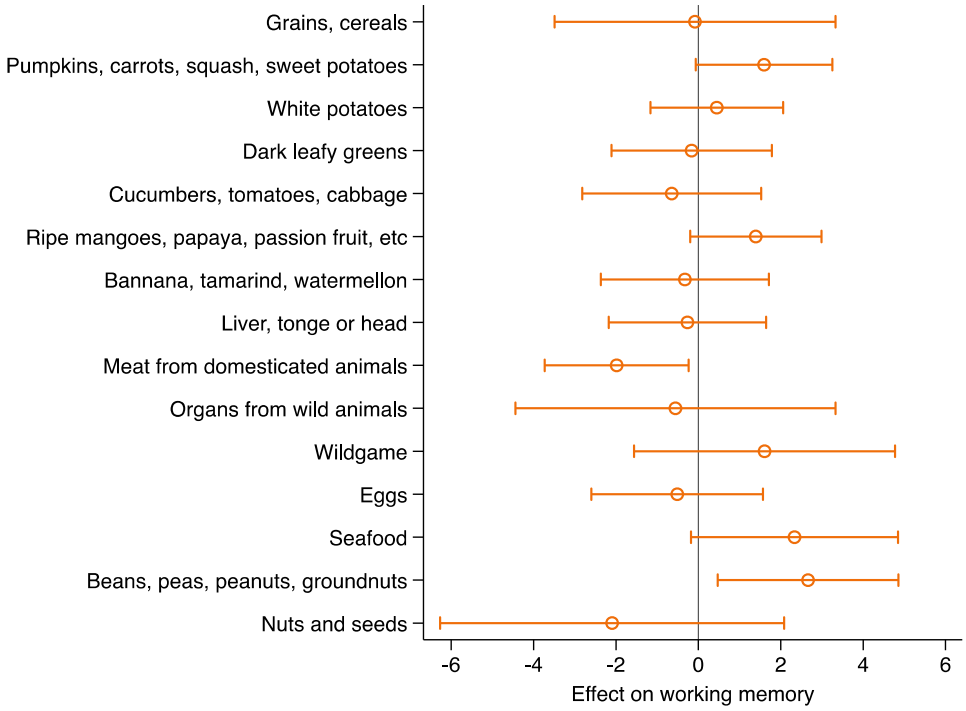
FIGURE 23: HOUSEHOLD FACTORS AND WORKING MEMORY



Food and nutrition appear to be the consistent determinants of working memory even if they are not significant. As seen in Table 28: Working memory difference and difference regression coefficients, working memory is affected by diet. Diet is partially controlled by household stability and wealth. Poor roof quality as well as having a pastoralist head of household is associated with poverty, therefore they understandably are negatively associated with working memory.

The only variable with a significant effect is the presence of a school feeding program. A school feeding program, likely reduces hunger, increases nutrition, and aids girls focus and learning. A child that is hungry during the working memory assessment is more likely to try to rush through it or not even try.

FIGURE 24: WORKING MEMORY AND DIET



We collected diet diversity by asking about 19 different food groups. In Figure 24 we regressed 15 of those food groups against working memory. A diverse diet appears important for a good working memory but certain food groups have greater effect on working memory than others. Beans and meat have significant coefficients at the 95% confidence level. Several other food groups had nearly significant correlations such as ripe fruit, squash and root vegetables, and seafood.

Meat had a negative association with working memory while beans have a positive one. Somali food is heavy with goat and camel, and some families might substitute a more diverse diet for higher quantities of meat. While protein is good for child development heavy meat consumptions might also be correlate with more rural or pastoralists families where nutrition is poorer.

Working Memory and Somali and English Literacy

A scientific consensus has yet to be determined if working memory can be taught or learned. Working memory is valuable tool for learning. Because working memory is a valuable tool for learning it expected that higher scores in Somali literacy, English literacy, and Numeracy are associated with higher working memory scores.[85] As seen in Table 29 higher scores in all three assessments were associated with higher working memory scores. Higher grades can also associated be with higher grades, but even after controlling for grade level, higher numeracy and literacy scores still had associations with higher working memory scores.

[85] We are using the word, association, to describe the relationship between working memory and assessment scores because using the word affect implies a directional causal relationship, one that we are not sure there is between the assessment scores.

TABLE 29: WORKING MEMORY, LITERACY, AND NUMERACY SCORES

Binned Score	Score Association	Association Controlling for Grade
Somali Literacy Score		
25 - 50	5.4	2.7
50 - 75	8.4	6.7
75 - 100	10.9	9.2
English Literacy Score		
25 - 50	6.3	4.0
50 - 75	7.0	4.7
75 - 100	7.9	5.9
Numeracy Score		
25 - 50	5.8	2.7
50 - 75	9.2	5.7
75 - 100	12.9	8.6

Numeracy scores have the largest association with working memory scores. Girls who have a numeracy score above 75 have on average have a 12.9 point increase in working memory score. The equivalent association for Somali literacy is a 10.9 point increase. When we control for grade, this increase declines dramatically for numeracy score associations but less for for Somali literacy. Numeracy could have a closer connection to analytical thought processes while Somali literacy may require more memorization.[86] The working memory assessment is not analytically difficult. The images are easy to decipher – it is only a test of the current level of working memory, not effort or thought process.

It is expected that controlling for grade reduced assessment associations. The effect of grade on girl's working memory score generally increases the higher the grade. The last year of secondary was associated with the highest working memory scores. It was not significant but girls in their last year of secondary were associated with an additional 4.2 working memory points. Controlling for the effect of grade to further isolate the relationship between literacy, numeracy and working memory helps eliminate the effects grade level could have on working memory.

The association with English literacy decreased when we controlled for grade. English literacy, as second language is difficult, only 2% of girls in the three rounds scored between 75 and 100. It is expected that high English literacy scores would have the strongest associations with working memory.

Regressing each normalized literacy and numeracy scores against working memory produces similarly strong associations between assessment scores and working memory. For each one-point increase in Somali literacy score working memory increased 0.14 points. For each additional point in English literacy working memory

[86] Numeracy is potentially more dependent on grade level, while Somali literacy is likely more dependent on individual.

increased 0.15 points. Numeracy has the largest association, a one-point increase in numeracy is associated with a 0.17 point working memory increase.

Conclusions

Isolating what most directly affects working memory is difficult. There is no hard consensus on what increases working memory. Working memory has a close association with learning and SOMGEP-T assessments scores. This association is not because working memory is an assessment, like Somali literacy, English literacy and numeracy. The working memory assessments is not hard, the girl either remembers or she does not, and the enumerators record the answer and move on. The working memory assessment measures a cognitive ability, short-term memory, but some research indicate working memory assessments failing to correlate with other cognitive abilities like long term memory.[87] Memory decay could be a better determinant of educational success than short term memory. Memory decay is harder to measure, therefore working memory has to be the functional proxy.

A causal relationship between working memory and educational attainment is indiscernible but the strong association between the two indicate further research is necessary. If working memory increases educational attainment, the most likely way to increase working memory appears to be through diet and nutrition. Inland regions had lower working memory assessments scores than coastal regions, a difference of 4.1 points potentially attributed to decrease in access to seafood. While increasing working memory was not an outcome for SOMGEP-T, establishing the reasons for cognitive ability differences will help target future interventions.

Transition

The second top-level outcome for all GEC-T projects, including SOMGEP-T, is transition, which generally consists of enrolment in formal or non-formal education, advancement through grade levels, and shifting into productive and non-exploitative employment. Transition is a complex outcome largely due to the many different individual-level outcomes that mu2st be captured to accurately assess it. The complexities arise from two characteristics of the outcome: first, because the programme supports non-formal education – e.g., ALP and ABE programmes – and is open-minded regarding the practical educational and employment pathways available to the marginalised girls it seeks to support, there are a wide set of outcomes the programme considers a "success" in terms of transition. These include enrolment in formal or non-formal education and age-appropriate, gainful employment.

Second, the programme rightly views "success" as dependent on one's starting point. In other words, enrolment into vocational training is a success if the girl has finished primary school or was out-of-school to start, but would not necessarily be considered a success if the girl dropped out of primary school to begin vocational training. This fact multiplies the number of pathways girls can and do follow. In practice, most of the many possible pathways – over 30 in the current evaluation round – are seldom used, and the vast majority of girls are defined as successful or unsuccessful, in transition terms, based on whether they re-enrolled in

[87] Cowan N. 2008. What are the differences between long-term, short-term, and working memory?. Progress in brain research, 169, 323–338.

school, dropped out, or remained in school and either did or did not advance a grade level from the preceding year.

In this round of data collection and analysis, we expanded our methods for capturing information on transition, responding to difficulties in previous rounds and to the shrinking size of our panel samples tracked over time. Much of our analysis relies on the same data utilised in previous rounds, in which we collect information on current status (enrolment, employment, etc.) and a girls' status one year prior. We then compare girls' status at present to her status one year prior to determine her transition status. An identical measure of transition was captured in previous rounds, allowing us to use difference-in-differences models among a panel of girls successfully re-contacted across rounds to study changes in transition outcomes between intervention and comparison communities.

The drawback of this approach is that – as with the learning samples studied in the previous section – girls fall in and out of the sample over time, reducing the strength of our analysis. For instance, our panel sample of girls tracked successfully from baseline to endline consists of 955 girls, reduced from the original baseline sample of 1741. Moreover, this panel is only suitable for analysis of overall changes from baseline to endline, as it includes girls who did not appear in the ML1 and ML2 rounds.[88] While changes from baseline to endline are our central interest in this evaluation, lacking information from the intermediate evaluation rounds obscures important trends and reduces our ability to explain how the programme impacted intervention communities, why the programme was or was not successful, and how programme impact was shaped by the COVID-19 pandemic and other exogenous shocks.

To improve the information available for studying transition, we made two important adjustments to the endline methodology. First, as noted in Section 2, we made extra efforts to capture transition information even for cohort girls who were unavailable to complete the full household survey and learning assessments. If a girl moved from her community to a new village or a regional town, we attempted to visit her and complete data collection as usual. In total, we contacted and interviewed dozens of girls who had moved from their villages to regional urban centres. In some cases, girls had moved to an area for which a visit by a field team was logistically or financially prohibitive, such as Mogadishu or locations abroad. In these cases, we called the girl or a close family member and collected information on her enrolment and transition status. In total, phone-based collection of transition information expanded the endline sample by 237 girls (148 in-school and out-of-school cohort girls).

Second – and linked to the first adaptation – we collected retrospective transition data for all girls contacted at the endline. For each respondent interviewed at endline, we collected enrolment status and grade level for every year from 2017 to present (2021). The advantage of this data is that it is captured for a slightly larger overall sample – 1310 cohort girls – than either the baseline-to-endline or ML2-to-endline panel samples, as shown in the table below. Moreover, while the data includes less detail about transition outcomes – if girls are not enrolled in school, ALP, or ABE, they are considered unsuccessful transition cases, because we did not collect information on employment – the data is much richer in terms of temporal coverage.[89] Using

[88] Panel samples with more complete temporal coverage include panels of girls who were successfully contacted in all four rounds (n = 399) and in the baseline, ML2, and endline rounds (n = 740).

[89] The lack of information regarding employment and other alternative pathways is less problematic than it may appear, because relatively few girls make use of these pathways, as we demonstrate in Section 5.2.

retrospective self-reports of this kind allows us to analyse transition status over five years for a set of 1310 girls.

TABLE 30: SAMPLES OF TRANSITION DATA AND THEIR BENEFITS

Sample	Sample Size	Advantages	Disadvantages
BL-to-EL Panel	955	• Captures life of programme • Relatively large sample size	• Obscures trends between 2017 and 2021
ML2-to-EL Panel	953	• Relatively large sample size	• Short time period captured
BL-ML2-EL Panel	740	• Captures life of programme • Reveals pre-COVID trends	• Smaller sample size
Retrospective EL Data	1,310	• Captures life of programme • Reveals pre-COVID trends and trends between BL and EL	• Relies on accurate retrospective self-reports from respondents

In the next section, we describe the transition pathways girls can follow and define the various pathways that are considered successful transition. In the section that follows, we analyse transition rates among the primary cohort of in-school girls and out-of-school girls – the cohort originally recruited at baseline, not including ALP and ABE girls. Our first analysis uses the panel datasets that have been the staple of our analysis in each evaluation round and in the learning results presented in the previous section. We then turn to the retrospective data collected at endline to check the robustness of our results and gain greater insight into trends in transition rates over time.

Transition Pathways

SOMGEP-T takes a broad view of what constitutes successful transition, tailoring the goal outcomes to match the realities in which rural Somali girls live. The table below describes the pathways that the program considers successful or unsuccessful transition outcomes, depending on the girl's "starting point." To be clear, transition as an outcome is measured over time – at baseline, successful transition was measured by comparing girls' enrolment and employment status at baseline to their status in the year prior to baseline. In this round of data collection, we define transition based on a comparison of enrolment and employment status at ML2 backward to baseline.

The definition of transition as change over time can complicate our conceptual discussion; our outcome is defined by change over time, and we are analysing changes in that outcome over time. This approach is necessitated, in part, by the fact that we must capture baseline transition outcomes, which are dependent on the pre-baseline status of girls, and by the similar fact that we must capture endline transition outcomes, which depend on the pre-endline (2020) status of girls, during which no evaluation was conducted. The critical point to remember is that the majority of our analysis is focused on a binary successful/unsuccessful classification of a wide array of transition outcomes, and that – while this classification encapsulates change over time in its definition – the analysis of how that outcome has changed in a difference-in-differences framework is identical to that of learning and other outcomes in this report.

The table below illustrates many of the possible pathways applicable to SOMGEP-T girls. As noted above, the starting points defined in the table are one year prior to data collection; for baseline transition rates, the "starting point" is the year prior to the baseline, while for endline transition rates, the same logic dictates that the starting point be defined as a girl's status in 2020.

TABLE 31: TRANSITION PATHWAYS

	Starting Point	Successful Transition	Unsuccessful Transition
Lower primary school	Enrolled in Grade 3, 4	• In-school progression • Drops out but is enrolled into alternative learning program	• Drops out of school • Remains in same grade
Upper primary	Enrolled in Grade 5, 6, 7, 8	• In-school progression • Moves into secondary school • Moves into ALP	• Drops out of school • Moves into work, but is below legal age (under 15 years)
Secondary school	Enrolled in Grade F1, F2, F3, F4	• In-school progression • Enrols into technical & vocational education & training (TVET) • Gainful employment • Moves into ALP	• Drops out of school • Moves into employment, but is unpaid or otherwise exploited

Out of school	Dropped out	• Re-enrol in appropriate grade level in basic education • Enrol in alternative learning program • Engages in wage/self-employment	• Remains out of school

While much of the table above is self-explanatory, we do note how we define different employment outcomes. Girls are classified as entering under-age employment if they are employed outside the home, in any capacity, but are under the age of 15 years. Gainful employment refers to employment among girls who are 15 years or older and who have a permanent, formal, paid position. Girls who are in "casual" or "informal" employment, even if they are of the proper age, are classified as being non-gainfully employed. Only gainful employment is considered a positive transition outcome for SOMGEP-T.

Overall Transition Outcomes

Although our primary interest in this section is in assessing the programme's impact on transition rates since the baseline, we begin our analysis by considering the transition pathways taken by girls since the ML2 round in 2019. The motivation for this approach is to illustrate – in a slightly less complex setting, owing to the shorter time period covered – the distribution of pathways that girls follow, in practice. While we coded over 30 distinct transition pathways, the vast majority of girls fall into just a few. For girls who were previously out-of-school, these include: re-enrolment; transition into informal education, vocational training, or employment; and remaining out-of-school.[90] For girls who were previously in-school, the most common pathways include: continued enrolment and advancement in grade level; continued enrolment and being "held back" a grade; dropping out. Most girls enrolled in school do not later transition into informal education or employment, at least within the period studied here.

The table below reports transition outcomes among a sample of girls (n = 765) who were enrolled in school at ML2, in 2019. The top panel reports the girls' transition status in 2020, essentially one year after the ML2 data collection was completed. The girls in the table all began from the starting point, having been enrolled in formal school at ML2. By reading across the columns in the top panel, we can compare the paths taken by enrolled girls from ML2 into the next year, across intervention and comparison groups.

[90] Transition into ABE and ALP is also included as a pathway. However, because this analysis is focused on the original cohort girls – rather than the girls who were recruited from ABE and ALP centres in later rounds – the share of our sample who transitioned into these specific programmes is relatively low. With that said, many girls who are coded as moving into informal education are actually entering ABE and ALP programmes, but our survey questions and response options were insufficiently nuanced to distinguish between different non-formal programmes.

Overall, girls in intervention communities tended to have higher transition rates than those in comparison communities. This difference is most stark when comparing the share of girls who remained enrolled and "advanced a grade"; among intervention girls, 82.2 percent remained in school and advanced a grade, compared to 76.8 percent among comparison girls. The overall gap in successful transition is slightly smaller, because a very small number of comparison girls moved into vocational training or ALP education; moreover, although the gap is not statistically significant ($p = 0.17$), the magnitude of the difference in transition rates is notable. An additional finding of interest is that the share of girls held back a grade is lower in intervention communities, a pattern which emerges in later results as well.

TABLE 32: TRANSITION OUTCOMES FOR GIRLS ENROLLED AT ML2 (2019), OVER TWO SUBSEQUENT YEARS

	Intervention	Comparison
Outcomes – 2020 Transition Status		
Dropped out	7.9%	11.7%
Dropped out into non-gainful employment	0.2%	0.0%
Dropped out into gainful employment	0.0%	0.0%
Dropped out into vocational training	0.0%	0.3%
Dropped out into informal education	0.0%	0.0%
Remained enrolled, held back a grade	9.7%	10.8%
Remained enrolled, advanced a grade	82.2%	76.8%
Dropped out, enrolled in ALP	0.0%	0.3%
Dropped out, enrolled in ABE	0.0%	0.0%
Successful transition rate	82.2%	77.4%
Remain-enrolled rate	91.9%	87.9%
Outcomes – 2021 (Endline) Transition Status		
Dropped out	11.5%	15.4%
Dropped out into non-gainful employment	0.7%	0.9%
Dropped out into gainful employment	0.5%	0.0%
Dropped out into vocational training	0.0%	0.0%
Dropped out into informal education	0.7%	1.2%
Held back a grade	4.2%	6.0%
Advanced a grade	82.2%	76.2%
Dropped out, enrolled in ALP	0.2%	0.3%

Dropped out, enrolled in ABE	0.0%	0.0%
Successful transition rate	83.6%	77.7%
Remain-enrolled rate	87.3%	83.7%

The bottom panel of the table above repeats the analysis of transition outcomes among girls enrolled at ML2, but looking further ahead, to the girls' transition status at the endline, in 2021. The sample of girls is identical across the two panels, but the time elapsed since ML2 has increased in this analysis. Unsurprisingly, dropout rates are higher over this longer time horizon, as girls have additional time to drop out and girls have aged an additional year. The latter point reflects the fact that dropout rates increase with age and that dropping out becomes more common in higher grade levels. As before, successful transition is more common among intervention girls, with this difference being driven almost entirely by differences in dropout rates and the rate at which girls who remain enrolled manage to progress across grades.

The sample discussed above includes only the set of girls who were enrolled at the time of ML2. We performed a similar assessment of girls who were out-of-school at ML2, to understand whether intervention girls who were out-of-school were more likely to re-enrol or transition into the alternative education options available to them through SOMGEP-T. The table below reports transition outcome in 2020 (top panel) and 2021 (bottom panel) among this set of girls. The available sample is significantly smaller – just 137 girls, in total, divided between intervention and comparison groups – but is, nonetheless, informative.

In both periods considered, intervention girls are more likely to re-enrol in school. Re-enrolment rates are low – just 28.6 percent among the intervention group, one year – but this is consistent with relatively low re-enrolment rates among out-of-school girls in general, as documented in previous rounds. Also, it is almost certainly a function of the impact of COVID, as girls' enrolment in 2020 would have been directly impacted by the months-long closure of schools.

TABLE 33: TRANSITION OUTCOMES FOR GIRLS OUT-OF-SCHOOL AT ML2 (2019), OVER TWO SUBSEQUENT YEARS

	Intervention	Comparison
Outcomes – 2020 Transition Status		
Remained out of school	69.4%	72.7%
Transitioned into non-gainful employment	0.0%	0.0%
Transitioned into gainful employment	2.0%	1.1%
Transitioned into vocational training	0.0%	0.0%
Transitioned into informal education	0.0%	0.0%
Re-enrolled in formal school	28.6%	23.9%
Enrolled in ALP	0.0%	1.1%
Enrolled in ABE	0.0%	1.1%

Successful transition rate	30.6%	27.2%
Outcomes – 2021 (Endline) Transition Status		
Remained out of school	61.2%	62.5%
Transitioned into non-gainful employment	8.2%	5.7%
Transitioned into gainful employment	2.0%	2.3%
Transitioned into vocational training	0.0%	3.4%
Transitioned into informal education	0.0%	0.0%
Re-enrolled but held back a grade in 2021	10.2%	13.6%
Re-enrolled and advanced a grade in 2021	18.4%	11.4%
Enrolled in ALP	0.0%	0.0%
Enrolled in ABE	0.0%	1.1%
Successful transition rate	20.4%	18.2%
Successful transition rate, including held-back students	30.6%	31.8%

Enrolment rates remained stable from 2020 (top panel) to 2021 (bottom panel). The bottom panel separates girls who remained enrolled from 2020 to 2021 into two groups – those who advanced a grade (18.4 percent of intervention girls) and those who remained in school in 2021 but did not successfully advance to the next grade level (10.2 percent among intervention girls). While the overall share of girls who moved from non-enrolment in 2019 to being enrolled in 2021 remains higher among intervention girls, the table also reveals that intervention girls are less likely to be held back a grade than their comparison group peers.

On one hand, this is a positive outcome, as enrolment without the learning typically implied by successful grade progression misses the point of encouraging enrolment in the first place. In other words, the goal of enrolment is not simply to have children in school, but to help them learn. Insofar as grade progression is associated with positive learning outcomes and grade repeating is associated with stagnant learning outcomes, we are encouraged by the lower rate of grade repeating among intervention girls. On the other hand, grade advancement and repetition are not necessarily reliable measures of how much a student has learned, as standards can be relaxed to promote grade advancement. It is not possible to determine whether relaxed standards or higher achievement account for higher rates of promotion among intervention girls. Although the programme has not had a discernible impact on learning outcomes in our evaluation, learning outcomes are systematically higher in intervention schools, and have been since the baseline. This would imply that grade promotion rates should be higher among intervention schools, even if equivalent standards are applied.

Leaving grade promotion and grade repeating rates aside, the bottom panel of the table above demonstrates, again, that programme schools are more effective at re-enrolling out-of-school girls, at least over the period

between ML2 and the endline. At the time of the endline, 28.6 percent of intervention girls who had been out-of-school at ML2 were enrolled in school; among comparison girls, this rate was 25.0 percent.

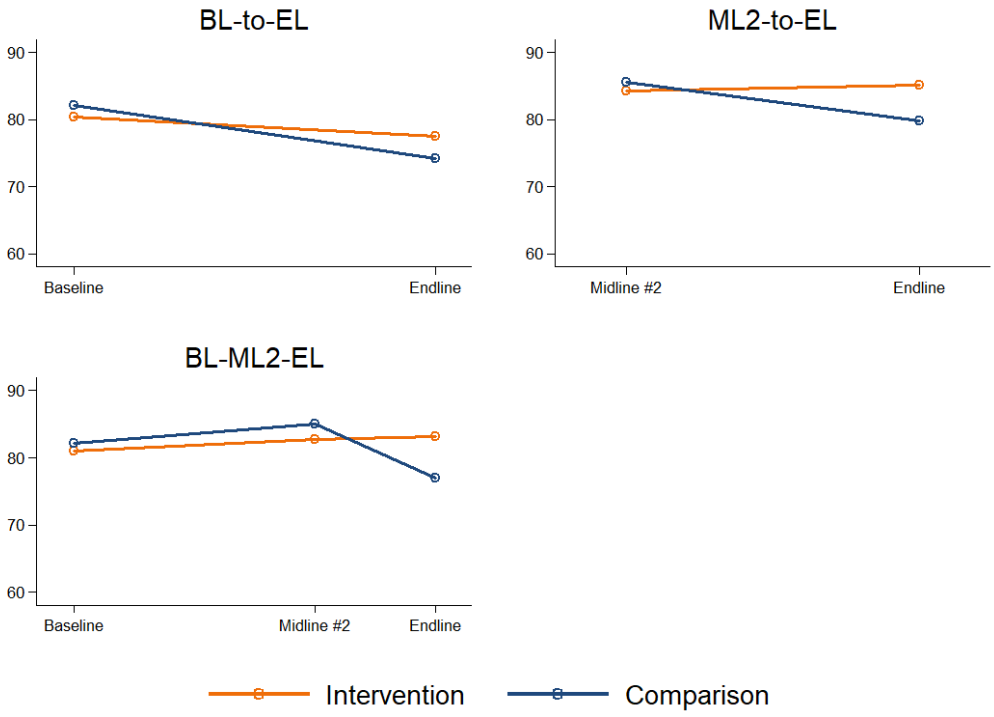
The last finding of note from the table above concerns transition of girls, by the time of the endline, into employment and non-formal education. Among intervention girls, a higher share (8.2 percent) had moved into non-gainful employment, which we define as employment or labour for which they are not paid. This typically includes household or, less commonly, agricultural work outside their home. This rate is worrying, especially considering the availability of other options, including vocational training and ALP education. However, it is important to note the small sample size available for analysis and the relatively small number of girls (4 out of 49 among the intervention group) who followed this path – not to dismiss the finding altogether, but to highlight the fact that small levels of sampling variation can drive large differences in relatively rare outcomes.

Thus far, analysis of transition paints a fairly positive image of the programme, though these findings are limited to the period from ML2 to the endline. We now turn to a more robust analysis of transition rates, using the panel samples described in the introduction to this section. We define transition as a binary outcome, with re-enrolment, grade advancement, transition into gainful employment, and transition into non-formal education all defined as a success in terms of transition. The figure below reports transition rates in intervention and comparison communities for the in-school girl cohort, across three alternative samples.[91] An important consideration regarding the sample is that inclusion is defined by *cohort*, i.e. the group a girl belonged to when originally recruited. The figure below concerns the in-school girl cohort. However, when viewing the graph in the top-right panel – transition from ML2 to endline – this does not mean the girls were all enrolled at the outset of the analysis period, in ML2. Many were not enrolled at that time but *were* enrolled when they were initially recruited.[92]

[91] For overall sample sizes for each panel, see the introduction to this section. However, because this analysis focuses exclusively on the in-school girl cohort, the sample sizes are 604 (BL-to-EL), 711 (ML2-to-EL), and 498 (BL-ML2-EL), respectively.

[92] "Initial recruitment" occurred across the first three evaluation rounds. Most cohort girls – in-school and out-of-school girls alike – were recruited at baseline. However, when cohort girls could not be located or otherwise fell out of the sample, their replacements were selected from within schools (for replacements of in-school girls) during both ML1 and ML2.

FIGURE 25: TRANSITION RATES AMONG THE IN-SCHOOL COHORT



The results in the figure present a consistent image, despite the different samples employed and the different time periods over which the analysis occurs. In the aggregate, from baseline to endline, the programme is associated with a 5.1 point improvement in transition rates among in-school girls. The intervention group started from a lower starting point, with transition rates 1.7 points lower than those of the comparison group at baseline. By the time of the endline, this gap had reversed, and transition among the intervention group stood at 77.6 percent, compared to 74.2 percent among the comparison group.[93]

The graphs in the top-right and bottom-left panels, above, illustrate the temporal variation in how the programme impacts transition rates. The panel tracking girls across three evaluation rounds strongly suggests that the programme had no impact from baseline to ML2 among this cohort, with the trends in transition rates paralleling each other closely. However, between ML2 and endline, comparison communities experienced a sharp decline in successful transitions, while transition rates in intervention communities held steady.

One possible explanation for differential trends from ML2 to endline is that the worsening drought in the last several months has affected transition rates, by driving families to migrate or reducing their financial means to pay for school. However, there is no evidence – based on an informal coding of drought-exposure and severity – that intervention schools are more or less impacted by the current drought than are comparison schools. This analysis is based on a rough coding of schools as particularly drought-affected based on the

[93] In terms of a regression-based difference-in-differences model, the baseline to endline panel sample produces a programme impact estimate of 5.1 points, though this effect is not statistically significant. Within the ML2-to-EL panel, the programme appears to have produced a 6.9 point increase in transition rates, a finding that is significant at the 10 percent level (p = 0.08).

region they are located, reports from the Food Security and Nutrition Analysis Unit (FSNAU) and their drought mapping efforts from December 2021. This rough coding is far from perfect, and our data allows us to assess the observable implications of drought relatively directly, through information on diet and food consumption over the previous 24 hours. These alternative approaches produce broadly similar results with regard to intervention and comparison schools. We employed measures that included a lack of good sources of protein in the last 24 hours, and the consumption of two or fewer, one or fewer, and no meals in the last 24 hours. For each of these measures, intervention schools are either less marginalised than, or virtually indistinguishable from, the typical comparison school. In order for drought to explain away differential trends, schools would need to be differentially exposed to the drought.

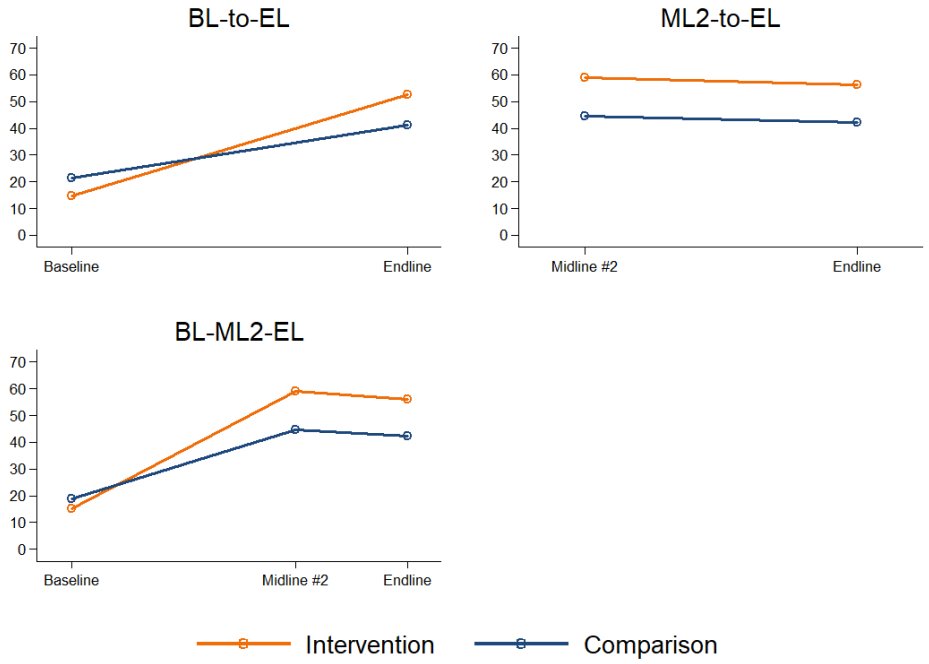
A more likely explanation is that one or both of the two major shocks that occurred between ML2 and endline – drought and the COVID-19 pandemic – while distributed relatively evenly across intervention and comparison groups, had differential impacts on the two groups *because* of the programme's activities. In other words, we believe it is possible that the programme's activities made enrolment and transition in intervention communities more resilient to such shocks, with girls more likely to remain in school despite disruptions and exogenous forces. While this may be viewed as a generous interpretation of the evidence, it aligns generally with how we view the COVID-19 shock, in particular. The impact of COVID-19 on schooling was relatively uniform across zones. Where variation in the severity of the pandemic occurred – if there was variation at all – was between urban centres and rural areas, though poor health surveillance and tracking data mean that we can only speculate about the extent of COVID-19 infections and mortality rates across geographic space.

More importantly, though, the pandemic's impact on schooling outcomes occurred primarily through the closure of schools, which was uniformly instituted and enforced, at least within the three main geographic zones of this study. For this reason, we are confident that the pandemic does not violate the parallel trends assumption underlying difference-in-differences; while it represents an exogenous shock, it was not applied differentially to intervention and comparison groups. As a result, one explanation for the declining transition rates in comparison communities and stable transition rates in intervention communities from ML2 to endline is that girls in intervention communities were better able to weather COVID-19, remaining enrolled and re-enrolling in the face of the disruptions it caused.

Unpacking the trends in transition rates among in-school girls, it appears that a portion of the programme's impact is caused by higher rate of grade promotion in intervention schools, among the subset of girls enrolled in school. We noted this pattern earlier in this section in a different context, but it appears in the broader estimates of programme impact as well. For instance, while the programme improved transition rates by 5.1 points from baseline to endline among the cohort of in-school girls, when we classify grade repetition as a successful transition – an alternate coding – the programme's impact diminishes somewhat, to 3.9 points. To be clear, this finding is not a critique of the programme, because we cannot determine whether grade-level standards have been relaxed or tightened in different schools. Grade promotion, if it signals knowledge acquisition, *should* be viewed differently than grade repetition. Nonetheless it is worth noting that at least some of the programme's impact is in reducing grade repetition, even if it is not the dominant source.

Interestingly, the programme's impact within the out-of-school girls cohort is similar in magnitude but concentrated in a different period. The figure below reports transition rates among the out-of-school cohort, whose sample size varies from 351 for the BL-to-EL panel, to 242 in the ML2-to-EL and BL-ML2-EL panels. The cohort of out-of-school girls was smaller than the cohort of in-school girls even at the outset of the evaluation, comprising 43.3 percent of the baseline sample. However, it shrank dramatically over time, as out-of-school girls were not replaced when they fell out of the sample in the intervening rounds.

FIGURE 26: TRANSITION RATES AMONG THE OUT-OF-SCHOOL COHORT



Despite the limited sample sizes available, the figure above paints a clear picture of impact over the life of the programme. From baseline to endline, the programme produced an 18.0 point improvement in transition rates among the cohort of out-of-school girls ($p = 0.04$). As the graph in the lower-left panel shows, the programme's impact was concentrated in the first two years of implementation, between the baseline and ML2, after which the evolution of transition rates among intervention girls mirrors that of comparison girls through the endline. This contrasts with the timing of impacts seen among the cohort of in-school girls; it also fits with our understanding of programme implementation because, by the time the ML2 evaluation was conducted, both ALP and ABE centres had been established and were active.

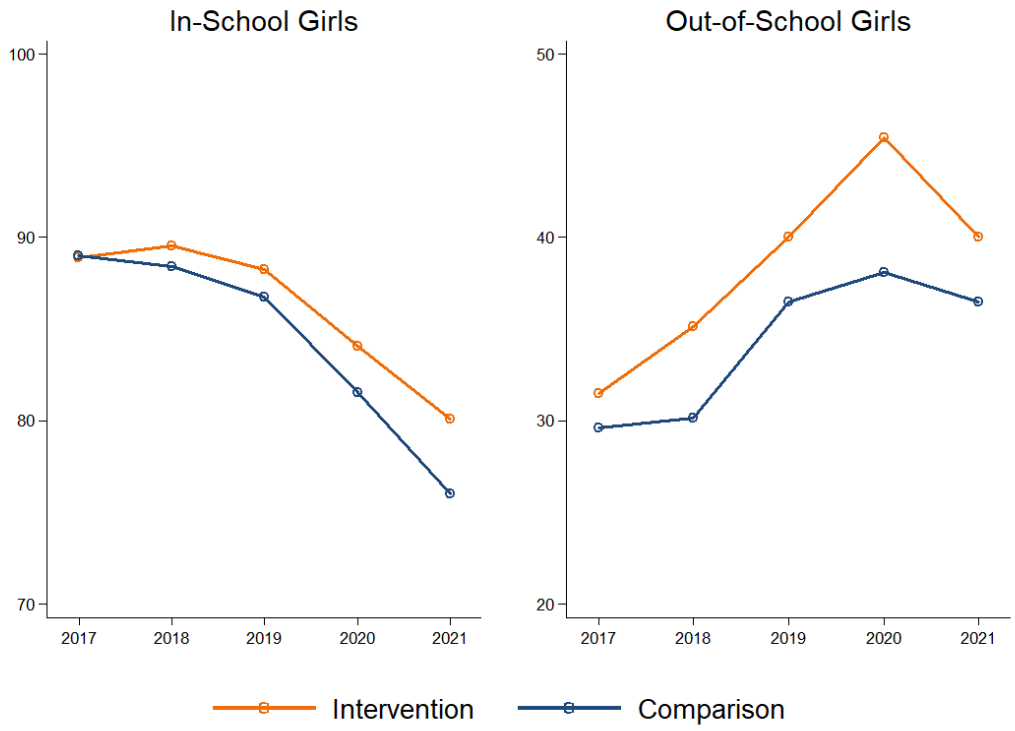
The primary methodological criticism of our analysis to this point centres on the sample sizes available. In small samples, difference-in-differences estimates can be highly variable (subject to significant sampling variation), especially when coupled with a clustered design in which girls are clustered within schools in the sample. The clustered design further reduces statistical power and precision, already a problem with a small sample.

To check the robustness of our findings to a larger sample, we employ the retrospective data collected during the endline evaluation round, described in the introduction to this section. Each girl interviewed at endline was asked to report her enrolment status and grade level for each year, beginning in 2017, yielding a sample of 1,310 girls for whom we have enrolment rates dating to the start of the programme. While this data can

also be questioned in terms of its accuracy – due to issues of recall for periods so far back – it provides a useful secondary check on the programme's impact.[94]

In the figure below, we report enrolment rates from 2017 to 2021 using data reported retrospectively at endline. Note that we focus on enrolment and not transition, because the latter requires knowing whether the girl advanced a grade level year-on-year. While we captured this information from each girl, we recognise that, if recall bias is a concern, it is most likely to occur when recalling a specific fact that may be less memorable than overall enrolment status.

FIGURE 27: ENROLMENT RATES, 2017-2021, RETROSPECTIVE REPORTS



The left panel of the graph reports enrolment rates among in-school girls. On one hand, the fact that enrolment rates are not 100 percent at baseline is worrying, because in-school girls were all recruited from within schools at the time of the baseline in 2017. A number of girls who were reportedly enrolled at baseline (either self-reported or reported by their caregiver) now say they were not enrolled in 2017. Although this does cast doubt on the veracity of either self-reports or retrospective self-reports, it is also possible that girls

[94] Concerns regarding recall accuracy are justified, but it is easy to overstate them, depending on the circumstances. In this case, we asked girls to report whether they were enrolled in school and their grade level in each year. These are mostly objective outcomes – they have a true answer that is only partially open to interpretation – and they are memorable, by which we mean that girls are likely to recall whether they attended school in a given year, because it shapes so much of their life at that time. Most concerns about recall are focused on outcomes that are easy to forget, such as what one ate over the last week or month. Enrolment in school is less likely to be remembered incorrectly.

temporarily attended school and were considered enrolled at the time, but no longer view their level of engagement in 2017 as equivalent to actual enrolment.

The trend shown in the left panel of the figure is telling. While both intervention and comparison communities have seen a decline in enrolment rates over time, the decline is steeper among comparison communities. From a nearly identical starting point in 2017 – just 0.1 percentage points difference – enrolment rates fell by 13.0 points through 2021. In contrast, enrolment rates by just 8.9 points over the same period in intervention communities. As with our analysis of transition rates among in-school girls, reported previously, the biggest gaps between intervention and comparison groups, among the in-school girls cohort, emerges between the ML2 (2019) and endline (2021) rounds.

The right panel also buttresses some of the results reported earlier among out-of-school girls. First, we see a sharp increase in enrolment rates among the intervention group, relative to the comparison group, over the first three years of the programme. However, these gains are partially wiped out by the time of the endline, with the intervention group experiencing a significant decline in enrolment. Even with this recent decline, the gap between out-of-school girls in intervention and comparison communities widened from 1.9 points in 2017 to 3.5 points in 2021. At its largest, in 2020, the gap was 7.4 points.

Our retrospective measures of enrolment are imperfect, and the results using them are not as conclusive – in a statistical sense – as those employing the more traditional measure of transition we have employed in previous rounds. However, by providing a secondary check on trends between intervention and comparison communities over time, they confirm and reinforce the earlier findings. It is clear that the programme has had a substantial impact on transition rates in SOMGEP-T communities, with higher re-enrolment of out-of-school girls, and more in-school girls remaining enrolled and advancing in grade level over time.

Subgroup Programme Impact

In this section, we extend our analysis of SOMGEP-T's impact on transition rates to consider the possibility that the programme had differential impacts within different subgroups. Heterogeneous impacts are of particular interest because the programme's beneficiaries vary in terms of their relative marginalisation, the specific barriers to education they face, and their starting points with regard to learning and even enrolment status. If the programme was able to bring girls into school, but this effect was limited to less marginalised girls, this is an important nuance to the overall finding – discussed in the previous section – that the programme had a positive impact on transition rates, writ large.

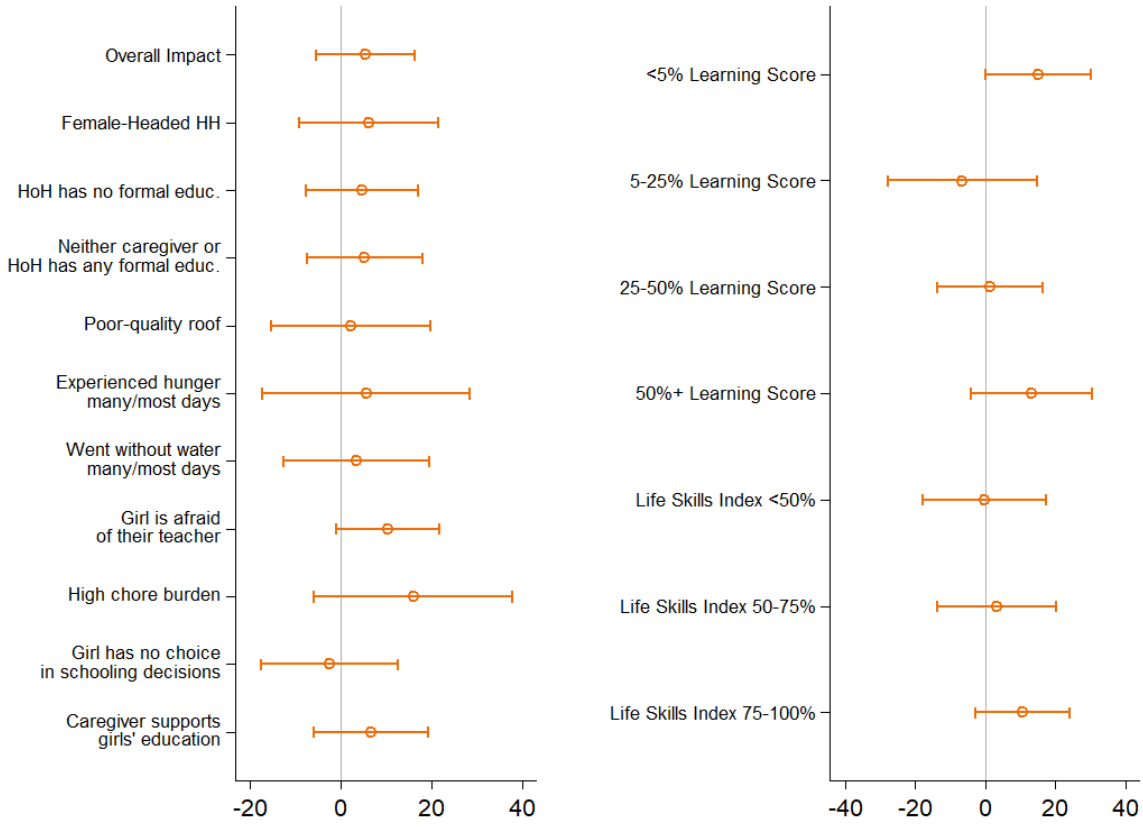
To study heterogeneous programme impacts, we classified girls into subgroups on the basis of their responses to survey questions and performance on learning assessments in earlier rounds. This is a subtle, but important, point: many of the variables we use to construct subgroups can and do change over time – e.g., learning scores, domestic chore burden, fear of one's teacher, etc. – and they may specifically change in response to programme interventions. We are not interested in the correlation between these factors and transition rates; if we were, we would study the relationship between, e.g., home roof quality, and transition rates over time and across groups, while taking advantage of variation in home roof quality within the same individual over time. Instead, we identify and classify individuals on the basis of their baseline characteristic (or, when studying impacts from ML2 to endline, their characteristic at ML2) and then assess programme impact using difference-in-differences within that subgroup. In each case, we employ a model identical to those used earlier for the analysis of aggregate transition rates – a linear (OLS) regression predicting success on a binary transition variable, while controlling for pre-existing baseline differences between intervention and

comparison groups within this particular subgroup, and comparing trends over time between intervention and comparison groups.

We start by considering heterogeneous impacts since baseline, employing the baseline-to-endline panel of 965 girls who were tracked over this period. Each subgroup analysis employs a subset of this data, sometimes as small as 95 girls (n = 190 across the two rounds), in the case of the subgroup who reported having insufficient food to eat many times over the previous 12 months. The figure below reports impact estimates within individual subgroups; we plot the treatment estimate in orange, with a bracketed bar around it representing the 95 percent confidence interval for the estimate. Each estimate comes from a different regression, estimated on a sample consisting only of those respondents who meet the definition of the subgroups listed along the y axis.

It is important to note the first finding in the left panel, which reports aggregate transition impact from baseline to endline. This is important for comparison purposes, because the programme, overall, had a positive impact on transition rates. Therefore, for us to find heterogeneous impacts of the programme, the impact would not need to differ from the vertical line at zero; it would need to differ from the aggregate or "overall impact."

FIGURE 28: HETEROGENEOUS PROGRAMME IMPACTS ON TRANSITION RATES, BY SUBGROUP, BASELINE TO ENDLINE



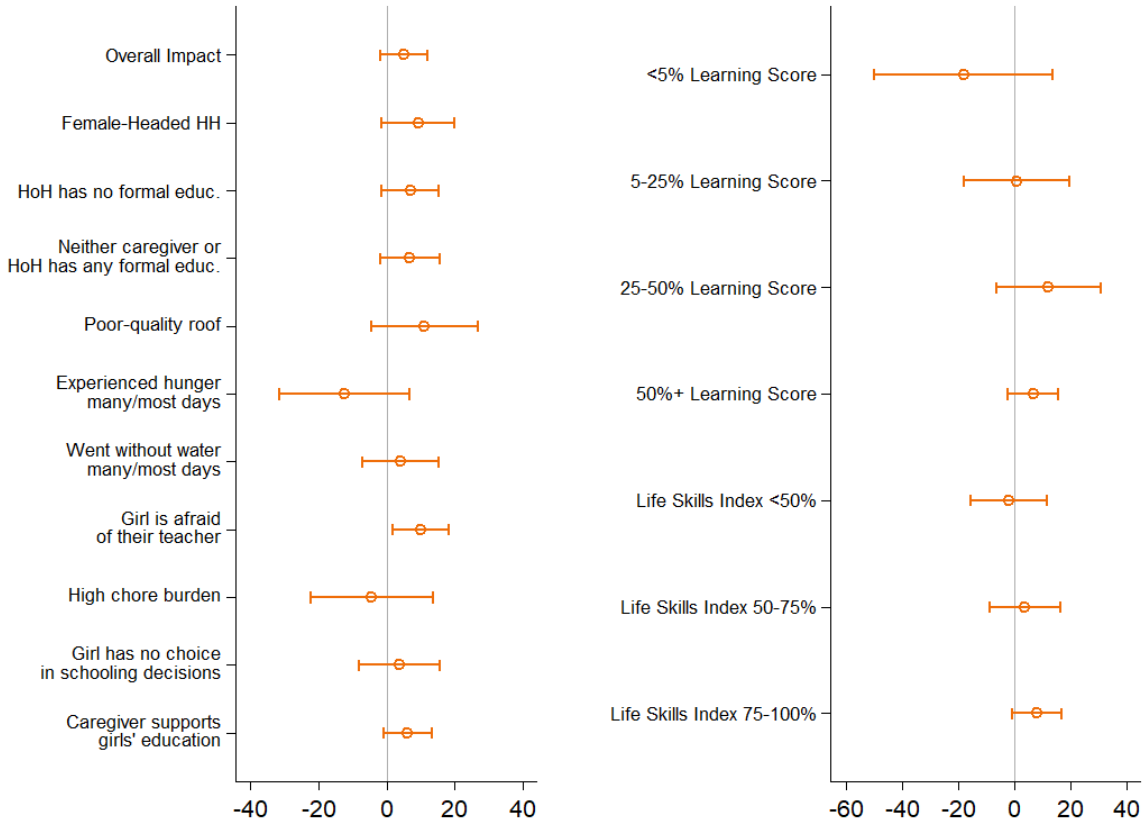
As the results show, household characteristics, including the educational attainment of a girls' head of household and caregiver, do not seem to alter the small positive effects the programme had on transition rates. Household level poverty, in the form of a substandard roof, household hunger, and lack of consistent household access to water, also does not seem to shape how the programme affects transition rates. On the other hand, among girls who reported being afraid of their teacher at baseline, the programme's impact on transition is slightly higher than the overall result. And girls who were subject to a heavy chore burden (a half-day or more of household work, on average, as reported by their caregiver) at baseline see particularly positive effects from the programme. The latter two findings align with other findings in this report, especially regarding the prominent role domestic work and household chores play in preventing girls from attending school. They are also consistent with the programme's own design and the logic of interventions meant to improve the gender-sensitivity of teachers, reduce their use of corporal punishment, and reduce the burden placed on girls within their households.

The findings regarding chore burden and fear of one's teacher are slightly different within the ML2-to-endline analysis, reported in the figure below. In this sample, we can see that girls who report being afraid of their teacher at ML2 reap slightly greater benefits from the programme, in terms of transition outcomes, than the typical girl. But in this sample, domestic chore burden is no longer associated with a larger-than-average programme impact. This may be because the programme – through community sensitisation – sought to reduce this burden. Girls who at ML2 were still tasked with a heavy chore burden may belong to a family or household that is more resistant to change in intra-household labour divisions and even the importance of girls' education. Among such a subset, we might speculatively assume that transition outcomes would be less responsive to programme interventions. This would explain the disjuncture between a relatively large positive estimate among this subgroup at baseline but a negative estimate in the ML2 analysis among a set of girls that, on this single characteristic, are similar to those at baseline.

The analysis beginning with baseline also yields an interesting pattern with regard to baseline learning scores, with girls who scored very low or very high (approximately the bottom and top quartile of the distribution) responding especially positively to the programme's efforts to increase transition rates. High-scoring girls might be expected to have higher transition rates in general, because better-performing girls are more likely to stay enrolled and – more importantly for the purposes of driving up their transition rates – advance to the next grade in school, rather than being held back. However, it is less clear why the programme's impact – comparing changes in intervention and comparison schools – on transition rates would be systematically higher among this group of girls.

The programme also had relatively greater impact on transition among girls who scored below 5 percent on the aggregate of the numeracy and Somali literacy assessments at baseline. This finding is readily explained: this subgroup is made up largely of out-of-school girls, who were not enrolled during the baseline. The programme implemented specific interventions – including alternative education options and outreach to out-of-school girls to encourage them to enrol in formal school – that would have had a particularly strong effect on transition rates among this subgroup, an argument that appears to be borne out in the data. Interestingly, when we consider the ML2-to-endline analysis in the figure below, this pattern no longer holds. This actually helps confirm the argument above, because many of the gains in enrolment among out-of-school girls would have been obtained by the time the ML2 evaluation was conducted, as ALP and ABE programmes had already been started and community-based efforts to promote enrolment had been running for at least two years by that time. Therefore, the "low-hanging fruit" that out-of-school girls – and, by association, low-scoring girls – represented at baseline had already benefitted from the programme's efforts and there were fewer gains to be made over the ML2-to-endline period.

FIGURE 29: HETEROGENEOUS PROGRAMME IMPACTS ON TRANSITION RATES, BY SUBGROUP, ML2 TO ENDLINE



ALP Girls Transition Outcomes

Recruitment of the ALP girls' cohort took place over two evaluation rounds, as described earlier. The first group of ALP girls was recruited during ML1; these girls were re-contacted at ML2 and again at endline. During ML2, some new ALP girls entered the sample for the first time as replacements for ALP girls – initially recruited at ML1 – who could not be re-contacted at ML2. Below, unless otherwise noted, all analysis refers to girls who were first recruited in ML1, and followed up in ML2 and the endline. The below table shows the transition outcomes for all ALP girls. A successful transition is defined as achieving gainful employment, informal education (including ABE and ALP), and formal education. Overall, 56% of ALP girls had a successful transition in the endline. This is 27 percentage points lower than the in-school girls in the overall cohort but 26 percentage points higher than the out-of-school girls. Similar to the findings in learning outcomes, girls in ALP had improved outcomes compared to out-of-school girls but were unable to "catch-up" to in-school girls.

TABLE 34: TRANSITION OUTCOMES FOR ALP GIRLS IN ENDLINE

	Enrolled at ML1 (n=209)	Enrolled at ML2 (n=213)
Total transition status		
Successful Transition	53%	59%
Specific transition status		
Not Enrolled	44%	37%
Non-Gainful Employment	3%	4%
Gainful Employment	1%	1%
Informal Education	22%	24%
Enrolled, Formal Education	15%	19%
Enrolled, ALP	13%	15%
Enrolled, ABE	1%	0%

We also looked at whether girls from marginalized households were more or less likely to successfully transition. However, being from a marginalized household did not appear to impact transition rates. Unsurprisingly, girls from pastoralist households were slightly less likely to be enrolled in any type of education. In addition, VSLA participation was not shown to be correlated with overall transition rates. However, girls from households with VSLA participation were slightly more likely to be enrolled in formal education than informal education.

TABLE 35: TRANSITION OUTCOMES FOR ALP GIRLS BASED ON HOUSEHOLD CHARACTERISTICS

	All (n=163)	Female HoH (n=79)	Pastoralist HoH (n=6)	HoH – No Education (n=71)	HoH – No Job (n=78)	VSLA Participation (n=33)
Total Transition Status						
Successful Transition	54%	58%	50%	55%	52%	52%
Specific Transition Status						
Not Enrolled	42%	37%	50%	41%	42%	44%

Non-Gainful Employment	4%	5%	0%	4%	5%	4%
Gainful Employment	1%	3%	0%	1%	0%	4%
Informal Education	22%	23%	0%	23%	22%	11%
Enrolled, Formal Education	16%	16%	17%	11%	14%	26%
Enrolled, ALP	14%	15%	33%	18%	15%	11%
Enrolled, ABE	1%	1%	0%	1%	1%	0%

Next, we looked at girl characteristics, including age, GwD, and participation in GEFs. Younger ALP girls, ages 14 to 15, were more likely to be enrolled in school (both formal and informal education), with almost 90% reporting enrolment in formal education. Approximately 40% of all older girls (16+) were not enrolled in school and not employed. 11% of older girls were married. The differences in age may indicate increasing family or community pressure to drop out of school once reaching a certain age. Also, older girls may have faced more pressure to drop out to help their family during the COVID-19 pandemic whereas families supported younger girls going back to school. Of ALP girls, only 6 were GwD and 7 were involved in GEFs so no conclusions could be drawn about how disabilities or GEF participation impacted transition.

Lastly, we asked ALP girls to report their activities in 2020. Girls overwhelmingly stated that they were either enrolled in formal education or ALPs (98%), whereas their reported activities during the endline were much more varied. This calls into question the accuracy of their self-reporting but it could be that high enrolment was reflective of a pre-COVID-19 environment. Girls could have dropped out once schools shut down and not returned once they reopened. This mirrors trends described elsewhere in the report where enrolment fell during the endline, possibly due to COVID-19 or the ongoing drought. This is also supported by the low numbers of girls reporting employment of any kind; it is unlikely that if girls had to drop out due to COVID-19 or the drought, they would easily be able to find paid employment elsewhere.

TABLE 36: RETROSPECTIVE TRANSITION OUTCOMES FOR ALP GIRLS

Transition Outcome	2020	2021
Not Enrolled	0%	42%
Non-Gainful Employment	0%	4%
Gainful Employment	0%	0%
Vocational Training	2%	1%
Informal Education	0%	22%

Enrolled, Formal Education	49%	16%
Enrolled, ALP	49%	14%
Enrolled, ABE	0%	1%

ABE Girls Transition Outcomes

Girls in ABE programmes were first recruited in ML2, limiting the ability to compare from baseline to endline. Successful transitions for ABE girls between ML2 and endline include enrolment in formal or informal education or gainful, age-appropriate employment. Unsuccessful transitions include non-enrolment or non-gainful employment.

347 girls who were enrolled in ABE programmes during ML2 were successfully followed up with in the endline. Of those 347 girls, 291 (84%) had successfully transitioned. The majority of those who successfully transitioned were enrolled in formal education (73.8%). These rates are similar to those reported by in-school girls in the overall cohort. This finding is interesting because ABE girls typically did not attend school before ABE or had only attended early primary. After enrolling in ABE, however, they were able to successfully transition. This may indicate that ABE is successful in either teaching girls enough to allow them to enrol in formal education or motivating them to pursue other successful transition pathways.

TABLE 37: TRANSITION OUTCOMES FOR ABE GIRLS IN ENDLINE

Transition Outcome	All (n=347)
Total transition status	
Successful Transition	84%
Specific transition status	
Not Enrolled	16%
Non-Gainful Employment	0%
Informal Education	11%
Enrolled, Formal Education	62%
Enrolled, ALP	1%
Enrolled, ABE	10%

We also looked at whether ABE girls from marginalized households were more or less likely to successfully transition. Over half of ABE girls come from marginalized households, specifically households where the HoH does not have an occupation or does not have any education, which is aligned with the programme's target ABE population. It does not appear that marginalization of the household impacted ABE transition as all girls had relatively similar rates of successful transition. Girls from pastoralist households were slightly less likely to transition into formal education which is similar to other findings that show girls from pastoralist

households having more barriers to engaging with formal education. We also analyzed whether household protein consumption could have impacted transition outcomes. However, transitions did not differ between households with and without protein consumption (83% successful transition vs. 87%, respectively).

TABLE 38: TRANSITION OUTCOMES FOR ABE GIRLS BASED ON HOUSEHOLD CHARACTERISTICS

Transition Outcome	All (n=347)	Female HoH (n=148)	Pastoralist HoH (n=56)	HoH – No Education (n=176)	HoH – No Job (n=171)	VSLA Participation (n=33)
Total Transition Status						
Successful Transition	84%	82%	79%	88%	83%	88%
Specific Transition Status						
Not Enrolled	16%	17%	21%	13%	16%	12%
Non-Gainful Employment	0%	1%	0%	0%	1%	0%
Informal Education	11%	6%	7%	10%	9%	6%
Enrolled, Formal Education	62%	66%	50%	67%	64%	70%
Enrolled, ALP	1%	2%	0%	0%	2%	0%
Enrolled, ABE	10%	8%	21%	10%	8%	12%

We also looked at whether girl characteristics impacted transition outcomes. There was a steady decrease in enrolment as age increased among ABE girls. Interestingly, even though older girls were less likely to be enrolled in school, they were also unlikely to be employed. Similar to ALP girls, this may be a result of COVID-19 where older girls faced increased pressure to not return to school once it reopened. GwD had similar outcomes to other girls, which should be taken as a positive outcome given that GwD often face additional barriers to continuing education. Interestingly, while girls in both ABE and GEFs represented only 12% of the total ABE population, all ABE girls who were in GEFs had successful transitions. This may be due to the small sample size but is similar to other findings that show that girls in GEFs do better than other girls.

TABLE 39: TRANSITION OUTCOMES FOR ABE GIRLS BASED ON GIRL CHARACTERISTICS

Transition Outcome	All (n=347)	GwD (n=21)	Participation in GEFs (n=43)
Total Transition Status			

Successful Transition	84%	86%	100%
Specific Transition Status			
Not Enrolled	16%	14%	0%
Non-Gainful Employment	0%	0%	0%
Informal Education	11%	10%	0%
Enrolled, Formal Education	62%	67%	91%
Enrolled, ALP	1%	0%	5%
Enrolled, ABE	10%	10%	5%

As with the primary cohort of girls discussed above, ABE girls were asked to retrospectively provide information on their transitions during 2020. In 2019, the first year that ABE girls were included in the evaluation, all were enrolled in ABE. In 2020, 64% of ABE girls reported that they had now enrolled in formal education, indicating a successful transition from ABE to formal education. The majority of remaining girls were still in ABE. While the percentage of ABE girls in formal education did not change significantly between 2020 and 2021, 22% of those enrolled in ABE in 2020 either transitioned to another type of education or dropped out of school. This may indicate that the school closures due to COVID-19 impacted ABE girls more than girls enrolled in formal education, and that they were less likely to come back to ABE once the schools were reopened.

TABLE 40: RETROSPECTIVE TRANSITION OUTCOMES FOR ABE GIRLS

Transition Outcome	2019	2020	2021
Not Enrolled	0%	0%	16%
Non-Gainful Employment	0%	0%	0%
Informal Education	0%	0%	11%
Enrolled, Formal Education	0%	64%	62%
Enrolled, ALP	0%	4%	1%
Enrolled, ABE	100%	32%	10%

Qualitative Analysis

In the qualitative data, respondents listed many of the same barriers to enrollment and continued enrollment that have been mentioned since the baseline. Financial constraints remain one of the most commonly mentioned barriers; many families cannot afford school fees for either any of their children or – in the case of some families – cannot afford school fees for all of their children and must therefore be selective and send

only some of their children to school. Additionally, families continue to rely on girls' contributions to the household, particularly when the family is facing economic hardship. Qualitative interview respondents were presented with a hypothetical scenario of a girl whose family is experiencing economic hardship due to drought and were asked what the likely outcome would be for that girl. In most cases, respondents expressed serious doubt that the girl would be able to continue her education, given the financial resources needed for school fees and to pay for uniforms, books, and other supplies.

However, it was also clear from the qualitative interviews that various aspects of the program have contributed to addressing these barriers. Although respondents were doubtful over the ability of girls whose families are facing serious economic hardship to continue school, they were also likely to mention that if these girls lived in their community, which has received support from the program, these girls would be able to access support from various groups in the community and potentially continue their education. Teachers, CECs, and GEFs appear to be particularly active in raising awareness on the importance of education among community members, providing special attention to vulnerable girls, monitoring the status of struggling students in school, fundraising to provide financial support when needed, and following up with unenrolled girls and drop-outs to encourage enrollment. These are the groups respondents indicated struggling girls could turn to for support during hardship. There appears to be a particular focus on proactively providing support to girls whose families are facing economic hardship, pastoralists, and children with disabilities. "There are families like Haawa's family whose livestock had finished, and their children were helped by the agencies, teachers, and the school administration."[95]

Of course, there are cases in which respondents feel it is outside their ability to provide all necessary support – for example, when a girl's family moves and there is nowhere suitable for her to stay, when there is no secondary school nearby, when the community is unable to raise the necessary funds to cover fees, or when the layout of the community and the school infrastructure are unsuitable to the needs of the child (e.g. very rough terrain or long distances, lack of necessary equipment to meet needs of students with certain disabilities). Regardless, the proactive engagement of the aforementioned groups in encouraging girls' enrollment and continued enrollment among groups that are traditional marginalized represents an important shift in community attitudes that should be noted and can be directly attributed to the program.

In addition to the apparent success of the GEF, CEC, and teacher-focused activities in encouraging enrollment and re-enrollment, ABE and ALP activities have provided an important alternative route to continuing education, particularly for girls whose schedules are not compatible with normal schooling hours. As one respondent notes,

> *"CARE International has constructed new classrooms at the school. In the afternoons and evenings, additional teaching classes have been offered to the school. And, after reopening, the school is always two months ahead of the other schools. That helps pupils who arrive at school later in the day, and the institution has power, allowing pupils to enroll late at night."*
>
> -FGD with CEC members, Int. 106

[95] FGD with Mothers, Int. 308.

These lessons have had an important positive impact on learning outcomes, which in turn encourages girls to stay in school: "Evening lessons aided pupils in math and English studies, which boosted the girls' educational morale."[96] These classes are particularly important for girls who are older, married, have children, or whose home circumstances otherwise do not allow them to attend or make them feel comfortable attending regular school hours. We know that these girls face higher barriers to enrollment and re-enrollment given the ease with which they fall behind and the stigma they face from other children. As one respondent explains, "There are girls that have enrolled the education while they were old and then leave after they have difficulties for reading and finally get married."[97]

Unfortunately, the qualitative results also suggest that external factors have reversed or nullified some of the successes mentioned above. First, during COVID, schools were mandated to shut down, and despite efforts in some communities to continue education remotely, many fell behind or were unable to access education at all during the shut-downs. There are reports in the qualitative data of teachers attempting to continue teaching via Whatsapp or online resources, children attempting to continue learning from one another, and parents paying for private schooling to keep their children up-to-speed on their education. However, financial resources would clearly have been needed to access private schooling, internet connection, electricity, and phone services, meaning already disadvantaged children were further disadvantaged during the school shut-downs. As one student explains, "There are students who are close to the teachers like me. I had taken advantage of them. And those who are far away from teachers didn't get any help."[98] A mother explains, "People are not always on the same financial state. I was not able to pay a home teacher to help my daughters."[99]

As a result, many students fell severely behind. It appears at least some schools administered an exam to all returning students to place them in the appropriate level once schools re-opened, a practice which may have deterred some from returning. As one respondent explains, "Many students have dropped out, although some have been able to study at home. Others got disappointed with how they will pass the exam when the school reopens."[100] For those who did return, difficulties keeping up with the content may have further deterred students from remaining enrolled: "When the school opened, the lessons could not be understood, since we were off for a long time. This resulted that student drop out of school."[101]

Another external factor that may have affected transition results is the severe drought experienced in some communities, though the effect the drought had likely differed from community to community. Some areas experienced an influx of new individuals: "Yes, due to the drought, another community has moved in, which has resulted in an increase in the number of pupils."[102] Others would have experienced an out-flux of community members, leading to drops in enrolment and re-enrollment.

[96] FGD with CEC Members, Int. 106.
[97] Vignette FGD with Girls, Int. 607.
[98] Vignette FGD with Girls, Int. 604.
[99] FGD with Mothers, Int. 301.
[100] FGD with Teachers, Int. 209.
[101] Vignette FGD with Girls, Int. 604.
[102] FGD with CEC Members, Int. 106.

Sustainability

This section presents an analysis of project sustainability at each of three levels: community, school, and system. The results of this analysis are presented in narrative form in the subsections that follow. The triangulated analysis was used to generate a qualitative sustainability score (ranging from 0 to 4) for each of the key sustainability indicators identified in the SOMGEP-T Logframe. Sustainability scores for each indicator can range from 0 to 4, in line with the FM's MEL Guidance:

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

Scores for indicators in the same level – community, school, or system – are aggregated into a single level score; in turn, those levels are averaged to produce a single sustainability score. These scores, for baseline, midline, and endline are reported in the scorecard table below.

While additional rationale for each indicator's score is provided in the more detailed narrative sections that follow, it is important to note that the scores are admittedly subjective. We assigned scores based on the totality of information available – in some cases, this was limited to a few qualitative interviews, while in other cases it included quantitative data from multiple respondent groups and a range of qualitative data. As a broad rule of thumb, where noticeable but not dramatic changes have been observed since the baseline, we tended to score indicators as one grade above the baseline. Where two indicators in the same level (community, school, or system) both had marginal improvements, we split the difference, again relying on our best, but subjective, judgment.

TABLE 41: SUSTAINABILITY SCORECARD

	Community	School	System
Indicator 1:	***Percentage of CECs actively engaged in mobilizing for girls' education through fundraising for payment of additional teachers' salaries and school supplies***	***Percentage of project target schools adhering to implementation standards for ALP, ESL, Numeracy and Remedial classes.***	***Inclusion of ALP in the national non-formal education frameworks***

	Community	School	System
	Baseline Status:	**Baseline Status:**	**Baseline Status:**
	9.5% of CECs provide support for teacher salaries (18.8% intervention; 0% comparison).	39.7% of teachers in intervention schools reported using formative assessments. This assessment is limited to the same sample of schools employed at midline, to facilitate comparisons. (Comparison schools: 62.5%)	Not applicable because activities not established yet
	31.4% of parents report CECs provided financial support of some kind (fundraising, infrastructure, buying materials, financial support to students) to schools in the last 12 months (31.3% intervention; 31.6% comparison).	32.8% of teachers in intervention schools could show records of using formative assessments. (Comparison schools: 44.6%)	
Indicator 1:	**Midline round 1 Status:**	**Midline round 1 Status:**	**Midline round 1 Status:**
	30.2% of CECs provide support for teacher salaries (25.8% intervention; 34.4% comparison).	69.0% of teachers in intervention schools reported using formative assessments. (Comparison schools: 72.4%)	Some mention of ALP in REO interviews. Awareness among Ministry officials is clearly growing with increased discussion of alternative learning opportunities.
	The panel for household survey at Midline round 1 is exclusive of out-of-school girls. At Midline round 1, the wider panel between BL and ML2 is thus used.	34.5% of teachers in intervention schools were able to show records of the use of formative assessments. (Comparison schools: 24.1%)	

	Community	School	System
	47.6% of CECs, per head teachers, raise funds for school improvements (59.4% intervention; 35.5% comparison).		
Indicator 1:	**Midline round 2 Status:**	**Midline round 2 Status:**	**Midline round 2 Status:**
	15.9 % of CECs provide support for teacher salaries (21.9% intervention, 9.7% comparison)	91.2% of teachers in intervention schools reported using formative assessments. (Comparison schools: 91.2%)	Talk of ALP is very limited in interviews with MoE respondents. Many mentions of special efforts for including the nomadic population and disabled children. But most respondents do not view these as possible at this time, due to financial constraints and other operational challenges.
	54.9% of parents report CECs provided financial support of some kind (fundraising, infrastructure, buying materials, financial support to students) to schools in the last 12 months (57.9% intervention; 50.5% comparison).	35.1% of teachers in intervention schools were able to show records of the use of formative assessments. (Comparison schools: 22.8%)	
	46.3% of CECs, per head teachers, raise funds for school improvements (56.3% intervention; 35.5% comparison).		
	Endline Status:	**Endline Status:**	**Endline Status:**

	Community	School	System
Indicator 1:	12.1 % of CECs provide support for teacher salaries (16.7% intervention, 7.1% comparison)	81.5% of teachers in intervention schools reported using formative assessments. (81.5% intervention; 81.1% comparison)	All REOs interviewed in the endline had knowledge of ALP implementation within their schools and most had awareness of ABE programmes. Barriers to success during the evaluation period included COVID-19 closures and low attendance. There are attempts to reach GwDs and pastoralist girls through alternative education, but ALP and ABE programs do not currently consistently reach GwDs and pastoralist girls. Reasons cited include both resource constraints and difficulties addressing the unique barriers of these groups (e.g. pastoralist children migrating).
	Parents report CECs provided financial support to students to schools in the last 12 months (24% intervention; 17% comparison).	11.4% of teachers in intervention schools were able to show records of the use of formative assessments. (11.4% intervention; 14.0% comparison)	
	66.7% intervention and 71.4% comparison CECs, per head teachers, raise funds for school improvements.		
	Percentage of parents in intervention schools indicating that CECs are functional	***Not applicable***	***No. of MOE departments engaged in support of girls' education from National to regional and district levels.***
	Baseline Status:	**Baseline Status:**	**Baseline Status:**

	Community	School	System
	71.9% of parents in intervention communities report a functional CEC, compared to 58.5% of parents in comparison communities.	***Not applicable***	Interviews reveal a systemic lack of funds. Local schools do not have enough money to maintain facilities and pay their staff partly because CECs do not have the funds to support their schools. The CECs do not have funds because little to no financial support is given to them by the MoE. Some in-kind support from MOEs was provided. 31% of schools had CEC Management plan, taken as proxy of MOE involvement.
Indicator 2:	**Midline round 1 Status:**	**Midline round 1 Status:**	**Midline round 1 Status:**
	The panel for household survey at Midline round 1 is exclusive of out-of-school girls. At Midline round 1, the wider panel between BL and ML2 is thus used.	***Not applicable***	Two out of six REOs explicitly confirmed the existence of gender units. Ongoing difficulties recruiting and retaining female teachers and high-quality teachers. Mixed evidence on mainstreaming of improved teaching practices, but these appear to be emergent at this phase. Minimal evidence of MOE efforts to encourage female pre-service graduates.
	96.8% of schools have a functioning CEC (head teacher survey) (96.9% of intervention schools, 96.8% of comparison schools)		
	Midline round 2 Status:	**Midline round 2 Status:**	**Midline round 2 Status:**

	Community	School	System
Indicator 2:	66.5% of parents in intervention communities report a functional CEC, compared to 59.2% of parents in comparison communities.	***Not applicable***	Gender units seem to be more active than at previous Midline. Their financial sustainability, however, cannot be assessed. Various activities were detailed relating to teaching quality and addressing retention and transition concerns. However, lack of funding complicates the efforts and their uniformity cannot be assessed at this stage. When looking at quantitative measures of CEC activity as a proxy for MoE activity, all signs are positive.
	According to the head teacher survey 96.8% of schools have a functioning CEC (100% of intervention schools, 93.6% of comparison schools)		
Indicator 2:	**Endline Status:**	**Endline Status:**	**Endline Status:**
	70.6% of parents in intervention communities report a functional CEC, compared to 68.9% of parents in comparison communities.	***Not applicable***	All except one REO were aware of the Gender Department within their MOE, but in most cases, gender departments had lost significant funding, been completely defunded, or been absorbed into another part of the ministry by the time of the evaluation. When active, Gender Departments focused most heavily on female training, re-enrolment of OOS girls, school monitoring, and awareness raising.
	According to the head teacher survey 94.6% of schools have a functioning CEC (94.6% of intervention schools, 93.6% of comparison schools)		

	Community	School	System
Baseline Sustainability Score*	1	1	0.5
Basleine Overall Sustainability Score**		0.8	
Midline Round 1 Sustainability Score*	1.5	2	1.5
Midline Round 1 Overall Sustainability Score**		1.7	
Midline Round 2 Sustainability Score*	1.5	3	2.0
Midline Round 2 Overall Sustainability Score**		2.2	
Endline Sustainability Score	1.5	1.5	1.5
Endline Overall Sustainability Score		1.5	

	Community	School	System
*Range 0-4		**Average of the three level scores	

Indicator 1 – CEC Mobilization

Community level - Percentage of CECs actively engaged in mobilizing for girls' education through fundraising for payment of additional teachers' salaries and school supplies

Community Education Committees (CECs) are local structures that oversee the operation of a primary school in a given community and typically consist of the school's head teacher, community leaders, and parents. CECs support the operation of schools in a number of ways, by monitoring student and teacher attendance, promoting enrolment and attendance, raising awareness of the importance of education, liaising with religious leaders and other individuals of influence in the community, and providing material support to the school, among other tasks.

CECs are important for school governance and management, as they are involved in monitoring enrolment, attendance and teaching quality. In terms of the sustainability of project efforts, CECs contribute to sustainability of results by securing materials and in-kind support for the schools that they operate in. CECs are expected to raise funds from their own resources, by mobilizing community members to donate, and by seeking support from outside organizations, including NGOs and the government. After the conclusion of SOMGEP-T, the sustainability of schools will require continued material support sourced by CECs.

CEC Contribution to Teacher Salaries

One aspect of the CEC's financial obligations is support of teacher salaries. This is critical as teacher salaries are often delayed or not paid at all, as reported through qualitative interviews in each round of the evaluation. Head teachers reported whether the CEC provided at least a partial salary to male and female teachers in their schools. While intervention schools had consistently higher CEC support than comparison schools, CEC support fell from 31.6 percent in BL to only 16.7 percent in the endline. Despite reporting no CEC support in the baseline, comparison schools reported between 7 and 10 percent support for the other rounds. These results could be troubling as they suggest that CECs are not capable of supporting teacher salaries long-term. However, it could also indicate an increase in the support of teacher salaries from the MOE, decreasing the need for CEC support.

TABLE 42: CEC SUPPORT FOR TEACHER SALARIES, BY ROUND

CEC provides support for teachers' salaries	BL	ML1	ML2	EL
Intervention	31.6%	16.1%	21.9%	16.7%
Comparison	0.0%	10.3%	10.7%	7.1%
Overall	24.0%	13.3%	16.7%	12.1%

The qualitative results differ somewhat from the quantitative results – in the majority of the qualitative interviews with MoE officials and CEC members, respondents reported that CECs raise funds to cover teachers' salaries. In a minority of locations, the ministry is actually covering salaries (as suggested above) or CEC members are reportedly unable to raise funds from community members because the community is unable to provide financial support. As one MoE official explains, "The committee can support or donate when 85% or 75% of the community have something (money), but when the whole community has 2% and

they cannot support, you can feel that."[103] Challenges raising funds are also likely to reflect a decrease in remittance frequency and volume, as many of the schools in the regions where SOMGEP-T operates are heavily dependent on remittances for their operation.

However, in the majority of locations, respondents report CECs are able to raise funds, primarily from community members and diaspora. There were multiple accounts across interviews of CEC members constructing collection boxes in the community to raise funds to cover teachers' salaries and school maintenance projects:

> *Some of the teachers' payments that the diaspora community had been supporting us had been suspended, so we created a payment box that we use to pay those teachers so that they can continue teaching at the school.*
>
> -FGD with CEC, Int. 107

In rare cases, CEC members have even been successful in raising funds to provide an additional salary over and above what the ministry is already providing: "I have observed [redacted] primary and secondary schools that have been paid $400 per high school teacher by the community diaspora and other communities that originated from this district, which is in addition to the school payroll that the Ministry of Education has paid. Since the CEC is in charge of providing, they also hired teachers from Camuud and Hargeisa because of their good salaries."[104]

It is unclear from the qualitative interviews whether enough funds are raised to consistently provide full support or whether support is instead sporadic. It is similarly unclear whether COVID or other factors affected the ability of CECs to consistently raise funds in the last year in particular, which could explain the diverging results of the quantitative and qualitative data. However, the qualitative data does suggest that most CECs consider it part of their role to cover gaps in teachers' salaries when possible, and communities and diaspora are increasingly engaging with CECs to meet these community-level education needs.

In addition, we examined the percentage of girls in schools who receive scholarship support as reported by caregivers. These are girls who received scholarship from any source, not solely SOMGEP-T or the CEC. The scholarship support for girls in intervention schools increased over the programme period by 0.1 percentage points, it increased for girls in comparison schools 0.7 percentage points. This small increase in comparison schools could be a result of scholarships from a separate program ("Girls Education Empowerment Project in Somalia", or GEEPS), which CARE began providing in 2021, including in SOMGEP-T comparison communities. The increase of scholarships during ML1 and ML2 are consistent with expectations – the program was winding down at the EL, schools were closed, and budgets were tight.

TABLE 43: SHARE OF GIRLS WHO RECEIVE SCHOLARSHIP SUPPORT

	BL	ML1	ML2	EL
Intervention	12.2%	19.3%	18.23%	12.3%
Comparison	5.5%	3.5%	14.1%	6.9%

[103] KII with ROE Officials, Int. 710.
[104] KII with REO Officials, Int. 706.

Overall	9.1%	12.2%	16.4%	9.8%

This jump in share of girls who receive scholarships from ML1 to ML2 in comparison schools could be the result of SOMGEP-T's positive influence spilling over into comparison communities or other programs providing support to CECs, such as GEEPS. x

We also asked the head teacher to report CEC's share in scholarship support. As seen below, the overall support from the CEC was quite low and decreased in all schools over the programme period. Similar to the teacher salary support described above, this may indicate that CECs are less able to financially support schools than previously. This may be due to financial implications of the COVID-19 pandemic or the drought.

TABLE 44: SHARE OF CEC IN SCHOLARSHIP SUPPORT

	BL	ML1	ML2	EL
Intervention	-	14.3%	22.2%	7.1%
Comparison	-	37.5%	25.0%	0.0%
Overall	-	19.4%	22.7%	4.0%

On support for school fees, the qualitative results are mixed. In some cases, the qualitative data suggests that CECs are unable to provide or raise financial support for girls' school fees, but in other cases, CEC members and other community members report that CECs take an active role in raising funds to cover school fees. When asked about the activities of the CEC, one CEC member explains, "We provided free education to individuals who could not afford to pay their school fees. Educational advancement was also one of the key factors that aided the education of girls. We've also given them books and pens for those who can't afford them."[105]

Again, the timelines are unclear, and it is entirely feasible that COVID or the drought directly affected the ability of CECs to raise funds as they had in the past. For example, COVID's effect on the global economy may have affected the ability of diaspora to provide support, and on a more local level, the fact that schools were closed may have made CEC members reluctant to ask for funds and community members reluctant to give them when their girls were not even attending school. The drought has had a significant negative impact on the financial situation of pastoralist families and has changed the composition of targeted communities. As one MoE official explains, "When there is a drought, the pastoralist and rural populations in this region will all shift to another site in search of pasture and water for their animals, and you can see the detrimental effects of this."[106] Many of those who were receiving support and whose families would have been sensitized by the CECs moved as a result of the drought, further hampering the abilities of CECs to effectively engage with community members to raise funds. "There were about 500 people who benefited from the SOMGEP-T

[105] FGD with CEC Members, Int. 107.
[106] KII with REO Officials, Int. 706.

project, even some of them were studying the school in the afternoon. Such people were affected by the drought and all of them moved out."[107]

Overall, the programme was negatively correlated with girls receiving scholarships and CEC raising funds for school improvement. However, head teachers from both comparison and intervention schools reported similar rates of funds for school improvement at the endline. The programme was positively correlated with share of CEC in scholarship support. It is important to note that even though share of CEC scholarship support decreased in both intervention and comparison schools, it decreased 9.9 points less in intervention schools than comparison schools.

TABLE 45: IMPACT OF PROGRAMME ON CEC FINANCIAL SUPPORT FOR EDUCATION

	Intervention			Comparison			Diff. in Diff.
Indicator	ML2	EL	Diff.	ML2	EL	Diff.	
Girls who receive scholarship	56.3%	43.8%	-12.5	13.3%	36.7%	+23.4	-35.9
Share of CEC in scholarship support	22.2%	7.1%	-15.1	25.0%	0.0%	-25.0	+9.9
CEC raises funds for school improvements	59.4%	66.7%	+7.3	39.3%	71.4%	+32.1	-24.9

We also asked parents whether they thought the CEC in their school had taken on specific activities: 1) raising funding, 2) improving infrastructure, 3) providing financial support for students, and 4) buying learning materials. While parents may not have as deep of an insight into how CECs work, they are less likely to report positive results because of social desirability bias. Therefore, this gives us a better insight into both how public the CECs' work is and how they are using funds to benefit the school. At BL, intervention schools had mixed results compared to other schools. However, the programme was positively correlated with increases in raising funding, improving infrastructure, and providing financial support. The purchase of learning materials was only slightly negatively correlated with the programme; intervention and comparison households reported relatively similar rates for purchase of learning materials. These findings largely contradict the head teacher findings where programme impact was negatively correlated with CEC funding for school improvements.

[107] KII with REO Officials, Int. 705.

FIGURE 30: FINANCIAL SUPPORT OF SCHOOLS BY CECS OVER THE LAST YEAR

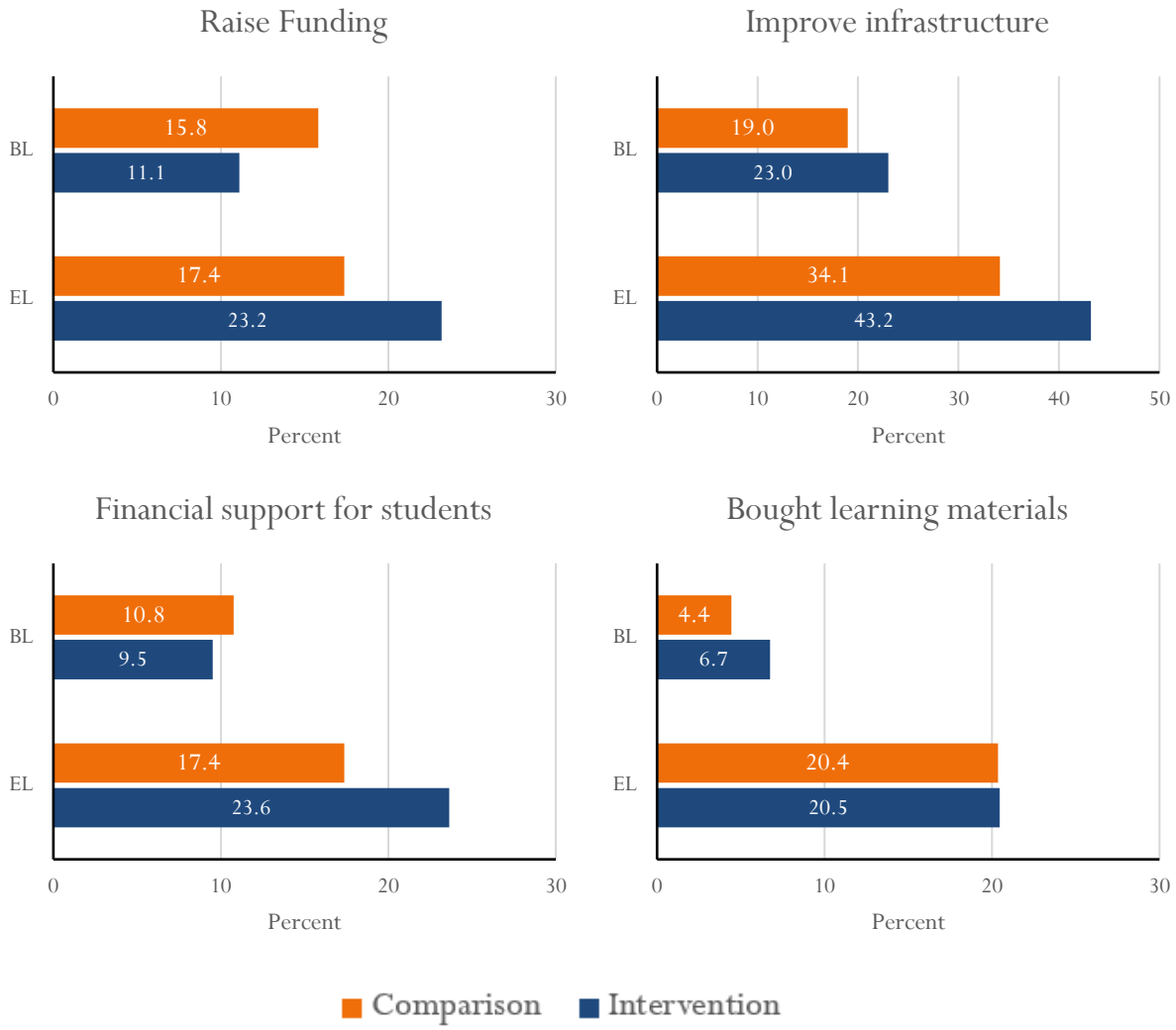


TABLE 46: IMPACT OF PROGRAMME ON CEC ACTIVITY LEVELS

	Intervention			Comparison			Diff. in Diff.
Indicator	BL	EL	Diff.	BL	EL	Diff.	
Raise funding	11%	23%	+12	16%	17%	+2	+11
Improve infrastructure	23%	43%	+20	19%	34%	+15	+5
Financial support for students	10%	24%	+14	11%	17%	+7	+8

Bought learning materials	7%	21%	+14	4%	20%	+16	-2

There is evidence in the qualitative data that CECs play a role in raising funds for school improvement projects, at least in some communities. In others, as has been stated above, CECs feel they are unable to contribute or raise funds and instead focus their attention more heavily on awareness raising and school monitoring activities, which will be discussed later in the sustainability section of this report. When asked whether the CEC has changed in the past years, one CEC member explains, "Yes, we have changed a lot, such as selecting the poor from the community for help from the agencies. We have also built these two rooms and laid the foundation stone for the school yard. We have also done a lot of awareness and there are many students in the school."[108]

Indicator 2 – Implementation of ALP, Numeracy, and Remedial Classes

School level - Adherence to Implementation Standards for ALP, Numeracy and Remedial Classes

CARE held trainings for intervention teachers. These trainings were intended to increase teaching quality and therefore ultimately increase learning outcomes. One of the targets of teacher trainings was to increase the capacity of teachers to help struggling students by giving them remedial support. Teachers need to be able to identify struggling students to be able to help them. CARE taught teachers how to use formative assessments to identify struggling students. Formative assessments are not required by the MOE or school administrators. Teachers do not have to submit formative assessments results and the assessments do not have to be in written form. Formative assessments help teachers understand students' current academic standing and knowledge.

Formative assessment use serves as a proxy for increased teaching quality. Teaching quality is hard to define and measure but formative assessments use is measurable. We have measured formative assessment use in a multiple of ways. Enumerators observed the use of formative assessments during classroom observations. Enumerators recorded teachers' reported use of formative assessments. Enumerators also asked teachers for records of formative assessments. Enumerators, during household surveys, asked caregivers if their daughters have received remedial support and if so, by whom. These effect of CARE's intervention on these eight variables are shown below.

TABLE 47: PROGRAMME IMPACT ON FORMATIVE ASSESSMENT USE AND RECORDS OF FORMATIVE ASSESSMENTS

Full Panel		
	Intervention	Comparison

[108] FGD with CEC Members, Int. 109.

Indicator	BL	ML1	ML2	EL	BL to EL Difference	BL	ML1	ML2	El	BL to EL Difference
Report using formative assessments	34.6%	70.4%	90.6%	81.5%	46.9%	49.1%	70.4%	90.6%	81.1%	28.7%
Records of formative assessments shown	53.3%	52.6%	37.5%	11.4%	-42.0%	53.9%	31.6%	25.0%	14.0%	39.9%

BL to EL Panel							
	Intervention			Comparison			Diff and Diff
Indicator	BL	EL	Difference	BL	El	Difference	
Report using formative assessments	42.3%	82.4%	40.1%	50.0%	81.4%	31.4%	8.7%
Records of formative assessments shown	59.1%	10.7%	-48.4%	55.6%	16.7%	-38.9%	-9.5%

Teacher's reported use of formative assessments has increased over and above comparison schools. The BL to EL sample of classroom observations indicate an 8.7% increase in teacher reported formative assessment use over and above comparison schools. This increase is not significant but it is large. The programme impact from BL to EL on records of formative assessment appears to be negative, at -9.5%. This effect is also not significant. It is possible that intervention teachers are not keeping records of formative assessments because they are only using them to give remedial support. Comparison teachers might not know exactly what comprises a formative assessment and therefore provide records of formal summative assessments, passing them off as formative. The BL to EL sample, while having a larger sample, covers up some fluctuations that are captured in the two midlines.

The full sample, consisting of schools that have been observed all four rounds, shows an apex of teacher reported formative assessment use in the ML2 for both comparison and intervention schools. The full sample also shows that records of formative assessments have been on a monotonic decline since the BL for both intervention and comparison schools. This decline might be the result of teachers and enumerators understanding better what formative assessments are, or it could be a decline in formative assessments in their written forms.

TABLE 48: CAREGIVER REPORTED REMEDIAL SUPPORT

	Intervention		Comparison		Diff and Diff
	BL	EL	BL	EL	
CEC provided remedial support	5.16	8.64	3.16	7.78	-1.1
Girl received remedial support last 12 months	30.46	27.96	20.83	30.20	-11.9
Provider of remedial support					
Student	1.8	3.3	0.00	2.60	-1.2
Family member	16.04	20.65	14.55	25.97	-6.8
School staff	10.38	5.43	9.09	3.90	0.2
Teacher	69.81	70.65	74.55	66.23	9.2

Enumerators asked caregivers in the household survey if their daughter has received remedial support from the CEC. Rates of CEC provided remedial support for both intervention and comparison girls remained low from the BL to the EL, each increasing slightly. Comparison caregivers reporting a slightly larger, 1.1 percentage point increase of intervention caregivers.

Remedial support for girls in the last 12 months, as reported by caregivers, declined for intervention schools and increased for comparison schools. The non-significant difference and difference coefficient is -11.9 percentage points. This is an unexpected finding; intervention teachers were expected to provide increased remedial support after trainings. When we regressed the giver of remedial support, as reported by the caregiver, we find intervention teachers did provide more remedial support over and above comparison teachers from BL to EL, a difference and difference of 9.2 percentage points.

Remedial support given by other students and family members to caregiver daughters increased more for comparison daughters than for intervention daughters. This is the inversely related to teacher given remedial support. One possible explanation is that intervention teachers increased their remedial support reducing the need for family members or other students to provide the support.

CARE's intervention appears to practically, but not conclusively, increase formative assessment use and remedial support when considering teacher reported formative assessments use as well as caregiver reported remedial support.

Indicator 3 - ALP in NFE Frameworks

System level - Inclusion of ALP in the national non-formal education frameworks

To measure system-level sustainability of inclusion of disadvantaged children in education, we interviewed ten REOs throughout all project regions. Inclusion of disadvantaged children takes many forms, including ALP for girls transitioning out of primary school, ABE for out-of-school girls, and specific initiatives for GwD

and girls from pastoralist households. We asked REOs about all of the above programmes, their successes and challenges, and the ability of the MOE to sustain these initiatives post-SOMGEP-T.

Unlike the two midline evaluations, where few REOs were aware of ALP or ABE within their region, all REOs we interviewed in the endline had knowledge of ALP implementation within their schools and most had awareness of ABE programmes. Often, REOs would use ALP and ABE almost interchangeably or describe them together with little nuance on the difference between the two programmes. At times, this made it challenging to completely parse the successes and challenges of each individual program.

REOs reported that girls who had graduated from the ALP were now teachers at the ALP, indicating that, despite a lack of awareness from the REOs in previous evaluations, the ALP had been active long before the endline.[109]

REOs cited several specific marketable skills that the ALP taught to girls, including cooking, tailoring, computer skills, and henna.[110] Many REOs viewed ABE positively, with one saying the following:

> *This approach has aided drop-out students in achieving parity with their peers. Second, this effort aided vulnerable students who were unable to pay their school fees and facilities, including male and female students. Finally, it provided encouragement to the teachers who had accepted renumeration as part of this effort.*
>
> -KII with MOE Official, Int. 702

In fact, REOs who spoke on renumeration said that teachers who worked in ALP and ABE were paid for the extra hours.[111]

However, REOs in all regions did indicate some barriers to the full success of both ALP and ABE. At least some of the ALP and ABE in schools were suspended or delayed during the first waves of the COVID-19 pandemic.[112] In addition, a REO pointed out that the ALP classes were supposed to be held in the afternoon, when many students attended Islamic school, leading four schools in their region to stop ALP due to low attendance.[113]

While ALP and ABE are not specifically intended for GwD or girls from pastoralist households, non-formal education initiatives and alternate learning experiences for those girls are also often needed. A REO said that ALP was used for GwD: "[Those with disabilities] have dropped out of school, and there is no school for the disabled people here. They will adapt ALP education for the students who have dropped out of school or are not able to continue due to their circumstances".[114] Often, REOs reported that there were separate schools for GwD but at least one reported that there were no separate facilities in their region for GwD.[115] Several REOs mentioned that community awareness had been done to reduce stigma around disability.[116] In addition,

[109] KII with MOE Official, Int. 707; KII with MOE Official, Int. 702.
[110] KII with MOE Official, Int. 705; KII with MOE Official, Int. 702.
[111] KII with MOE Official, Int. 704; KII with MOE Official, Int. 702; KII with MOE Official, Int. 703.
[112] KII with MOE Official, Int. 701; KII with MOE Official, Int. 708.
[113] KII with MOE Official, Int. 706.
[114] KII with MOE Official, Int. 707.
[115] KII with MOE Official, Int. 702; KII with MOE Official, Int. 710; KII with MOE Official, Int. 701.
[116] KII with MOE Official, Int. 707.

some REOs described teacher training, either on stigma reduction, or on specific assistance for individuals with disabilities (such as sign language).[117] Despite these efforts, large barriers still face GwD:

> *We do not have qualified teachers to teach pupils with disabilities. We include disabled children in our classes, but our goal is to educate every disabled child in every area.*
>
> -KII with MOE Official, Int. 706

> *Except the deaf children we don't have yet supported any other disabled people and this is because we don't have the materials needed to support people with disabilities.*
>
> -KII with MOE Official, Int. 710

Some systems have been shown to work successfully for both GwD and girls from pastoralist households. A REO reported that their region's schools had implemented an attendance registration system to improve the attendance of GwD and pastoralist girls. The system tracks attendance and if a girl is reported absent, the teachers and the school principal are in charge of contacting parents to encourage them to send the girl to school. The REO reported that this system has greatly improved attendance in their schools, even for girls from pastoralist households and GwD.[118]

While the REO above stated that the registration system worked for girls from pastoralist households in their region, many other REOs said the primary barrier for pastoralists is consistent access to education because of the need to migrate for water and pasture:

> *When there is a drought, the pastoralist and rural populations in this region will all shift to another site in search of pasture and water for their animals, and you can see the detrimental effects of this. As students move with their families, this has an impact on their school attendance.*
>
> -KII with MOE Official, Int. 706

Several REOs mentioned solutions for pastoralist households, including limiting the requirements needed for enrolment and utilizing mobile classrooms:

> *We encourage the pastoralist community to enrol their children in the nearest school to where they are moving without any additional requirements.*
>
> -KII with MOE Official, Int. 702.

> *There is a team who has everything that the children need and can move without the children dropping out. The temporary schools are built where the nomads reside mostly.*
>
> -KII with MOE Official, Int. 707.

[117] KII with MOE Official, Int. 710; KII with MOE Official, Int. 702.
[118] KII with MOE Official, Int. 702.

However, those REOs with concrete examples of working with pastoralist households were few and most REOs stated that their region did not provide additional services for pastoralist households. Some REOs stated that they, or others, had visited pastoralist households and encouraged them to not let their children drop out of school.

A REO said that the MOE is planning to start mobile classrooms for pastoralist families and is currently training teachers to run this program.[119] In addition, the same REO said that, during COVID-19, the MOE provided students with radios to continue learning from their homes – a potential approach to better reach and accommodate pastoralist households. Another REO said that the MOE would like to start a boarding school specifically for pastoralist children so that they can have continuity in their education.[120] REOs identified funding as the primary barrier to implementing these programmes or others that would target disadvantaged children. Several REOs also expressed some distress at SOMGEP-T ending and at least one asked for the programme to be expanded into other schools soon.[121]

Indicator 4 - CEC Functionality

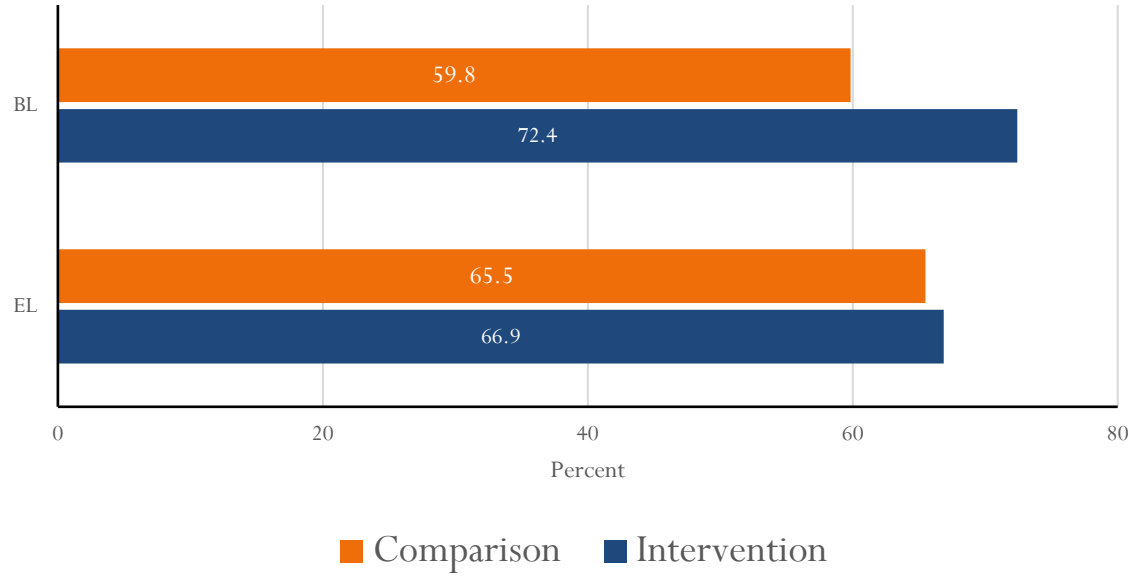
Community Level - CEC Functionality and Activity Levels

This community-level sustainability indicator focuses on the extent to which CECs are perceived as functional by parents and community members. As mentioned previously, CECs are fundamental in ensuring the sustainability of education efforts given they provide financial support, help manage the school, and serve as an intermediary between the MoE and the community. Parents should be aware of the activities of CECs given their involvement in school oversight and management.

First, we asked parents whether their child's school has a "CEC that helps with school-related matters". This indicator does not provide a sense of how active the CEC is, but whether parents are aware of a CEC within their school.

[119] KII with MOE Official, Int. 702.
[120] KII with MOE Official, Int. 705.
[121] KII with MOE Official, Int. 707.

FIGURE 31: SHARE OF SCHOOLS WITH CEC



Upon a comparison between baseline and endline, it appears that CECs within programme schools declined during the programme period by 5.5 percent while CECs in comparison schools increased 5.7 percent. At endline, the proportion of intervention and comparison schools that had CECs were almost identical. Head teachers are likely to be a better informant on CEC activities, particularly as they are often members of the CEC. Starting in ML1, head teachers were asked if their school had a functioning CEC in place. We compared the household and head teacher results in the below table. Please note, that because the head teachers were not asked this question at baseline, we compared the results of the ML2 surveys to the EL surveys for both households and head teachers.

TABLE 49: IMPACT OF PROGRAMME ON CEC FUNCTIONING, PER CAREGIVERS AND HEAD TEACHERS

	Intervention			Comparison			Diff. in Diff.
Indicator	ML2	EL	Diff.	ML2	EL	Diff.	
School has functioning CEC - HH survey	66.0%	70.6%	+4.6	56.1%	68.9%	+12.8	-8.2
School has functioning CEC - Head teacher survey	100.0%	94.6%	-5.4	93.6%	93.6%	0	-5.4

It is clear that head teachers are much more likely to report the existence of a CEC at their school in both intervention and comparison schools. It should be noted that head teachers are aware that the programme

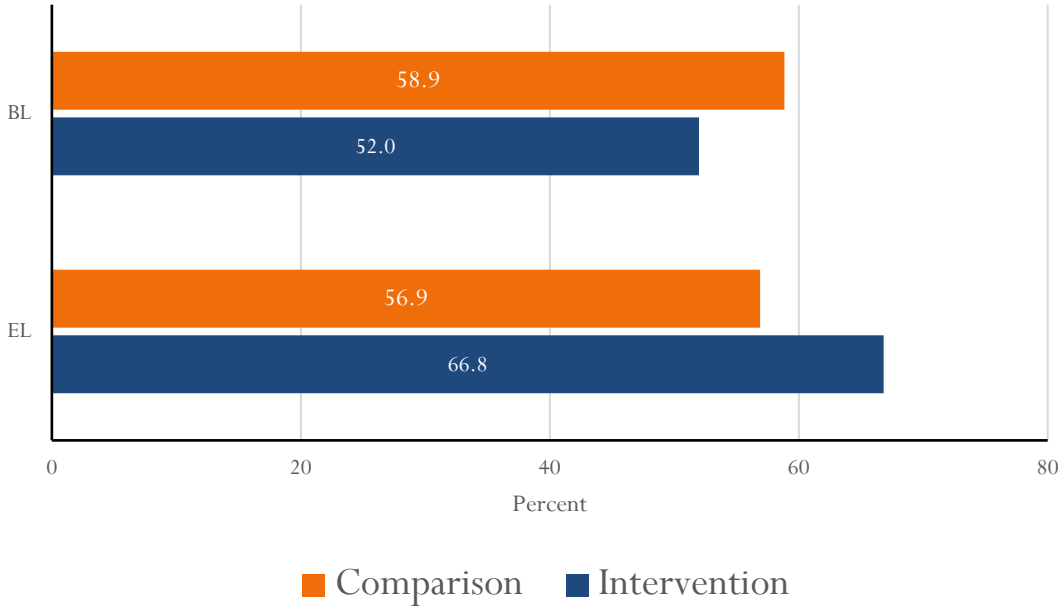
outcomes involve CECs and thus may be more likely to report their existence (social desirability bias). When looking at the change between ML2 and the endline, all households are more likely to report a CEC at the endline. It may be that CEC levels had fallen, or at least, community awareness of CECs had fallen throughout the programme period and had recently become more active; perhaps as a result of COVID-19 and their role in re-enrolling children once schools reopened. This is supported in the qualitative data, which suggests CECs played a pivotal role in bringing girls back to school and following up with those who did not return as expected.

> *Yes, as a committee, when the school reopened we made an effort to bring back the absent students, and the result was that they returned to school, and these actions had a massive effect, which helped both boys and girls.*
>
> -FGD with CEC Members, Int. 103

Although they provide insight into whether CECs exist in target areas, the above measures do not accurately reflect the specific activities of CECs. Below, we have analysed measures based on specific activities that are expected of a CEC.

For those households that reported a CEC in their schools, we also asked them how frequently the CEC communicated with parents. In this instance, the programme's CECs had a marked increase from BL to endline (14.8 percent) compared to a decrease in the comparison CECs (-2 percent).

FIGURE 32: CEC COMMUNICATION (MONTHLY OR WEEKLY)



These findings are supported by the findings in the qualitative interviews, where mothers report that the CECs serve as a key link between parents and teachers/school administration. CECs primarily interact with parents during awareness raising, in monitoring student attendance, and when monitoring conflict between students and challenges between teachers and students. CECs are also involved in the selection of vulnerable

community members for support, a role which involves becoming intimately familiar with the needs of all community members.

> *The council connects parents and teachers. They also solve disputes between them. Furthermore, they link agencies and parents.*
>
> -FGD with Mothers, Int. 308
>
> *This committee is in touch with the families of the girls attending school. Therefore, there are girls whose education is paid for by the CEC after learning of their family circumstances.*
>
> -FGD with CEC Members, Int. 108

In addition, we asked head teachers whether their CEC had a school management plan and whether the CEC had a monitoring visit within the last year. In both intervention and comparison schools, these indicators increased dramatically between the BL and endline. While the difference-in-differences model shows a negative programme effect, it is important to note that intervention schools reported much higher agreement in the BL and maintained an overall higher level of agreement in the endline. In other words, it would have been challenging to have an overall positive programme effect because of the already higher rates in intervention schools. That being said, at endline, only 64.9% of head teachers at intervention schools reported a school management plan and 78.4% reported a monitoring visit within the past year so there was clearly more room for growth on both of these measures. It is also important to note that the 56.1 percentage point increase from baseline to endline in head teachers reporting a monitoring visit by CEC members in comparison schools reflects the trainings conducted by the Education Sector Programme Implementation Grant (ESPIG), a programme funded by the Global Partnership for Education (GPE). CEC training has been a critical component of these ESPIG programming since 2019.

TABLE 50: IMPACT OF PROGRAMME ON MEASURES OF SCHOOL MANAGEMENT

	Intervention			Comparison			Diff. in Diff.
Indicator	**BL**	**EL**	**Diff.**	**BL**	**EL**	**Diff.**	
School management plan	45.9%	64.9%	+18.9	12.9%	54.8%	+41.9	-23.0
Monitoring visits by CEC	64.9%	78.4%	+13.5	22.6%	74.2%	+51.6	-38.1

Whereas at the time of the midline, many CECs were playing a passive role in engaging with schools and parents, the qualitative results suggest that CECs have become more proactive due to their involvement with the program. As one CEC member explains, "In the old days this committee used to come to school only

when there were problems, but now every week someone from within the committee comes and evaluates what has been improved and what hasn't."[122]

CECs are taking an active role in coordinating with religious leaders, monitoring the attendance of both students and teachers, monitoring teacher/student relations and curbing corporal punishment, establishing and enforcing codes of conduct at schools, reporting issues that need to be escalated to the MoE, and resolving conflicts. Whereas in the midline, REOs were calling for additional training of CECs, in the endline, the majority noted significant positive changes in CECs, mostly as a result of the training they have received. The CECs the REOs describe today contain many of the characteristics of an "ideal" CEC, one that is "responsible for the supervision of teachers,"[123] "encourages the parents in the village to educate their children,"[124] "knows the needs of the school,"[125] and more.

> *Previously, the CEC did not function properly. They are now coordinating events between professors and students on a regular basis. As a result, this has aided in the encouraging of girls' education, and their activities have increased during the last four years."*
>
> -KII with REO, Int. 704

> *Yes, before they had been the only board in the school but when they were trained they did a lot of work such as monitoring the teachers as well as raising awareness for the parents at their homes to educate their children, both boys and girls. They have also reported people with disabilities."*
>
> -KII with REO, Int. 710

> *The CEC was not aware how to work four years ago, but later on they have understood the way they need to work with the schools, which they are doing coordination for the teachers both morning and afternoon. They also contribute some distributions in the school and they work to make good communication between the teachers and the parents of the students. They also coordinate the community to encourage the students school enrolment."*
>
> -KII with REO, Int. 707

[122] FGD with CEC Members, Int. 108.
[123] KII with REOs, Int. 710.
[124] KII with REOs, Int. 710.
[125] KII with REOs, Int. 707.

Indicator 5 – MOE Support for Girls' Education

System level - Ministries of Education Engagement in Girls' Education Initiatives

Long-term sustainability of SOMGEP-T's gains largely depend on the ability of the MOEs to implement and maintain girls' education initiatives. SOMGEP-T provided support to the MOEs to improve sustainability in the following four areas: (1) Strengthening the capacity of Gender Departments to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points, particularly in rural areas, (2) Providing support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/transition issues (3) Working closely with MoEs' TVET/NFE Units to explore opportunities for vocational training, and (4) Advocating for employment of female pre-service graduates in target schools.

We interviewed REOs in the baseline, two midlines, and now, the endline evaluation. In the baseline, REOs reported varied performance, with inconsistent financial support to teachers and staff and few REOs reporting Gender Units within the MOE. While we reported improved results in both midline evaluations, there was still regional inconsistency and almost complete dependence on external assistance. In the endline evaluation, there was once again regional inconsistency, even, at times, within regions, and mixed findings with regards to the ability of the MOE to sustain programmes started in SOMGEP-T.

Unlike the previous midline, conducted in 2019, where some REOs regularly reported negative perceptions about CARE and SOMGEP-T with regards to communication and funding, feedback specific to SOMGEP-T in the endline was largely positive:

> *"SOMGEP-T has increased the number of students enrolled in the school, the training of the CEC, and finally, the understanding of the community for the importance of education".*
>
> -KII with MOE Official, Int. 706

Some REOs requested that SOMGEP-T be repeated and be expanded to other schools because of the improvements they have seen over the project period. There were few mentions of the communication or financial gaps discussed in the previous midline, but school fees were a common complaint. Some REOs said that girls were forced to drop out recently because they could no longer afford the school fees. A REO supported this, saying: "The [school] fees for the girls encouraged the other girls who were not studying in school to come in because the parents could not afford to pay the fees so the parents brought their daughters to the school. These were the effects; now the project is over and everything is gone".[126] This was not universal as other REOs did not identify the stop of SOMGEP-T as a factor in girls dropping out of school or say that girls were dropping out due to school fees. More often, REOs said that girls were dropping out to get married, although that may also be for financial reasons. REOs did identify numerous challenges to sustaining SOMGEP-T's progress including current financial shortages at the MOE (this being the most common obstacle), the COVID-19 pandemic, and a lack of staff and transportation to implement work. Despite these obstacles, REOs reported that they were able to perform their duties including supporting re-enrolment of OOS girls, providing training to teachers and to CECs, engaging in community awareness raising, particularly

[126] KII with MOE Official, Int. 707.

about the importance of girls' education, monitoring school progress and attendance of girls, and providing support to increase the number of female teachers in schools.

Strengthening the Capacity of Gender Departments

All except one REO were aware of the Gender Department within their MOE. However, those who were able to speak about their Gender Departments almost universally said the Gender Department has since lost significant funding, been completely defunded, or been absorbed into another part of the Ministry. There was inconsistent reporting among REOs as to whether their Gender Department staff were still working without pay or whether they had stopped working or been reassigned. A REO reported that a Gender Department representative had stopped working due to lack of funds.

During the period where the Gender Departments were operational, they provided varied services to schools. Most often, REOs said that Gender Department staff were focused on female teacher training and re-enrolment of OOS girls. One REO said that Gender Department staff were focused on protection measures for girls at school, specifically rape and gender-based violence.

> *The Gender Department is one of the most important departments in the ministry, and they do a lot to improve the education of girls and during the project continued the quality [of education] and number of girls [enrolled] increased significantly.*
>
> -KII with MOE Official, Int. 709

REOs also credited Gender Departments for the increase in female teachers across all regions. A REO specifically highlighted the Gender Department's role in increasing the number of female teachers: "[The increase in female teachers] was caused by the collaboration of departments, school administration, and [REOs]… as well as the gender department who worked specifically with us… [The Gender Department] has done a lot for this region including having 90 female teachers for 2 consecutive years in 2018 and 2019".[127]

Gender Departments may also have increased the overall awareness of REOs and other staff in the MOE of challenges specific to girls' education. Certainly, the REOs were able to speak more clearly about the barriers that faced girls, including GwD and girls from pastoralist households, in this evaluation round than in previous rounds. On one hand, this suggests that – combined with the increase in female teachers – the effects of the Gender Department are likely to continue beyond the programme, because awareness among MOE staff outside the Gender Department will remain within the MOE and changes in their awareness may translate into more gender-sensitive policy and implementation of policy.

On the other hand, there are also reports of fiscal difficulties affecting Gender Departments, including downsizing, which likely reflects a failure to prioritise the department and its work in an environment of resource scarcity. Gender departments in some locations do not appear to have sustainable funding, as one REO said, "it is closed these days due to lack of support".[128] They are also not paying staff salaries. One MOE official said their gender department needs additional support but "exists so far", while another stated that the gender department head was previously not paid but is now paid a salary. Finally, one REO noted that they

[127] KII with MOE Official, Int. 710.
[128] KII with MOE Official, Int. 710

need more support but they are able to continue with their current level of funding.[129] Some MOE officials indicated that their gender departments do not have sustainable funding but report that gender departments currently exist, despite reports – from one REO – that funding has ended.[130]

Improving Teaching Practices and Retention/Transition

Improvement in teaching practices and retention was tackled through a number of different interventions discussed in the interviews, including but not limited to: 1) teacher training, 2) increased parent and community communication, 3) monitoring of the schools, and 4) exams. REOs had overall positive remarks about the improvement in teaching practices and more mixed comments on the ability of the MOE, school, and community to retain girls in school.

Almost all REOs said that teaching quality had improved during the project period. An REO exemplified this when he described how teachers' styles had changed:

> *In the past, only the teacher would stand there and student should only listen but now the student must be involved in the lesson in which the teacher can ask questions and make suggestions… So we changed a lot of the ways of teaching.*
>
> -KII with MOE Official, Int. 703

The improvement in teaching quality was largely attributed to training by either SOMGEP-T or the MOE. Most REOs indicated that the teachers in their region had received training. A REO stated that untrained teachers were referred to the MOE to receive training.[131] Trained teachers were more likely to plan lessons in advance and arrive on time; the same REO said that attendance has increased as a result of improved teaching quality.[132] Some REOs also said MOE staff received training, including on disability, teaching quality, and corporal punishment. On the one hand, the REO reported connection between access to trainings and teacher behaviours such as improved attendance and preparedness is a positive finding. However, its translation into sustainability of outcomes is unclear, potentially suggesting that dips in resources or programme funding may just as quickly translate into worsening teaching performance or practices.

Awareness raising among both parents and the larger community was another effective tool in improving enrolment, according to participating REOs. Respondents cited parent communication as one of the most effective methods to increase enrolment. Many REOs suggested that one of their most important roles was talking to parents to encourage them to either enrol their children into school or to prevent them from dropping out. As described in the previous section, one REO described an attendance system where parents were contacted after a student absence; the REO claimed that "there were no more absentees at this time".[133] Another REO said that they were responsible for coordinating student attendance at each of their schools.[134] REOs also spent time raising awareness for both parents and community members on girls' right to education. Some did this awareness raising directly and others encouraged the CEC to raise awareness within their

[129] KII with MOE Official, Int. 709; KII with MOE Official, Int. 707; KII with MOE Official, Int. 708
[130] KII with MOE Official, Int. 702; KII with MOE Official, Int. 702
[131] KII with MOE Official, Int. 705
[132] KII with MOE Official, Int. 705.
[133] KII with MOE Official, Int. 702.
[134] KII with MOE Official, Int. 701.

community. Several also mentioned raising awareness about the importance of enrolling children with disabilities in schools and encouraging students to not discriminate against a classmate with disabilities.

With REOs reportedly taking an active role in and responsibility over issues such as attendance and retention, the sustainability of improvements in this area appears positive. While the described methods – including individual outreach efforts and awareness raising – are time intensive, they are less resource intensive than – for example – teacher training. REOs can – and appear to be motivated to – effectively take on these responsibilities with the infrastructure and resources available.

While REOs did play a large role in both improving teacher quality and raising awareness, REOs often saw their primary job function as monitoring schools: "My job is to monitor and find out the real situation of education, and also to report it to the administration of the Ministry of Education to deal with it".[135] The focus of monitoring differed between REOs: some reported monitoring CECs, others focused on schools and teachers, and a few said they monitored student progress. Many REOs felt that teacher work improved when they knew that REOs or other staff were monitoring their work. Others also looked at monitoring as a type of needs assessment:

> *If the schools are not supervised and their educational situation is not monitored, their needs will not be implemented and not much be achieved.*
>
> -KII with MOE Official, Int. 709

Teaching quality and the resulting student progress was also measured through student exams. All REOs indicated that either monthly or weekly exams took place in schools to assess student progress and be able to quickly identify which students may need additional support. At least one REO stated that the exams were designed by the teachers and that the MOE did not oversee the content of the exams, leading to low comparability and inability to assess standards across the region.[136] The exams may be especially critical for ABE, where at least one ABE uses a weekly exam to transfer students to a "level that is on par with his educational test".[137]

Supporting Vocational Training

Only one REO specifically mentioned vocational training in the interviews, speaking positively about the increase in training through SOMGEP-T. SOMGEP-T did not offer vocational training instead connecting graduates with opportunities. As described above, REOs mentioned henna, cooking, computer skills, and tailoring in the context of ALP, ABE, or non-formal education, a sign that graduates are pursuing opportunities sourced but not offered through SOMGEP-T. One REO reported that 40 women had graduated from the non-formal education programme set up in their region. There were no specific questions about vocational training in the research tool and so these programmes may have been ongoing but not captured in the interviews.

Advocating for Post-Graduation Female Employment

[135] KII with MOE Official, Int. 705.
[136] KII with MOE Official, Int. 710.
[137] KII with MOE Official, Int. 702.

Unlike the two previous midline evaluations, the hiring of female teachers was thoroughly explored in the endline. While partially discussed above, as part of the role of the Gender Departments, most REOs reported that the number of female teachers in their region has increased dramatically – potentially leading to better outcomes for girls as they have aspirational role models.

> *The change happened when the support of girls was established and now the teachers who are working in the region are female teachers and it can be said that it has doubled as before.*
>
> -KII with MOE Official, Int. 707

Multiple REOs also commented that the teaching quality of female instructors was higher than that of male instructors.

> *Female teachers are the one of the best teachers to convey the lessons well. More female teachers are being hired by the Ministry of Education followed by agencies because they want to improve the quality of teacher education. This change is due to the fact that women have become more aware of their role and role in contributing to the development of social education, all this contributed to the growing number of female teachers.*
>
> -KII with MOE Official, Int. 709

However, REOs in other locations had more inconsistent views about female teachers. At least one REO said that there had not been any encouragement from either the MOE or SOMGEP-T to hire more female teachers and that there was a lack of female graduates in the teaching field.[138] Contrarily, another REO said that their region required schools with more than 20 children to have at least one female teacher and that there were at least six schools in their region with female principals.[139] Some REOs also said that female teachers were more common in the ALP and ABE programmes as graduates of those programmes were often hired to be teachers. REOs across zones identified motherhood as a barrier to becoming a teacher: "There were no female teachers in the last four years who have applied. It is also possible that women are busy with the life and raising of children".[140] Just as marriage was described earlier as one of the primary reasons for girls to drop out of school, it also appears to be a barrier in recruiting more female teachers. This has a cyclical effect on girls' enrolment as female teachers can act as role models to young girls, demonstrating how to utilize their education and have a successful career either by delaying marriage or by balancing marriage, motherhood, and a job successfully.

While overall results were mixed, and it was clear that certain programme areas had slowed or stopped post-SOMGEP-T, it appeared that REOs in all regions were largely continuing their role in maintaining school and teaching quality. Despite the defunding of Gender Departments across all regions, the social awareness brought by the Gender Department's activities was present in many of the comments that the REOs made. REOs also intended to continue monitoring schools and improving teacher quality through MOE trainings and in-classroom support. Teacher trainings may be less likely in the future as they were often supported by SOMGEP-T or other international agencies. While information on vocational training was limited, the little

[138] KII with MOE Official, Int. 704
[139] KII with MOE Official, Int. 703.
[140] KII with MOE Official, Int. 709.

information gathered indicated that non-formal education and vocational training had allowed some girls to pursue certain skill-based careers. Lastly, almost all REOs reported an increase in female teachers in their regions, a change that is hopefully sustained for years to come.

Overall Findings

Indicator 1

Results for indicator 1 suggest external factors such as drought and the COVID-19 pandemic have had a negative impact on the ability of CECs to provide support to schools over the period from midline to endline. The trends that emerged from the quantitative and qualitative results on this indicator diverged. From baseline to endline, CEC support fell from 31.6 to 16.7 percent, scholarship support for girls in intervention schools decreased by 9.38 percent, and CEC share in scholarship support decreased in all schools. However, the qualitative results suggest communities and diaspora have increasingly engaged with CECs to meet community-level education needs over the life of the programme but may simply be unable to provide their usual level of support at this time. There were also conflicting results between interview groups. The results of the head teacher survey show the programme was negatively correlated with girls receiving scholarships and CECs raising funds for school improvement, whereas the survey with parents shows the programme was positively correlated with increases in raising funding, improving infrastructure, and providing financial support. The qualitative data provides support for both conclusions in that CECs appear to be more heavily engaged with schools but may currently be focusing more heavily on awareness raising and school monitoring activities while their communities and usual sources of support face financial constraints.

Indicator 2

The results for indicator 2 are not conclusive and show mixed results across groups and sub-indicators. The programme appears to have increased the use of formative assessments, as teachers' reported use of formative assessments increased in intervention schools from baseline to endline over and above comparison schools. Although records of formative assessments decreased from baseline to endline, this may be due either to 1) enumerators and teachers having a better understanding of what a formative assessment actually is by the time of the endline, or 2) intervention teachers not keeping records of formative assessments because they are using them specifically to give remedial support.

Remedial support results varied by group. CEC remedial support remained low from baseline to endline but increased slightly for girls from both intervention and comparison schools. On the other hand, remedial support for girls in the last 12 months, as reported by caregivers, decreased for intervention schools. However, teachers at intervention schools did provide more remedial support over teachers from comparison schools, which is the result we would expect from the teacher trainings. Although remedial support given by other students and family members to caregiver daughters increased more for girls in comparison schools than girls in intervention schools, this may be due to the increase in teacher support just mentioned.

Indicator 3

Awareness of ALP and ABE has clearly increased drastically from midline (when they were established) to endline, as all REOs interviewed in the endline had knowledge of alternative education implementation within the schools in their areas. Opinions on the effectiveness of the programs was largely positive, and REOs reported instances of participating girls becoming teachers and learning skills that make them highly marketable. There were some barriers to success faced during the endline period, including COVID-19 closures and issues with low attendance.

Despite a clear increase in awareness among REOs of the issues facing pastoralists and GwDs, REOs appear to feel that the ALP and ABE programs are not necessarily able to consistently target girls from each group. REOs mentioned efforts to engage with both groups, but it was not always clear how these efforts were or were not related to the ALP and ABE programs, and REOs also remained vocal about the numerous challenges they face in engaging GwDs and pastoralist girls. Regardless, this increased awareness of alternative education as an option for girls with unique challenges and the increase in efforts to engage with GwDs and pastoralist girls represents a notable improvement from baseline to endline.

Indicator 4

The results for indicator 4 suggest that while CEC levels have not changed drastically, the quality of CEC engagement has improved as a result of the programme. According to the results from the household survey, the existence of a CEC has increased slightly over the life of the programme, and according to the results from the head teacher survey, it has decreased slightly. It should be noted, however, that head teachers were much more likely overall to report the existence of a CEC, suggesting community awareness of CECs may have fallen during school closures (as a result of the pandemic) but CECs perhaps recently became more active when schools re-opened.

The quality of CEC engagement has clearly improved over the life of the programme. There has been a marked increase from baseline to endline in CEC communication with parents, and CECs are also increasingly forming school management plans and monitoring schools. The qualitative results suggest CECs also raise awareness, monitor attendance, handle conflicts that arise between students and between students and teachers, and take an active role in identifying and following up with vulnerable children in the community.

Indicator 5

Although the results of the endline confirmed that Gender Departments were formed as a result of the program, they also brought into question the sustainability of activities related to strengthening gender departments because almost all had lost significant funding, been completely defunded, or been absorbed into another part of the ministry by the time of the endline. While they were operational, they focused most heavily on female teacher training and enrolment of OOS girls, activities which – at least according to the REOs interviewed – had a positive effect on targeted communities.

On teaching quality, the widespread belief among REOs is that teaching quality has drastically improved as a direct result of training provided through the programme. Some positive effects REOs cited include improved teacher and student attendance, improvements in teaching methods, and improved teacher preparedness. REOs have also taken on a more active role in attendance and retention through their school monitoring visits and community awareness raising.

Awareness of TVET efforts was minimal. This is not entirely unexpected given the program connected graduates with vocational opportunities but did not provide them directly.

There has reportedly been a dramatic increase in the number of female teachers as a result of the program, though there are still some challenges REOs cited to hiring female teachers, such as motherhood.

Sustainability - Project Response

CARE agrees with the analysis provided by the external evaluator. It is clear that many gains have been realized at school level, both in terms of CEC functionality and teacher capacity. On the other hand, the gains

at system level were far more limited: changes in practice at DEO and REO levels mainly in school supervision, Gender Focal Points support to girls' forums and coaching of teachers. The impact of COVID-19 on state revenue means that in the medium term, MOE budgets will not be able to fully absorb Gender Units and the frequent school visits to support teachers that have been supported by the project.

Still, some 'legacy effects' at system level may have a longer-term impact. SOMGEP-T has not only trained CECs, but redesigned CARE's approach to CEC training. The SOMGEP-T CEC training model was scaled up through three system strengthening projects (Education Is Light/ *Waxbarashadu Waa Iiftin* and *Horumarinta Elmiga*, funded by the EU; and the Global Partnership for Education-funded Education Sector Program Implementation Grant/ESPIG, at Federal level). Furthermore, SOMGEP/SOMGEP-T's results demonstrating the impact of CECs have contributed to their prioritization in sector planning.

It is likely that Gender Units/ Focal Points and their activities will continue to be dependent on funding from development partners (for instance, through the Global Affairs Canada-funded GEEPS program), and may not continue to exist as independent units within Ministries. Their legacy remains through the increased prioritization of gender-focused programming in sector plans; the enhanced awareness of gender-related issues among education officials; and mostly through the increased gender capacity at regional and district levels.

One of the most critical contributions of SOMGEP-T was the increased decentralised capacity at regional and district levels. The implementation of activities such as Girls' Empowerment Forums, teacher coaching, remedial learning, ALP and ABE in partnership with regional and district officials has contributed to increased awareness of needs, support for gender-related issues (including but not limited to the need for female teachers) and enhanced capacity to coach teachers and CECs to support the ultra-marginalised. These shifts in perceptions and practices are contributing to the success of other initiatives, most notably system-led CEC trainings.

Most importantly, the shifts at individual girl / household level are likely to contribute to a broader process of social norm change. As discussed above, the participation in GEFs was one of the most important factors associated to positive impact on learning. Qualitative data indicates that this impact was not only observed at the individual girl level, but also among their families and communities, as girls sought to pass along the skills acquired, and engaged in multiple forms of girl-led action in their villages. The girls' own narratives of impact, shared through a network tool, highlight their increased capacity to seek and provide help; to work collectively in initiatives; and to dare to take new roles at the classroom and household. This less tangible, but powerful legacy of the GEFs is likely to be one of the most sustainable impacts of SOMGEP-T. The GEF model in itself and the demonstration of its return on investment is potentially one of the project's main contributions to the Somali education system.

Intermediate Outcomes

Attendance

Attendance is one of four primary intermediate outcomes targeted by SOMGEP-T programming. School attendance is critical to the achievement of improved learning outcomes for the simple reason that girls who do not attend school consistently will not learn, regardless of what improvements to their school – such as

teaching quality – are made. Attendance not only impacts learning outcomes but also indirectly influences transition rates, because low attendance rates produce students who are *unable* to advance into the next grade level, especially into secondary school, because they are not qualified to do so.

In theory, attendance is a relatively straightforward outcome to assess, because there are objective measures of attendance. In short, a student is either in school or not. Unfortunately, assessment is complicated by several methodological problems, including issues of practical measurement and "selection" into enrolment that can bias analysis of attendance rates.

Our analysis includes three measures of attendance. The first are attendance headcounts conducted in programme schools during site visits by our researchers. These measures are reliable, in terms of determining the number of students present on a particular day, though they are limited to a one-day snapshot of attendance rates. More problematically, they rely on knowledge of enrolment levels in each class, which can be influenced by differential enrolment rates across schools. One component of SOMGEP-T programming has been to promote enrolment, especially of marginalised girls. To the extent that enrolment has increased, it increases the "denominator" in assessments of attendance, which can bias our analysis *against* intervention schools. Consider an example, in which a girl who would not otherwise enrol in school was encouraged by their CEC: in the absence of the intervention, she would not have enrolled; due to the intervention, she enrolled in school but did not attend consistently. To the extent that SOMGEP-T brought more girls into school and these girls had lower than average attendance, this would bias our analysis of how SOMGEP-T affected attendance.

Our second measure is drawn directly from school records, which provides a wider picture of attendance across the entire school year. However, these measures are contingent on accurate enrolment data being maintained by head teachers and accurate attendance data being recorded daily by teachers. In practice, attendance records tend to be incomplete and often disorganised. For instance, 35.3 percent of classrooms we visited for physical headcounts during the endline did not have a record of attendance from the previous day. Similarly, our team leaders classified attendance records as "extremely complete" in just 29.4 percent of classrooms they visited. As the table below shows, the maintenance of enrolment records seems to have improved since the baseline, while the share of classrooms where attendance is taken and recorded consistently has stayed relatively stable.

TABLE 51: QUALITY OF ENROLMENT AND ATTENDANCE RECORD-KEEPING, BY ROUND

	Baseline	Endline
Number of days attendance records completed, out of last 5		
Intervention	3.92	3.90
Comparison	3.96	4.2
No enrolment record for classroom		
Intervention	5.2%	1.2%
Comparison	10.0%	4.4%
No attendance taken or recorded that day		
Intervention	20.2%	17.7%

Comparison	24.3%	24.0%

In addition, school attendance records are subject to the same methodological limitation as headcounts, that promotion of enrolment may actually bias attendance rates downward in programme schools, if children who enrol in response to programme interventions have lower average attendance. Given the generally marginal quality of record-keeping, we consider the analysis of school-based attendance records helpful, but of secondary importance.

Our third measure is derived from interviews with girls' caregivers, in which we ask how many days a girl has missed over the previous month and whether the girl has attended most, more than half, about half, or less than half of school days in the previous year. This measure has the advantage of not being contingent on accurate record-keeping, but relies on accurate recall by caregivers. Ultimately, no measure of attendance is perfect – even direct, daily observations of classroom attendance are impacted by differential enrolment rates. However, by triangulating across multiple measures, this section offers relatively rigorous evidence regarding attendance outcomes in sampled schools.

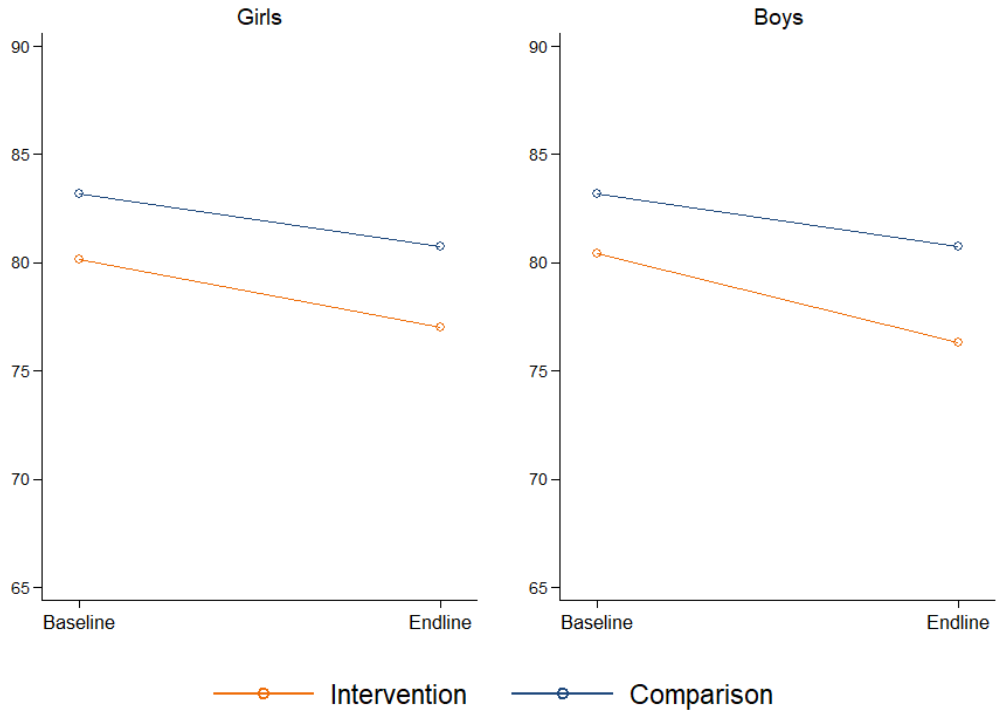
Classroom Headcounts

We start by analysing the results of physical classroom headcounts. In each evaluation round, we attempted to conduct headcounts in every classroom in each school visited. In reality, some schools were closed for examination preparations at the time of our visits, slightly reducing the overall set of headcounts conducted. To illustrate, out of 69 formal primary schools visited at endline, three were not holding classes, preventing headcounts; another three schools entirely lacked enrolment records, rendering it impossible to calculate attendance rates. To guard against bias due to changing sample composition across time, we assess changes in attendance rates among the "panel sample" of schools – schools in which we completed headcounts at both baseline and endline.[141]

In the aggregate, SOMGEP-T programming does not appear to have had any impact on attendance rates among girls since the baseline. We analysed 824 classroom headcounts in the baseline and endline evaluation rounds using difference-in-differences; without controlling for additional factors, such as geography and grade level, our estimate is that girls' attendance fell 0.3 percentage points in response to SOMGEP-T programming. The left panel of the figure below illustrates the trend over time among this sample, with combined attendance rates falling from 81.5 to 78.5 percent from baseline to endline. Intervention schools experienced a very slightly larger decline, though this effect is far from statistically significant.

[141] It is important to note that we cannot construct a precise panel of headcounts across time, insofar as the same schools may have a different number of classes or levels of classes in different rounds. Our analysis is of a panel of schools, rather than a panel of classes, with the idea that using a panel sample helps to control for school-specific factors that drive attendance that might otherwise influence our results.

FIGURE 33: ATTENDANCE RATES, BASELINE TO ENDLINE, BASED ON CLASSROOM HEADCOUNTS



More extensive regression analyses did not change these results appreciably. For instance, accounting for district and the grade level of the classes monitored suggested the programme's yielded a 0.06 point increase in girls' attendance. Controlling for region and weighting observations to ensure each school received the same weight across baseline and endline rounds also did not alter the findings.

Interestingly, boys' attendance also declined overall since baseline, but declined more quickly in intervention schools, as shown in the right panel of the figure above. Across rounds, boys in intervention schools experienced a decline – relative to boys in comparison schools – of between 1.8 and 2.3 percentage points. In other words, our best estimate of the programme's impact on boys' attendance is that it reduced boys' attendance by between 1.8 and 2.3 percentage points, depending on the precise specification of the model.

The table below reports estimates of programme impact for girls and boys. As above, the analysis focuses on changes from baseline to endline, including only schools that were sampled in both rounds.[142]

[142] Excluding the midline rounds from the analysis provides a slightly larger sample size for this analysis, because schools with missing data from the midline rounds would require us to remove some data from the baseline and endline samples to maintain the panel structure of the analysis.

TABLE 52: PROGRAMME IMPACT ON ATTENDANCE RATES

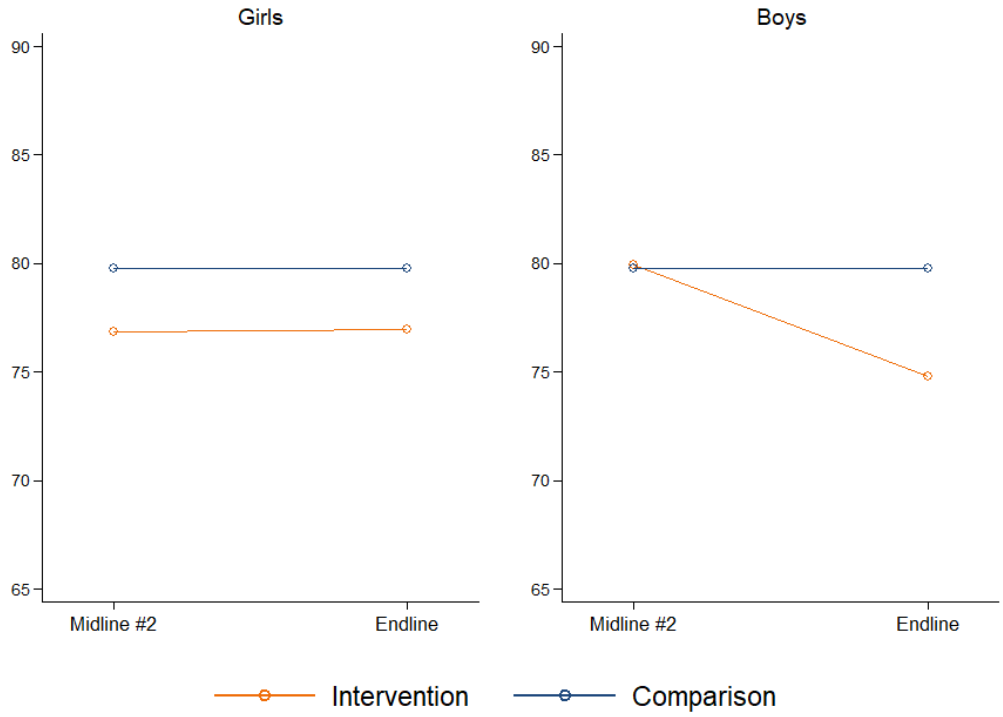
	Girls	Boys
Total	0.06	-1.8

In addition to changes from baseline to endline, we also consider shorter-term shifts in attendance rates, from ML2 to endline. This analysis still covers two years of programming – from late 2019 to late 2021 – though this period includes a time in 2020 when the COVID-19 pandemic resulted in several months of school closures. As before, the sample for this analysis consists of schools in which headcounts were completed at both ML2 and endline.

The figure below shows the trends in attendance rates across the two rounds for intervention and comparison areas, with girls' and boys' attendance reported in the left and right panels, respectively. For girls, the programme has not produced any change in attendance rates over this period, with attendance in intervention schools increasing by just 0.1 percentage points. However, boys' attendance in programme schools has declined precipitously relative to comparison schools, dropping 7.3 points over the same period.[143] It is important to note that falling boys' attendance in programme schools appears to be a retrenchment from gains made during ML1 and ML2 – boys' attendance in intervention schools improved, relative to comparison schools, during these earlier rounds, with these gains erased during the endline round. As noted above, boys' attendance in intervention schools has fallen 2.3 points, relative to comparison schools, since the start of the programme.

[143] This difference is statistically significant at the 5 percent level.

FIGURE 34: ATTENDANCE RATES, ML1 TO ENDLINE, BASED ON CLASSROOM HEADCOUNTS



There are two possible interpretations of the sharp decline in boys' attendance since ML2 and the smaller – but still meaningful – decline since baseline. The first is that the programme has *caused* a decline in boys' attendance, though there is no clear theoretical mechanism by which this could have occurred. A second interpretation is that intervention schools were differentially affected by shocks – COVID-19 and worsening drought in recent months – that hampered school attendance, but that girls' attendance was buttressed through support from SOMGEP-T. Our argument in this case is that, in the absence of the SOMGEP-T intervention, girls' attendance in programme schools would have declined on a similar trajectory as that of boys; in short, the programme arrested a fall in girls' attendance.

The results from physical headcounts are mirrored by those based on teachers' attendance counts. This is not surprising, because official attendance taken by teachers was taken during the same school day in which physical headcounts were completed, so attendance rates tend to be similar.[144] Programme impact is more negative when using teachers' attendance records, though the pattern of impacts are qualitatively similar to those using headcounts: boys' attendance in intervention schools declined at a steep rate relative to

[144] At endline, attendance based on headcounts was 80.2 percent, while attendance based on teacher records *on the same day* was 78.7 percent, among those classrooms in which the teacher had recorded attendance information that day. This gap is almost certainly a function of teachers taking attendance at the start of class or the school day, while our researchers completed headcounts later, allowing late-arriving students to be captured.

comparison schools; girls' attendance in intervention schools also fell from baseline to endline, but less so than boys' attendance.[145]

While we also collected data on teacher-recorded attendance from the previous day, this data is both less reliable and encompasses a smaller sample size. It covers a smaller sample size, empirically, because fewer teachers had attendance information for the day prior to our team's visit than on the day of the visit.[146] It is less reliable because teachers may have retroactively recorded attendance values. Prior-day attendance values are markedly higher – 2.4 points higher for girls and 3.0 points higher for boys – than same-day attendance values recorded by the same teacher, suggesting that some teachers retroactively record attendance and may, speculatively, have inflated values for prior-day attendance because it cannot be directly checked by our researchers. Regardless of the exact mechanism driving higher attendance rates for prior days, the suspect data quality and smaller sample size motivate our decision not to analyse attendance data recorded for the day prior to school visits.

School Attendance Records

School attendance records provide a second option for assessing the impact of SOMGEP-T on attendance. As we noted in the introduction to this section, school attendance records have the advantage of providing a broader picture of attendance, over an entire school year, than headcounts can provide. We collected attendance from school records for only cohort girls, which also means that this analysis is particularly focused on the girls most likely to be impacted by SOMGEP-T programming. Unfortunately, school records are often of poor quality and – in particular – there is significant missing data due to incomplete records.

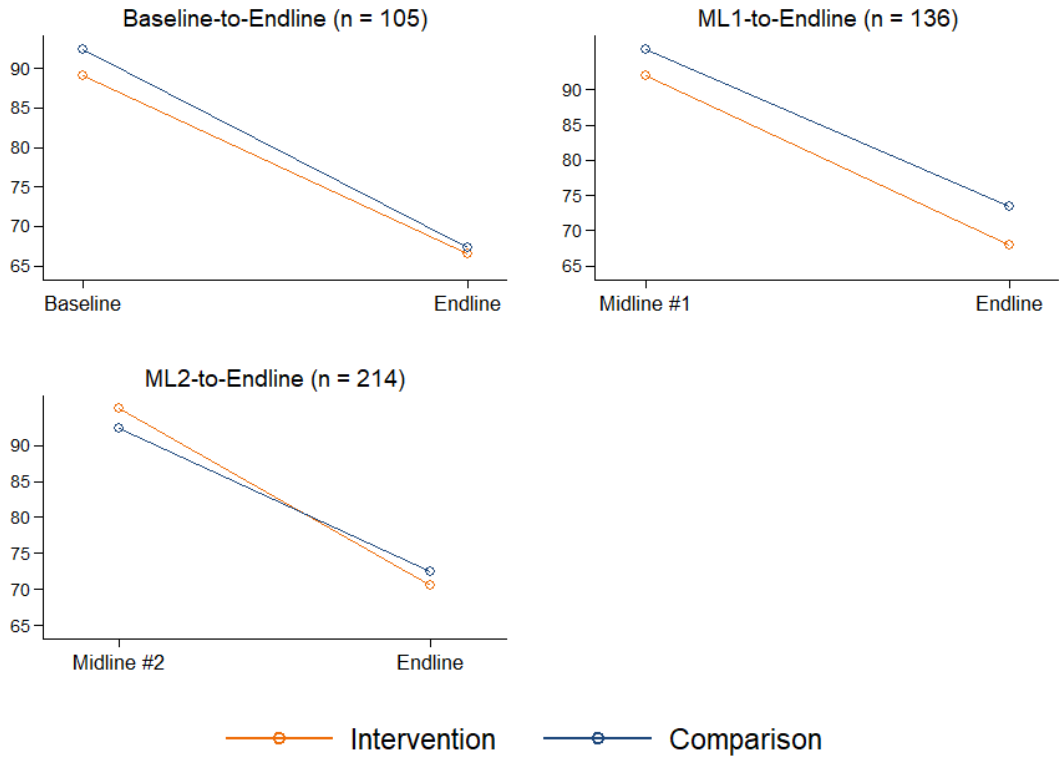
In the figure below, we analyse changes in attendance rates over three different time periods and using three different samples. In each panel of the figure, the sample contains the set of girls for whom there were complete attendance records in both periods – e.g., both ML1 and the endline, in the top-right panel. As the title of each panel makes clear, the sample sizes are often small: for instance, just 105 girls had complete attendance records in both baseline and endline.[147]

[145] While we also collected data on teacher-recorded attendance from the previous day, this data is both less reliable and encompasses a smaller sample size. It covers a smaller sample size, empirically, because fewer teachers had attendance information for the day prior to our team's
on the day prior to our team's visits than on the day of the visit – evidence that the team's visit prompted

[146] This may reflect the impact of our team's visit, with teachers more likely to record attendance because they know a team is arriving to monitor their school. However, it also reflects the fact – documented by team leaders over multiple evaluation rounds – that many teachers record attendance temporarily on a chalkboard at the front of class, erasing the attendance count at the end of the day. As a result, for some classes, attendance for previous days is not recorded anywhere, even if the teacher took attendance.

[147] The small number of girls for baseline-to-endline analysis, in particular, is not surprising. In order to appear in this sample, girls must be successfully tracked and re-contacted across both rounds, must be enrolled in school in both rounds, and must have complete attendance records in both rounds. The intersection of all three characteristics encompasses a small number of girls, especially when analysing change over a longer time period.

FIGURE 35: COHORT GIRL ATTENDANCE RATES, BASED ON SCHOOL ATTENDANCE RECORDS



The results of this analysis are mixed. Among the small set of girls in the baseline-to-endline analysis, attendance rates rose by 2.6 percent for girls in intervention schools, relative to those in comparison schools. Changes from ML1 and ML2, respectively, to the endline were more negative, however: in the ML1-to-endline analysis, the programme produced a -0.9 point change in attendance among intervention girls; in the ML2-to-endline analysis, programme impact was -4.5 points. In the latter case, attendance among 86 comparison girls fell by 19.9 points, while attendance among 128 intervention girls fell by 24.5 points.

While collecting data on school attendance in the latest round, we also compiled attendance records from last year, the 2020-2021 school year. As expected, this information was only available for a subset of girls in the endline evaluation; however, it provides an opportunity to check for changes in attendance rates between the 2020-2021 and current school years, despite the fact that no evaluation was conducted during the 2020-2021 school year. In total, attendance records were sufficiently complete for both years for 151 girls. Overall, attendance rates fell dramatically this year, from an average of 70.4 percent last year to 58.9 percent in the current year. Attendance fell more dramatically among girls in intervention schools, however, broadly consistent with the other findings in this section.

Caregiver Reports

Our third source of quantitative measures of attendance come from girls' caregivers, who were asked to report on their girls' attendance rates during the household survey. Caregivers responded to a series of questions; as with our other data sources, the measures provided by caregivers are imperfect, as they can be

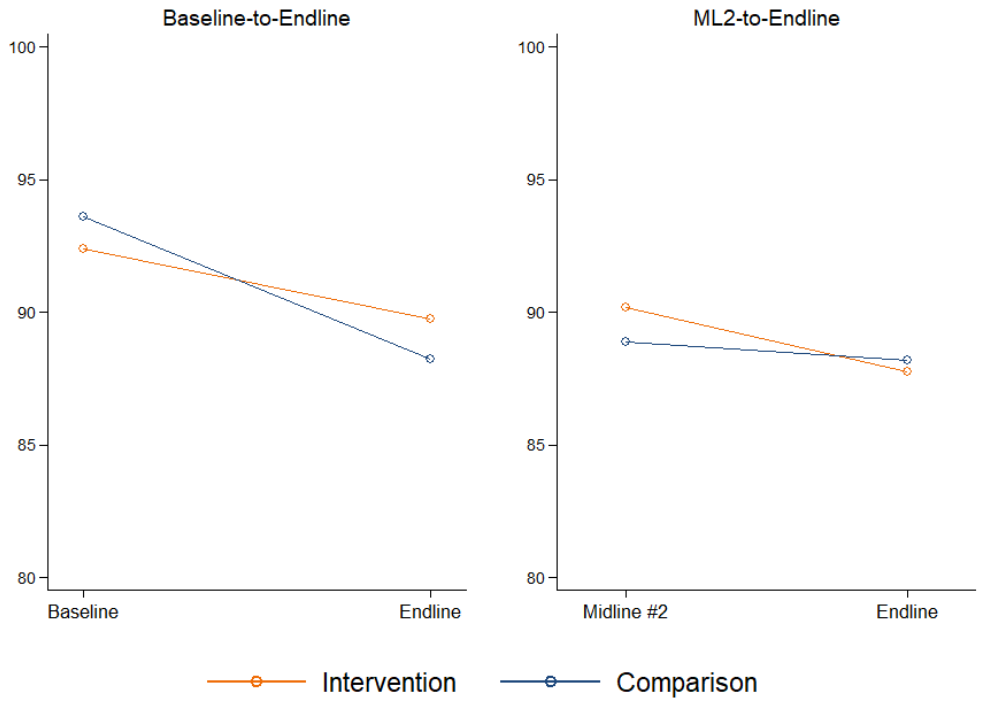
subject to recall bias and caregivers may not be reliable sources, especially when their girls are older.[148] The discussion in this section centres on cohort girls – those who were initially recruited as in-school girls or out-of-school girls – rather than ALP or ABE girls. The results are also limited to those girls who were enrolled at the time of data collection.

The first measure is relatively straightforward: the share of caregivers who reported that their girl attended school "most days" during the school year thus far. Naturally, the sample available for analysis includes only girls who were enrolled in school in each round of data collection. To illustrate, the baseline-to-endline analysis includes only girls who were successfully contacted and were enrolled in school in both baseline and endline. The figure below reports the share of girls who attended school "most days" for our two main analysis samples. The left panel shows that attendance rates, according to this measure, fell in both intervention and comparison schools over the life of the programme, but that attendance rates fell more quickly in comparison schools. Our estimate of programme impact from baseline to endline is 2.7 points; girls were 2.7 points more likely to attend school most days as a result of the SOMGEP-T intervention.

Analysis of changes over a shorter time period – and using a slightly different and larger sample – are less positive. The right panel of the figure above analyses a sample of 646 girls who were successfully tracked from ML2 to endline and who were enrolled in both rounds. In this sample, girls' attendance fell by 1.7 points as a result of the programme. Notably, the qualitatively different estimates of programme impact – positive since baseline, but negative since ML2 – are maintained when we control for additional factors, such as geographic zone and age of the girls in the sample.

[148] Recall bias is relevant because both measures of attendance derived from caregiver reports are based on a long recall period: the previous month and the current school year (approximately four months long at the time of data collection). Survey respondents have notoriously unreliable memories over a long period, especially on topics that are not extremely memorable; individual absences from school are unlikely to be memorable, outside the context of a serious illness. In addition, caregivers may not be entirely aware of their girls' attendance at school. Older girls are likely to be more autonomous and caregivers are less likely to be aware if they missed school, reducing the accuracy of caregiver responses.

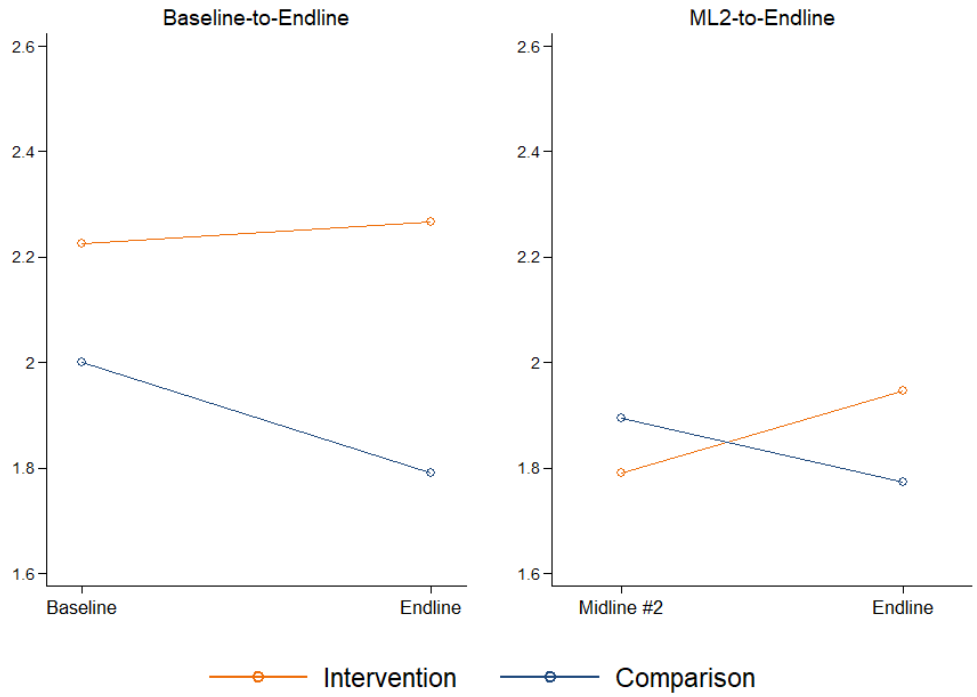
FIGURE 36: SHARE OF STUDENTS ATTENDING SCHOOL MOST DAYS, ACCORDING TO CAREGIVERS



A binary measure of whether girls attended "most days" is somewhat blunt, although even a more nuanced coding – which utilises a follow-up question targeted at caregivers who reported their girl had *not* attended most days – produced very similar results. A secondary measure using caregiver reports also provides a more nuanced metric, by asking caregivers to report the number of days of school their girl missed in the last month. Across all rounds of data collection, cohort girls missed an average of 2.34 school-days in the month prior to being surveyed. In the aggregate, attendance improved across rounds, though this may be a function of girls with a low propensity for attendance selecting out of enrolment. In the endline, 55.7 percent of caregivers reported that their daughter had not missed a single day in the past month, while 6.9 percent had missed 10 or more days over the same time period.

The figure below shows trends in the number of days missed from baseline to endline (left panel) and from ML2 to endline (right panel). Over both time periods, the mean days missed by girls in intervention schools rose, while simultaneously falling in comparison schools. From baseline to endline, girls in comparison schools shifted from missing a mean of 2.0 days to 1.79 days per month. In contrast, girls in intervention schools went from missing 2.23 days per month to 2.27 days. Findings over the period ML2 to endline are substantively similar: absenteeism among girls in intervention schools rose while falling among girls in comparison schools. Over this shorter period, our best estimate suggests that the programme reduced girls' attendance by 0.28 days per month from ML2 to endline, though this result does not approach conventional standards for statistical significance.

FIGURE 37: DAYS OF SCHOOL MISSED IN PREVIOUS MONTH, ACCORDING TO CAREGIVERS



While these findings are subject to caveats surrounding recall and reporting accuracy, they are noteworthy, in part, because they are drawn from a relatively large sample. The baseline-to-endline analysis is based on 382 girls tracked across the two rounds, while the ML2-to-endline sample utilises two rounds of data, each, from 595 girls. The findings are also robust to approaches that account for the exponential distribution of the outcome variable. The number of days missed is heavily right-skewed, with most girls missing 3 or fewer days of school, but a long right tail, with some girls missing 20+ days of school.[149] To ensure a small number of outliers were not driving our results, we experimented with caps of 2 and 10 days on the number of days a girl could miss. For instance, in one analysis, we re-coded all girls who missed more than 10 days to a value of 10 days missed, censoring the right tail of the distribution. Neither cap altered the substance of our findings – attendance in intervention schools dropped, relative to comparison schools, over time in both samples studied.

Subgroup Attendance Rates

Nearly all the analysis in this section has concerned aggregate differences between intervention and comparison schools and the girls that attend them, with the goal of assessing programme impact. In this section, we draw on the relative richness of our data to study the ways in which school- and household-level characteristics influence attendance. Much of this analysis is driven by insights from qualitative interviews, as

[149] Caregivers were not asked how many possible days of school there were in the previous month. However, most schools are open 6 days per week, yielding an average of 26 possible school-days in a typical month.

teachers, CEC members, and others we interviewed highlighted a number of specific barriers to consistent attendance. In most cases, it is possible to evaluate these qualitative arguments against quantitative measures.

Starting at the broadest level, two macro-level factors can be expected to influence attendance rates: exposure to conflict locally, and the experience of drought. In the communities where SOMGEP-T works, conflict is generally localised and short-duration. At the same time, localised conflicts do occur – typically along clan or subclan lines – and these conflicts can be disruptive, driving displacement, temporarily forcing a school to close, or making it temporarily unsafe for children to travel to school.

Our measure of conflict is based on parental reports that their village was "affected by conflict" in the past 12 months. At endline, conflict was somewhat less common that during ML2 – there were only three schools with multiple parental reports of conflict, compared to 13 such schools during ML2. Within the endline data, conflict-affected schools are associated with significantly lower attendance outcomes: for instance, caregivers are 14.0 percentage points less likely to report that their girl attended school most days if they live in a village that was conflict-affected. This analysis is strictly correlational, of course, and can only hint at the impact of conflict on attendance rates. However, it is strongly suggestive that schools which were conflict-affected in ML2 and not in endline experienced a net increase in the share of caregivers reporting their girl attended school most days. On the other hand, schools that were peaceful in ML2 but conflict-affected at endline experienced a precipitous drop – 13.9 percentage points – in the same measure.

In contrast to conflict, drought has a less stark negative correlation with attendance. Indeed, in some of our cross-sectional models using only the endline data, drought-affected schools had slightly higher attendance rates – including in both classroom headcounts and reports of caregivers. Our measure of drought exposure is a binary metric derived from Food Security and Nutrition Analysis Unit (FSNAU) reporting from December 2021. Using their reporting on rainfall levels, we identified schools as drought-affected if they were in areas with 75 percent or less of long-term average rainfall for the period October 1-December 5. This time period of reporting is useful because it covers around one month prior to fieldwork and the month of fieldwork, a time period during which the drought's effects would start to be felt. It is important to note, though, that the drought has continued since the end of data collection and this analysis is based on the impacts of drought that are less severe than at the time of writing. In total, 45 of 99 site locations in southern Togdheer, southern Sool, and Mudug were classified as drought-affected.

Making use of ML2 data and the fact that the drought worsened significantly around the time of the endline, we also analysed how attendance in drought-affected schools changed over time. This analysis more closely fits our theoretical expectations: in non-drought schools, caregivers reported an increase in absenteeism of just 0.09 days over the last month from ML2 to endline; in drought-affected schools, the data show a much larger increase of 0.44 days missed over the same period. Though this result is not statistically significant, it represents a 31.1 percent increase in absenteeism from ML2 to endline. A similar analysis utilising attendance headcounts resulted in a smaller estimated effect size, but a compelling trend: schools affected by the worsening drought experienced a small decline in girls' attendance from ML2 to endline, while those outside the main drought zone experienced the opposite.

It is also important to note that the most deleterious and widespread impact of drought is unlikely to be on attendance, but on retention, as droughts drive displacement to urban areas, and movement of pastoralists in search of water and pasture for their livestock. Impacts on attendance are likely to be subdued, because the students most directly impacted by drought are those leaving their communities, not those who remain. The evaluation data show that out-migration has increased from drought-affected communities in the last 12

months. On average, the regions affected by the drought reported lower out-migration of adolescent boys and girls in the previous three rounds of data collection. But thi pattern has shifted – as shown in the table below, households in drought-affected areas have "caught up" and now report almost equal levels of out-migration. In essence, while the number of girls and boys leaving their homes from non-drought areas declined in this round, the opposite trend occurred in those areas affected by drought.

TABLE 53: MEAN NUMBER OF ADOLESCENTS AGED 11-21 MIGRATING AWAY FROM HOUSEHOLD IN PAST 12 MONTHS

	Girls – ML2	Girls – EL	Boys – ML2	Boys – EL
Non-Drought Areas	0.44	0.32	0.45	0.32
Drought-Affected Areas	0.28	0.40	0.33	0.38

At a narrower level, several school characteristics are related to individual-level attendance outcomes. In line with the programme's Theory of Change, a number of qualitative interviewees emphasised the role of the CEC in prompting enrolment, retention, re-enrolment following the end of COVID-related school closures, and attendance. Although CECs tended to describe their activities as critical for promoting school attendance, the extent of this relationship is not clear from the quantitative data. The mere existence of a CEC, for instance, was insufficient, actually leading to lower attendance rates overall. The frequency of CEC meetings and whether they actively monitor student attendance have more direct theoretical links to increasing attendance, and this is partially borne out in the data. The issue we face is in distinguishing which aspects of an "active CEC" matter most – linear models controlling for a wide range of other factors seem to suggest that a CEC which plans and achieves bi-weekly or monthly meetings and monitors attendance can reduce student absences by around 0.4 to 0.5 days per month – but models focused on other measures of attendance are less clear-cut and suggest that active CECs are correlated with higher attendance, but without being able to say which dimension of "active" matters most.

In the qualitative data, respondents discussed a number of CEC activities that may influence attendance, either directly or indirectly. First, the qualitative interviews confirm that there are CECs taking an active role in school monitoring, which includes monitoring attendance of both students and teachers, monitoring the relationships between teachers and students, monitoring lessons and quality of teaching, and enforcing a school code of conduct. When asked whether the CEC has changed recently, one CEC member reports, "Yes, we have because in the old days this committee used to come to school when there was only problems but now every week someone from within the committee comes and evaluates what has been improved and what hasn't, such as teachers attendance as well as the students attendance and the hygiene of the school."[150]

In addition to directly tracking and following up with students and teachers on their attendance, CECs also play a role in maintaining the school environment, reporting issues with school infrastructure to the MoE and CARE, supporting school sanitation, and raising funds for school projects (e.g. establishing a designated school yard, bringing chairs to the school).[151] Although it was not possible to confirm CEC fundraising takes

[150] FGD with CEC Members, Int. 108.
[151] FGD with Teachers, Int. 202.

place in every sampled community, CECs in some cases are also able to raise funds for teachers' salaries and girls' school fees, which directly promotes attendance among both groups.

Lastly, CECs play a role in raising awareness among community members and with religious leaders on the importance of girls' education and the need to reduce the chore burden to allow girls to attend and focus on school. As one CEC member explains, "We met with the parents to encourage them to reduce the home tasks for the child to have a chance that they can read their lessons."[152] Although the qualitative interviews as a whole suggest the chore burden is still a major barrier to attendance, they also suggest that community attitudes toward the importance of girls' education and of the need to reduce the chore burden have improved over time, and CECs have played an active role in contributing to those attitude shifts.

In the quantitative data, a more consistent story emerges from analysis of schools' infrastructure and services available to female students. Qualitative interviews from the current and previous evaluation rounds highlighted the discomfort and fear female students feel when using toilets that lack privacy, which are shared, or which are of poor quality or have poor hygiene.[153] In the risk mapping exercises, girls commonly cited toilets as places they do not feel safe and comfortable. As one girl explains, "I feel fear to use the toilet in the school because the boys are there."[154] Poor sanitation facilities can affect girls' learning most directly through their impact on attendance – if a girl is ill but is capable of attending school that day, she might choose not to, knowing that she will need to use facilities that are uncomfortable, unhygienic, or lack privacy. Even more common are cases in which girls leave the school to use facilities elsewhere, resulting in missed instructional time.

A similar logic applies to the distribution of sanitary towels through schools. Absences as a result of menstruation – and the embarrassment and stigma that surrounds it in many communities – are widespread and have been documented in prior evaluation rounds. Qualitative interviewees again referenced the effect of menstruation on school attendance.[155] Distribution of sanitary towels does not entirely reduce embarrassment and shame felt by female students, but ensures they have access to products that reduce it and allow them to attend school while menstruating.

Both arguments are borne out in our cross-sectional analysis. In schools with separate toilets available for girls' use, caregivers are 4.3 points more likely to report that their girl attended school most days in the past year. In schools that provide sanitary towels to all female students, caregivers report 0.43 fewer absences during the past month. This latter finding, in particular, is consistent with the idea that distribution of sanitary towels makes it possible for a subset of female students – those who would otherwise lack access to these products – to attend school for the multiple days during which they are menstruating. It is also possible that schools which distribute sanitary towels widely are systematically different from other schools – perhaps they have institutionalised greater gender-sensitivity in their teaching and administrative practices, or have stronger support from the community for girls' education. While schools that distribute sanitary towels are somewhat more likely to employ at least one full-time female teacher, and slightly more likely to be rated as

[152] FGD with CEC Members, Int. 110.
[153] FGD with Girls, Risk Mapping, Int. 505; Int. 509.
[154] FGD with Girls, Risk Mapping, Int. 505

"extremely well-managed" by caregivers, these differences are relatively small; tentatively, it appears the distribution of sanitary towels promotes girls' attendance.

The final school-level characteristic that qualitative evidence suggests shapes attendance outcomes is the existence of and girls' participation in the Girls' Empowerment Forums. Participation in a GEF can widen a girls' social support network, provide her with access to female mentors, and give her influence and ownership over some aspects of school management. Even if a girl does not participate in a GEF, we might expect the presence of a GEF in her school to promote a more female-friendly environment, especially as GEF participants often report actively reaching out to other female students who are struggling in school or who face barriers to enrolment and attendance.[156] Naturally, participation in a GEF should magnify these impacts, but girls who choose to participate in the GEF may also be more motivated or have greater commitment to schooling, which could account for higher attendance rates.

Both aspects of this argument are supported, tentatively, in the quantitative data. Girls who attend schools with a more active GEF – defined based on the share of girls who report being aware of GEF activities – miss slightly fewer (0.09) days of school per month. On top of that, girls who are, themselves, active in the GEF, miss another 0.10 fewer days per month. The magnitude of these effects is small, but the findings are robust to a wide set of control variables for geography, age, grade level, and other factors that predict attendance rates.

The qualitative data suggests GEF girls are very active in their schools and communities and play a similar role to CECs in raising awareness, solving disputes, raising funds for schools fees and projects, and following up with other students on attendance and enrolment. They appear to be particularly supportive of marginalized and disadvantaged groups, such as pastoralists and children with disabilities. As one girl describes, "We encourage the uneducated people to come to school, and those who have dropped out we also go to the parents and ask why they did not bring the child back."[157] GEF girls are seen as the "link between female students and the administration of the school,"[158] due to their ability to gain the trust of other female students and bring issues forward to the administration that the girls would normally be too shy to discuss with a teacher or principal, such as girls' sanitation.

> *One of the things they have done is connect girls that aren't part of the girls' empowerment forum and teachers in touch about matters that those kind of girls are embarrassed or too shy to speak of with their teachers.*
>
> -FGD with CEC, Int. 108

When qualitative interview respondents were posed with scenarios of struggling girls and asked to whom those girls could turn for support, the GEF was one of the most commonly mentioned groups.

The last factor we consider demonstrates arguably the most consistent effect on girls' attendance: the assignment of household work to girls. In the previous evaluation report, we documented the relationship between domestic "chore burden" and attendance rates for girls, and this finding is actually stronger in the

[156] FGD, Vignettes, Int. 604; Int. 607; FGD with Mothers, Int. 308; FGD with CEC Members, Int. 108.
[157] FGD, Vignettes, Int. 604.
[158] FGD with Mothers, Int. 308.

endline data. Qualitative interviewees repeatedly highlighted the importance of domestic work as a barrier, either implicitly – as when mothers said that they made an effort to reduce their daughter's chore burden so she could focus more on school – or explicitly, as when adolescent girls discussed actual cases of their friends and acquaintances dropping out due to their domestic chore burden. In fact, when we presented female students with brief stories and asked them what would happen to the girls in the hypothetical stories, tales of girls with a heavy chore burden tended to elicit the greatest consensus. Interviewees were emphatic that the girl in question would have to drop out of school because of her responsibilities at home, and that the only alternative outcome would be contingent not on her own actions, but on those of her parents or family members – in short, if her chore burden were not reduced, her fate was sealed.

For our primary analysis, we defined a heavy chore burden as typically spending an entire day caring for children or older family members, fetching water, cleaning, or doing agricultural or other labour. The first critical point is that girls with a high chore burden are exceedingly unlikely to be enrolled in school – among the endline sample, they have a 34.4 point lower likelihood of being enrolled. Second, girls with a high chore burden, if they are enrolled, attend school considerably less often: according to caregivers, they miss an extra 1.21 days of instruction per month. Given that the typical girl in our sample missed 2.20 days of school in the previous month, 1.21 additional days represents a 55 percent increase in absenteeism.[159] Moreover, this likely understates the impact of chores, insofar as girls with significant responsibilities for their younger siblings are more likely to arrive late or leave school early.

Although the qualitative data suggests the awareness raising efforts of CECs, GEFs, teachers, and others have led to shifts in community attitudes toward the chore burden, it also suggests the chore burden is still a main barrier to attendance, particularly among families facing financial difficulties. One girl from a GEF explains this is a common reason families cite when asked why a girl has dropped out or is not attending school: "Some people tell us about circumstances such as housework or livestock and some others say the girls refused to go."[160] Additionally, during the COVID closures, both girls and mothers interviewed for the endline admitted girls shifted their focus back to housework. When asked how the shutdown affected her girls' education, one mother explains, "My daughters were not totally studied anything during lock down. They were only helping me for house cleaning."[161] As has been noted in other sections, girls whose families could not afford a private tutor were more likely to be affected in this way during the school closures, meaning already disadvantaged girls faced further setbacks after schools re-opened.

Many of the factors that we would expect to influence attendance rates – and which are embedded in SOMGEP-T's Theory of Change – appear to predict attendance rates in the endline sample. In general, the strongest vindication of the programme's Theory of Change comes in the role of domestic workload, provision of sanitary towels, and the importance of sanitary infrastructure, all of which show a consistent relationship with attendance rates, across multiple measures of attendance and across statistical models.

[159] The strength of this finding is reinforced in two ways. First, when we alter our coding to capture greater variation in chore burden – analysing the impact of light, medium, and heavier chore burdens – the negative effect of chores on attendance rates increases monotonically. In other words, a heavier chore burden results in lower attendance, exactly as we would expect if chore burden were exerting a direct effect on attendance. Second, when we use an alternative measure of attendance – caregivers' assessments of whether girls attend "most days" – the impact of chores on attendance is even more sharply demonstrated. Caregivers whose girls complete a "quarter-day" of chores on a typical day are 8.9 points less likely to claim that their girl attends school on most days; when chore burden increases to a full day, caregivers are 18.7 points less likely to claim their girl attends school most days.

[160] FGD, Vignettes, Int. 604.

[161] FGD with Mothers, Int. 301.

Conflict and drought also have important effects on attendance, though it is difficult to draw strong conclusions regarding conflict because few SOMGEP-T villages were impacted by conflict prior to the endline round, according to the caregivers surveyed. Meanwhile, the quantitative evidence linking GEF participation to attendance is much weaker, but buttressed by the strength of the qualitative evidence provided by GEF participants themselves.

Programme Impact on Enrolment Rates

As noted at the outset of this section, systematic changes in enrolment rates in intervention and comparison schools would bias our analysis of attendance rates if intervention schools saw an influx of students with a low propensity for consistent school attendance. Changes in enrolment levels are subtly different from changes in transition outcomes, because improved transition rates cover a wider range of outcomes, including movement into other schools, alternative education, and employment. Much of our attendance analysis is based on attendance of all girls or boys, not only those in the evaluation cohort or those directly benefitting from programme interventions (e.g., girls in lower grades, who were too young to be eligible for inclusion in SOMGEP-T). Therefore, even if SOMGEP-T has not caused an increase in positive transition outcomes, if enrolment has increased among students with a low propensity for attendance, it would artificially reduce our estimates of the programme's impact on attendance rates.

We assess whether enrolment levels have increased using data collected from each classroom during physical headcounts, and also from the head teacher's enrolment records. In general, the number of distinct classes has increased in both intervention and comparison schools, suggesting that there has been a generalised increase in enrolment numbers.[162] Total enrolment numbers of both boys and girls have increased as well; for instance, in intervention schools, we documented a total of 2,973 girls across 32 schools at baseline and 3,238 girls across the same 32 schools at endline, an 8.9 percent increase in enrolment.

Importantly, however, comparison schools have seen a larger increase in girls' enrolment since baseline. The table below reports the change in total enrolment in intervention and comparison schools across different time periods. From baseline to endline, girls' enrolment increased at a faster pace in comparison schools; from ML2 to endline, girls' enrolment in intervention and comparison schools both increased considerably, but at a similar rate. As such, differential increases in enrolment are unlikely to be producing bias in our analysis of attendance rates, because a greater influx of students into comparison schools should, if anything, drive attendance rates in comparison schools downward. The only way changes in enrolment could produce downward bias on our impact estimates of attendance would be if the increased enrolment in intervention schools was disproportionately concentrated among girls with low attendance rates – a possibility we cannot assess empirically and one which would be fairly coincidental.

TABLE 54: CHANGES IN ENROLMENT LEVELS, BY INTERVENTION STATUS

	Intervention	Comparison	Difference
Girls, Baseline to Endline	8.9%	17.2%	-8.3
Girls, ML2 to Endline	18.3%	16.7%	1.6

[162] Researchers were instructed to complete a headcount in every available class, though this direction may not have been followed as closely in prior rounds.

Boys, Baseline to Endlne	3.8%	21.7%	-17.9
Boys, ML2 to Endline	18.2%	3.1%	15.10

Conclusions

The analysis in this section has attempted to incorporate multiple data types and sources to draw conclusions regarding SOMGEP-T's impact on attendance rates. Unfortunately, most of the individual analyses produced inconclusive findings, with weak or null effects or effects that were more substantial but statistically insignificant. Further, many of the findings were inconsistent across data sources, complicating the conclusions we would otherwise be able to draw if consistent trends were observed in statistically insignificant results.

Overall, there is very little evidence that SOMGEP-T improved attendance rates, aggregating over the life of the programme. Classroom headcounts – in many ways our most reliable measure of attendance – showed no improvement in intervention schools, vis-à-vis comparison schools, since either the baseline or ML2. While differential enrolment or dropouts across intervention and comparison groups could skew attendance figures, there is also little evidence that enrolment in SOMGEP-T schools increased more quickly than comparison schools. It is still possible that the *type* of student brought into SOMGEP-T schools has a lower propensity for consistent attendance. However, this is not a question that can be answered satisfactorily with existing data, and differential enrolment of this kind would need to occur on a large scale to alter our core results.

Analysis of school attendance records and caregiver-reported attendance is similarly inconclusive. School records show a positive impact since baseline but only among the extremely small sample of students who were enrolled and successfully contacted at both baseline and endline; a larger sample of students tracked from ML2 to endline show a negative – though not statistically significant – programme impact. Among caregiver reports of attendance, our preferred metric is the number of school-days missed in the previous month, because this measure avoids the uncertainty inherent in our other survey question, in which caregivers were asked whether their girl attended school "most days." With regard to absences over the previous month, the programme seems to have had a null or negative impact over both the full programme lifecycle and since ML2.

It is important to note that none of the negative trends described in this section are statistically significant, and the conclusion with the greatest validity is that the programme did not have an effect on attendance rates that is distinguishable from zero. Although there are potential caveats that could explain away many of the null or slightly negative findings regarding attendance, overturning all of them would require a highly implausible mix and number of specific circumstances – systematic differences in the nature of students enrolling in intervention versus comparison schools, systematic differences in how survey respondents interpreted questions, and systematic bias in recall, among others.

The endline evaluation took place in the context of two overlapping shocks: COVID-19, which prompted mandatory school closures, and a worsening drought that is particularly severe in approximately half of the sample communities. While COVID-19 school closures have passed, they almost certainly affected attitudes toward school attendance, reshaped the lives of students and their families, and potentially undermined the small, but meaningful, gains in attendance rates that had been documented in the ML2 evaluation, a study conducted prior to the COVID-19 outbreak and in a non-drought environment.

School Governance and Management

The next intermediate outcome is improved school management and governance. In this section we will investigate the extent to which SOMGEP-T interventions have impacted indicators related to school governance and management. The improvement of this outcome in schools not only increases the sustainability of the project but also creates an environment allowing for improvements in students learning. To measure the impact of the programme on the selected indicators we will utilize a utilise a difference-in-differences approach. The change in intervention schools during the implementation period will be viewed against the change in comparison school during the same period.

We will utilise several different data sources for our evaluation of school management outcomes. First, we will use two different quantitative surveys. One conducted with head teachers and the other carried out with student's primary caregivers. Qualitative data collected from Focus Group Discussions (FGDs) will also be utilised. This data was collected from CEC members, teachers, and caregivers and will allow us to draw better inferences and have a deeper understanding of the quantitative data.

The analysis utilises the same key indicators established in the quality of school governance baseline and the two previously conducted midline evaluations. Based on these indicators and the data available, this section will be divided into two sub-sections. First, we will look at school management from the perspective of head teachers. Next, we will look at primary caregiver's perceptions of school management. The strengths and weaknesses of each of these approaches will be discussed below. Both quantitative and qualitative data will be used throughout both sub-sections.

Community Education Committees Assessment from Head Teachers Perspective

First, we will investigate school management based on the perspective of head teachers. This section focuses primarily on the performance of the CECs as they are the most important body of school management at the local level. The committees consist of local volunteers including parents, religious leaders, head teachers, and members of women's and youth groups. The body has many responsibilities, including acting as a liaison for the relationship between the school and the local community. Members of the CEC also monitor various metrics at school such as teacher and student attendance and performance. The CEC is also involved in creating or reviewing school development plans. These include elements that are of relevance to the sustainability of the SOMGEP-T project, such as plans for enrolling out-of-school girls. The CECs often raise money from the community to cover teacher salaries or to make improvements in school facilities and resources. The material and immaterial support that the body provides is essential for the sustainability of the project. Moreover, an active and effective CEC may positively contribute to learning outcomes through its influence on teaching quality, practices, and the resources available at the schools.

The primary data source for this section is the survey of head teachers. One head teacher was interviewed at each of the intervention and comparison schools. Additionally, FGDs were conducted with teachers as well as members of the CECs. The main benefit of utilising head teachers to evaluate the performance of the CECs is that their position gives them unique insight into the functioning of the school and its administration. As many CEC activities may be conducted out of view of parents and students, interviewing head teachers should allow us to have a full picture of the impact of CECs. However, the use of this survey does come with a significant drawback, the sample size. Due to there only being one head teacher per school means that we have a significantly reduced sample size to draw inferences on programmatic impact from. In the usable panel

dataset, meaning teachers that were surveyed at baseline and recontacted at endline, there are 31 head teachers in intervention schools and 37 head teachers in comparison schools.

Where possible we will look at the change in indicators between the baseline survey and the endline survey. However, several of the indicators utilised in this section were not incorporated into the baseline survey. Therefore, for these indicators we will investigate the programmatic impact between the second midline survey and the endline survey. We will begin by looking at the indicators that are able to be evaluated starting at the baseline survey. Results are presented in the table below.

TABLE 55: PROJECT IMPACT ON SCHOOL MANAGEMENT – BL TO EL

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	BL	EL	BL	EL		
School has management plan	45.9%	64.9%	12.9%	54.8%	-23.0	0.101
CEC member has visited school	64.9%	78.4%	22.6%	76.7%	-40.6	0.005
Monitored: teacher attendance	83.3%	58.6%	57.1%	52.2%	-19.7	0.399
Monitored: facilities	50.0%	69.0%	28.6%	56.5%	-9.0	0.692
Monitored: teaching quality	54.2%	51.7%	57.1%	43.5%	11.2	0.642
Monitored: student attendance	58.3%	72.4%	28.6%	73.9%	-31.3	0.186
Monitored: student retention	37.5%	20.7%	28.6%	30.4%	-18.7	0.428

First, we look at if a school has a management plan, which is a product produced by the CEC. This metric only indicates if the management plan exists, later we will investigate the contents of the plan. Both intervention and comparison schools saw large increases in the percent of schools that had a management plan at endline compared to baseline. As of the endline nearly two-thirds, 64.9 percent of intervention schools had a management plan in place. Comparison schools saw a similar increase; however, they had a much lower starting point, 12.5 percent of schools versus 45.9 percent of intervention schools. We then look at if a CEC member had visited a school. Similarly, to the findings for school management plans, both groups saw an increase with comparison schools starting from a much lower level. This difference in starting level has a major impact the findings for these metrics. The negative results of our difference-differences regressions are heavily influenced by the large gains seen in comparison schools and are seen despite the large gains in intervention schools.

One interpretation of these findings is that the programme has not improved CEC management practices at all, but that CECs in the area, more generally, were improving their practices and this included improvements in intervention schools. However, it is far more likely that this reflects a concerted effort by other programmes operating in the area to train CECs and their members. For instance: Save the Children Federation trained 330 Somaliland CEC members in 2020 and another 330 in 2021 through the ESPIG programme, while the Education is Light trained 63 CECs in Puntland from 2019 to 2021. As such, it is not surprising that CEC management practices have improved rapidly even in comparison schools. Combined with the lower starting point for comparison CECs, this explains the negative programme impact estimates across most of the outcomes reported in the table above.

Next, we look at if a CEC member had monitored various school management functions during their visits. These included teacher attendance, facilities, teaching quality, student attendance, and student retention. Both intervention and comparison schools saw decreases in the monitoring of teacher attendance and teaching quality. The largest decrease was in the monitoring of teacher attendance in intervention schools, dropping from 83.3 percent to 58.6 percent from baseline to endline. Conversely, both intervention and comparison schools saw an increase in the monitoring of student attendance and facilities. For the monitoring of student retention, intervention schools saw a decrease while comparison schools saw a slight increase. These findings of changing patterns of monitoring by the CEC is interesting. There has clearly been a shift among both groups of schools from monitoring teaching related indicators (quality and attendance) to student indicators (attendance). This was reinforced by one of the teachers during an FGD.

> *School administration sometimes tells us how to prioritize or coordinate lessons but no one else helps us improve our teaching quality, either the Ministry of Education or the agencies.*
>
> -FGD with Teachers, Int. 201

This could be due to a change in priorities communicated by the MoE, or challenges in monitoring teaching related to the COVID-19 pandemic. However, none of these findings were statistically significant, most likely due to the small sample size of the head teacher survey. To help further our understanding of the trends in these variables the table below presents the change in the same indicators from the second midline to the endline.

TABLE 56: PROJECT IMPACT ON SCHOOL MANAGEMENT – ML2 TO EL

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	ML2	EL	ML2	EL		
School has management plan	73.0%	64.9%	38.7%	54.8%	1.2%	0.078
CEC member has visited school	86.5%	78.4%	67.7%	76.7%	18.7%	0.237
Monitored: teacher attendance	68.8%	58.6%	71.4%	52.2%	-28.9%	0.660
Monitored: facilities	62.5%	69.0%	38.1%	56.5%	3.0%	0.517
Monitored: teaching quality	53.1%	51.7%	38.1%	43.5%	18.0%	0.678
Monitored: student attendance	68.8%	72.4%	57.1%	73.9%	-18.2%	0.434
Monitored: student retention	37.5%	20.7%	28.6%	30.4%	0.0%	0.302

Interestingly, intervention schools had higher values during ML2 than at either baseline or endline for having a school management plan and having a CEC member visit a school. Or in other words, the values for these variables peaked during the second midline survey and then decreased again by the time of the endline survey. However, this trend did not occur in comparison schools, who display a more linear trend of growth from baseline to midline to endline. This finding helps confirm our theory of diminishing returns. Intervention schools seemed unable to maintain the high levels of success seen during the ML2 survey in contrast to comparison schools who started from a lower value and were able to continue their progress.

Next, we will turn to indicators where our comparison is only possible from the second midline survey to the endline survey, beginning with the functioning of the CEC.

TABLE 57: PROJECT IMPACT ON CEC EXISTENCE AND ACTIVITY

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	ML2	EL	ML2	EL		
Does this school have a functioning CEC?	100.0%	94.6%	93.5%	93.5%	-5.4	0.479
School management good or very good	91.9%	85.7%	89.7%	89.7%	-6.2	0.524
CEC meetings once a month or more	94.6%	94.1%	75.9%	71.4%	4.0	0.763
In past year, # visits by CEC	8.4	11.4	7.8	9.3	1.5	0.661

Here we will look at some of the basic functions of CECs and general perceptions of the management of the school. We first look at if the school has a functioning, CEC, an obvious prerequisite for any of the other competences of the CECs. During the second midline survey 100 percent of head teachers indicated that their school had a functioning CEC during the endline this decreased to 94.6 percent. However, this only represents a real decrease of 2 head teachers changing their response and reporting that their school does not have a CEC. The comparison group saw no change in this metric. Similarly, intervention schools saw a decrease in head teachers reporting that their school had good or very good management. Once again comparison schools saw no change in this metric. This led to small, negative, not statistically significant findings of impact for these two metrics. The downward trend witnessed is concerning, however when considering the very high result at the second midline for both findings among intervention schools as well as the small sample size of the head teacher survey it is difficult to draw larger conclusions from these findings.

We then looked at the frequency of CEC meeting as well as the number or visits by the CEC to schools during the past year. During the second midline evaluation 94.6 percent of head teachers reported that their school's CEC meet at least once a month. There was very little change in this number at endline with 94.1 percent of headteachers now reporting that their school's CEC met at least once a month. Comparison schools saw a slight decrease of 4.5 percent leading to a modest positive but not statistically significant finding. Another modest positive finding was seen in the number of visits by CECs to schools. Intervention schools saw 3 more visits per year from their CECs, bringing the total at endline to 11.4 visits, or almost once per month. This show an increase in activity by CECs, so although there may not be more CECs at schools, the ones that do exist seem to be operating more frequently.

Next, we will further investigate the school management plans mentioned earlier in this section. Once again, this comparison is only possible between the second midline round and the endline.

TABLE 58: PROJECT IMPACT ON SCHOOL MANAGEMENT

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	ML2	EL	ML2	EL		
Does plan include info on monitoring school	77.8%	87.5%	66.7%	76.5%	-0.1	0.996
Does plan include info on child protection policies	77.8%	83.3%	58.3%	82.4%	-18.5	0.266
Does plan include info on encouraging school enrolment	88.9%	91.7%	66.7%	94.1%	-24.7	0.169
Does plan include info on plans to follow-up with dropouts	74.1%	83.3%	66.7%	76.5%	-0.5	0.980

Each of the regressions in the table above explore if a school's management plan includes information on a specific subject. The results in this section all follow a similar pattern. Both intervention and comparison schools saw increases in the inclusion of all four types of information. monitoring, child protection, enrolment encouragement, and follow-up with dropouts from the midline survey to the endline survey. However, as seen with previous indicators the intervention schools had a much higher starting point, meaning that as of the second midline evaluation their school management plans included more of the pertinent information. This led to negative findings for the difference-in-differences regressions in this table, particularly for the inclusion of child protection policies and encouraging school enrolment. The large gains seen by the comparison schools offset the gains of the intervention schools, leading to negative findings. None of the findings in this table were statistically significant. It is also important to note that despite the negative findings there were positive changes in all of the metrics for intervention schools, meaning that the school management plans in all intervention schools did see improvement in content.

Once again there are several possible explanations for the negative findings. First there may be some limit where program activities are effective at creating change, meaning that there are diminishing returns and the higher starting point of intervention schools means that they can only see a certain level of increase. A second explanation, which could also be happening alongside the first is that, as discussed during the second midline evaluation, the MoEs of each zone have also worked on improving CECs. This could help explain the large increases seen in comparison schools in metrics related to school management.

The next table presents the findings for a closer look at the impact of the project on CEC monitoring.

TABLE 59: PROJECT IMPACT ON CEC MONITORING

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	ML2	EL	ML2	EL		
Does CEC follow-up with or contact dropouts	70.3%	64.9%	67.7%	67.7%	-5.4	0.678

Does CEC monitor teacher attendance	70.3%	64.9%	64.5%	48.4%	10.7	0.494
Does CEC take action against teachers on attendance	67.6%	64.9%	48.4%	41.9%	3.7	0.790
Does CEC raise funds for school improvements	51.4%	54.1%	35.5%	64.5%	-26.3	0.044
CEC reinforces non-violent discipline vs corp. punishment	67.6%	62.2%	54.8%	45.2%	4.3	0.778
Does CEC address child protection issues	78.4%	70.3%	74.2%	67.7%	-1.7	0.900
Does CEC promote enrolment of OOS children	78.4%	67.6%	61.3%	64.5%	-14.0	0.371

Between the second midline survey and the endline survey head teachers in intervention schools reported a drop in CECs performing all but one of the functions in the table above. CECs following up with dropouts fell from 70.3 percent of intervention schools at the second midline survey to 64.9 percent of schools at endline. CECs promoting enrolment of out of school children fell from 78.4 percent of intervention schools during the second midline survey to 67.6 percent of schools during the endline survey. The only indicator where intervention schools saw an increase was in raising funds for school improvement, increasing from 51.4 percent of schools at the second midline to 54.1 percent of schools during the endline survey. Comparison schools saw similar results with two indicators having positive trends over the implementation period of the programme. In comparison schools the CEC promoting enrolment of out of school children increased from 61.3 percent of schools during the second midline survey to 64.5 percent of schools at endline. The CEC raising funds for school improvement also saw an increase from 35.5 percent of head teachers reporting their school's CEC fulfilled this function as of the second midline survey to 64.5 reporting the same during the endline survey.

Our difference-in-difference regressions found mixed results with 3 positive indicators and 4 negative indicators. It is also interesting to note that intervention schools had systematically higher second midline averages of all the indicators in the table above. With intervention schools seeing consistent decreases across almost all the indicators the result of the regression was mostly determined by how the comparison schools fared. Where comparison schools saw the largest increase, the CEC raising funds for school improvement, we found the largest negative impact and the only statistically significant result. A possible explanation for seeing an increased focus on raising income and a reduction in conducting other activities is that without revenue the CECs struggle to perform other functions. This was mentioned by one teacher during an FGD.

> *They support the school as they can while they do not have also have an income.*
>
> - FGD with Teachers, Int. 205

Several comments by CEC members are also enlightening as they point to communication, and not the activities listed in the table above, as their main function. This could show that while the CECs are more

active than prior to program implementation they are taking on slightly different roles than the indicators we are measuring here. The first respondent also once again discusses the lack of funding for CECs.

> *The CEC serves as a conduit for communication between the Ministry of Education and the general public. So, following the observation, we report any difficulties in the school, such as broken toilets, to the Ministry of Education or international non-governmental organizations. If there is a problem between the teachers we contacted and the local administration, we also resolve it. However, we do not have any funds to improve any of our activities.*
>
> -FGD with CEC, Int. 106

> *We have a system in place for instructors and parents to communicate with one another. We resolve difficulties whenever instructors or parents report a child with a problem. We normally pay a visit to the school to assess its current status.*
>
> - FGD with CEC, Int. 104

Community Perceptions of School Management

The next sub-section will focus on student's primary caregiver's perceptions of school management at the school that their child attends. As with the section above we will once again look at the functioning of the CECs, however in this section we will also look at several indicators looking at a broader understanding of school management as well as the performance of the head teachers. Due to the survey utilised in this section being of student's caregivers we have a much larger sample to draw on in this section. In the usable panel dataset, meaning caregivers that were surveyed at baseline and recontacted at endline, there are 447 caregivers with girls enrolled at intervention schools and 512 caregivers with girls enrolled at comparison schools. A drawback of the data used in this sub-section is that caregivers may not have a full understanding of the functioning of their child's school. This is particularly true of some CEC functions that only apply to a small subset of the student population, notably student retention and remedial support. However, by comparing the views of head teachers and caregivers we believe that the strengths and weaknesses of both datasets will be balanced and provided for a clearer understanding of school management and governance.

The table below reports the first set of difference-in-differences results focusing on overall school management, the existence of a CEC, and the impact of the CEC if present.

TABLE 60: PROJECT IMPACT ON SCHOOL MANAGEMENT – HH SURVEY

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	BL	EL	BL	EL		
Does the school have a CEC that helps with school related matters	78.3%	74.1%	63.5%	74.9%	-15.6	0.088
School managed extremely well	47.5%	47.5%	39.1%	41.8%	-2.7	0.637
School management improved	51.4%	49.8%	44.3%	46.5%	-3.8	0.440

School head teacher excellent	50.0%	49.2%	42.7%	43.4%	-1.4	0.764
Frequency of CEC comms (3=weekly, 0=never)	1.57	1.72	1.62	1.62	0.22	0.369
The CEC initiatives improved the quality of schooling	67.0%	74.7%	64.4%	67.1%	5.0%	0.615

We begin by looking at several indicators related to how caregivers perceive the CECs at their children's schools. First, is if caregivers are aware of an active CEC. For interventions this has decreased slightly, with 78.3 percent of caregivers reporting a functioning CEC at baseline and 74.1 percent reporting a functioning CEC at endline. Conversely, comparison schools saw a 11.4 percent increase between the baseline and endline surveys. This led to a statistically significant negative finding for this indicator. Caregivers with girls at interventions schools also reported a modest increase from the baseline survey to the endline survey of CEC initiatives improving the quality of schooling. Comparison schools also saw an increase, although it was smaller leading to a positive finding of programmatic impact.

The two previous findings point to CECs that improving but not expanding. Caregivers believe that less schools have CECs that help with school related matters but the CECs that do exist are having a greater impact on improving the quality of schooling. This could point to the programme improving the functioning of CECs rather than helping CECs expand to other schools. We will further investigate this assumption later in this section when reviewing the findings for caregiver perceptions of CEC activities.

TABLE 61: PROJECT IMPACT ON CEC ACTIVITY – HH SURVEY

Regression Details	Intervention		Comparison		Diff-in-Diffs	P-Value
	BL	EL	BL	EL		
CEC Monitored student attendance	56.0%	65.9%	53.2%	54.5%	8.6	0.353
CEC Monitored teacher attendance	37.3%	51.4%	38.6%	44.3%	8.4	0.413
CEC Raised funds	11.1%	23.2%	15.8%	17.4%	10.5	0.069
CEC Improved school infrastructure	23.0%	43.2%	19.0%	34.1%	5.0	0.527
CEC Supported students financially	9.5%	23.6%	10.8%	17.4%	7.5	0.301
CEC Bought learning materials	6.7%	20.5%	4.4%	20.4%	-2.2	0.730
CEC Promoted enrolment of out-of-school children	19.0%	35.0%	17.1%	31.7%	1.3	0.906
CEC Provided remedial support	5.2%	8.6%	3.2%	7.8%	-1.1	0.762

CEC Reinforced the use of non-violent disciplines	4.4%	16.4%	3.8%	15.6%	0.2	0.964
CEC Monitored student retention	13.5%	14.5%	12.0%	14.4%	-1.3%	0.812

Both intervention and comparison schools saw increases in all metrics of CEC activity between the baseline and midline surveys as reported by girl's primary caregivers. Among intervention schools the CEC monitoring teacher attendance increased from 37.3 percent at baseline to 51.4 percent at endline. The CEC improving school infrastructure increased from 23.0 percent during the baseline survey to 43.2 percent during the endline survey. Similarly, comparison schools saw an increase in the CEC supporting students financially from 10.8 percent at baseline to 17.4 percent at the endline survey. Comparison schools also saw an increase in the CEC improving school infrastructure from 19.0 percent to 34.1 percent as of the endline survey. Overall, our difference-in-differences regressions found broadly positive programmatic impact, with the impact on CECs raising funds being the only statistically significant result.

It is important to note that these findings do not represent a direct measurement of an increase in CEC activities, instead it represents a change in caregiver's perception of caregiver's activities. That is not to say that CECs have not completed more infrastructure projects or raised more funds, only that the findings in this table do not tell us that directly. Instead, they show that caregivers in all schools are more aware of the functioning of the CECs at their child's schools.

This perception of increased CEC activity was seen in the comments of several mothers during focused discussion groups.

> *Teacher's discipline has gotten a lot better than before because any teacher who is late for his job the committee warns them and then anything after that they get reported to the ministry.*
>
> - FGD with Mothers, Int. 306

> *Our school's CEC are excellent. They are employed at the school and are responsible for the construction of new classrooms. The teachers are likewise great, and the students' education is improving.*
>
> -FGD with Mothers, Int. 307

The vast majority of the mothers interviewed during the focused discussion groups had high praise for their local CECs. Most stated that the CECs were working to improve the school through various mechanisms. The mother quoted below is representative of many of the respondents.

> *The council always works to improve the quality of the school. At the same time, they also make an effort to connect parents with the school.*
>
> - FGD with Mothers, Int. 308

School Management: Overall Findings

Overall, there are mixed results for the findings on school management and governance. This is true not only between indicators but also between how head teachers and caregivers report on the progress of CECs. When looking at the survey of head teachers they reported a drop in the number of CECs functioning as well as a reduction in the number of functions being carried out by CECs. This trend was reversed for the survey of caregivers, with them reporting that the CECs in their daughters' schools had conducted more functions. One possible explanation for this is that these results are more about visibility of the CEC than its actual performance. The programme and its emphasis on the CEC may have made more caregivers aware of the committee's activities during the implementation period, therefore increasing the number of caregivers reporting that the CEC is conducting various activities. It is possible that the CECs were initially less active in intervention schools in ways that would be visible to community members. The absence of such activities from comparison schools could explain the initial distinction between them and the intervention schools. As the programme has progressed, CECs have increased their outward facing functions making the parents of enrolled girls more aware of their activities. Meanwhile, the MoEs in each region have worked with the CECs, increasing their presence across the board, potentially explaining why the comparison schools are narrowing the gap to the intervention schools.

Another interesting finding of this section is that intervention schools had systematically higher baseline values for many of the indicators than comparison schools did. It is difficult to tell whether this is due to some of the programme activities at the time of the baseline, to randomness, or to a result of some form of bias in the school selection. Regardless, the higher baseline values in intervention schools have several important impacts on our findings. The first, as discussed in previous sections is that we could be seeing a form of diminishing returns of the CECs. Meaning that they are a very good mechanism for increasing a metric from 20 to 80 percent but for some reason unable to increase above that level. For example, when looking at the first table of this section, comparison schools had a baseline level of less than 30 percent for half of the indicators, and that of less than or equal to 60 percent for all of them. Meanwhile, intervention showed a level of more than or equal to 50 percent for five of the seven indicators. In addition, while it may not be possible to attribute higher CEC activity to the program per se, this point is not entirely relevant. For sustainability of the program may in fact be better that the CEC activity is not dependent upon the programme. The more independent the CEC, the more sustainable the programme. As such, the fact that the intervention schools continue to have a much higher level of CEC activity as reflected in the responses to the questions analysed above, bodes well for the programme.

Teaching Quality

Outcome	Indicator	BL	EL Target	EL	Target achieved	Change

					? (Y / N)	
Teaching Quality	Percentage of teachers who apply improved teaching practices in literacy and numeracy	39.4 %	60% increase in the proportion of teachers with self-reported use of formative assessments from baseline	82.4%	Y	Intervention teachers, from the BL to the EL, had a 100% increase in use of formative assessments

Main qualitative findings

Teachers, during FGDs, showed significant support for formative assessments. A number of teachers describe the utility of formative assessments for evaluating student understanding in real time so that one can adjust the lesson accordingly. As one teacher described it, "at the end of the week I give a quiz to assess whether the students understand the unit or not, so if I find out that the students understand [the lesson] I will teach on, but if they don't, I will have to change the whole process".[163] Some teachers are still confusing formative assessments with summative assessments, because when they elaborate on the type of formative assessment used, they describe monthly tests and exams.

Classroom evaluators also have difficulty discerning formative assessments from other assessments. While 39.4% of teachers report using formative assessments in the final block of observation during the BL, classroom evaluators report observing formative assessments 58.5% of the time.

[163] FGD with Teachers, Int. 208

Teacher Perceptions	Shifts in teachers' awareness of quality education	N/A	Teachers express awareness on the importance of child protection and gender-equitable practices and describe how they are using strategies to support struggling students, including children with disabilities.	82.4% of intervention teachers disagree that corp. punishment is sometimes necessary. 83.8% of intervention teachers agree corporal punishment slows down learning. 85.3% of intervention teachers agree they have the same expectations of girls and boys. 94.1% of intervention teachers agree strongly that they adjust lessons to help struggling students. 80.9% of intervention teachers agree it's important to design lessons that are gender-sensitive.	Y	50.7% of teachers were observed physically disciplining girls at the BL as opposed to 0.0% at the EL. The respective proportions for boys are 47.9% and 0.0%.

Main qualitative findings

The qualitative findings indicate an attitude change regarding corporal punishment. While teachers can be biased regarding their abuse therefore reporting lower than actual rates of corporal punishment, establishing corporal punishment as taboo and counterproductive is beneficial – even a decrease in reported use of physical discipline is an improvement over the BL. One CEC member described a belief "students' skin belongs to the instructor, and their bones to the parents", but an effort to "educate teachers on not hitting pupils [has resulted in] aggressive behaviour no longer [being] tolerated".[164] While the attitude towards physical disciplinary action has changed, the lack of corporal punishment for both boys and girls in the classroom EL observation may partially be a result of a teacher's awareness of their observation (teachers are likely aware corporal punishment is viewed as unacceptable therefore the classroom observations biased by the Hawthorne effect).[165] Teachers want to be socially desirable therefore they adjust their classroom behaviour in the presence of an outside observer.

During FGDs teachers described moving students with visual impairments to the front of the classroom as well as well working to protect disabled students from classmates. They also mention working more closely with individual students, even students who are struggling – "I teach students who have difficulty understanding, I help people with special needs like those with hearing impairments, and I encourage girls

[164] FGD with CEC Members, Int. 107

[165] The Hawthorne Effect can be described as "awareness of being observed or having behaviour assessed engenders beliefs about researcher expectations. Conformity and social desirability considerations then lead behaviour to change in line with these expectations". See: McCambridge J., Witton J., and Elbourne DR. 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol. 67(3): 267-77.

to study".[166] Other teachers describe using games or objects to teach students struggling in math.[167] This individualized approach to teaching is a form of remedial teaching.

Students in FGDs noted remedial education can be sourced through family members, teachers and private schools. Many mentioned that struggling students should try to get help outside the classroom. One girl noted a struggling student "can tell her friends … that the subject is difficult to her, and they [can] advise her to go to a private school to study".[168] A few interviewees reported that if a student faced enough learning difficulties they would drop out.

The quality of an education, from any institution, is dependent on the pedagogy of the teachers. SOMGEP-T aims to increase both the quantity and the quality of education. SOMGEP-T can increase the quantity by increasing enrolment and improving attendance whereas increasing quality requires teacher training, improved curriculum, and inclusive teaching practices. This particular section will focus on the classroom, from the perspective of an outside observer (our enumerator), the teacher, and from the student. It will also cover teaching quality as perceived by the caregiver of the child.

While teaching quality is difficult to measure, we have established a number of outcomes that serve as proxies for teaching quality. These proxies include:

- An activity index. Active teaching, whether that is engaging students or creating a classroom environment that fosters exchange, is important to child development and learning. Giving students opportunities to lead, present, or debate can teach them tangible skills they can leave school with. Traditionally classes revolved around repetition and memorization. Engaging students so that they actively participate in the learning process help students develop leadership skills as well as motivation and determination. An engaged student who participates and receives positive enforcement will likely develop a stronger sense of agency and self-advocacy.
- Formative assessments. SOMGEP-T intended to increase teacher use of formative assessments. Formative assessments in their simplest form are a low risk assessment that measures a student's knowledge and understanding of a topic that helps inform both the teacher and the student of gaps in learning.[169] The advantage of a formative assessment is that it allows the teacher to adjust the lesson plan in ways the summative assessments do not since summative assessments often take place that the end of terms. Formative assessments are important because they allow for dynamic teaching as well as remedial teaching. Sometimes regardless of student ability, a class or concept

[166] FGD with Teachers, Int. 209
[167] FGD with Teachers, Int. 202
[168] FGD with Students, Int. 608
[169] The term "Formative Assessment" refers to a range of methods employed by teachers to elicit feedback from students, during learning, for the purpose of adjusting instruction to meet students where they are in the learning process. As opposed to formative assessments where the goal is to collect information at the end of some unit *of learning*, formative assessments are used to guide teacher and student actions in real time *for learning*. See: Black, P., and D. William. 2009. Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability 21.1: 5–31.

might be too difficult in the form that it is originally presented. A formative assessment can identify student disabilities as well as issues in teaching pedagogy.

- Corporal punishment. Corporal punishment is a detriment to not only enrolment and attendance but teaching quality as well. Physically disciplined students lose motivation to learn.[170] Students who face physical discipline in the classroom are likely to focus on fear instead of learning. Positive reinforcement is a much motivator. Physical discipline can silence a student or classroom and make it harder for a child to mature and develop normally.
- Teacher behaviour. A teacher's behaviour is a good determinant of teaching quality. Students in many ways emulate their teachers. Good teacher behaviour can nurture student behaviour, creating a productive environment where students are eager to live up to teacher expectations. Students are aware when teachers come prepared. They are also aware when teachers use appropriate language. The use of student names shows teachers are observing each and every student's development.

Many of these measures or proxies will be disaggregated by gender, as well as regressed against coached and non-coached teachers. The supporting data is sourced from the household survey, in which girls, caregivers, and boys were interviewed as well as well as the Classroom Observation tool where team leaders observed at least two classes in each school.

The views and perceptions of caregivers and girls are sourced from the household survey, a tool administrated by our enumerators in person, often at the girl's place of residence. The household survey captures girls' perceptions of their teachers' behaviour, as well as gender equity in the classroom. Caregivers are assessed on their perceptions of teaching quality as well as change in teaching practices. Classroom observation is led by the team leader, the most experienced enumerator. During the classroom observation the team leader notes classroom activities in three 15-minute blocks. Classroom activities include the use of formative interactions, presence of corporal punishment, and other teacher student interactions. The team leader also asks teacher a number of questions about their attitudes toward certain teaching practices, gender equity, and remedial methods.

Each survey interviewee has different influences affecting their responses. For instance, teachers may overstate their teaching quality for recognition, while girls may underestimate gender inequity in the classroom because they are not directly aware of its manifestation. Caregivers may rate teaching quality positively as it pertains to time in the classroom instead of learning outcomes. Analysing teaching quality through multiple lenses as well with many measures can help mitigate the effect any one of these influences because we can consider these biases actively, and weigh the value of different pieces of evidence to arrive at more a rigorous conclusion.

It is important to note a large exogenous factor that swept through communities globally in 2020-2021. COVID-19 is profoundly affecting education in SOMGEP-T and comparison schools. The pandemic induced school closures across all three zones. School closures make it harder for teachers and students to follow curricula as well as for schools to monitor students learning and teaching practices. The advantage of a

[170] Ahmad, I., Said, H., & Khan, F. 2013. Effect of corporal punishment on students' motivation and classroom learning. Rev. Eur. Stud., 5, 130.

differences and differences regression approach to analysing teacher quality is that it controls for Covid-19 and other exogenous factors.

Despite the ability to control for the ongoing pandemic with our regression model it is discouraging to see learning and teaching measures decline so much since the ML2. As one teacher put it, "during covid-19 there was a decline in the knowledge of students, although not all students are the same, some students did not have access and some had access to WhatsApp and TV at home, and some did not have nothing and they are poor".[171] Students re-entered schools after closures at different levels. Some students had studied at home, listened to lessons over the radio, or read books.[172] Other students primarily did chores. Many teachers were only able to encourage students to continue reading the lessons over the phone.[173] When schools were back in session the increase in variation of student ability likely made it harder for teachers to follow their planned curriculum, and introducing the tough choice of either teaching to the top or to the bottom of the class. While COVID-19 impacted all students, it did not do so equally, and we can see for certain teaching quality measures intervention schools were less impacted.

Caregiver perceptions of teaching quality

Caregiver perceptions of teaching quality are important for a few reasons. First, they are an indicator of the quality of an education in their child's school. Second, their perception can be indicative of broader community perceptions of education. Lastly, caregiver perceptions of teaching quality likely influence student attrition, attendance, and school support.

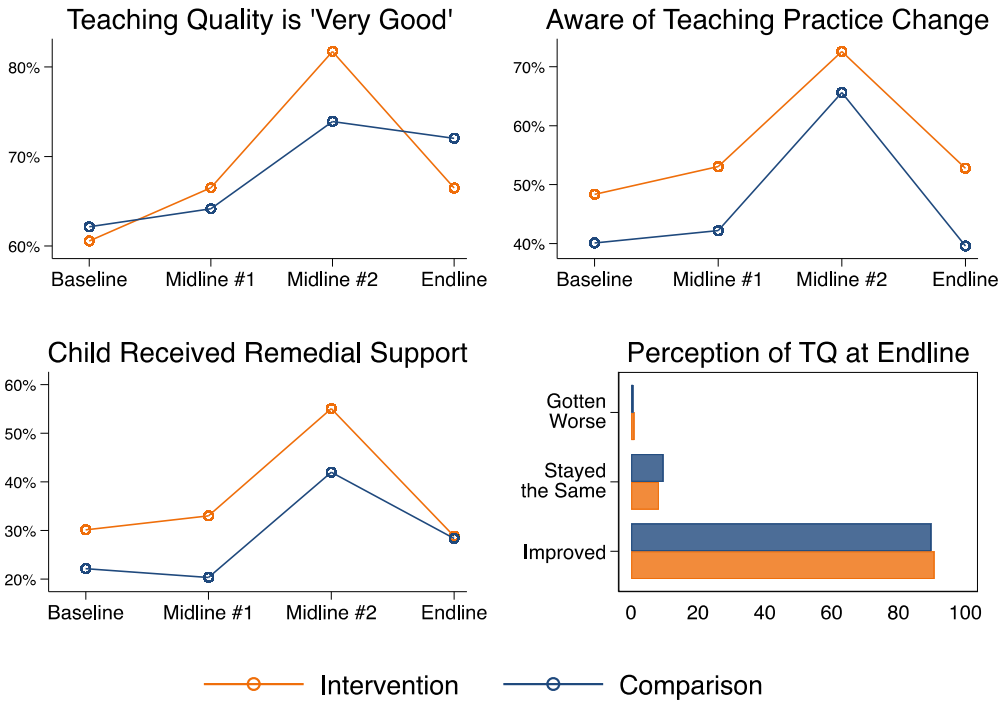
The rate of caregivers at the EL who believe teaching quality has improved in the last 12 months is nearly equivalent for intervention and comparison groups. Both groups overwhelmingly report teaching quality has improved. To better delineate perceptions of teaching quality improvements a binary variable was created from the categorical variable seen in Figure 38 that defined a positive perception as 'improvement' and a negative perception as 'stayed the same' or 'gotten worse'. This new definition indicates a 1.86 percentage point difference in positive perceptions of improvement as the endline of the intervention sample over the comparison sample.[174]

Caregivers on average perceive the quality of teaching in both intervention and comparison schools as 'very good'. Only a handful of caregivers reported teaching quality as poor. Out of 4,723 caregiver responses to this question, only 88 responded that teaching quality was very poor. Due to this skewed distribution, we created binary variable for caregivers' perception of teaching quality. We put all 'very good' responses into the first bin, and all other responses into the second. Intervention schools fared slightly worse from BL to EL, an insignificant difference in difference result of 2.5 percentage points. What is more concerning is the decline of caregivers reporting teaching quality as 'very good' from ML2 to EL. This negative trend from ML2 to EL is indicative of significant exogenous shock. One culprit for this decline could the increase in remote teaching during COVID-19. Caregivers' perceptions of teaching quality may have decreased because they believe remote teaching is lower quality. They could also be voicing frustration because they have had to provide

[171] FGD with Teachers, Int. 208
[172] FGD with Mothers, Int. 307; FGD with Teachers, Int. 201
[173] Vignettes with Girls, Int. 602; FGD with Teachers, Int. 205
[174] Intervention caregivers have an 87.99% perception of improvement and comparison caregivers had an 86.13% perception of improvement. Sometimes we need to redefine what is a positive response to a question because respondents could be afraid or embarrassed to admit outright their beliefs.

more childcare during school closures. SOMGEP-T schools are associated with a 7.1% additional decline in 'very good' teaching quality, as reported by caregivers, over and above that of comparison schools. And while this difference is not statistically significant, one possible explanation is that caregivers in intervention schools may have begun to expect higher quality teaching over time, as seen by a gradual but positive derivative seen from BL to ML2. The positive derivative of intervention schools is visually greater than that of comparison schools during that time. When schools closed the caregivers sending their children to intervention schools could have internalized higher quality teaching as the new normal, causing them to report worse teaching quality in the EL than their comparison caregiver counterparts.

FIGURE 38: CAREGIVER PERCEPTIONS OF TEACHING QUALITY



Caregiver perception of remedial support drastically decreased from the ML2 to EL for both comparison and intervention schools. This decline is inverse to the sustained increase seen by both groups from the BL to ML2. This decline is again likely due to school closures. As one teacher noted, "lessons were delayed, some of the students left to the countryside, [and] some others moved to other places, this caused a severe delay".[175] Remedial support is harder to offer remotely since it is difficult for teachers to present lessons visually as well as to monitor students' body language and understanding of material. There will be a greater need for remedial support as students go back to school because at home learning effects students unequally. Despite caregivers' perceptions more girls reported they get the necessary support from their teacher at the EL than they did at the BL. Intervention girls saw a 3.9 percentage point difference-in- difference result but the p-value was only 0.607, this shows that parents and children can have inverse perspectives and that intervention teachers could be better at giving support remotely than comparison teachers. This difference could be the

[175] FGD with Teachers, Int. 208

result of a difference in definition of support, because caregivers were asked about 'remedial support' whereas girls were asked only about 'support'.

Changes in teaching practices, as perceived by caregivers, has declined from ML2 to the EL. Caregiver perceptions of changes in teaching practices are important because they are often deciding if a girl is to enrol in school or not. An awareness of change in teaching practices is positive because it shows parental engagement in a child's education. Caregivers' awareness of changes in teaching practices from BL to EL improved for intervention schools and degraded for comparison schools. The difference and difference result is a 4.9 percentage point difference.

While Figure 38: Caregiver Perceptions of Teaching Quality shows a steep decline in caregiver perceptions of change from the ML2 to the EL this is inconsistent with the changes described in FGDs. School closures forced a huge change in in teaching practices and while caregivers are aware of these changes, they also likely to view them as ephemeral. This decline could be due to the lack of traditional classroom time for caregivers to evaluate for change.

Learner-centred pedagogy

TABLE 62: INDICATORS OF LEARNER-CENTRED PEDAGOGY IN INTERVENTION AND COMPARISON SCHOOLS FROM BL TO EL

	Intervention		Comparison		Diff-in-Diffs	P-value
Outcome	BL	EL	BL	EL		
Did NOT spend most of the time copying from the board	54.9%	63.2%	43.3%	52.5%	-0.1%	0.942
Did NOT spend most of the time repeating words aloud	71.8%	82.4%	55.0%	89.8%	-24.3%	0.021 *
Use of student-centred games/activities	21.1%	10.3%	25.0%	6.8%	7.4%	0.413
Students instructing each other	31.0%	16.2%	28.3%	8.5%	5.0%	0.616
Teachers ask open-ended questions	43.7%	38.2%	28.3%	28.8%	-0.06%	0.621
Teacher asks for student opinions	45.1%	30.9%	28.3%	22.0%	-7.9%	0.496
Sought to involve student who was not participating	50.7%	57.4%	35.0%	44.0%	-2.4%	0.845
Students worked in groups	22.5%	5.9%	21.7%	0.0%	5.0%	0.534

The indicators in Table 62 are taken from the classroom observation survey. The team leader recorded instances of each teacher or student action in three 15-minute blocks. These instances were then aggregated

into one dummy variable equalling one if the action was observed 2 or more times. Only two classroom observations out of 465 had all activities observed 2 or more times.

These indicators are focused on learner-centred pedagogy. Traditionally classrooms have been focused on the teacher and rote learning by students. Learner-centred pedagogy places the focus on the student, making the student an active participant of their own education. Copying from the board and repeating words aloud are examples of rote learning. Student-centred games, teamwork, active participation, and discussions are examples of student-centred pedagogy.

The coefficients in Table 62 do not share a common vector – there is a great variation in magnitude as well as direction. The inconsistency in coefficients could be a result of the wide variations in measurements. While intervention schools were associated with a larger increase in use of games and activities over and above comparison schools, they are also associated with a decrease in teachers seeking student involvement. While all of these active teaching methods are targeted by SOMGEP-T, the program is not associated with all of them in the same way. Some of the reasons that could cause this variation in coefficients is the small sample size. There are only 34 intervention schools and 30 comparison schools in EL to BL panel. These samples barely meet the sample size recommendations for a normal distribution.[176] Outliers in these small sample sizes can have large effects on coefficient magnitude and direction.

The only significant regression intervention coefficient is repeating words aloud. There was a smaller decrease of word repetition as a teaching method in intervention schools as compared to comparison schools, from the BL to the EL. Comparison schools started with a much higher rate of repetition in the BL, 45.0% as compared to 28.2% in intervention schools.[177] While comparison schools had lower absolute rates in the EL of word repetition than intervention schools the large difference and difference coefficient can be partially attribute to the pareto principle.[178] Reducing word repetition from 40% to 20% of class time could be easier than reducing it from the 20% to 10%.

Overall, three difference-in-difference coefficients are positive, two are practically zero, and three are negative. There were declines in active teaching methods from the BL to EL in a number of categories. Use of student-centred games, intra student instruction, asking student opinions, and group work all decreased from the BL to the EL for both comparison and intervention schools.[179] The decline was less for intervention schools as compared to comparison schools, as seen by their three positive difference and difference coefficients. A plausible explanation for these positive difference and difference coefficients is that intervention teachers are more resilient, due to SOMGEP-T's training, to changes in curriculum or other exogenous shocks.

Narrowing the lens to just ML2 to EL panel sample over that period, all active teaching methods, except for asking open ended questions, had positive intervention difference-in-difference coefficients. The coefficients range from 1.3% to 10.7%. While none of the differences were significant, it shows that due to the variability,

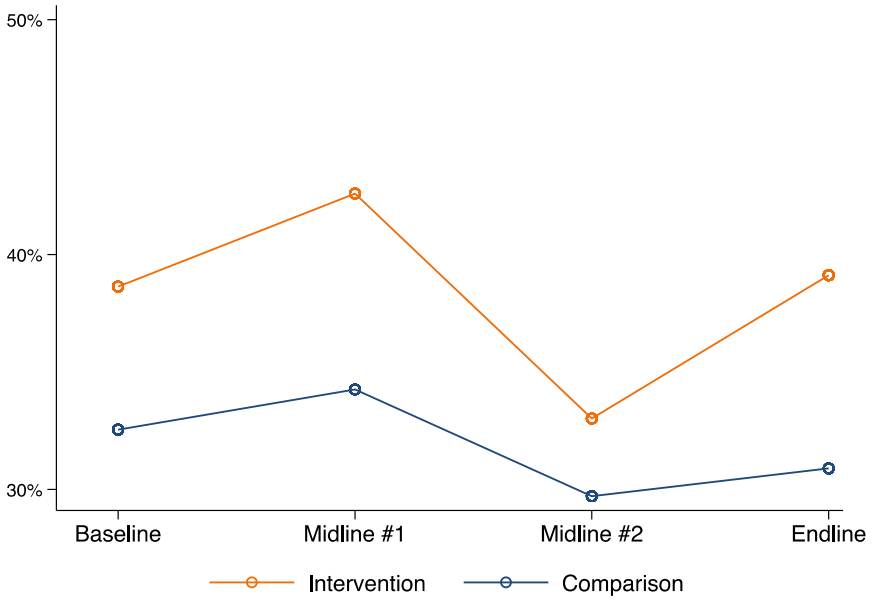
[176] N=30 is commonly recognized as the start of a fairly normal distribution.
[177] To avoid double negatives and ease understanding, the metric is recoded in this paragraph. The inverse of 'Did not spend most of the time repeating words aloud' is 'spent most of the time repeating words aloud'.
[178] The pareto principle in its simplest form states that 80% of the output is due to 20% percent of the input. Stated another way that last 20% percent of output requires 80% of the input.
[179] We can reject that the order of the questions in the survey tool caused the decline, since the order has stayed the same from BL to the EL.

and small sample size, there is little significance difference in either direction. The difference-in-difference positive coefficients for the ML2 to EL panel sample of 59 schools are as follows:

- 8.2% Did NOT spend most of the time copying from the board
- 2.6% Use of student-centred games/activities
- 5.4% Students instructing each other
- 10.7% Teachers ask open-ended questions
- 1.3% Teacher asks for student opinions
- 9.6% Sought to involve student who was not participating
- 6.8% Students worked in groups

While these coefficients are not significant, they are sign of intervention teaching methodology moving in the right direction from the ML2.

FIGURE 39: AGGREGATED INDICATORS OF LEARNER-CENTRED PEDAGOGY IN INTERVENTION AND COMPARISON SCHOOLS (ACTIVE INDEX)



To reduce the effect of a small sample size and variability in each metric we created an index.[180] Enumerators record the presence of each teaching method in three 15-minute time blocks. If an active teaching method is observed more than once (two times) the active teaching method dummy variable equals one. The dummy variables are then aggregated by observation into one index, the active teaching index. The index measures

[180] The index was created by taking the mean of each of eight indicators of classroom activity. The indicator ranges from 0 to 1 with zero meaning that none of the 8 indicators was observed and 1 being all eight indicators were observed. It is not realistic to expect that all eight of the classroom strategies would be used in a single classroom and therefore the score shouldn't be interpreted in an absolute sense. Instead, the index captures the relative diversity of active approaches being employed in comparison and interventions schools.

the number of active teaching practices observed (in a minimum of two observation blocks per classroom observation) out of eight distinct practices. It is scaled to range from zero to one, such that one represents a block in which all eight active practices are observed, and zero represents a block with no active teaching practices. Aggregating the methods into one help mitigates the effect of subject taught as well place in term period because both can change which teaching methods are used. The index better captures whether the teacher is an active instructor with learner centred methods, instead of whether or not they used a particular method on a given day.

As seen in the figure above, both comparison and intervention schools are recovering from a steep decline in the quality of teaching that was documented in ML2. Only intervention schools have surpassed their active index starting point of 38.6%; at endline the mean score among intervention school is 39.1%. In contrast the index score among comparison schools have declined slightly over the same period, from 32.5 % to 31.0%. Looking at just the ML2 to EL panel sample we get a 4% difference in difference active index coefficient. this difference is insignificant with a p-value of 0.377, but as shown by the figure above, intervention schools are recovering faster since ML2 than comparison schools.

TABLE 63: COMPARING COACHED INTERVENTION TEACHERS TO NON-COACHED INTERVENTION TEACHERS

Independent Variable	Effect on Active Index	P-value
Endline	- 2.4%	0.626
Coached	13.2%	0.013
Coached*Endline	-6.1%	0.417
Constant	37.2%	0.000***

Teachers were coached by CARE on classroom best practices, inclusive education, and boosting numeracy. This coaching is intended to increase equity in the classroom, as well as foster more productive student-teacher interactions and boost student numeracy skills. CARE's coaching overlapped to some extent with other training programs such as the MOEs but CARE's coaching is intense, broad, and recent.

Given the programme's emphasis coaching in their intervention, we should expect the programme's impact on teaching practices to be most concentrated among those teachers who received direct coaching, instead of assuming that all teachers will improve their teaching practices simply as a function of being in an intervention school. To assess this argument, we compared index scores between coached and non-coached intervention teachers as well as comparison teachers (all non-coached).

Coached teachers start with higher index scores for active classrooms than non-coached intervention teachers. This difference, as seen by the coached coefficient in the table below is large at the baseline – a finding that is surprising, given that teachers were selected for coaching in response to observed shortcomings in their teaching practices. Teachers selected for coaching had a 13.2% significant difference over non-coached intervention teachers at Baseline. The difference-in-difference coefficient, negative 6.1%, is not expected but also not significant. Coached teachers' active indices declined more than non-coached intervention teachers since the BL. Despite this decline, at the EL coached intervention teachers still employ 7.1% more active teaching methods than non-coached intervention teachers.

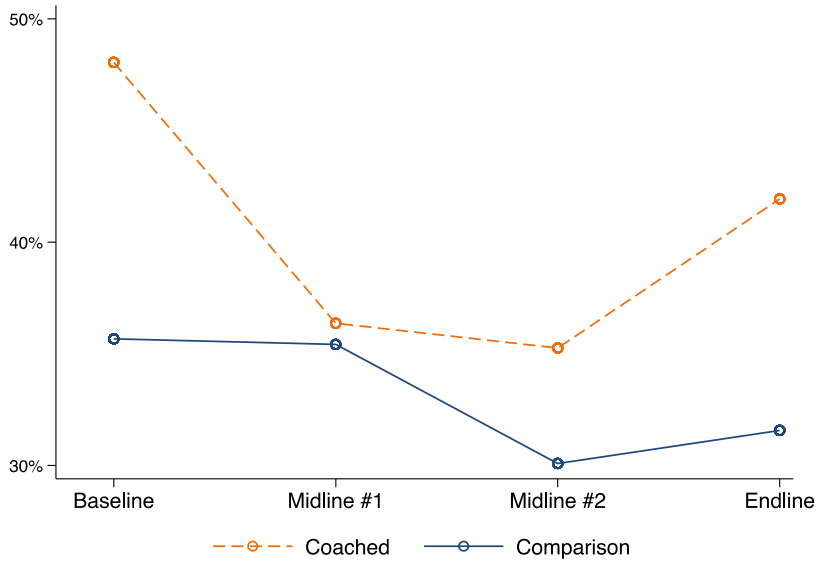
TABLE 64: COMPARING COACHED INTERVENTION TEACHERS TO COMPARISON TEACHERS

Independent Variable	Effect on Active Index	P-value
Endline	- 1.6%	0.685
Coached	17.3%	0.000
Coached*Endline	-6.9 %	0.296
Constant	33.1%	0.000***

Comparing coached teachers to comparison teachers produces similar results. No individual difference-in-difference active teaching metric coefficient was significant when isolated from the index. The coached intervention teachers had an even higher significant difference in active index at the BL over comparison teachers. At the EL coached teachers still employed 10.4% more active teaching methods than comparison teachers.

CARE's coaching does not appear to have a significant effect on active teaching methods. The effect of coaching on these eight teaching methods remains to be seen.

FIGURE 40: ACTIVE TEACHING METHODS, COACHED TEACHERS VERSUS COMPARISON TEACHERS



As seen in the figure above, coached teachers have a positive difference-in- difference regression coefficient result over comparison teachers from the ML2 to the EL. The difference and difference result from the ML2 to EL sample is 6.4% percent over and above comparison teachers. It is plausible that teacher coaching does not take full effect until ML2 which could justify the positive difference-in-difference regression result from the ML2 to EL but not from BL to ML2.

Formative Assessments

Formative assessments are useful because they allow teachers to restructure classes to cater to the level of understanding in the classroom. Formative assessments can be used often and spontaneously. They help students evaluate their understanding before it is too late to catch up. As one teacher put it:

> First of all, there are two books, one taught by the teacher and the other written by the student, so when I finish the lesson I ask if the students understand the lesson, if they understand I will go on and at the end of the week I give a quiz to assess whether the students understand the unit or not, so if you find out that the students understand and will teach more than that, but if they don't, you will have to change the whole process, for example in Arabic or English you will say explain in Somali to me, and then have to translate it back to the language it was.
>
> - FGD with Teachers, Int. 207

Our main measure of formative assessment use is based on three questions asked of teachers following classrooms observations. Self-reports of formative assessments can be problematic because teachers may report higher than actual usage but teachers were also asked if they have records of formative assessments and to describe the formative assessment they have used. During the BL evaluation the question structure was slightly different. These three questions were asked three times, once during each observation block. To be able to compare formative assessments across rounds, we aggregated the blocks from the BL into one dummy variable to mimic the variable in subsequent rounds.[181]

54.6% of teachers reported they used a formative assessment at least once during the BL. Because we combined new dummy variable in the EL with the dummy variable used in ML1, ML2, and EL we can then regress using the BL to EL panel. The intervention schools had an 11.7% increase of formative assessments over and above comparison schools but it is not significant with a p value of 0.297. This is a substantial difference above comparison schools and the lack of significance may be because the method of measurement changed from the BL to the EL therefore the outcome may not fit well in a regression.

Narrowing the panel to just the ML2 to the EL is a may be more accurate determination of the effect of the intervention on formative assessment use because the question was asked in a different way but despite narrowing the window, the intervention effect is even less significant and the difference-in-difference intervention coefficient is even slightly negative.

Teachers and enumerators may be confused on what formative assessments are. Some teachers understand formative assessments can be as simple as "questions for the students to evaluate their understanding".[182] Other teachers report using formative assessments but when are asked to describe the formative assessment 17.2% of teachers said they administered monthly assessments. One teacher in the survey mentioned they administer a formative assessment "every three months".[183] Monthly assessments can be formative but teachers are unlikely to be able to adjust the lesson back a month to cater to struggling students.

[181] If at any point during the three observation blocks an enumerator observed formative assessment the new dummy variable is set to one.
[182] Teacher FGD, Int. 210
[183] Teacher in Classroom Observation Survey

Enumerators asked for records of written formative assessments during each round. From ML2 to EL intervention teachers' records of formative assessments decreased 18.6% over and above comparison schools. One hypothesis for this decrease is intervention teachers might have a gained a better understanding of formative assessments than comparison teachers since the ML2. Therefore, intervention teachers are no longer providing summative assessments as records to enumerators that might not be able to tell the difference. Intervention teachers also could be using oral formative assessments instead of written formative assessments so they do not have paper records. One teacher noted, "I make sure the students understood the previous lesson because I ask questions".[184] An increase in oral formative assessments could possibly explain this substantial decline seen in the intervention schools.

Despite a decrease in records of formative assessment, one teacher noticed a change in their behaviour, mentioning that they used to just write on the board but now they ask students to write so each can understand their writing proficiency.[185] Another teacher understands the motive for giving formative assessments, indicating when "we find out the results we can do a revision for the students to improve".[186] The FGDs with teachers established that many understand the utility of formative assessments, a sign they likely use them.

Corporal Punishment

Corporal punishment is not a productive teaching method. While it may be effective in controlling a classroom it does not aid learning. Fearful students have a harder time engaging as well are more likely to drop out. Corporal punishment does not fix the root cause of student behaviour issues, and in many cases can exacerbate it. A focus in SOMGEP-T is to reduce corporal punishment as a form of discipline in the classroom.

TABLE 65: CORPORAL PUNISHMENT IN THE CLASSROOM

	Intervention		Comparison		Diff-in-Diffs	P-value
Outcome	BL	EL	BL	EL		
Observed use of physical punishment (toward any gender)	67.6%	1.7%	88.1%	1.8%	20.4%	0.011 *
Observed use of physical punishment (toward girls)	66.2%	1.5%	86.2%	1.7%	19.9%	0.013 *
Student report – Teacher corporally punished girl in last week	43.2%	23.0%	25.9%	18.3%	-12.6%	0.221
Student report – Teacher's discipline or punish students who get things wrong in a lesson	76.9%	54.0%	76.9%	52.0%	2.0%	0.740

[184] Teacher in Classroom Observation Survey
[185] Teacher FGD, Int. 210
[186] Teacher FGD, Int. 209

Student report – Teacher corporally punished other students in the last week	65.2%	54.0%	49.1%	52.1%	-14.3%	0.274
Student report – feels afraid of the teacher, agrees a lot	37.3%	45.3%	36.0%	52.0%	-7.9%	0.364

The top two outcomes in the table above come from the classroom observation survey. The bottom four outcomes come from the household survey. The two outcomes measured in the classroom observation were recorded in three 15-minute observation blocks as described previously in this section. The four outcomes from the household survey were posed as questions to intervention and comparison girls. The outcomes were then placed on a binary scale – any corporal punishment is unacceptable.[187]

The two significant intervention effects are observed use of physical punishment on all genders and observed use of physical punishment on girls. The programme impact is 20.4% for physical discipline on all genders, and 19.9% for girls. This difference is large and is possibly aided by a high level of observed corporal punishment in comparison schools during the BL as compared to intervention schools. Enumerators observed 20.5% more physical punishment on any gender at comparison schools over intervention schools at the BL. Both outcomes have a lower limit of 0.0% therefore this intervention effect could be attributed to the large difference in observed rate of corporal punishment in the BL as well as the programme effect on intervention teachers.

Students reported much higher rates of corporal punishment in the EL than were observed by team leaders during classroom observations. It is plausible that teachers limit their physical discipline in front of outside observers therefore the observed rate of physical discipline in the EL is not representative of routine classroom discipline.

While the consensus in teacher FGDs is that corporal punishment is detrimental to a student's education a few teachers are still unsure. One teacher indicated use of corporal punishment is acceptable depending on the reasons and another said "the teacher is meant to hit the student with a small stick [so] that it does not hurt" to make the student behave.[188] Despite these outliers the majority of the qualitative data supports a changed perception of corporal punishment. Multiple mothers suggest corporal punishment has declined in schools because teachers do not beat students like they used to.[189] However, some mothers also still rationalize the use of corporal punishment in the classroom. One mother believes "if the teacher does not beat the students, then they will not learn well".[190] For a girl to admit she has been physically punished might be an admission of guilt or wrongdoing, so rates of self-reported abuse could be biased particularly with a caregiver present.

[187] Experiencing corporal punishment once or twice in the last week or every day were binned together for ease of interpretation of regression results.
[188] FGD with Teachers, Int. 208; FGD with Teachers, Int. 207
[189] FGD with Mothers, Int. 301; FGD with Mothers, Int. 301
[190] FGD with Mothers, Int. 310

The number of intervention girls experiencing corporal punishment in the last week declined more for intervention girls than it did for comparison girls; the difference-in-difference coefficient is negative 12.6%. Girls overall did not report the same large decline in corporal punishment on their peers. For comparison schools rates of corporal punishment on peers actually increased from the BL. Intervention girls again over and above comparison girls reported fewer instances of corporal punishment on peers, an intervention effect of 14.3%. Girls could be embarrassed to admit they have been physically punished in front of their caregivers but will candidly report physical punishment as received by their peers.

Girls also report they are more fearful in the EL than they were in the BL. The difference-in- difference result are negative 7.9% but this finding is not necessarily causal due to the coefficients p-value of 0.364. Notably, feelings of fear increased considerably among girls in both intervention and comparison schools, which could reflect a structural or systemic cause. One explanation may be that the disruption of schools due to COVID-19 has increased stress on teachers, pressured them into moving more quickly in their lessons, or increased misbehaviour (or perceived misbehaviour) among students who have not been in class as consistently over the two years prior. Given that school closures temporarily eliminated many teachers' livelihoods, this may have produced a situation in which teachers are less accommodating or more stressed in general.

Of these six metrics across both intervention and comparison groups from BL to EL, there are only two instances of increasing rates of corporal punishment – student reported rate of physically disciplined classmates and girls reported feelings of fear. The only increase for intervention schools from BL to EL is in girls' feelings of fear. SOMGEP-T may not have completely eliminated corporal punishment in intervention schools but the explicit public perception of corporal punishment has changed. Teachers, girls, and caregivers are less accepting of corporal punishment as a standard teaching practice and while the trends vary across the different metrics, the overall trend indicates physical discipline is decreasing, and decreasing at a greater rate at intervention schools, a sign of progress and improved teaching quality.

Classroom Demeanour

TABLE 66: CLASSROOM ENVIRONMENT

	Intervention			Comparison			Diff-in-Diffs	P-value
Outcome	BL	ML2	EL	BL	ML2	EL		
Harsh language not observed	84.5%	-	87.9%	89.5%	-	90.9%	2.0%	0.815
Very respectful language observed	-	85.3%	77.4%	-	70.9%	87.3%	-24.2%	0.02*
Teacher used student names	-	48.3%	56.5%	-	54.6%	38.2%	24.5%	0.064
Student report – feeling welcome by teachers, agree a lot*	75.1%	-	89.1%	74.3%	-	89.4%	-1.0%	0.931

*Zone is controlled for

The classroom environment is important as it needs be safe enough for students to step out of their comfort zones and mature. We measured classroom environment during classroom observations as well as by asking girls if they felt welcome.

The most significant difference and difference result between intervention and comparison classrooms is the observed use of respectful language from ML2 to EL. Comparison school teachers significantly increased their use of respectful language over and above intervention schools. The difference is large – 24.2%. This is consistent with the use of harsh language. Harsh language increased from the BL to EL more for intervention schools than it did for comparison schools.

The next most significant difference and difference result is with a practically significant p value of 0.064 is 24.5% increase in use of names in interventions schools over that of comparison schools. Names show teachers care, are prepared and support individuality.

Harsh language use is inversely related to corporal punishment at a nearly significant level. The presence of harsh language at the EL is associated with an 8.1% decrease in corporal punishment at 0.056 significance. Teachers may resort to harsh language as an alternative to physically disciplining students. One enumerator reported a teacher yelled "Hey big head put your head on the table, I don't want to see you not be able to answer that question again," at a control school.[191]

Intervention girls reported a 4.1% increase in shouting over and above comparison girls from the BL to the EL. This could be due to the decrease in corporal punishment in intervention schools as described in the previous sections. Despite an increase in shouting students reported feeling safer from the BL to the El. Intervention girls reporting feeling unsafe at school decreased 5.6% from the BL to the EL.

Intervention classrooms had a noticeable increase of observed name use from ML2 while comparison name groups had a major decrease. While this difference does not meet the 0.05 significance cut off it is close enough to draw a few inferences. The first is that despite the interruptions from covid induced closures, intervention teachers were able to remember and use student names, indicating higher teacher engagement and preparation than comparison teachers.

The most important metrics – girls reporting feeling safe and welcome – are positive indicators demeanour has improved since the BL. It is not possible to derive a significant causal relationship between CARE's program and classroom demeanour increasing positive perceptions of safety is progress towards increased teaching quality.

[191] Enumerator 301, School 221

Gender Equity in the Classroom

TABLE 67: STUDENT ENCOURAGEMENT BY GENDER

	Intervention		Comparison		Diff-in-Diffs	P-value
Outcome	BL	EL	BL	EL		
Teacher provides encouraging feedback to boys	59.2%	64.7%	63.3%	62.7%	6.2%	0.613
Teacher provides encouraging feedback to girls	60.6%	64.7%	61.7%	61.0%	4.8%	0.695
Student report – teacher asks the same number of questions to girls and boys	93.2%	91.5%	88.1%	92.9%	-6.6%	0.043 *
Student report – teacher asks boys harder questions than girls	8.4%	6.8%	8.1%	8.3%	-1.8%	0.615

A classroom with equal emphasis on girls and boys' education has gender equity. Girls are cognizant of inequity in teaching practices, and can get discourage if teachers encourage boys noticeable more than girls. One of SOMGEP-Ts goals is to help teachers increase gender equity in education.

We measure gender equity through classroom observations of encouragement as well as girls reports of teacher questions. Girls were asked if the quantity of questions were the same for both genders as well as the difficulty of the questions.

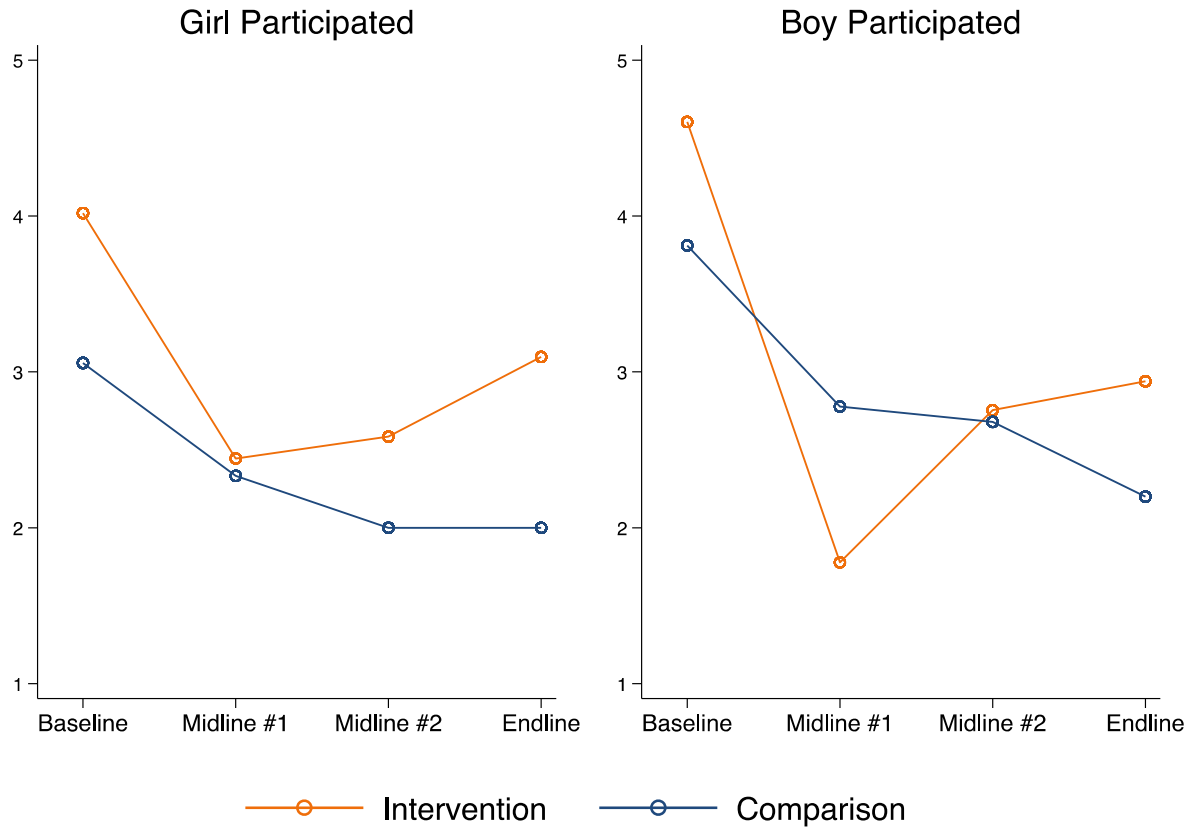
Observed encouragement in intervention classrooms increased 6.2 percentage points for boys, and 4.8 percentage points for girls, over the comparison group. These difference and difference results are not significant but they are consistent with each other. Enumerators also observed teachers encourage boy and girls at roughly the same rate at both the EL and BL and at both intervention and comparison schools.

Encouragement is an important part of any education. One girl mentioned that teachers encouraged her to continue learning even after schools closed due to covid.[192] Encouragement can increase drive and ambition. One mother said, "if you encourage the child and give them an award, they might be encouraged to continue their education".[193] Encouragement can also increase participation, as seen in **Error! Reference source not found.**. Another mother indicated schools are doing more training to encourage girls to speak in front of the class and to join discussions.[194] This increased encouragement, a form of positive reinforcement, is likely a benefit to the classroom and learning outcomes.[195]

[192] Vignettes with Girls, Int. 602
[193] FGD with Mothers, Galmadug, Int. 310
[194] FGD with Mothers, Galmadug, Int. 310
[195] Homme, L., 1970. How to use contingency contracting in the classroom.

Girls in intervention schools reported a decrease in gender equity in quantity of questions posed to each gender from the BL to the EL. Girls in comparison schools saw an increase. The difference-in-difference result is negative 6.6%. The difference and difference result for girls reporting girls are asked more questions is 1.8%. The equivalent result for girls reporting boys are asked more questions is 4.9% which is significant at the 95.4% confidence level. This is an unexpected programme impact but it is possible that teaching quality has improved in intervention schools, increasing the total number of questions posed to students but an increase in questions could exasperated a gendered divided.

FIGURE 41: STUDENT PARTICIPATION BY GENDER



SOMGEP-T target gender equity in the classroom, doing so both through teacher trainings and girls' trainings. One teacher noted they are assessing students on how they participate in the classroom, and one mother indicated the school is encouraging girls to "participate in every activity in school as well as in the community".[196] Increasing confidence in girls could increase girls' participation, another SOMGEP-T target. Girls were asked if they felt confident answering questions in class in the household survey. Intervention schools had a 10.6 percentage point larger gain of girls who stated the felt strongly confident over and above

[196] FGD with Teachers, Int. 202; FGD with Mothers, Int. 304

comparison schools. This difference is nearly significant with p-value of 0.077 and it is also a substantial increase over comparison schools.

The trend for intervention schools appears to be in the direction of pre-ML1 participation levels, as seen in the figure above, which depicts the average number of times a student participated, disaggregated by gender and intervention. As seen by the figure, intervention schools fared better after ML1 than comparison schools. Programme impact on student participation looks to have started after the ML1. Girls, more than boys, are observed participating more.

TABLE 68: TEACHER PERCEPTIONS OF GENDER IN THE CLASSROOM

		Intervention		Comparison		Diff-in-Diffs	P-value
Outcome	Answer	ML2	EL	ML2	EL		
I have the same expectations for boys and girls in class	Agree Strongly	80.3%	77.4%	80.0%	70.9%	6.2%	0.577
Boys and girls are better at different subjects	Disagree Strongly	11.5%	21.0%	16.4%	18.2%	7.7%	0.436
It's important to design lessons that are gender-sensitive	Agree Strongly	63.9%	61.3%	63.6%	54.6%	6.4%	0.617

A teacher's perception of gender in the classroom is important because there is a power dynamic between teachers and students. Teachers have ability to control some of the gender dynamic in the classroom. While teachers can always be explicitly gender sensitive in their answers to our questions, but implicitly not, setting a standard of what is expected of teachers is progress towards the positive outcome of gender equity.

The percentage of teachers who agree strongly they have the same expectations for boys and girls has decreased from ML2 to EL for both intervention and comparison schools. Intervention schools managed better with a difference and difference coefficient of 6.2 percentage points. We cannot know for sure why there was a decline – it could be due to teachers expecting less of boys.

The number of teachers that disagree strongly boys and girls are better at different subjects increased from ML2 to EL. Intervention schools increased 7.7 additional percentage points over and above comparison schools. Despite this increase the percentage of teachers at the EL that disagree strongly in both groups is still quite low. Boy and girls could be better at different subjects because they are nurtured by caregivers and teachers to focus on different subjects. Teachers that disagree strongly may be disagreeing with a difference in innate ability, neither boys nor girls are born with learning abilities better than the other.

The percentage of intervention teachers that agree strongly lessons should be designed to be gender-sensitive decrease for both comparison and intervention schools from the BL to the EL. The difference-in- difference result is 6.4%. Comparison schools suffered a decline over and above intervention schools but not in a

significant way. Regressing teachers who received inclusive pedagogy training against comparison teachers also did not yield significant results on any of these three metrics.

FGDs support gender equity has improved over time. Mothers perceive that equity has increased in the classroom. One mother noted, "[teachers] do not choose pupils based on gender and instead ask all students for their opinions without taking gender into account".[197] While a mother told the focus group she had unenrolled her daughter because the teacher was discriminating against the child, most describe improvements since the BL.[198] An increased awareness of gender inequities in education may increase the reported rate of inequity initially but overtime will likely have positive effect.

The metrics in Table 68: Teacher perceptions of gender in the classroom might not tell the whole story. Teachers who do not disagree strongly that boys and girls are better at different subjects could be openly evaluating girls' and boys' current knowledge of different subjects instead of evaluating their intelligence or ability to learn. They could also be indicating girls are actually better at some subjects than boys. One teacher told the team leader during an FGD:

> We have received training from CARE about educating and improving of girl's education. Such as advising girls on how to reach high level as boys and consulting us on challenges.
>
> FGD with Teachers, Int. 208

Therefore, as the quote above indicates, it is plausible that teachers are agreeing there are inequities in education instead of different innate learning abilities between boys and girls.

Equal access to classroom materials is another important step towards gender equity but regressing another measure, equal access to desks and learning materials, does not yield a significant difference. Regressing the presence of mixed gender seating arrangements in the classroom yielded a difference and difference result of 5.7%. Not a significant result but an additional sign that gender-equity could be improving in intervention classrooms.

Teacher Effort

Teacher effort is important part of teaching quality. Teacher preparation and communication are key indicators of teacher effort. Students are likely to reciprocate teacher effort in the classroom therefore high teacher effort motivates students to work hard. A high amount of teacher effort indicates they care about their job as well as the education of their students.

TABLE 69: TEACHER PREPARATION

	Intervention			Comparison		
Outcome	ML1	ML2	EL	ML1	ML2	EL

[197] FGD with Mothers, Int. 305
[198] FGD with Mothers, Int. 309

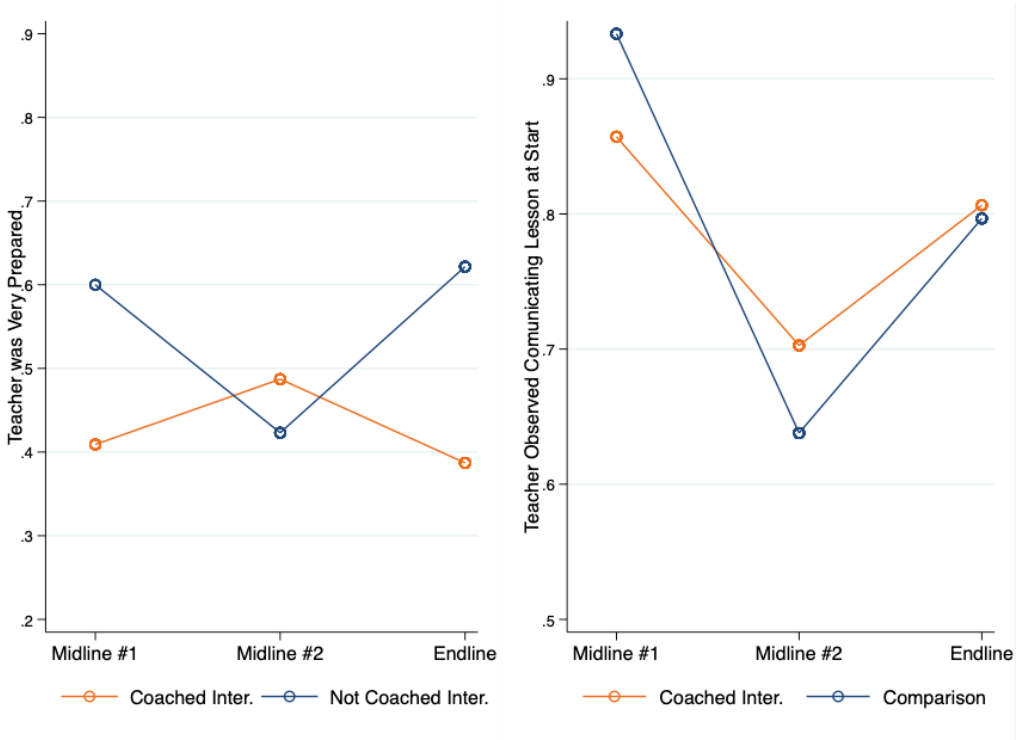
Was teacher very prepared and had a plan for the lesson you observed?	48.2%	52.8%	46.3%	59.3%	37.7%	43.4%
Did teacher communicate lesson objective at beginning of class?	80.8%	76.5%	74.1%	96.3%	65.4%	79.3%

Effort was measured by the team leader during classroom observations using two indicators – teacher preparedness and teacher communication. Communication was also observed by the team leader, in a binary – was the teacher was observed communicating the lesson plan or not? Communicating the lesson plan is important to setting expectations in the classroom. These two questions were not added into the tool until ML1 so the analysis will be on a smaller panel. To aid in the analysis, teacher preparation was made into a dummy variable – one for a very prepared teacher and zero for a somewhat or not prepared teacher.

Teacher preparation declined from the ML1 to the EL. The decline was not as great for intervention schools as it was for comparison schools. Neither a regression model comparing intervention schools and comparison, coached versus not coached produced significant results. As seen by Figure 42, coached teachers' preparation appears inversely correlated to non-coached intervention teachers and coached teachers' communication is marginally correlated to comparison teachers. These differences are not significant indicating coaching has not had a clear effect on either communication, preparation, or teacher effort.

Teacher effort can take form in many other ways such as diligent grading, following up with struggling students, and adjusting lesson plans. Another metric is whether a teacher reports they can identify struggling students before they take the exam, a good indicator they have taken the time to distinguish each student's knowledge and learning ability. Intervention teachers were associated with an additional 9.9 percentage point increase in agreement they can identify struggling students over and above comparison teachers from ML2 to. It was not significant with a p-value of 0.140, but it demonstrates there are other metrics for measuring teacher effort. Along with the increased ability to identify struggling students, intervention teachers also started class on time 1.7 percentage points over and above comparison teachers since ML2 but the p-value is large at 0.795. The two additional metrics support that intervention teachers were better than comparison teachers, but a cause has not been isolated.

FIGURE 42: TEACHER PREPARATION AND COMMUNICATION



Teacher effort is also contingent on school quality. Mothers during FGDs indicate school quality has increased since the BL, one saying "a lot of things have changed in the last few years, including improvement in the quality of education and school maintenance".[199] A teacher also mentioned renovations and a new water tank.[200] Teacher quality and effort likely goes up with the quality of the school.

Student reported teacher absenteeism has decreased since the BL. The decrease is slightly greater for comparison schools over and above intervention but it is not significant. Low pay can increase absenteeism. Some teachers in FGDs discuss their poor pay, one suggesting "sending the money to the teachers instead of the students, while also supporting the school facility".[201] Paying teachers more to show up is important, but what they do when they show up is as important. It is their pedagogy that influences students' education.

Progress to Teaching Quality Targets

CARE, through SOMGEP-T, emphasised two teaching quality targets. The first is the use of formative assessments to aid numeracy and literacy outcomes in the classroom. The target, a 60% increase in the proportion of teachers using formative assessments, was achieved. At the endline, 82.4% of teachers reported using formative assessments. While the story is more complex than the reported increase in formative assessment use from the EL to BL, the qualitative data supports this finding. Formative assessments are now a teaching best practice. Now that the majority of teacher have applied formative assessments to their teaching

[199] FGDs with Mothers, Int. 308.
[200] FGDs with Teachers, Int. 208
[201] FGD with Teachers, Int. 210

pedagogy the remaining teachers are likely to follow. Teachers also appear to be using formative assessments not just because they are told to do so but because they recognize the assessments utility. This recognition will help the sustain and potentially compound the teacher trainings after they end.

The second teaching quality target, a shift in teacher perceptions of teaching quality, was measured in four ways: perceptions of corporal punishment, equity in expectations of girls and boys, remedial teaching, and gender-sensitive lessons. These metrics were measured from the ML2 to the EL. At the endline, each measure for intervention schools are as follows:

- Increased to 82.4% of teachers disagree strongly that corporal punishment is sometimes necessary.
- Increased to 83.8% of teachers agree strongly corporal punishment slows down learning.
- Decreased to 76.5% of teachers agree strongly they have the same expectations of girls and boys.
- Increased to 94.1% of teachers agree strongly that they adjust lessons to help struggling students.
- Decreased to 80.9% of teachers agree strongly it is important to design lessons that are gender-sensitive.

For the two outcomes that decreased from the ML2, expectations of girls and boys as well as gender sensitive lessons, the difference and difference results had positive coefficients – the decline for intervention schools from ML2 to the EL was less than the decline for comparison schools.

A shift in teacher perceptions of teaching quality is hard to measure. This target is not as conclusive as formative assessments. Perceptions, as measured by the perceiver, are explicit. They might not be a completely accurate determine of how teachers feel, teachers may hold implicit perceptions that they do not share with enumerators. Despite the possibility of this bias, in general intervention schools fared better than comparison schools in these four metrics from the ML2 to the EL.

The qualitative data supports an increase in teaching quality. The consensus in teacher FGDs is training is positive – "Students' education has improved greatly as a result of their teachers receiving training".[202] When some teachers were asked to elaborate, they simply talk about their increased knowledge and experience, not a clear sign of a causal effect of the intervention.[203] Other teachers gave specific ways they have adapted and evolved their teaching methods, one stating,

> Firstly, teachers have developed lesson preparation, and secondly, they have been able to teach lessons equally to boys and girls.
>
> FGD with Teachers, Int. 208

Another teacher responded that teacher supervision aids teacher preparation.[204] Some teachers may only be come prepared to class with continued supervision other teachers may benefit from training, including sustainably improving "the quality of lesson explanations".[205]

Teaching quality standards have changed since the BL. While the causal effect of CARE's intervention is hard to discern, if teachers, students, and mothers are noting a change, supported by positive difference and difference coefficients in numerous metrics, practically speaking, the intervention has improved teaching

[202] FGD with Teachers, Int. 205: FGD with Teachers, Int. 210;
[203] FGD with Teachers, Int. 209
[204] FGD with Teachers, Int. 202
[205] FGD with Teachers, Int. 205

quality. An additional evaluation of teaching quality in a few years could define the sustainability of this change as well potentially reveal lagged effects of teaching training that were not captured in our BL to EL panel sample.

Life Skills

Outcome	Indicator	BL	EL
YLI	Median Youth Leadership Index Score[206]	50.6%	59.6%
Life Skills	Life skills Index Score for ISG Intervention Girls over 12	71.7%	78.2 %
GEF	Girls Empowerment Forum– have girls ever participated in GEF activities[207]	13.5%	74.7%

Main qualitative findings

The qualitative data supports a positive trend in life skills as well as youth leadership. Intervention girls respond positively to girls' empowerment programs, training, and encouragement. Girls also report they are passing on what they have learned to other girls in the school, as well as boys. One girl noted, "We are the best students. Now I give training to the lower-level girls".[208] GEF girls are also responsible for spreading awareness on issues ranging from sanitation and self-hygiene to school preparation, which several girls described as a core focus of their GEF membership.

Girls report feeling they need continued support, admitting they can feel aimless without additional training. The effect of initial training and encouragement appears to wane over time as girls matriculate out of the forum.

Some girls noted GEF participation in activities such as picking up trash, planting trees, and painting, to improve their school's environment. These appear to be the primary GEF activities after raising awareness. A few hoped to receive support from boys and teachers cleaning and sanitizing the schools.[209] Hygiene and sanitation improvements were also mentioned by several girls, with one girl stating, "we help the community and the school with cleaning because when everything is not hygienic in the communities and schools, a lot of diseases can be born".[210] At times, girls mentioned performing more general cleaning; this was not the intent of the GEFs, and it may have occurred in only one or two schools. In some cases, discussions of cleaning may have been conflated with more specific hygiene enhancements in the school.

Many girls also note positive effects of the GEF – increased enrolment, fundraising, school pride, inter-student instruction, and conflict resolution. Many girls also talk positively of cleaning up the environment, from gardening, painting, and fixing furniture – all examples of environmental improvements that are likely empowering.[211]

Developing life skills during childhood and adolescence increases the chances of living a happy and productive life. Increasing life skills, particularly for disabled students, has positive associations with better transition and life outcomes.[212] Increasing agency and desire to learn, can increase transition rates, and motivation. Life skills is a fairly broad category, but CARE has specifically targeted certain life skills for intervention and evaluation.

SOMGEP-T established Girls' Empowerment Forums (GEFs) in intervention schools. Girls received training in leadership, decision making, agency, school governance and financial literacy. Along with GEFs, CARE

[206] The YLI score is created from the index. These observations are taken from the intervention group in the full panel.
[207] Intervention Girls in BL to EL Panel
[208] Girls Networking Exercise, Int. 403
[209] Girls Networking Exercise, Int. 408; Girls Networking Exercise, Int. 410; Girls Networking Exercise, Int. 403
[210] Networking, Int. 405; Interview with Fieldwork Team Leader, Team 5; Networking, Int. 406
[211] Girls Networking Exercise, Int. 407; Girls Networking Exercise, Int. 404; Girls Networking Exercise, Int. 410; Girls Networking Exercise, Int. 405
[212] Patton, J.R., Cronin, M.E. and Jairrels, V., 1997. Curricular implications of transition life skills instruction as an integral part of transition education. *Remedial and Special Education*, *18*(5), pp.294-306.

introduced Alternative Basic Education programming (ABE). The target of ABE is to increase out of school (OOS) girls' foundation skills and competencies to either smooth transition into formal school settings or to better life outcomes. These girls were OOS girls at the BL and were generally pastoralists or poor girls. The last sample that we will analyse is Alternative Learning Program (ALP) girls. These girls were added into the panel at the first midline from 35 ALP centres in intervention communities. There is no logical comparison group for ALP girls.

SOMGEP-T can directly increase girls' life skills through training, encouragement, and soft skills development. SOMGEP-T can also indirectly increase girls' life skills through training of teachers, changing negative perceptions of gender, as well as increasing gender equity. Changing girls' attitudes, such as increasing their self-confidence, can have broad effects on life outcomes.

CARE has an established logframe indicator, Youth Leadership Index (YLI), that measures self-perceptions of five leadership competencies (self-confidence, vision, organization, voice and decision-making). We have also used an additional metric for life skills, the Life Skills Index (LSI) standardized across the GEC. These two indices will be the primary measures we compare intervention and comparison schools against.

This final analysis will evaluate the change in life skills from BL to EL focusing on the 1,930 intervention and comparison girls that were contacted in both the BL as well as in the EL. This group of girls will be disaggregated in numerous ways, comparing intervention to comparison girls, ISG girls to OOS girls as well breaking girls up by age.

The BL to EL analysis will be supplemented by and ML2 to EL analysis for ABE girls and a ML1 to EL analysis for ALP girls. Girls participating in GEFs will also be disaggregated to isolate the effects of these forums. The overlap of the Youth Leadership Index and the life skills index will be addressed and rationalized in the subsequent sections.

Youth Leadership Index

The Youth Leadership Iindex is derived from 21 questions enumerators asked during the household survey. All but four questions are asked of every girl interviewed in the EL. In the BL we administered the YLI only to ISG girls, therefore much of the YLI analysis will focus on ISG girls. The choice structure and order of these questions has stayed the same over all four rounds. This measure was curated by CARE as an indicator of self-perceptions of leadership and a proxy for girls' agency.

> The Youth Leadership Index, YLI, was designed specifically by CARE to longitudinally measure changes in self-perceptions of leadership among youth, specifically those aged 10-17. The questions in the YLI ask youth about their self-confidence, their decision-making, problem solving and organizational skills, their sense of voice, and their ability to motivate others. The YLI also measures cooperation, diligence, independent thinking, personal responsibility, and leadership interest.
>
> -www.care.org[213]

[213] https://www.care.org/wp-content/uploads/2020/05/CARE-YLI-Toolkit-FINAL-WEB.pdf

Girls are given the option to answer each question, as written below, on a four-point scale: rarely, sometimes, most of the time, and almost always. The answers do not need to be recoded since "almost always" is a positive response to every question in the YLI.

TABLE 70: YOUTH LEADERSHIP INDEX QUESTIONS

Questions in Order
1. I like to try new activities that I may not know how to do.
2. My friends ask me for advice.
3. I recognize when people have different skills to contribute to a task.
4. I am comfortable when my teacher calls on me to answer a question. **
5. I contribute ideas to discussions at home even if they are different from others' ideas.
6. I ask questions at school when I don't understand something. **
7. I can describe my thoughts to others.
8. The things I do set a good example for my peers.
9. I consider possible outcomes of my decisions before making them.
10. I accept responsibility for the outcomes of my decisions.
11. I recognize when choices I make today can affect my life in the future.
12. I can show what is important to me with my actions.
13. If someone does not understand me, I try to find a different way of saying what is on my mind.
14. I encourage others to join together to help my community.
15. I cooperate with others to get things done at home.
16. If someone treats me unfairly at school, I am comfortable telling an adult. **
17. I am willing to work hard to achieve my dreams.
18. I am better able to finish a task when I plan ahead.
19. When I have the opportunity, I can organize my peers to do an activity.
20. I am interested in being a leader at my school. **

21. I try to understand the cause of a problem before trying to solve it.

** indicates these questions where only asked of in-school girls. Bold questions are ones that closely overlap with a life skills index question.

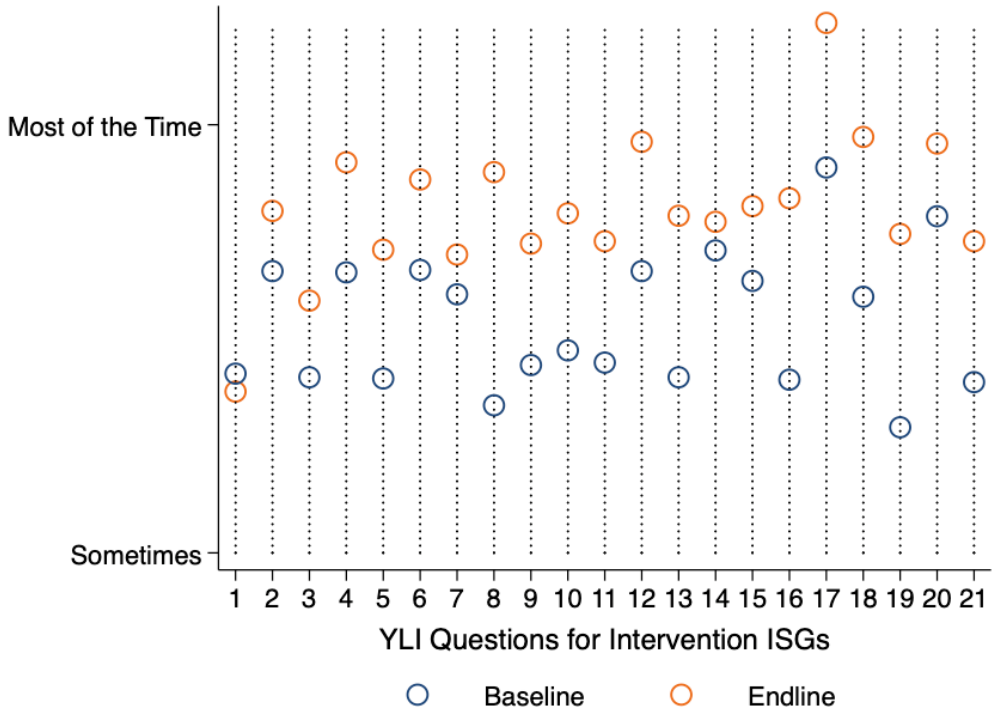
Giving girls reason to learn, to delay gratification, or disobey regressive norms can better life outcomes. One CEC member described their reason for supporting girls' education saying "in the past, many believed that girls did not have the right to education but now we understand women who are educated can better support their children in the future".[214] This member exemplifies how even positive changes in attitudes can be based in antiquated and gendered expectations. The expectation in this case that most women will bear the 5.9 children, the average in Somalia.[215] Education should not only be regarded as a tool for child rearing. Increasing girls' desire to learn independent of societal expectations, can more directly affect outcomes.

The score created from these 21 questions ranges from 21 to 84, but adjusting it to a 0-100 point scale, without changing the ordering or spacing, helps interpreting scores. There are 455 girls that were asked all 21 questions at the BL and the EL. Four questions, as indicated in Table 70, were only asked of ISG girls in the final round. To make sure the index is not biased because some girls were asked only 17 questions, we created a panel of the 150 OOS girls who answered the YLI questions in the EL, traced them back to the BL, and removed those same four questions. After the questions were removed the score was normalized on the 0-100 point scale and those girls were added to sample who answered all 21 questions. The panel BL to EL sample consists of 605 girls. The same was done with the full panel, the smaller samples contain 75 OOS girls and 324 ISG girls for a total of 399 girls.

[214] FDG with CEC Members, Int. 107

[215] World Bank. "Fertility rate, total (births per woman) - Somalia." *World Development Indicators*. The World Bank Group, 2022, data.worldbank.org/indicator/SP.DYN.TFRT.IN. Accessed 24 Jan. 2022.

FIGURE 43: YLI QUESTIONS FROM BL TO EL



The mean YLI score at the EL for intervention girls was 59.6 on a 0-100 point scale. At the BL the equivalent score was 50.6 points. This is a 18.1% increase in points.[216] A substantive increase. The standard deviation also decreased during that time, from 18.9 to 15.9. The minimum score also increased from the BL to the EL, from 0 to 20.6 points. Figure 44 shows the difference between the mean question responses from EL to the BL for ISG intervention girls. Except for question one, "I like to try new activities I may not know how to do", the remaining 20 question means increased substantially from the BL. Question eight, "The things I do set a good example for my peers", had the largest increase from the BL.

The qualitative data supports GEFs increasing YLI scores. One girl in a Girls' Networking FGD talked about planning, being progressive, and asking for help.

> *One of the things that I have never done before includes being progressive with school and my courses. When my studies get hard, I have learned to go to those that know more than me.*
>
> - Girls Networking Exercise, Int. 405

The girl in the quote above is indicating she has learned determination, foresight and problem solving, three behaviours that are captured in questions 6, 11, and 18 of the YLI. These behaviours can be learned with positive reinforcement – they are not innate. Many girls in FGDs credit their GEF for giving them the

[216] 9 YLI score increase.

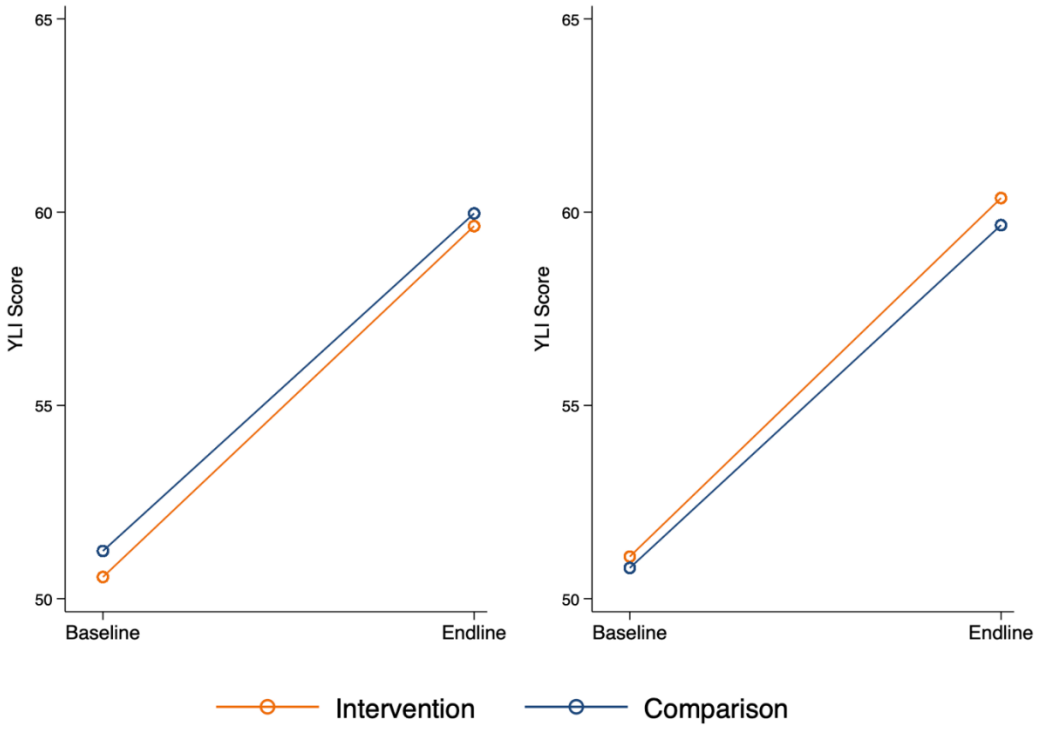
encouragement to develop skills and good habits as they pertain to school and life. One girl said the encouragement she received helped her study which helped with both her education and her family.[217]

When we isolate intervention girls that report participating in a Girls' Empowerment Forum (GEF) from the intervention sample we see a similar small difference from the comparison. Girls in GEFs score 51.1 on the YLI at the BL and 60.4 at the EL, a 9.3-point increase. Comparison girls not in the GEFs score 50.8 and 59.7 respectively, an 8.9-point increase.[218] It is possible that the effect of GEFs were reduced over time, as Fieldwork Team Leaders noted multiple GEFs that were no longer operating at the same capacity as previous rounds or at all.[219] Girls in GEFs noted four main activities during the girls networking FGDs. These activities are: cleaning and improving the environment, raising money, girl's encouragement, classroom improvement, hygiene/sanitation, and conflict resolution.[220] Conflict resolution – as stated in the beginning of the section – includes resolution of student-student and even teacher-student conflicts, though the focus seems to be on resolution of disputes between female students. Cleaning schools includes cleaning up rubbish as well as sanitization.[221] Sanitation awareness was specifically increased during COVID-19. One girl noted they cleaned student utensils as a team.[222]

GEF girls also participate in fundraising to support the school, raising money at the school as well as at home. Girls raise money from classmates to buy classroom supplies as well as from their parents to cover school fees for poorer children.[223] Overall a functioning GEF has a number of activities and initiatives in and out of the school. The high number of girls requesting continued GEF support and funding indicate Girl's investment and involvement in the forum.

[217] Girls Networking, Int. 405; Girls Networking, Int. 408
[218] We created a GEF intervention variable. Any intervention girl that reports participating in a GEF at the EL is included in this targeted intervention.
[219] To illustrate, of the 138 girls enrolled in intervention schools who – at the endline – reported participating in a GEF activity, 30.4 percent had last participated in a GEF activity "several months" prior, and just 50 percent had participated within the last month. While we cannot draw firm conclusions from this data, this is a slight decline in the reported frequency of activity levels from the ML2 round, and may help explain why GEF participation is not more closely related to improvements in YLI scores.
[220] Girls Networking Exercise, Int. 410; Girls Networking Exercise, Int. 404
[221] Girls Networking Exercise, Int. 406; Girls Networking Exercise, Int. 404
[222] Girls Networking Exercise, Int. 409
[223] Girls Networking Exercise, Int. 405; Girls Networking Exercise, Int. 409

FIGURE 44: YOUTH LEADERSHIP INDEX SCORES



The figure above shows the small positive difference in growth rate the intervention and GEF intervention had over the comparison from the BL to the EL. Relative knowledge of GEFs among enrolled intervention girls has decreased since the ML2, see Table 71, but more girls that are aware of GEF activities are participating in the EL. A smaller percentage of girls in GEFs report activities in the last week from the 64.5% in the ML2 to 50.0% at the EL. This could be an effect of covid-19 making it harder for girls to hold regular forums by closing schools, though one girl reported the forum read together during school closures, a sign of GEF activity outside the school setting.[224]

TABLE 71: GIRLS EMPOWERMENT FORUM PARTICIPATION BY ROUND

Percent of Enrolled Intervention Girls that Have Heard of GEF activities			
Percent of girls that have heard about a GEF			
BL	ML1	ML2	EL
27.5%	37.6%	45.6%	41.7%
Percent of Girls that Have Heard of GEF Activities and are Participating			
39.8%	72.3%	65.5%	73.8%

[224] Girls Networking, Int. 407

Participating Girls that have Attended a GEF in the Last Month			
-	-	64.5%	50.0%

Using the BL to EL panel of 1,210 girls that had received the YLI tool at the EL, we regressed the intervention against the comparison with no controls to get an insignificance intervention difference of 0.35 points. Controlling for age and starting grade also did not produce significant results.

Using GEF participation at the EL as the intervention and comparing them to non-intervention girls produced a difference and difference coefficient of 0.62 points at 0.856 significance. As seem below no coefficients are significant but all are positive.

TABLE 72: REGRESSION ESTIMATES OF PROGRAM IMPACT ON YLI SCORES

YLI Regressions on BL to EL Panel				
	Controls	Coefficient	Significance	Sample
Intervention	None	0.35	0.919	605
Intervention	Age Bins	0.47	0.890	604
Intervention	Starting Grade	0.35	0.919	605
GEF vs All	None	0.41	0.880	605
GEF vs Comparison	None	0.54	0.879	351

Table 72 shows the smaller intervention effects seen in the BL to EL panel as opposed to the BL to ML2 panel intervention effects see in Figure 45 and reported in the previous evaluation. While each difference and difference coefficient remain positive in the BL to EL panel, they are all small in magnitude.

One possible explanation for the small difference and difference YLI coefficients from the BL to the EL is the lack of sustainability of GEFs. Team leaders report many GEFs dissolving from the BL to the EL. One team leader told us one of the schools had a "GEF but it is operating well; the only thing they do is clean the school on Thursdays".[225] Another team leader said one of the GEFs had dissolved because the girls initiated into the GEF at the BL graduated by the EL.[226] Graduating is a better outcome than dropping out, but GEFs should turn over membership when girls complete school. If GEF activities are effectively creating youth leaders, each GEFs leadership should be able to incorporate new girls as girls leave. Given the erratic school closures over the last two years, recruitment could be difficult for GEFs without sufficient support.

The ML2 evaluation may be a better EL for measuring the effect of GEFs on the YLI since by the EL many GEFs had stopped meeting. Despite many GEFs dissolving, in the ones that did not dissolve, girls gave positive

[225] Interview with Fieldwork Team Leader
[226] Interview with Fieldwork Team Leader

feedback regarding the program pertaining to leadership, motivation to learn, and self-confidence. One GEF girl stated:

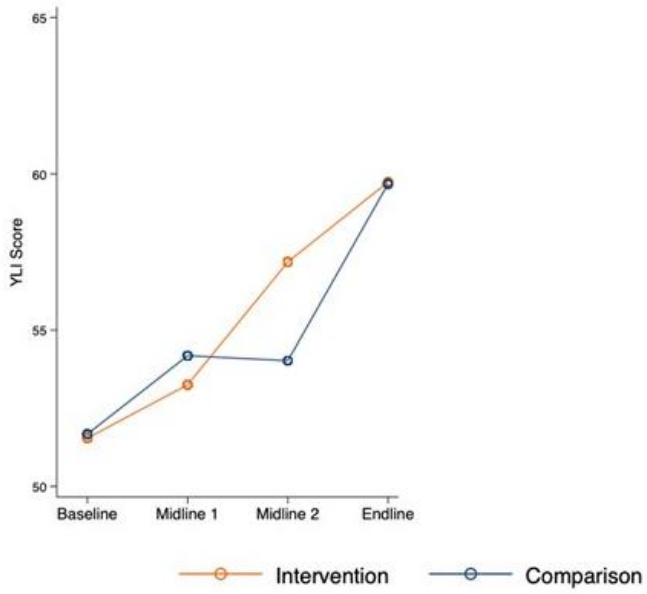
> *Ever since I joined, I have benefited by being able to stand in front of people and speaking in front of them, I used to be even of afraid of teachers or even answering questions in front the classroom.*
>
> - Girls Networking Exercise, Int. 405

This increase in self-confidence, as heard in girls FGDs, is consistent with the results in Table 72. A girl said that because GEF members are the best students they should help give training to girls in lower grade levels.[227] Training could incorporate 'lower-level girls' in GEF activities, thereby increasing or maintaining GEF membership overtime as older girls graduate. This incorporation of girls from lower grades may not show up in the data unless the girls were already in the panel. If the incorporated girls are in the panel, and indicate they are now participating in the GEF, GEF mean YLI scores could initially dip since newly incorporated girls could be more uneducated than the average. As one teacher said, "GEF girls successful invited and raised funds for uneducated girls to come to school".[228] These uneducated girls could negatively impact mean YLI scores until GEF trainings can take effect. Despite increasing the GEF intervention requirements (isolating girls that responded in the first round, all four rounds, and the last three), we could not discern a difference because the sample sizes are too small.[229]

The Cronbach's alpha is 0.9148 for the YLI index, a sign of good reliability, so questions are likely to be consistent with each other but it is still worth disaggregating by question. One question was practically significant at 0.052, intervention girls were associated with a 0.29 point increase for the statement "the things I do set a good example for my peers". This point increase is a 7.3% increase due to the small four-point scale.

[227] Girls Networking Exercise, Int. 403
[228] FGD with Teachers, Int. 207
[229] Girls were asked if they were participating in GEFs at each round. Only 12 girls responded yes on all four rounds. These samples are all very small – only 3 responded 'yes' all four rounds, and only 12 responded yes ML1 to EL – therefore it is expected we cannot reject the null hypothesis.

FIGURE 45: MEAN YLI SCORES BY ROUND, FULL PANEL



The graph on the right of illustrates why the BL to EL evaluation does not isolate discernible intervention effects. Comparison girls and intervention girls end up in nearly identical positions despite taking drastically different paths as seen in the ML1 and ML2 YLI assessments. While both intervention groups and comparison are trending upwards, their paths are not mirrored showing that there are other factors at play that we are not able to control for in our regressions.

Life Skills Index

The life skills index (LSI) has been measured using 40 different questions. We have slightly modified the index questions over time therefore some questions are merged with others in previous rounds to allow comparisons between each. Not all questions are asked of all girls, there a number of questions that are only relevant to a subgroup, either OOS girls, ISG girls, girls over 12, or a combination of two of the three. The focus will be on a BL to EL change and the indices will be built to aid that analysis. The LSI is split into three indices, as defined by the UK FCDO. Learning to learn is comprised of seven questions about the feelings of self confidence in the classroom. Learning for life consists of 11 questions about personal perceptions of education, and sources of motivation. The last section, agency, included 8 questions about a girl's control of decision making and agency.

TABLE 73: LIFE SKILLS INDEX QUESTIONS AND APPLICABLE SAMPLE AT BL AND EL

	Questions	Baseline Respondents	EL Respondents
Learni	I am able to do things as well as my friends	All	None

	I can read as well as my friends	None	All
	I get nervous when I have to speak in front of an adult	OOS girls	ALL
	I get nervous when I have to speak in front of a group of people my age	OOS girls	ALL
	I get nervous when I have to read in front of others	ISGs	ISGs
	I get nervous when I have to do maths in front of others	ISGs	ISGs
	I feel confident answering questions in class	ISGs	ISGs
	I feel confident answering questions when I'm in a group of people	OOS girls	ALL
Learning for Life	I would like to continue studying/ attending school after this year	ISGs	ISGs
	I would like to continue learning by going back to school, learning a vocation or trade	OOS girls	OOS girls
	I recognise when choices I make today about my studies can affect my life in the future	ISGs over 12	All over 12
	I recognize when choices I make today can affect my life in the future	OOS girls over 12	All over 12
	I can describe my thoughts to others when I speak	All	All
	I can work well in a group with other people	All	All
	When I have the opportunity, I can organize my peers or friends to do an activity	All	All
	I often feel lonely	OOS & ISG >=12	OOS girls >12
	I often feel lonely at school	None	ISGs over 12
	I ask an adult if I don't understand something	OOS girls	OOS girls
	I ask the teacher if I don't understand something	ISGs	ISG
Agency	Who decides: Whether or not you will go to school	ISGs	ISG
	Who decides: Whether or not you can go back to school or vocational training	OOS girls	OOS
	Who decides: Whether or not you will continue in school past this grade	ISGs	ISG
	Who decides: If you will work after you finish your studies	ISGs	ISG
	Who decides: How often you spend time with your friends	All	All
	Who decides: When/ at what age you will get married	All	All

	Who decides: Who decides what type of work you will do after you finish your studies	All	ISG
	Who decides: What type of work you will do	None	OOS
	Who decides: How you spend your free time	OOS & ISG >=12	OOS & ISG >=12

The subset of girls that were asked each question, the questions structures, as well as the number of questions have changed from the EL to the BL. While this evolution make analysis harder on the backend, we decided it is better to fit the indices and questions to the girls and adjust for these changes in the analysis than try to keep the questions consistent and ignore relevancy. For example, "I am able to do things as well as my friends" is only asked in the BL but "I can read as well as my friends" is asked in the EL in place of it.

TABLE 74: QUESTION COUNT BY GROUP

Life Skills Questions		
	Old at BL	Young at BL
ISG	25	17
OOS	24	18

The number of questions for each group are held constant over time which means the same girl regardless of her age or group is asked the same number of questions at the BL as she is at the EL. ISG girls, older than 12 were asked the greatest number of questions of any group. The question number after the EL increased because current enrolment was added as a relevancy for some questions, so questions split into two similar questions, one for enrolled girls and one for unenrolled girls. This is an important condition because keeping the question number the same for each girl across all four rounds allows her to get the highest score. We normalize scores, but a decreasing number of questions from BL to EL would likely increase variation when normalized.

The advantage of a difference and difference regression model is that small changes in wording between questions from BL to EL are applied to equally to both intervention and comparison girls, so even if the new wording skews the response in the EL this will be controlled in the regression.

TABLE 75: LIFE SKILLS INDEX MEANS AND COEFFICIENTS

Life Skill Index Means and Coefficients								
Sample	Intervention			Comparison			Difference and Difference	
	BL	ML2	EL	BL	ML2	EL	Coefficient	P-Value
ISG all	69.83	-	78.36	70.41	-	76.66	2.28	0.299
OOS all	-	68.69	73.38	-	68.48	71.20	1.97	0.415

ISG Old	70.19	-	78.63	72.10%	-	77.54	2.92	0.225
OOS Old	-	70.04	74.98	-	68.77	70.65	3.06	0.233

The sample sizes, down the rows of the table are, 604, 254, 397, and 187 girls. Disaggregating by cohort created smaller sample sizes. Each coefficient in the table above indicates an increase in LSI points.

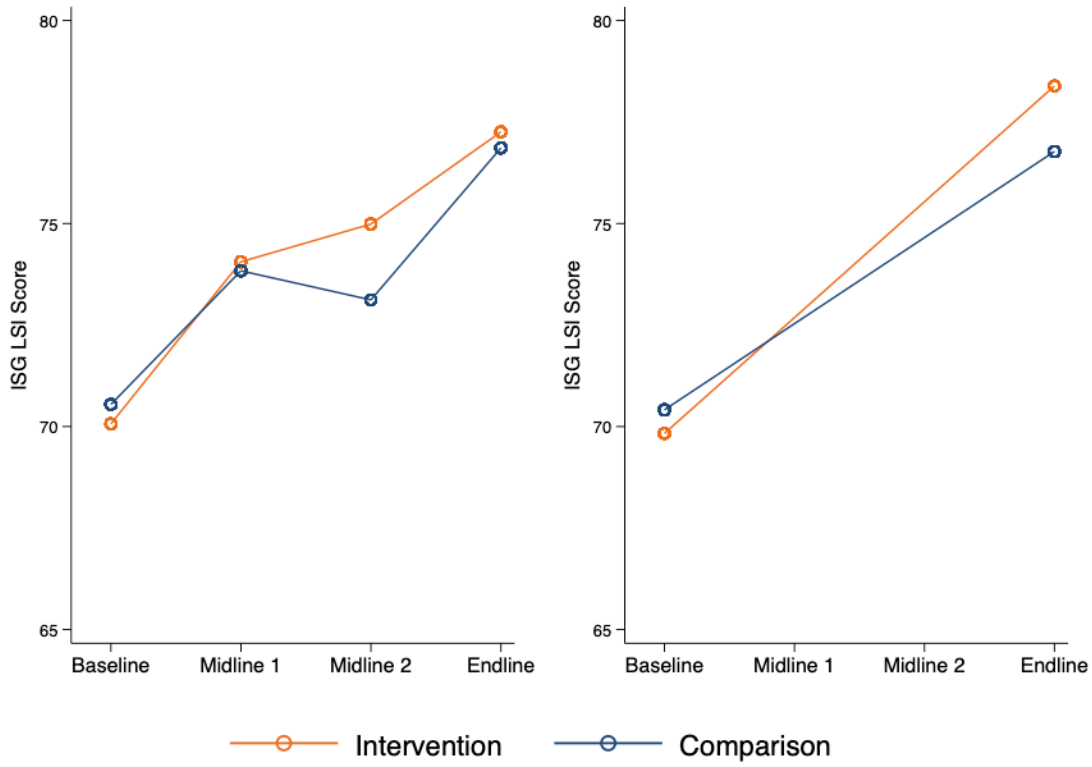
The largest difference is in OOS girls (who were over 12 at the BL) in the ML2 to the EL panel. While none of the differences are significant, they are all positive. Practically speaking the correlation between the intervention and LSI scores is a positive one; the lack of a causal significance could have more to do with the small sample sizes and large exogenous factors that influence learning in the three regions. As seen in Figure 46, following a sample from the BL to the EL visually depicts the exogenous factors that influence LSI scores. Like YLI scores, a comparison from the BL to the ML2 of LSI scores would produce larger difference and difference coefficients than the comparison from BL to EL.

From the ML2 to the EL comparison girls have raised their LSI scores much faster than intervention girls, as seen in Figure 45, but intervention ISG girls over 12 at the BL, have increased their scores dramatically over comparison girls by the EL as seen in Table 75. One girl spoke plainly about gaining self-confidence, stating:

> *I benefitted by gaining confidence which helped me within the community, my household, and my school because I speak my mind and share my opinions on what I think is right or wrong.*
>
> - Girl's Networking Exercise, Int. 405

While the EL to BL evaluation didn't produce significant results, two out of three evaluations, EL to ML1, EL to ML2, and EL to BL show the intervention as positively associated with the LSI index. An increase in sample size or timeframe could be enough to discern significance.

FIGURE 46: LIFE SKILLS INDEX SCORES, BY INTERVENTION STATUS AND ROUND



The ISG LSI score full panel sample consists of 400 girls while the ISG BL to EL sample consists of 605 girls. The few girls that had LSI scores in some rounds but not others were removed to create Figure 46.[230]

Subgroup Analysis of Life Skills

Life skills questions have three subgroups, learning to learn, learning to live, and agency. The questions in each subgroup are curated to work as proxies for each group. In totality they make up the LSI but analysing each subgroup allows a closer look at intervention effects.

[230] There is a trade-off between sample size and comparability. A full panel sample, where every girl was contacted in all rounds, is smaller than the EL to BL sample but shows a more detailed trend due the four additional means from the addition of ML1 and ML2.

TABLE 76: LIFE SKILLS, LEARNING TO LEARN, LEARNING TO LIVE, AND AGENCY

Life Skills Index Sub Groups								
	Intervention			Comparison			Difference-in-Differences	
	BL	ML1	EL	BL	ML1	EL	Coefficient	P-Value
Learning to Learn								
ISG	67.26	-	75.24	68.06	-	73.41	2.36	0.329
OOS	-	64.70	66.69	-	61.92	65.21	-1.30	0.598
Learning to Live								
ISG	73.17	-	76.94	74.17	-	76.65	1.25	0.601
OOS	-	71.61	74.37	-	72.12	72.62	2.25	0.328
Agency								
ISG	55.57	-	69.71	53.97	-	65.89	2.48	0.476
OOS	-	59.35	67.85	-	61.47	66.20	3.78	0.313

As seen in Table 76 the coefficients range in magnitude. Due to how the household survey was formatted we had to break up ISG and OOS girls because OOS girls were not asked any LSI questions during the BL. OOS girls were evaluated from the ML2 and the EL therefore the coefficients are expected to be smaller due the shorter timeframe, but for learning to live and agency they are larger.

ISG intervention girls are associated with a 2.36 increase in learning to learn points from the BL to the EL over comparison girls. Overall, each subgroup's coefficients, except for ISG and OOS girls and learning to learn, were close in magnitude and direction. Due to the closeness of the subgroup coefficients the analysis of the index as whole is not covering up a variety of intervention effects on the three subgroups. The LSI index coefficient as seen in Table 75, closely resembles a median, not just a mean, of the three subgroups.[231]

Agency for OOS girls increased the most for intervention girls at 3.78 points over comparison OOS girls from the ML2 to the EL. OOS girls' agency scores also started at a low point compared to the intervention starting points for Learning to Live, and Learning to Learn. Selecting for the OOS girls that were enrolled in the EL, the intervention coefficient increased to 3.94. This is consistent with our expectations. An OOS girl that re-enrols likely feels a greater sense of agency by going to school instead of staying home and doing household tasks.

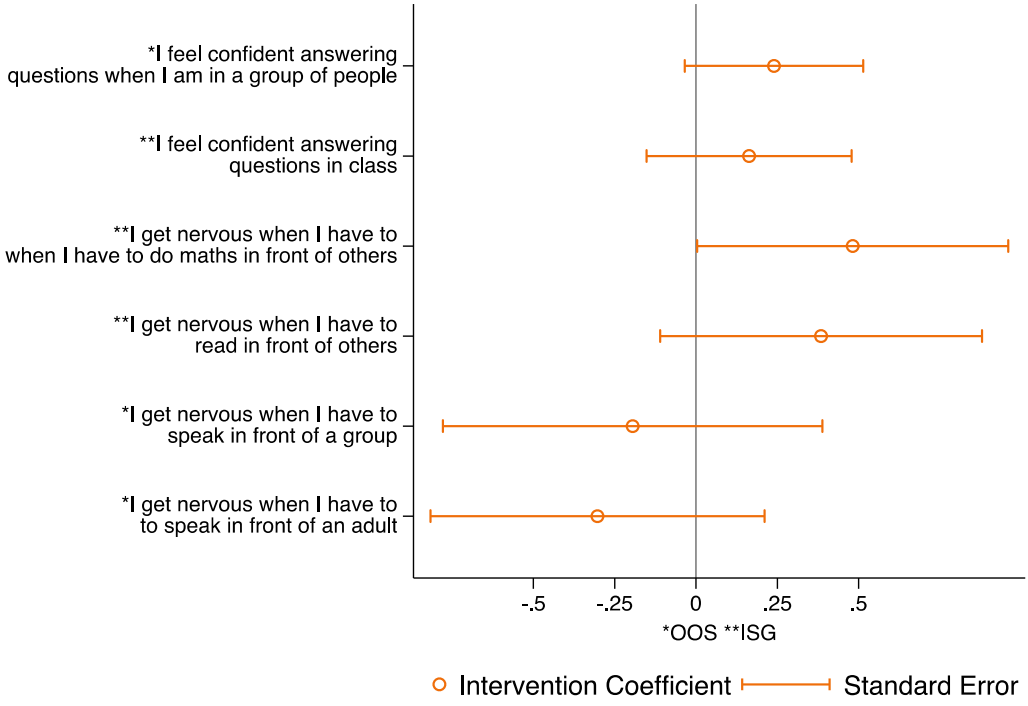
Girls' Confidence in and out of the Classroom

One of SOMGEP-T goals is to instil self-confidence in girls. The six questions in the LSI index directly asked girls about their confidence in different scenarios. The questions used the 5-point Likert scale. Two of the

[231] A median is a good comparison to the mean. The three subgroups in the LSI index do not contain the same number of questions therefore one subgroup can affect the LSI coefficient more than the others.

questions, "I feel confident…" were recoded to reverse the point scale to standardize the inter question vector. The sample sizes for the ISG and OOS regressions marked in Figure 47 are 828 girls and 698 girls. The regressions are clustered by school and do not have any controls. The coefficients are representative of the difference between the intervention over and above the comparison group from the BL to the EL.

FIGURE 47: OOS AND ISG GIRLS SELF-CONFIDENCE



As seen in Figure 47 only one ISG question, "I get nervous when I have to do maths in front of others" is significant at the 94.7 percent confidence level.[232] The coefficient of 0.49 equates to a 9.8 percent increase, over and above comparison schools from the BL to the EL. This is a substantial and significant increase and likely a result of the SOMGEP-T intervention, since CARE targeted girl's numeracy as well self-confidence.

OOS intervention girls felt more confident answering questions in a group over and above comparison OOS girls but only at the 90.9 confidence level.

The qualitative data supports a causal relationship between CARE's intervention and girls' confidence.

> *They helped me not to be afraid to make presentation in front of the students, which was an encouragement to me, because in the past I used to be very scared, and it also helped other people in my family because I told them how I had benefited*
>
> - Girls Networking Exercise, Int. 410

[232] This is one of the questions that was recoded.

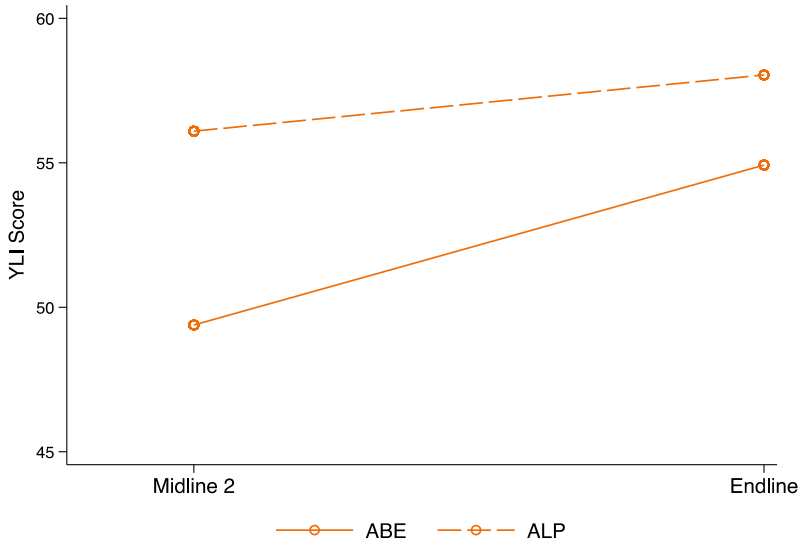
Girls have also stated that what they learn in GEFs they pass onto poor and illiterate girls.[233] Giving one girl self-confidence could have a multiplier effect in that she passes on that self-confidence to other girls.

ABE and ALP Girls

Alternative Basic Education (ABE) is an intervention implemented by CARE after the first midline to help OOS girls who have had little to no schooling learn basic skills, as well as try to prepare them for potential matriculation into formal school or transition into adulthood. The sample of ABE girls captured in the ML2 to EL panel is 347 girls.

CARE's Alternative Learning Programme (ALP) is a curated learning program for OOS girls who dropped out in mid- to late primary to improve transition rates as well as targeted, relevant life skills. The program was implemented with the help of the MOEs and is striving to sustainably uplift OOS girls that dropped out of formal schooling at an early age. ALP girls were first surveyed at the ML1, but to increase the panel sample size, the majority of this analysis will focus on the ML2 to EL panel sample of 215 ALP girls.

FIGURE 48: ABE AND ALP YLI SCORES



As seen in the figure above, ABE girls' YLI scores increased from the ML2 to the EL. ABE girls experienced an average 5.5 point increase in YLI scores from ML2 to EL. ALP girls also experienced an increase. The average ALP girl from ML2 to EL had a 2.0point increase in YLI score. ALP and ABE girls do not have good comparison groups to run robust difference-in-differences regressions.

ABE LSI scores increased from ML2 to EL. These 347 girls on average increased their LSI scores 4.0 points[234] This is a substantial increase for such a short period. Disaggregating ABE girls by age, shows that younger

[233] Girls Networking Exercise, Int. 406
[234] When adding ABE panel weights the average increase is 3.7 points.

ABE girls were associated with higher increases in LSI scores. Younger girls had a 5.2 point increase in YLI scores versus older girls' 2.8 point increase. This is an intuitive finding, younger children are more malleable, they have less culturally engrained attitudes towards themselves and their education therefore it makes sense that the ABE intervention could exert more positive influence.

TABLE 77: YLI AND LIFE SKILLS INDEX SCORES AMONG ABE AND ALP GIRLS, BY ROUND

Youth Leadership Index			
Cohort Type	ML1	ML2	EL
ABE Girls	-	49.3	54.9
ALP Girls	58.3	56.0	50.6
Life Skills Index			
Cohort Type	ML1	ML2	EL
ABE Girls	-	69.4	73.4
ALP Girls	78.0	77.0	76.2

ABE programmes were interrupted by school closures as indicated by one teacher.[235] Another teacher noted that some ABE girls matriculated into primary school, a finding that is supported by the data. [236] Sixty four percent of ABE girls reported being enrolled in school during the last year at ML2 which increased to 89.3% of ABE girls at the EL. [237] These ABE interruptions as well as overlap in other educational programs, such as primary school, make it hard to isolate the effect of ABE programs.

ALP girls saw their LSI scores decrease from ML2 to EL by 1.2 points.[238] This decrease is surprising considering OOS girls, during this same period, had an average increase in LSI scores of 3.5 points. Younger ALP girls were associated with an increase of 0.25 points LSI scores while older ALP girls were associate with a decrease of 2.1 points over the same period. This is consistent with the findings regarding ABE girls. Younger ALP girls are likely more influenced by the ALP programme than older girls.[239]

Programme Impact on Life Skills

Life skills are hard to measure. The two metrics, Youth Leadership Index (YLI) and Life Skills Index (LSI) are both measuring soft skills such as self-confidence, agency and planning as opposed to hard skills, like balancing a cheque book or computer skills. These soft skills are also not observed by enumerators but are instead

[235] FGD with Teachers, Int. 209
[236] FGD with Teachers, Int. 207
[237] FGD with Teachers, Int. 207
[238] If we add ALP panel weights when taking the mean this increases to 1.3 points.
[239] Lucas, Christopher G., Sophie Bridgers, Thomas L. Griffiths, and Alison Gopnik. "When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships." Cognition 131, no. 2 (2014): 284-299.

measured with two indices that are aggregates of survey questions. There are multiple points at which the data quality can begin to erode, therefore it is beneficial to have two metrics for life skills.

SOMGEP-T's programme is inconclusive. While the majority of difference and difference coefficients were not significant, see Table 75: Life skills index means and coefficients and Table 72: Regression Estimates of Program Impact on YLI Scores, the vast majority indicated positive associations between life skills and intervention groups, see Figure 44 and Figure 46. Analysing life skills question by question also yielded some significant difference and difference coefficients over and above comparison schools such as "the things I do set a good example for my peers" and "I get nervous when I have to do maths in front of others". Some key interventions, such as the ABE and ALP programmes, may not have reached maturation; their effect on life skills may have been blunted by school closures as well as the drought. A lack of significance, in some interventions on life skills outcomes, could be the result of the difficulty in isolating programme effects, instead of an ineffectual intervention.

Community Attitudes

Positive attitudes toward education are foundational to much of what SOMGEP-T is seeking to achieve in terms of improving enrolment rates, attendance, and enhancing girls' learning. Support from the community is critical to many different avenues of change. Within households, parents who do not support girls' education are unlikely to make the hard financial choices necessary to educate their daughters, nor will they prioritize schoolwork over the completion of chores. Girls whose parents value education but who live in a community where it is not valued may not feel it is worthwhile to continue their schooling, or may feel discouraged by the lack of support – or active discouragement – they receive from adults in the village.

But community attitudes are more critical to the proposed ToC than even this implies. Communities that are pro-education make the work of CECs easier, by contributing their time to engage in awareness-raising, by contributing funds to support students through the CEC, and by contributing materials for improving the school. They make efforts to recruit better teachers through the use of financial incentives. In reality, pro-education views in the community are, in many ways, a prerequisite for other aspects of programme impact.

Our assessment of community attitudes draws from a range of both quantitative and qualitative respondents, with the goal of shedding light – first – on whether the programme has improved community attitudes, and – second – on how attitudinal barriers to girls' education manifest in practice. In terms of quantitative indicators, we draw from responses given by caregivers, who are taken to represent the community more broadly; head teachers, whose opinions of community attitudes are indirect and subjective but, nonetheless, informative; and girls themselves, who report the extent to which they feel supported by their parents.

Following the second midline report, we first report attitudes among caregivers. As part of the household survey, caregivers were asked a short set of questions designed to understand their attitudes toward girls' education. First, they were asked whether girls' education is a worthwhile investment, even in a context of limited financial resources. Second, they were asked whether girls are just as likely to use their education as boys, a measure designed to determine whether parents see equal extrinsic value in educating their daughters and sons. Third, they were asked how decisions around their girls' education are made, and whether they consider the opinions of the girl in making those decisions.

In the ML2 round of data collection there was some suggestive evidence that caretakers of high-achieving girls put a higher value on their education. Those with high-achieving girls were 8.7 percentage points more likely

to state that investing in girls' education was worthwhile. To further examine whether this may have been a causal relationship, at the endline we decided to include a series of survey experiments intended to parse the drivers of community perceptions towards girls' education. In the first, caretakers were randomly assigned to be given scenarios where a girl, Nimco, was performing well or poorly in school, and then asked questions about the importance of her continued schooling as well as whether, and in what ways, that education would benefit her.[240] In the second survey experiment, the caretakers are asked not about a generic girl but about what they think their friend, whose daughter is in school, will do given that she is struggling to pay school fees and could use assistance at home. Here caretakers were also randomly assigned into scenarios where the girl was either high or low performing in school. [241]

As with our analysis throughout this report, we employ a difference-in-differences model to analyse programme impact, focusing on the baseline and endline data. We utilize the "true panel" of caregivers, based on the same logic we have discussed previously – by focusing on the exact same set of respondents over time, we avoid introducing bias due to the replacement process or differential attrition.

The figure below reports the trends in caregiver answers over time, disaggregated by intervention and comparison groups; this figure is a visual representation of the difference-in-differences models reported later in this section. The upper-left panel of the figure reports the share of caregivers who believe a girls' education is worth the investment. In many ways, this question addresses the crux of educational attainment in this region: while it is easy to value girls' education as a concept or in principle, real-world parents face hard trade-offs between paying for their daughter's education, paying for their son's education, feeding their families, maintaining their livestock herd or other source of livelihood, paying medical bills, and so forth. [242] As we discuss briefly below, financial hardship is a central refrain from qualitative interviewees, who suggest that financial constraints are the most important, or among the most important, barrier to enrolment.

As the figure shows, there has been a meaningful increase since baseline in the share of caregivers who believe girls' education to be a worthwhile investment. Both the final outcomes and improvement since baseline are larger for treatment communities than their comparison counterparts; treated (comparison) caretakers rose from 76% (80%) to 88% (86%). The areas who received CARE's programme therefore showed a 12

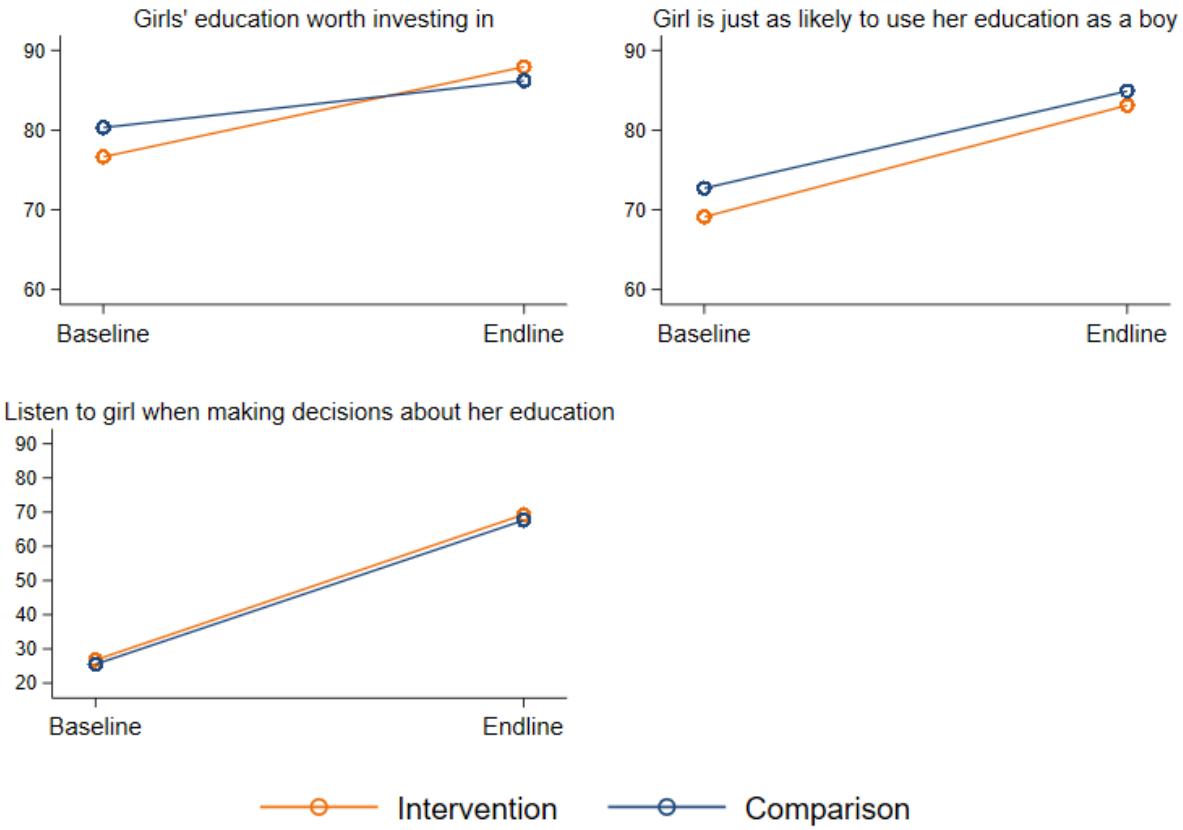
[240] The specific text of the vignette (question nimco_vig1a) was: "Now I would like to ask you a few questions about a hypothetical girl. Please imagine a girl named Nimco, who is 15 years old. She is in grade 8 and next year will attend secondary school. Her family is poor and it is always difficult to pay her school fees. This year it has become more difficult because Nimco's father has been sick and has had to go to hospital." The last sentence of the vignette was randomized between two options: "But Nimco is a good student, the best girl in her class" or "Nimco tries hard in school but is not a good student and had to repeat grade 7 twice." The purpose of varying the vignette in this way is to alert readers to the relatively different performance of Nimco in each story. Following the vignette, we asked respondents "How important is it for Nimco to continue her schooling?" and "If Nimco stays in school and completes secondary school, how likely is she to use her education?"

[241] The additional benefit of asking about this scenario is that it placed the hypothetical high/low-achieving girl in the context of the respondent's own community. Moreover, it is possible that when asking about a someone's own daughter they reply with an answer that they think is more socially desirable, i.e., that they do not discriminate when assigning value to educating children of different genders. But in this case, they can state that someone else discriminates in their weighting of educational value, in theory absolving them of bias.

[242] The question is also related to the economic view of education as a long-run investment that does not receive a higher level of investment because the returns are significantly delayed, are uncertain (the child's education may not improve their life prospects or those of their parents), and – in some cases – not shared with the individuals (parents or caregivers) making the investment.

percentage point increase, double that of non-CARE areas, albeit from a slightly lower baseline.[243] The upper-right panel also shows a general trend toward greater valuation of girls' education, but there is no evidence that the programme has contributed to this increase, given that intervention and comparison communities have evolved similarly. Moreover, the graphs display very large overall increases in the share of caretakers who report factoring their daughter's preferences into decision-making around her education, but here we also cannot determine that the programme created a differential impact relative to comparison communities.

FIGURE 49: EVOLUTION OF CAREGIVER ATTITUDES TOWARD GIRLS' EDUCATION, BY INTERVENTION STATUS



The results in the above graph are encouraging for two reasons. The first is that attitudes toward girls' education appear to be improving across all communities, not just those which have received benefits from SOMGEP-T programming. If this is part of a wider shift in attitudes, it will promise significant improvements for girls across the region, not strictly in communities targeted for educational interventions. Descriptively,

[243] In general, there are large improvements from baseline to endline in both intervention and comparison communities across all three metrics. In the most extreme case, the share of caregivers indicating their daughter has input into decision-making around her own education has risen from 26 percent at baseline to 68 percent at EL among the exact same set of caregivers.

the results also show that when considering the value of a educating a girl the programme may have had an outsized impact, over and above that observed in comparison communities.

There is reason to be cautious regarding the results of these measures, however. Elsewhere in this report, we have described how issues of social desirability bias attitudinal questions of this kind; we also note that attitudes do not translate directly into changes in behaviour, especially in a resource-scarce context. More importantly, the measures do not address differential attitudes toward high- and low-performing girls. Caregivers may believe that education is a worthwhile investment for girls, as a general rule, but wish to remove their daughter from school when she performs poorly – in which case, the justification might be that investing in this *specific* girl's education is not a good use of resources. This attitude is not explicitly expressed in any of the qualitative interviews. It is present, though, when CEC members, mothers, or teachers are asked about the performance of girls versus boys, and their answers fixate exclusively on the fact that the school's top students are often girls.[244] This positive statement about high-achievers does not reflect the reality for most girls and may overstate the gains made in girls' education. Along the same lines, focusing on the benefits that may accrue to a girl or her community when she is educated (e.g., she will succeed in life, or bring benefits to the community) is certainly a positive outcome, but is less desirable – as an end goal – than a community that intrinsically values education.

To better assess the magnitude of programme impacts, we estimated a series of regression models that provide a difference-in-differences estimate for each outcome. As the results in the table below show in the context of our aggregate models, the programme has produced small but positive impacts in two of the three indicators, though none of the effects can be distinguished from a null result.

In addition to aggregate findings, we also report findings for subgroup analyses – the impact of the programme on caregiver attitudes among the in-school and OOS girl cohorts (i.e. caregivers of girls who were in school or out-of-school at baseline). Interestingly, we find that, for the two outcomes in which aggregate impacts were highest, the impact is more positive among the cohort of OOS girls. This suggests that the programme has had its most substantial impact on caregiver attitudes among caregivers whose daughters were previously out of school, likely the best target for attitudinal change, as the goal of the programme is to increase enrolment and positive transition outcomes.

Unfortunately, none of the results reported in the table are statistically significant, but even if they were imprecisely estimated the results on the aggregate sample showed greater gains on each measure for the programme group relative to the comparison group. Additionally, there were very large effect sizes for caretakers of OOS girls on their likelihood of viewing girl's education as a worthwhile investment as well as whether girls are as likely to use their education as boys. The OOS cohort's influence in whether girls are as likely to use their education as boys appears to drive the aggregate sample impact for that outcome. This is a potentially encouraging finding – caretakers whose daughter was OOS at baseline may not have had them in school due to choosing to send a boy instead. To analyse this possibility further, we ran regressions on the OOS cohort while controlling for whether there is currently a school-age boy in the household but did not find any statistically significant effects.

[244] For instance, when asked about improvements in their school over the past three years, one respondent discussed how uplifting girls' education become essential once "it was understood that girls are better in school than boys" (FGD with CEC Members, Int. 108).

It should also be noted that, while changing perceptions of whether girls will use their education relative to boys likely advance the programme's overall objectives, this would only translate into higher levels of attendance for girls if the shrinking boy-girl gap was due to caretakers' perceptions of the value of girls' education increasing and not due to the perceived value of boys' education declining. With this in mind, it is possible that the differences-in-differences effect was driven by *net declines* in perception of whether boys would use their own education. As we do not currently have a quantitative measure of caretaker's perception of boys' education we cannot fully assess whether this is the result of a net increase in the perceived usefulness of girls' education or net decreases in the perceived usefulness of boys' education.

However, the observed trend that caretakers of OOS girls experienced greater programme impacts does not hold on average when estimating their willingness to include girls in conversations about her educational options, showing a 3.7% smaller improvement than the comparison group. We posit that this final finding should be taken lightly given the low baseline scores and enormous gains for both treatment and comparison groups over the programme's lifespan. Caretakers of comparison (treatment) communities moved from 26% (27%) strongly agreeing at baseline to 62% (59%) at endline.

The 36 percentage point shift, or 138% increase, in the comparison group suggests that there is a meaningful non-program driver of these perceptions between survey rounds. We hypothesized that OOS girls with a female head of household (HOH) may be more likely to include their daughters in the education conversation, be unevenly distributed between treatment and comparison communities, and consequently influence our results. Estimations of the effect female-led households had relative to male-led ones found that they were 15.3 percentage points more likely to report that girls' views were considered. This finding was statistically significant, and as such we believe it was the exogenous driver of our inconsistent results.

TABLE 78: REGRESSION ANALYSIS OF PROGRAMME IMPACT ON CAREGIVER ATTITUDES

Regression Model Description	Impact Estimate (Regression Coefficient)	Standard Error	Sample Size
Strongly agrees girls' education is a worthwhile investment, even when funds are limited			
Aggregate sample	5.4	4.4	1,805
In-school girls only	5.5	5.3	1,170
OOS girls only	**8.4**	6.9	635
Strongly agrees girls are as likely to use their education as boys			
Aggregate sample	1.8	7.0	1,805
In-school girls only	-0.9	8.7	1,170
OOS girls only	**7.5**	8.7	635
Girls' views are considered when making decisions regarding her education			
Aggregate sample	0.4	7.1	1,798
In-school girls only	1.2	8.4	1,165
OOS girls only	-3.7	9.5	633

It is worth noting that, insofar as the analysis above demonstrates small positive impacts of the programme, they seem to be driven by changes between baseline and ML1, with a lesser impact from ML1 to ML2 and ML2 to EL. For instance, when we limit our analysis to ML2 and EL only, we find that the programme has had smaller impacts of 2.2%, 0.5%, and -0.5% on the three caregiver attitudes outcomes respective to their order in the table. Likewise, when we analyse the full sample, across four rounds, we see a sharp improvement in caregiver attitudes in intervention communities, vis-à-vis comparison communities, from BL to ML1, and no declines in attitudes between ML1 to ML2 nor ML2 to EL. This is not a criticism of the project – the large gains made from BL to ML1 are still valuable, and we view the programme as having a net positive impact on caregiver attitudes, even if relatively small, since baseline.

As mentioned earlier, in this round of fieldwork we included a series of survey experiments aimed at generating evidence on whether parents weighed a particular girl's previous success in school when making the decision to enrol them for an additional year. The magnitude of this effect wields significant influence over SOMGEP-T's long-term impact on programme communities – girls who are newly enrolled in school because of the programme but are later withdrawn due to poor grades will not gain nearly as many marketable skills, and therefore will receive much less benefit, than those who stay in the long term.

The girls who have the greatest potential for educational gains are those who were OOS at the programme's inception. But even after they are successfully enrolled, they remain particularly disadvantaged relative to other girls their age because of their late start to schooling. As evidenced by them having been OOS at baseline, their parents are also either less likely to fully internalize the value of a girl's education or face greater hurdles to affording each subsequent year's school fees. So that SOMGEP-T can most effectively lower the burden of education it is important to fully understand how heavily parents weight different factors in the decision to enrol or re-enrol their daughter in school.

To parse out this information we randomly assigned caretakers to be given one of two scenarios about a hypothetical 15-year old girl, Nimco, whose parents are deciding whether to re-enrol her in secondary school the following year. All respondents were informed that her family is generally poor but that this year they will struggle financially even more than before due to the father being sick and needing treatment in a hospital. Half of the respondents were then told that she was either "a good student, the best girl in her class" and the other half that she "tries hard in school but is not a good student and had to repeat grade 7 twice."

TABLE 79: ANALYSIS OF STUDENT PERFORMANCE ON PERCEIVED VALUE OF EDUCATION

Question	Told Nimco is a bad student	Told Nimco is a good student
How important is it for Nimco to continue her schooling?	91%	93%
If Nimco stays in school and completes secondary school, how likely is she to use her education?	86%	89%
In your opinion, how will education benefit Nimco? It will benefit her:		
Economically (find a job, run a business)	97%	96.7%

With household management (manage a home, raise her children, be a better wife)	59.9%	59.4%
Generally (any of the above response options)	99.7%	99.2%
Education will not benefit Nimco	1%	1.4%

There are several takeaways from the data presented above. First, regardless of which of the extremes of educational aptitude they were given, parents overwhelmingly espoused the importance and usefulness of her continuing in school. This positivity is even more impressive considering that Nimco was presented to the respondents as a 15-year old girl; an age where 16% of Somali girls have already been married.[245] Second, if respondents were told that Nimco is a good student they were more likely to assert the importance and value of continuing her education. The difference between the 'good' and 'bad' student responses on the second question, how likely is Nimco to use her secondary education, was found to be statistically significant. This provides some evidence that Somali parents do weigh their child's academic aptitude when making the decision to pursue further education. We consider that three percentage point shift to be quite meaningful, as it implies that the 'bad student' holdouts could be reduced by approximately 21% if they valued the education of a girl who was a 'bad student' equally to that of one who was 'the best in her class'.

Regarding the specific *ways* that Nimco would benefit from more schooling, respondents near-ubiquitously affirmed that secondary school would help Nimco find a job or better manage a business. This outcome is very encouraging if parents' decision to enrol their daughter in school is conditioned on an anticipated return on investment, even if it accrues to their daughter rather than themselves or that any 'return' is meaningfully delayed. That nearly half of respondents did not believe it would benefit her as a homemaker is somewhat discouraging. As of 2020 only 10% of Somali women were working or had in the past 12 months,[246] and as a result many Somali families probably do not intend for their daughter to enter the workforce. This means that their incentives for furthering her education are considerably limited if they do not value education's contribution to her ability to raise a family.

We also included a second survey experiment intended to induce the respondent to consider these trade-offs in the context of their own community. Caretakers were provided with a scenario where a friend was considering withdrawing their daughter from school because "she needs her help in the house and the school fees are expensive." Respondents were randomly assigned into two groups: the first group was told the girl was 14 and in 8th grade; the second group that the girl was 14, in 6th grade, and had failed 6th grade twice. The key distinction that separates the groups is the girl's educational performance in the past.

The information these survey questions provide that the preceding set does not is that, because this is asking the respondent what they think this parent *will* decide to do regarding the continuation of her daughter's education rather than the *importance or benefit* of doing so, we can assert an upper bound on how much the caretaker believes that someone in their community is constrained by the scenario's financial and labour constraints in their decision on whether to enrol their daughter in school. For example, if 90% of caretakers who are given the 'good student' scenario respond that they believe the hypothetical parent will re-enrol her

245 Somalia National Bureau of Statistics, Health and Demographics Survey 2020
246 Somalia National Bureau of Statistics, Health and Demographics Survey 2020

daughter in school this implies that, at most, 10% of households would be bound by the difficult financial situation and be consequently unable to re-enrol their daughter.

When given these scenarios, caretakers who were informed that the girl had failed 6th grade twice responded that they thought their friend would keep her daughter in school 75.6% of the time and those who were not given any information about her educational aptitude did so 79.3% of the time. That 3.7% statistically-significant difference is of a similar magnitude to the good/bad student differential in the previous set of survey questions. This reinforces our earlier assertion that parents do internalize their daughter's propensity for schooling to some degree when considering whether to re-enrol her, and therefore asserts one potential limitation on the extent to which CARE programming can influence enrolment.

Caretakers thought that other parents in their community would re-enrol their daughter approximately three-quarters of the time even though they themselves near-unanimously declared the importance and usefulness of her continued education. One possible reason for this discrepancy is that they perceive the community around them as being less supportive of girls' education than they consider themselves. To gain further insight into general perceptions of the community we took into account the viewpoints of head teachers, asking them a number of questions gauging their perception of attitudes among the community in which they serve. It is important to note the limitations of this approach, as it relies entirely on the subjective perceptions of head teachers, rather than the actual attitudes of community members. However, we consider head teachers well-informed observers of their communities, especially in relation to attitudes toward education. They are decidedly not objective observers – they likely have a viewpoint that education is unambiguously good – but their views of their communities (and, especially, changes in their views over time) are useful for triangulating community attitudes.

To start, we asked head teachers to rate the support for girls' education shown by fathers and mothers, respectively, in their communities. Head teachers were not faced with this question at baseline, so our analysis focuses exclusively on changes from ML1 to EL, as shown in the figure below. Head teachers rated support levels on a 1-5 scale, ranging from very unsupportive to very supportive. For the sake of this analysis, we treat this scale as a continuous, rather than ordinal, scale and analyse the mean support levels reported by head teachers.[247]

As the figure shows, head teachers in intervention schools viewed fathers in their communities as slightly more supportive of girls' education at ML2 than at ML1, an improvement which was surpassed by the rise seen among comparison communities. However, between ML2 and EL both community types converged at approximately the same point.[248] On the other hand, reports from head teachers imply a positive programme impact on support for girls' education among mothers: both groups of communities show a drop from ML1 to ML2, a decline which was less steep in intervention communities, and positive upwards movement for both community types between ML2 and EL. SOMGEP-T's impact is more clearly displayed when looking at the overall change between ML1 and EL. Over that period intervention communities exhibited a lower

[247] As with our other analysis, we study only the head teachers who were contacted at both ML1, ML2, and EL, limiting the sample size to 63 head teachers or schools in each of the three rounds. It is worth noting that head teachers are not static positions, and some head teachers may have been newly hired since the ML1 data collection round. This sample is consistent across rounds with respect to the schools visited, but not necessarily the respondents to which field teams spoke.

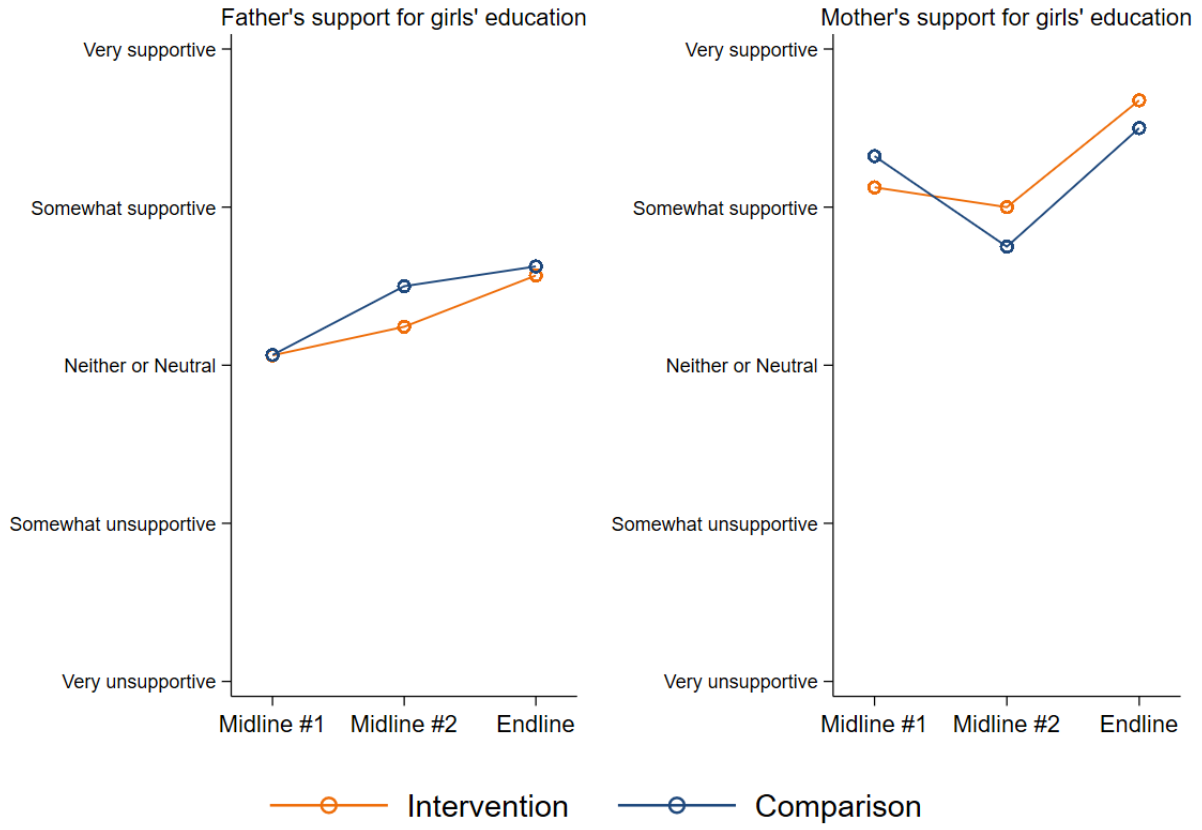
[248] Fathers in comparison communities exhibited an average score of 2.63 at endline and those in treatment communities 2.57 (midway between 'Neutral' and 'Somewhat Supportive'.

initial starting position for mother's attitudes and in absolute terms demonstrated a greater rise than comparison communities between ML1 and EL.

There is some mixed support for this in qualitative interviews. In multiple FGDs mothers expressed how their daughters were able to tell them that they had schoolwork or exams and therefore could not complete their housework. However, they responded slightly differently when asked what they thought would happen to a hypothetical student, Haawa, who was struggling to attend school or was arriving to school late in part because she was burdened with chores at home. Three mothers each asserted how in that scenario it was the parent's responsibility to inform the teacher their daughter would be late for school, rather than suggest it was the parent's responsibility to reduce her chores so she could attend school.[249] On a more positive note, this contrasts meaningfully with what mothers in another location said when describing how they were supporting their daughter's education, where three contended that they had personally taken over some of their house work so that she could attend school regularly and on time.[250] This presents us with a picture in which at least verbalized support for girls' education is improving across the board, but also where actions that reinforce this sentiment are unevenly spread.

249 FGD with Mothers, Int. 310
250 FGD with Mothers, Int. 309

FIGURE 50: HEAD TEACHERS' PERCEPTIONS OF COMMUNITY SUPPORT FOR GIRLS' EDUCATION, OVER TIME AND BY INTERVENTION STATUS



While the results for mothers are not statistically significant, this is primarily a function of the limited sample size, and not an indictment of the effect magnitude itself. We estimate that the programme has generated a 0.37 point change in perceived maternal supportiveness, on a 5-point scale. This represents a change of approximately 0.34 standard deviations, typically seen as a substantively meaningful effect size.[251]

Finally, we also presented head teachers with a hypothetical scenario describing a boy and a girl, in separate vignettes, who has achieved high marks in school and has been accepted to university. The child's family is presented as unable to pay the cost of university; head teachers were asked the likelihood that the community

251 On the other hand, we estimate a very small negative result (0.06) for fathers which may be taken as slightly offsetting the improvement among maternal attitudes. Of course, this is only a single measure of paternal and maternal attitudes. We also analysed data in which head teachers were asked whether mothers, fathers, or both parents would attend a meeting with their daughter's teacher, if one was called. Head teachers were asked to imagine a hypothetical set of parents and a hypothetical meeting, in order to gauge whether they think a typical father would make a tangible effort – in the form of sacrificing their time – to support their daughter's education. Very few head teachers (just 7.9 percent across all three evaluation rounds) believed that a father would attend such a meeting. However, the share who believe they would has increased in intervention communities, relative to comparison communities, suggesting a subtle and small shift in perceived support for girls' education among fathers. Again, this finding contradicts that reported above, suggesting that no major shifts in fathers' support – or head teachers' perception of it – have occurred.

would rally to support the child and raise the money necessary for them to continue their education. The fundamental idea of this question is that it moves beyond attitudes alone and asks head teachers to assess *which members* of the community other community members would be willing to make sacrifices to support.

The results for this set of questions are reported in the table below. Each head teacher was asked to rate the likelihood of community support in the context of a girl and a boy. As the table shows, head teachers increasingly believe that communities would be able to raise funds to support either a boy or girl to attend university. Gains are seen most clearly in the context of girls – the share who believe that their community is very likely to support a girl in this way has risen from 23.8 percent to 33.3 percent since ML1. There have been slight declines in 'very likely' to support since ML2, but we interpret this as a reduction in the *severity* but not the *broadness* of community support.

To display the continuous expansion of broad-based community support, the responses were aggregated into 'either somewhat or very likely' and 'either somewhat or very unlikely'. Resultingly, the growth in head teachers' perceptions of the community has continued apace. The sample is not adequately powered to detect any statistically significant effects but the reliable improvements over time suggest that these are genuine positive changes in perceptions. Of course, these measures do not imply communities actually *could* provide such support – we expect most head teachers overestimate the likelihood of such support – however, the trends are meaningful nonetheless.

TABLE 80: PERCEIVED COMMUNITY SUPPORT FOR BOYS' AND GIRLS' EDUCATION, ACCORDING TO HEAD TEACHERS

	Girls			Boys		
Likelihood of Raising Funds to Support Child's Education	ML1	ML2	EL	ML1	ML2	EL
Very Likely	23.8%	34.8%	33.3%	57.1%	65.2%	63.8%
Somewhat Likely	27.0%	24.6%	29.0%	22.2%	11.6%	24.6%
Sum of Very and Somewhat Likely	*50.8%*	*59.4%*	*62.3%*	*79.3%*	*76.8%*	*88.4%*

Unfortunately, the programme itself does not appear to have driven these changes, at least insofar as the difference-in-differences is able to isolate programme impacts. We calculated a measure of perceived "preference toward boys" – the extent to which head teachers believed their community favoured boys over girls – and studied the impact of the programme on this metric. Ultimately the programme appeared to have a small negative effect, where a negative effect refers to an increase in net preferential treatment toward boys. However, the sample sizes are small in the treatment and comparison groups[252] and the descriptive statistics show a large shift towards 'either somewhat or very likely' to support boys' education between ML2 and EL. Given that we are estimating differences-in-differences effects of girls support relative to boys, a large positive

[252] At endline, n=32 and n=37 for head teachers in comparison and treatment communities respectively

shift for supporting boys' education unevenly spread between treatment and comparison communities seems to have overshadowed the meaningful gains in support for girls in treatment communities.

Our final quantitative metric of community attitudes toward girls' education come from an entirely different source: the girls themselves. As with head teachers, girls are not entirely objective barometers of public opinion. Moreover, they are less astute observers of public opinion than head teachers. On the other hand, to some degree girls' perceptions of community attitudes is quite proximate to that which actually interests us. Specifically, if girls feel their community supports their education that is meaningful in and of itself; even if it does not accurately reflect community attitudes, perceived positive attitudes likely have benefits for girls' self-esteem and interest in continuing their education. And, in the end, we expect girls to have some insight into true public attitudes, at least among the set of community members they encounter and with whom they interact.

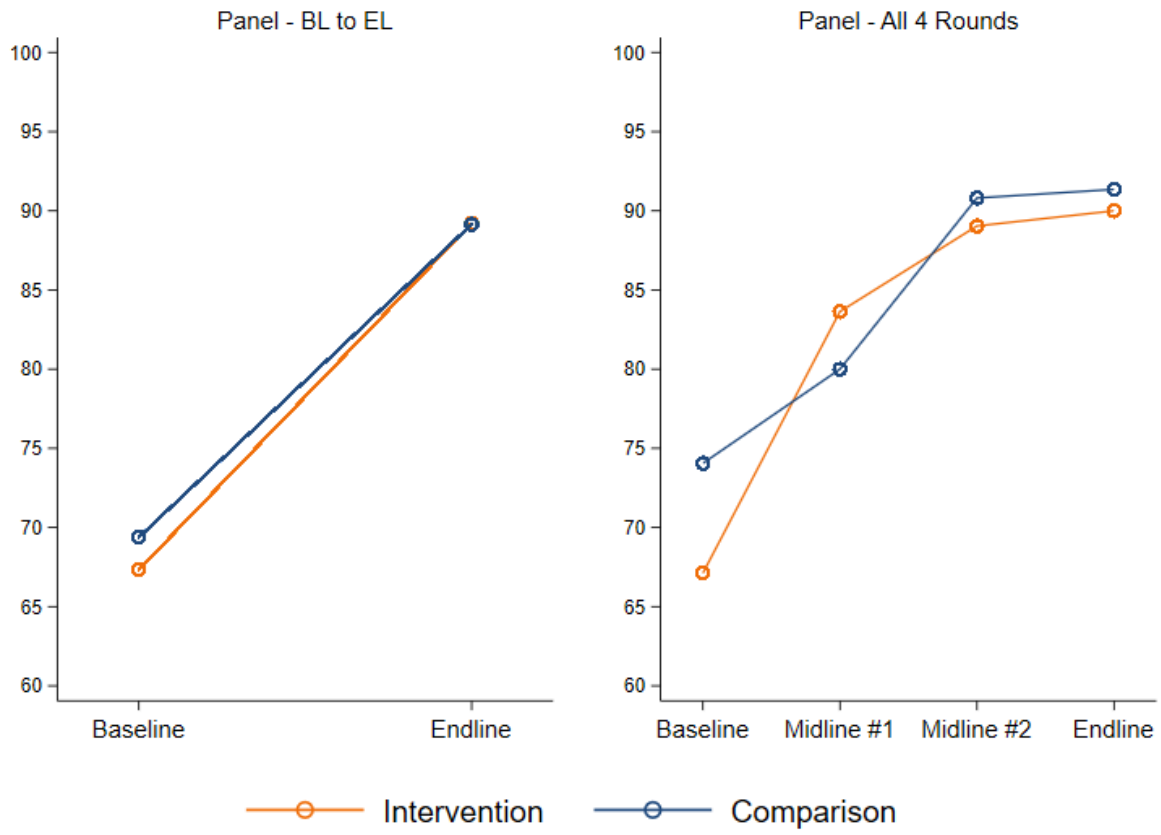
The indicator we study with regard to girls is the extent to which they feel they receive the support they need from their parents to continue to attend school and perform well. Girls were given four response options, ranging from "agree a lot" to "disagree a lot", with no neutral option provided ("don't know" was an option, though few girls availed themselves of it).

The results of our analysis are shown in the figure below. This question was originally targeted exclusively at girls who were enrolled in school, as the premise assumes the girl is enrolled, so our sample consists of in-school girls who were successfully contacted at both baseline and endline (left panel) or baseline, ML1, ML2, and endline (right panel). That is, the right panel includes only girls who appeared in all four rounds of data collection. The figure reports the share of girls who strongly agree that they receive the support they need from their parents.

When disaggregating by intervention and comparison communities, the share of girls who 'agree a lot' that they receive the support they need to stay in school and perform well in comparison communities increased from 69.4% to 89.1%, a 19.7 percentage point shift. In treatment villages these scores improved from 67.3% to 89.2%, a 19.9 point shift.[253] While the trends are not meaningfully different between comparison and treatment villages, this could be due to the topline outcomes clustering near the top of the potential distribution. As outcomes approach their upper bound the set of people whose opinions can be changed shrinks, so large effect sizes become more difficult to achieve. The levelling off seen between ML2 and EL provides some graphical evidence that a 'ceiling effect' is occurring. To further examine this possibility, we ran regressions on changes between the aggregated ML2 and EL survey data. That is, we did not contrast comparison and treatment groups but instead merged their data together to increase the sample size and thus statistical precision. Our finding was that girl's perception of whether they received sufficient academic support from their parents experienced a mere 0.03 percent improvement since ML2. While this is a test for the magnitude of changes between survey rounds, rather than for 'true' potential for further gains, it more rigorously asserts the flattening seen visually.

[253] These calculations restricted to the sample of girls who were interviewed in both the baseline and endline survey rounds.

FIGURE 51: SUPPORT GIRLS PERCEIVE FROM THEIR FAMILIES, BY INTERVENTION STATUS



Beyond this aggregate impact, we were interested in whether girls whose households were particularly disadvantaged financially saw similar gains. The logic of this analysis was that girls in disadvantaged households were less likely to feel supported, because their continued schooling may seem tenuous, dependent on their parent's ability to pay their school fees or otherwise continue to sacrifice to keep them enrolled. Our hypothesis was that the programme might have outsized impacts specifically among this group, because it might have eased the financial burden of marginalized families, resulting in increased perceptions, among girls, of support from their families.

We studied subgroups of girls whose caregivers reported that they had reduced their food expenditure sometime in the last three months, gone without food an entire day in the last 30 days, or reduced the number of meals they ate sometime in the last 30 days. In the end, we classified a family as economically marginalized if they had experienced any of these three metrics of hardship. When we analysed programme impact among a subsample of respondents who were economically disadvantaged then the estimated impact of the programme for this sub-group grew to 3.9 percentage points. While the difference between these estimates is not statistically distinguishable from zero, the gap is suggestive: tentatively, the programme may have increased the extent of support girls believe they receive from their families especially strongly in families who face more severe resource limitations.

The available qualitative data reinforce many of the themes that emerged in our analysis above. However, they also paint a picture of communities in which attitudes have shifted somewhat, but where much improvement is still needed.

The first theme that emerges from the qualitative interviews is the influence of female role models on the value communities place on a girl's education. In small communities there are generally few opportunities for even boys to utilize their education and establish a career outside of employment in agriculture, a limitation on the benefits of education that is a magnitude larger for girls. Resultingly, parents and leaders in these communities have limited if not zero interaction with women who have been able to employ their education to create opportunities for themselves to succeed. Members of the community who possess rigid beliefs about the extent of what girls can achieve with an education may revise those priors if faced with examples that show otherwise.

A CEC member described such a case, where a girl in their community who had completed their local primary and secondary schools and now taught there, with the result that the girls she taught were motivated to pursue more school themselves.[254] Another deliberated on their optimism that indirect exposure to successful women, even those with disabilities, could change community members' perceptions, stating that "they now believe they [disabled women] are more intelligent than the rest. They've seen on TV and in their phones several disabled females who are instructing others" before lamenting that this was not realized in their community, as "there is no support in this area for such girls."[255] Despite the current situation, their positive outlook toward the future may still be indicative of attitudes shifting, albeit slowly.

While attitudes may have changed, the drought that has been taking place across several project locations may have depressed local economic situations to a point where families are unable to send their daughters to school without assistance, regardless of the value they place on it or expected future benefit. When mothers were presented with a narrative about a girl, Haawa, whose family had lost their livestock due to the drought, their responses illustrated a community that strongly desires girls' education but which lacks the means to ensure attendance in the face of economic shocks. Their responses split between stating that the loss of her family's livelihood meant she could not continue school, and that she could continue, but only as long as the ministry of education, CECs, international NGOs, or the community continued to provide school feeding programs and tuition waivers. When asked if there were any students in their community whose families had been affected by the drought like Haawa's had, and what had happened to their education, they all stated that there were such students but that they had unfortunately dropped out of school.[256] This underlines how household finances are a core driver of enrolment, as well as the impact that programs like SOMGEP-T can have in increasing communities' educational resilience to economic shocks.

In addition to the retention of in-school students, a CEC's mandate also includes enrolling those who are currently out of school. With this objective in mind, CEC members espoused one strategy that had reportedly achieved significant success: a collaboration with community religious leaders. Religious leaders, who are looked up to for spiritual, personal, and in this case educational guidance, wield significant influence on whether girls go to school. This authority over educational matters is especially pronounced in communities where the only school may be a Quranic one, necessarily taught by those same religious leaders. For these reasons they worked in tandem with the CEC to "encourage both boys and girls to attend school and… this

254 FGD with CEC Members, Int. 102
255 FGD with CEC Members, Int. 101
256 FGD with Mothers, Int. 304; FGD with Teachers, Int. 209; FGD with Mothers, Int. 310

collaboration has yielded great results, raising awareness among students and inspiring girls to learn and pass on their knowledge to their children."[257] Religious leaders can both directly and indirectly impact girls' enrolment; they can influence the social value a community places on girls' education and directly permit their attendance or persuade their parents that they should attend.

However, it should be noted that while Quranic schooling does have its benefits – the curriculum generally includes reading, writing, and Quranic history – the relative (or potentially absolute) scarcity of lessons on topics such as numeracy and mathematics limits how effective additional years of school will be to SOMGEP-T's higher-order goals of promoting girls' economic prospects. In communities where both standard and Quranic schools are available, collaborations with religious leaders run the risk that parents who would have otherwise enrolled their daughters in standard education (or were already doing so) may be shepherded into Quranic schools instead. These girls will have then reduced their educational attainment in numeracy and other career-relevant skills, effectively lessening SOMGEP-T's overall impact.

While CEC members presented religious leaders as effective collaborators for increasing enrolment, they may have been biased slightly by their personal partnership in the matter. Religious leaders in different communities may also display meaningful variation in their attitudes towards girls' education. Mothers described how religious leaders in their villages exclusively valued Quranic education, stating that "religious scholars always encourage children or girls to learn religious knowledge, they say that religious education and standard education can't be taught at the same time, so let children get religious education."[258] Mothers in another location reported much of the same.[259] This provides some concerning evidence that religious leaders' incentives for getting girls to attend school are not fully aligned with that of SOMGEP-T's primary aims. If their promotion of Quranic schooling encroaches upon existing demand from parents of OOS girls for standard school or persuades parents of girls presently in standard school to switch, then the net impact of their assistance cannot be as assuredly positive as that portrayed by CEC members.

Beyond the view that Quranic learning is of higher importance, and that it and standard schooling are mutually exclusive, religious leaders also promote other views that are harmful to girls continued education, namely the promotion of early marriage. Religious leaders, fathers and young men in the community were identified as being reliable supporters of girls getting married "as early as possible."[260] The low average age of marriage interacts with the tendency to be enrolled in school in multiple ways. One mechanism is by moving forward the time where a girl would stop attending, usually once she is married or gives birth to children. It may also change parent's estimation of the benefits of even *beginning* school. Mothers emphasized how parents in nomadic communities believe they do not need to enrol their daughters in school *specifically because* she will get married at an early age.[261] The extension of this line of thinking is that if she is destined to complete little school then there is little benefit to attending at all. Unfortunately, given the sporadic nature of schooling for ever-moving pastoralist households, and depending on how long they anticipate the girl being in school before marrying, this reasoning may not be entirely unsound. Learning one half of the alphabet probably does not make the girl or her family perceivably better than not knowing any of it at all.

257 FGD with CEC Members, Int. 107
258 FGD with Mothers, Int. 306
259 FGD with Mothers, Int. 310
260 FGD with Mothers, Int. 303
261 FGD with Mothers, Int. 309

CEC members and teachers alike seem to recognize the detriment that early marriage has on a girl's tendency to continue school and were in agreement that they should be encouraged to forego marriage until they had completed their studies. But their opinions beyond that were divided along regional lines. In several locations, teachers and CEC members described their efforts to advocate for girls to continue their schooling even after marriage, and noted their success both generally "even the married girls enrol in the school now" as well as specifically "there was a girl here that got married while in 7th grade, and we reported to the school principal that they should help that girl continue her education"[262] and she completed secondary school as a result.

By encouraging their community to recognize marriage not as an impediment but as a complement to a girl's education, CECs are certainly directing their efforts in a positive direction. But unfortunately, even among proponents for girl's education this view is not universally held. CEC members described how they would encourage girls to delay marriage, but once done would intentionally keep them out of school. One interviewee reported that "we raise awareness to the children [that marriage would harm their education] and if this doesn't work, we try to keep the child out of school so as not to affect others."[263] The idea that married girls do not belong in school was reinforced by another respondent, who waxed philosophically about how "it's the nature of human beings, education and marriage cannot go together." Early marriage is clearly a high barrier to girl's continuing their education, but as the teachers and CEC members evidenced is surmountable and can be affected by those with educational influence in the community.

Narrowing our focus from community perceptions about girls' education to that of disabled girls specifically, the qualitative evidence suggests meaningful progress but also some mixed understandings of the hurdles faced by disabled children and what is needed to overcome them. CEC members discussed with us how "the community formerly assumed that a disabled child could not learn anything. They now realize, however, that with the help of the community, they can learn effectively."[264] This is meaningful progress, albeit given the low baseline view that disabled children have *no* potential for success in school. In the same interview, CEC members brought up how they were sharing best practices for teaching disabled children but did not deign to provide details of exactly what teaching techniques they considered to be the most effective. While taking the initiative to improve teaching practices for disabled children should be applauded, it must be emphasized that their barriers to education continue outside of the classroom, an understanding was lacking in our conversation with CEC members. Mothers showed a similarly narrow-minded idea of these learning difficulties by indicating that there were not people within their community who would *prevent* a disabled student from learning.[265]

Nevertheless, mothers presented a better understanding of the ecosystem of constraints faced by disabled girls after being presented with a vignette in which a disabled 13-year old student, Khadija, had trouble getting to school and was bullied while attending. Here, Girls described a community with mixed attitudes toward those with disabilities, describing how "some people would help her [Khadija] and others would ridicule her, whether they are at school or in the community."[266] Mothers gave a more optimistic take, acknowledging that Khadija's parents would need to "take her to school and picks her up after classes are over as well as protect her from other kids who would harass her."[267] In that same conversation another

262 FGD with CEC Members, Int. 110
263 FGD with CEC Members, Int. 109
264 FGD with CEC Members, Int. 104
265 FGD with Mothers, Int. 306; FGD with Mothers, Int. 308
266 Vignette FGD with Girls, Int. 604
267 FGD with Mothers, Int. 306

respondent suggested that Khadija could learn in that environment so long as her mother joined her for classes. Of course, this would be very difficult for a Somali parent to do and underlines the progress that still needs to be made if disabled students are to be enabled to learn.

Teachers, on the other hand, seemed to recognize the impact indirect support can have on disabled children's ability to succeed in their academics. They were given a hypothetical scenario about a girl named Barwaaqo who was teased at school because of her crutches and asked how a child like Barwaaqo would be treated in their community. One responded by describing the treatment of a boy in a wheelchair who had attended their school, first relaying how the other students would push him to and from school each day. They continued on to express how this community support enabled him to continue as far as 8th grade before going on to become a Quranic teacher.[268] It is not clear how much this experience would be emulated if it were a girl in the wheelchair instead of a boy, but the fact that in some communities' children with disabilities are given meaningful support suggest a positive shift in attitudes nonetheless.

Taken as a whole, this analysis has shown that perceptions by caretakers of the value of a girl's education, the likelihood that she uses it, and whether her opinions are an input to the decision to re-enrol her in school have improved generally and suggests that they have increased to a greater degree in treatment communities, with the caveat that our sample may not be adequately powered to causally identify some program effects. These attitudinal shifts primarily occur between Baseline and Midline #1, although girls' views of whether their parents sufficiently support their education continued into Midline #2 and head teachers' estimation of mothers' and the community's growing support for girls' education continued into Endline. That progress stalled on some metrics is not necessarily an area of concern – several of our quantitative indicators have risen to such an extent that we posit there is little potential improvement remaining.

Our survey experiments confirmed that parents highly value girls' education, but also asserted lesser support for girls who perform poorly in school. That finding, combined with only 59% of caretakers stating that secondary school would help a girl in her role as a homemaker, is worrisome. This concern is rooted in the fact that most Somali women are destined to be stay-at-home mothers, alongside qualitative reports which emphasise that religious leaders, highly influential within their community, promote girls' education only as far as it pertains to her Quranic learning and even then only up until the earliest point that she can become married.

The qualitative data shed only limited light on changes in community attitudes over time. But it is clear from the data that most community members support girls' education, although there are some misperceptions of what truly prevents girls from obtaining additional schooling, including early marriage and extensive household work. However, they do reveal that much of the community, including CECs, Teachers and Mothers, are avidly working to reduce these difficulties. Further effort needs to be made to communicate to religious leaders and fathers, the most significant proponents of early marriage and exclusively-Quranic education, that it is beneficial for everyone when women enter the work force, become better educated, and hold off on starting a family.

School-Related Gender-Based Violence

[268] FGD with Teachers, Int. 209

The final intermediate outcome we consider is gender-based violence. Safety and security at school and on the journey to school is an essential component of improving attendance and retention, because – as we will discuss in greater detail below – safety is a major motivating factor in girls remaining out-of-school.

Our focus in this section is on safety and security, somewhat broadly construed. In this, we mirror the previous midline evaluation, which was primarily concerned with gender-based violence, including harassment and other forms of violence, but which was not limited exclusively to violence that occurred in the school. We present results concerning general safety concerns, even if the violence is not explicitly gendered. Given the context in which SOMGEP-T is being implemented, a broader view of violence and security is useful for highlighting the types of problems girls face while trying to complete their education. We also cast a fairly wide net because it is difficult to ask girls direct questions regarding many forms of violence, and because indirect questions often elicit somewhat indirect answers. Using these responses requires some degree of contextual knowledge and interpretation; we tend to be cautious in our interpretation, but this means that our discussion is often around topics such as harassment or general feelings of insecurity, rather than very specific acts of violence or potential violence.

Our analysis is oriented heavily toward qualitative data, but we attempt to triangulate the information collected against quantitative data wherever possible. Often, this triangulation requires use of quantitative data in slightly indirect ways.

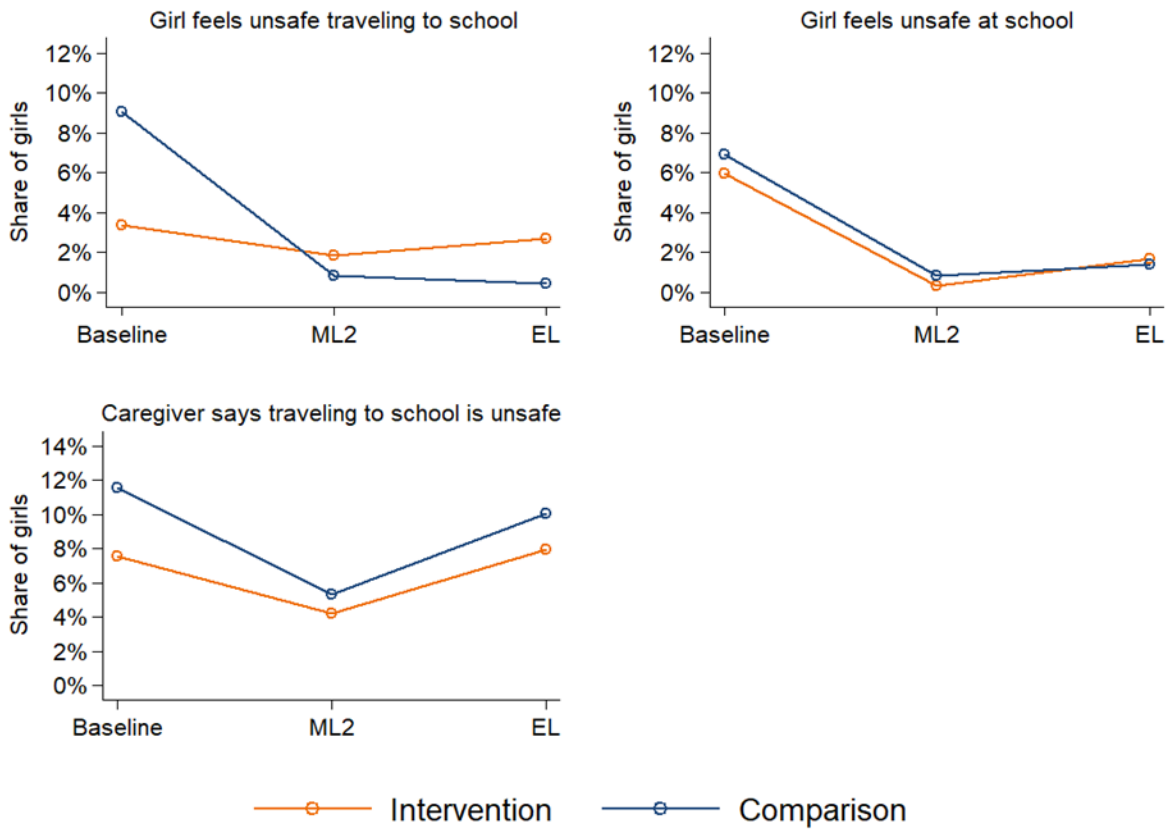
As shown in the graphs below, both girls and caregivers were asked questions on safety at school and when traveling to school. Girls were asked whether they feel safe traveling to and from school (yes/no) and whether they feel safe at school (yes/no). The percentage of girls who said they do not feel safe traveling to school is shown in the upper-left and the percentage who do not feel safe at school in the upper-right graph.

These graphs show a consistent, substantial decrease in the proportion of girls from intervention schools who feel unsafe either traveling to school or at school from baseline to midline, but then show either a slowdown of this trend or a reversal of this trend from midline to endline. As was noted in the midline report, any gains seen from baseline to midline cannot be attributed to the program itself because similar trends are seen across both intervention and comparison schools.

The proportion of girls from intervention schools who feel unsafe traveling to school has returned almost to baseline levels by the endline, whereas in comparison schools, the decrease observed from baseline to midline continues, but at a much slower pace, in other words suggesting that the majority of the comparison school gains were observed from baseline to midline. The final result is that whereas in the baseline, the proportion of girls who said they feel unsafe traveling to school was much higher in comparison schools, by the endline, the proportion of girls who say they feel unsafe is actually higher in intervention schools.

The proportion of girls who feel unsafe at school sharply decreases in both comparison and intervention schools from baseline to midline, but in both areas the proportion is beginning to reverse from midline to endline. By the endline, the proportion of girls from comparison schools who report feeling unsafe at school is slightly lower than the proportion from intervention schools but is roughly the same.

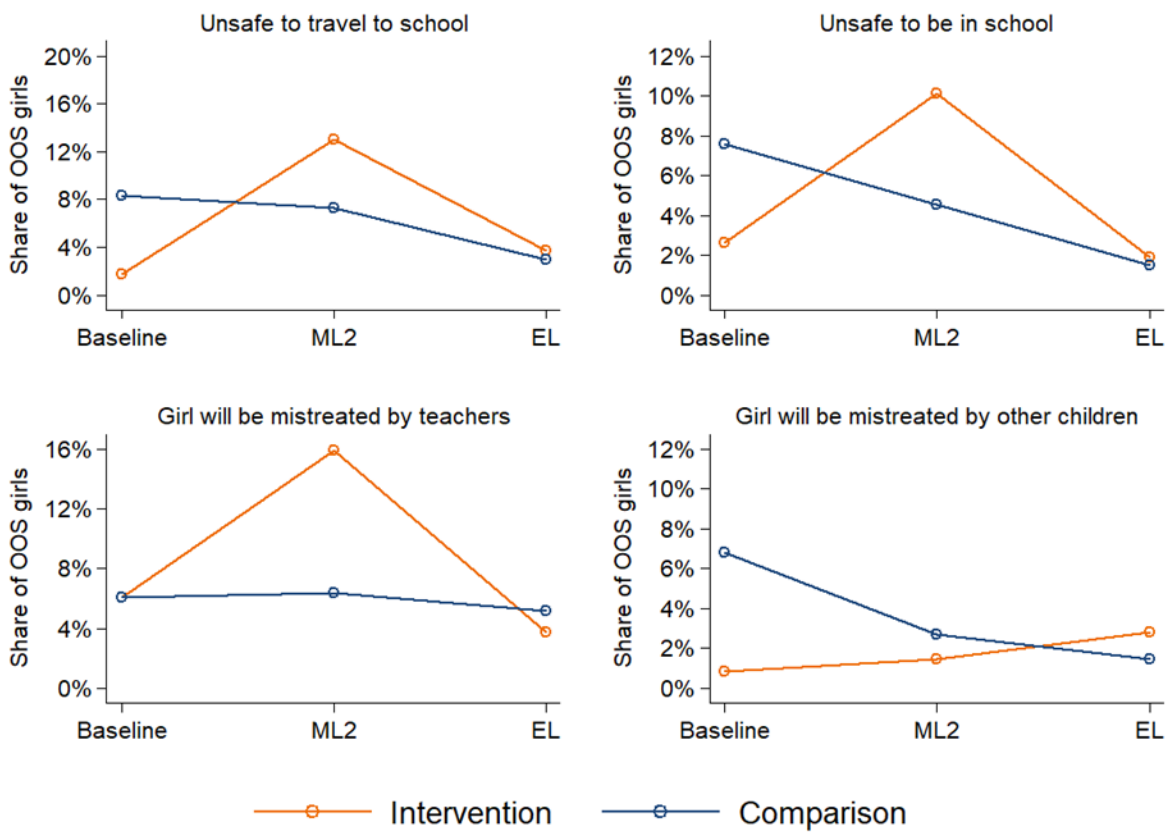
FIGURE 52: PERCEIVED SAFETY OF ATTENDING AND TRAVELING TO SCHOOL, OVER TIME



Caregivers were asked to rate the relative risk girls face when traveling to school, which we report in the lower-left panel. We interpret any response other than "very safe" as an implicit statement that it is not entirely safe for a girl to travel to schools in the area, and we report that share of caregivers who imply it is not entirely safe in the figure. Similarly here, the proportion of caregivers who say traveling to school is unsafe dropped steadily from baseline to midline in both comparison and intervention schools, but began to increase again from midline to endline, returning to roughly the same levels observed at baseline by the end of the programme. However, in this case, the proportion of caregivers from intervention areas reporting it is unsafe remained lower than the proportion from comparison areas during all three evaluation rounds.

The analysis above is based on a large sample of respondents, as all girls enrolled in school were asked these questions, and all caregivers – regardless of their girls' enrolment status – were asked about safety of traveling to schools in their area. However, we also analysed data from caregivers of out-of-school girls, and the reasons they gave that their girls were not enrolled in school. This analysis produced dramatically different results, as shown in the set of graphs below. Caregivers could give multiple reasons why their girl was not enrolled; we plot the share of caregivers who selected each option, focusing on responses that express concern about traveling to school, being in school, being mistreated by teachers, and being mistreated by other children. The figure below plots the share of OOS girls' caregivers who selected each response.

FIGURE 53: SAFETY-ORIENTED REASONS OOS GIRLS ARE NOT ENROLLED, OVER TIME



In two of the four cases (unsafe to be in school, girls will be mistreated by teachers), the share of caregivers from intervention schools who cite a given reason has decreased slightly from baseline to endline, and in the other two, the share has increased (unsafe to travel to school, girls will be mistreated by other children). In the cases where an increase was observed, the share of caregivers citing the reason in intervention schools was higher than the share in comparison schools by the endline, whereas the reverse was true at the baseline. Interestingly, despite the large differences between intervention and comparison results at either the baseline, the midline, or both, by the endline, the proportion of caregivers citing a given reason was roughly the same in intervention and comparison schools.

Our interest in this section is not exclusively in perceptions of safety among OOS girls. The qualitative data provides a wider view of safety concerns and gender-based violence that impacts key project outcomes, such as learning, attendance, retention, and so forth. The main qualitative tool that captures views on safety is a participatory risk mapping exercise, which CARE developed during the first-round midline evaluation. This information is supplemented with information from other qualitative interviews when possible.

As in both midline rounds, girls commonly mentioned they are afraid of the roads leading to and from school. There is concern primarily over fast moving cars, men or other risky individuals that could potentially be encountered, and robbery. As one girl described, "I am not afraid of places near shops, but I am afraid of places where there are many men, I am also afraid of open spaces, and the way I go to school I am afraid of

being robbed or harassed by boys when I pass by, as well."[269] Interestingly, a number of girls also cited being afraid of the police on the roads, with one explaining, "There are boys, also the police that I feel afraid of because they had stolen from girls their mobiles previously. Also the boys discriminate us there."[270]

There is also just a general fear of moving through empty or ungoverned spaces. In general, girls see empty spaces as areas where "anything can happen to you, and no one is there to help you."[271] Girls commonly mentioned fearing rape and fearing encountering wild animals in empty areas. Conversely, girls are also wary of crowded areas, including markets, because there are many men in these areas. In general, the gendered nature of the complaints girls raise has been observed consistently throughout all evaluation rounds – girls' concerns about safety still centre largely around fear of harassment or abuse from boys and men. As one girl explains, "On the road, there are many boys playing there and they may throw rocks at you."[272]

Whereas in the midline, stories of community-level violence emerged in the qualitative interviews, but in the endline, there were only a few mentions of conflict. There was mention of a charcoal-related conflict in one area and a "little war"[273] that resulted in school closures in another, but small-scale communal conflicts seem to have been less common in the last year in SOMGEP-T communities.

Within schools, there are a few main areas girls commonly cited being afraid of or feeling unsafe in. These are many of the same areas cited in other rounds. Girls still feel unsafe in toilets. Girls' reasoning for being afraid of toilets varies. In some cases, the toilets are poorly constructed and neglected and the girls are afraid they will collapse. As one girl explained, "I feel scared when I am going to the toilet because it was built long time ago, so I am afraid it collapse with me, and cockroaches are there too."[274] In other cases, girls are afraid of being seen because the toilets have no doors or are actively spied on by the boys: "I do not feel comfortable in the toilet because when we are using it we are afraid of the boys coming in on us."[275] Perhaps as a result, girls do not always feel comfortable passing the boys' toilets or going to the girls' toilet if it is close to the boys' toilet.

Secondly, girls feel unsafe near water tanks and wells. These areas tend to be crowded, and there are reports of boys harassing girls in these spaces. The overcrowding also causes students to be late, and it does not appear to be uncommon for students to wait for a drink without ever getting one before they have to leave for class. Girls are also afraid of falling into the wells.

Thirdly, girls feel unsafe within schools in the courtyard. As was mentioned above, girls generally fear open spaces and walking past boys. It is notable that whereas some explain they fear facing harassment from boys and men in public spaces, other responses suggest girls feel ashamed to even be seen in public spaces or find it inappropriate for girls to be in public spaces. One girl explains, "The courtyard and the toilet are public

[269] Risk Mapping FGD with Girls, Int. 510.
[270] Risk Mapping FGD with Girls, Int. 510.
[271] Risk Mapping FGD with Girls, Int. 504.
[272] Risk Mapping FGD with Girls, Int. 501.
[273] Risk Mapping FGD with Girls, Int. 507.
[274] Risk Mapping FGD with Girls, Int. 508.
[275] Risk Mapping FGD with Girls, Int. 502.

places that everyone can see you and I am ashamed of them."[276] Another explains something similar, "In the café most of the men drink tea and it is not suitable for girls so I am afraid to pass in front of it."[277]

Fourthly, girls are afraid of the front office, which they see as a space where teachers congregate. Many are afraid they will be yelled at or reprimanded and others are simply shy to pass the teachers and worry they might be interrogated by the teachers if they pass by the office.

As for areas girls feel safe or less afraid, toilets with locks were commonly cited as safe places, as were classrooms, the courtyard (which, as mentioned above, was also cited by some as an area where they feel unsafe), and in areas where girls and boys are separated. Lastly, some girls suggested that school conditions are safer with older and younger girls separated: "My classmate girls make me feel happy, because in the past old girls and young girls were same class but now we are separated and no one commits fights."[278]

The results up to this point have largely been similar to the results from previous rounds. However, it is worth noting that the girls interviewed in the endline made specific mention of programme activities in the context of making them feel safer and less afraid at school. The two most commonly mentioned programme activities were the increased involvement of CEC members and the increased involvement of GEF girls in school monitoring and conflict resolution. One girl shared, "the CEC comes to school every day to keep the security."[279] CECs may have also had a hand in separating the children into groups, including separating the younger and older children, which was cited by many girls as a reason incidence of conflict has decreased in school. As was mentioned not just in the risk mapping exercises but also across many of the qualitative interviews with various groups, GEF girls are also taking a very active role in resolving conflict between girls. As one girl explained, "The empowerment committee discusses reconciling students having conflict, so we improve security."[280]

Much of the programme's impact on school-based safety likely stems from two main activities. The first is the recruitment and incorporation of additional women into positions within the educational system, in the form of teachers, ALP teachers, gender focal points, and CEC members. While not a comprehensive solution, the presence of women can improve girls' feelings of safety and their ability to report issues to an adult. The second is the programme's focus on eradicating the use of corporal punishment, which – as discussed in the teaching quality section above – has declined markedly since baseline in both intervention and comparison schools. This decline is especially notable when we consider that girls occasionally report feeling unsafe or uncomfortable in the front office, presumably because this is a place where they may be disciplined.

Other actions may have had less direct, but still important, effects. Capacity-building with CECs appears to have increased their role in conflict resolution. At the same time, the share of schools with a school management plan that explicitly includes – as verified by our researchers through a review of the plan – information on child protection has increased somewhat since the baseline. A majority of CECs in intervention schools also reportedly reinforce the use of non-violent discipline, which is likely a result of the programme's capacity-building efforts with CEC members.

[276] Risk Mapping FGD with Girls, Int. 509.
[277] Risk Mapping FGD with Girls, Int. 505.
[278] Risk Mapping FGD with Girls, Int. 504.
[279] Risk Mapping FGD with Girls, Int. 504.
[280] Risk Mapping FGD with Girls, Int. 504.

It is difficult to draw firm conclusions based on the results in this section. The quantitative results suggest the situation in comparison and intervention schools is roughly the same at the time of the endline and some of the gains observed from baseline to midline have started to reverse. In the qualitative interviews, many of the risks mentioned in the baseline were again mentioned at the endline. However, there is also evidence of small improvements in some quantitative measures, and it is clear from the qualitative interviews that programme activities have had a direct positive impact on improving girls' feelings of safety in school. Given these improvements were mostly attributed to the involvement of CECs and GEFs, these results suggest that community-driven solutions for resolving conflict are often most effective and sustainable. Regardless, girls still face varied and multiple risks in public spaces in their communities, on the way to and from school, and in school.

Conclusions

Design and Methods Overview

- SOMGEP-T's endline evaluation built closely on previous rounds in methodological terms, tracking multiple cohorts of girls relatively successfully. In total, the evaluation included 69 primary schools, 32 ALP Centres, and 35 ABE Centres. Across all cohorts, the team successfully re-contacted 73.1 percent of the girls interviewed at ML2. And, among all the girls interviewed at any point in the previous three evaluations, 64.2 percent were successfully re-contacted and interviewed during the endline. The product of efforts at tracking girls is a strong quasi-experimental design with a panel of girls who have been tracked for up to four years.
- The evaluation's strength is hindered somewhat by differential attrition and re-contact, especially in the context of studying transition rates. Re-contact rates were lowest among older girls, many of whom have left their original villages to pursue secondary school educations, to get married, or to engage in paid work in urban areas. Re-contact rates were higher in intervention communities, which likely influenced our findings regarding transition rates, though the extent and direction of any bias this might cause is not immediately obvious.
- The evaluation has benefitted from repetition and learning across rounds, which has allowed for continual improvements in fieldwork procedures (especially around recontacting respondents), tool and survey question design (especially measures of community attitudes and school management practices), and triangulation across data sources. Combined with the varied set of respondents and tools employed since baseline, this has given the evaluation team access to rich data from many different voices -- teachers, CEC members, girls, caregivers, REOs, and others -- and from many different perspectives. This produces results that are often complex and contradictory, but also allows for adjudication between findings and nuance that would not otherwise be possible.

Cross-Cutting Findings and Trends

- SOMGEP-T is an integrated programme, which makes parsing the effects of individual interventions very difficult. However, specific interventions occasionally stood out in terms of their perceived impact – as reported by beneficiaries – and their estimated impact within quantitative data. GEFs, for instance, seem to be associated with both better attendance and improved learning outcomes. Attendance rates in schools with active GEFs are higher than those in other schools, and girls who participate in GEFs miss 0.1 fewer days of school per month, on average. GEFs are also valued by their participants and by teachers and CEC members, who highlight the role GEFs play in encouraging enrolment by out-of-schools and promoting retention.
- The programme's impacts were, surprisingly, not often concentrated among the most ultra-marginalised girls in the sample. Impacts with regard to learning and transition – where they were positive – tended to be broad-based. However, this can be misleading, because the programme's strategy for selecting communities in which to work emphasised particularly marginalised areas. In other words, "less-marginalised" girls in this evaluation nonetheless live in remote, high-poverty, and food-insecure areas with poor access to education. Where the programme showed outsized impact among the ultra-marginalised was in enrolment of out-of-school girls into formal school or alternative

learning avenues. Impacts on transition were also strongest among girls who faced individual-level barriers to schooling, including fear of her teacher and a heavy load of domestic work at home.

- More important, perhaps, is the fact that impacts – across almost any outcome where positive impacts were observed – were not concentrated solely among the best-off or least marginalised girls. Instead, they were broad-based, benefitting girls from the poorest households, the lowest-achieving learners, and girls with disabilities as well.

Learning

- The programme has not had an identifiable impact on learning scores in the aggregate. That is, while learning improved greatly among cohort girls in intervention communities over four years, this trend was almost exactly matched by gains made in comparison communities.
- Among specific subgroups, however, the programme has had important impacts. The programme has had positive impacts on both Somali literacy and English literacy scores among out-of-school girls, which is likely a result of the enrolment of out-of-school girls into ALP and ABE programmes, where they have rapidly gained literacy skills.
- Gains in learning among the in-school girl cohort seem to have taken place between baseline and ML2 but were largely erased since ML2. This is almost certainly a function of the combined shocks that have affected the region since the ML2, which was conducted in late 2019 – COVID-19, drought, and locust swarms, the combination of which has devastated the economy and forced school closures.
- The evidence supporting the programme's Theory of Change – with regard to learning – is relatively strong. Improvements in attendance, community attitudes, and girls' self-esteem and life skills are all associated with gains in learning scores, even after controlling for a number of other explanations.
- Non-formal education programmes – ALP and ABE – have had mixed success in improving learning. Among ALP girls, learning scores have improved very little since ML1, when ALP girls were first interviewed. But among ABE girls, there have been large gains in numeracy and Somali literacy over the last two years, relative to either in-school or out-of-school girls.

Transition

- Transition is the area where the programme has had the greatest verifiable impact. In the aggregate, from baseline to endline, the programme is associated with a 5.1 point improvement in transition rates among in-school girls. From baseline to endline, the programme produced an 18.0 point improvement in transition rates among the cohort of out-of-school girls.
- The programme's impact on transition outcomes among in-school girls was greatest in the final two years of the programme, with higher shares of intervention girls remaining enrolled through the school closures and disruptions caused by COVID-19. Impacts among out-of-school girls were greatest in the first two years of the programme, when the programme opened ABE and ALP centres and enrolled out-of-school girls into non-formal education. While the programme has not fundamentally changed the barriers to enrolment – financial shortfalls, the demand for girls to complete domestic chores, etc. – it is clear that targeted interventions can both increase retention and bring a large number of girls into non-formal schooling.

Sustainability

- Sustainability in the context of SOMGEP-T is largely focused on CECs and the incorporation of changes into MOEs and national policy environments. CECs have become considerably more active over the life of the programme – for instance, the share of schools with a management plan has increased from 45.9 to 64.9 percent in intervention schools, and CEC members complete monitoring visits with greater frequency.
- The programme's impact on CEC quality is difficult to determine, because other programmes have been active in training and supporting CECs, including in comparison communities. There has also been a slight downturn in CEC activity levels since ML2, potentially due to school closures and the economic dislocation stemming from COVID-19 and the worsening drought in rural communities. However, it is clear that CECs have become more active over time, and this suggests some aspects of school management improvements – and their knock-on effects on student attendance, retention, and teaching quality – may be maintained after the close of the programme.
- Regional Education Officers and other MOE staff have a better understanding of gender-sensitive education and non-formal education, and their respective importance. Efforts to recruit female teachers have also been put in place and are reportedly improving the gender balance of the teacher corps. Funding limitations continue to impede implementation of additional non-formal education programmes, and there is concern that ALP and ABE programming will falter without the direct support provided by SOMGEP-T and other programmes.

Intermediate Outcomes

- Surprisingly, given the impact the programme has had on transition rates, especially, the programme has had less verifiable impact on most of its intermediate outcomes. Attendance rates, in particular, have not improved in intervention schools, versus comparison schools, regardless of the metric employed to measure attendance. Attendance in intervention schools have been especially stagnant since ML2. One interesting note is that boys' attendance has dropped sharply in intervention communities; this suggests that stable girls' attendance in the same communities could be interpreted as positive impact of the programme – if boys can serve as a counterfactual for comparison, we might have expected a decline in girls' attendance in the absence of the programme.
- School management improvements have been widespread in intervention schools, but we cannot attribute these improvements to the programme itself, as comparison schools have actually experienced greater gains. As noted earlier, CEC-focused interventions from other programmes seem to account for the gains in management quality in comparison communities.
- Teaching quality has dropped sharply since ML2 in both intervention and comparison schools, a fact that is easier to understand when we consider the financial and other stresses teachers faced over the last two years of school closures. Many of the gains in teaching quality made between baseline and ML2 were wiped out by the time of the endline, though it is possible that these improvements will return naturally under more stable circumstances.
- Community attitudes toward girls' education have improved across both intervention and comparison communities. In the latter, the share of caregivers who believe girls' education is a worthwhile investment, even when funds are limited, rose by 6 percentage points from baseline to endline, while intervention communities saw an increase of 12 percentage points.

- Support for girls' education is stronger in cases where a girl is perceived to be high-achieving, suggesting that households view girls' education as a potential investment that will yield future earnings. Caregivers are less likely to value education for the benefits it provides in terms of household management, child-rearing, and other non-employment arenas.
- The programme's impact on girls' life skills and self-esteem was modest, but positive. On a 100-point index of life skills, the programme improved scores by 2.3 points among in-school girls and 2.0 points among out-of-school girls, over and above the improvements observed in comparison communities over the same period. Improvements on CARE's Youth Leadership Index (YLI) were smaller, but nonetheless positive.
- Gains in life skills and YLI scores among the intervention group were sharpest in the first two years of the programme, up to ML2 in 2019. Since that time, improvements have continued, but comparison girls "caught up" with rapid improvements. One reason for this trend may be that intervention girls had less exposure to schooling, GEF activities, and other life skills programming in the last two years of the programme – due to school closures – than in the first two years.

Annexes

Annex 1. Evaluation Approach and Methodology

Prior to Data Collection

Training

Training for SOMGEP-T EL took place from 31 October to 4 November 2021. The training was led by a team from Consilient. Training was led by varied staff members to allow breakout sessions for team leaders – separate from the enumerators on their teams – in which they discussed data collection tools specific to team leaders. Several CARE staff members were also in attendance; the CARE team introduced the project's goals and interventions, led training on child protection, COVID-19 mitigation steps, and provided valuable insight into the context and particulars of the project. They also participated in technical discussions regarding the evaluation, clarifying sampling issues and tool design.

In total, training included 54 field team members, with 11 team leaders overseeing 43 enumerators. Consilient also recruited three back up enumerators to replace enumerators that either failed the enumerator assessment or quit.

Often during the training, team leaders and enumerators were split into separate sessions. The logic of this decision was that team leaders are solely responsible for a number of tasks during fieldwork – conducting classroom observations, attendance headcounts, a survey with head teachers, most qualitative interviews, and monitoring and managing their team members – that require specialized training. Enumerators, on the other hand, need to fully understand the household survey and learning assessments, and time spent training them on the responsibilities that fall to team leaders detracts from time that could be better spent practicing the learning assessments.

The first four days consisted of both classroom-style training and practice. During this time, all team members participated in an overview of the project and a discussion of the methodology, including the types (i.e. respondent type, whether enrolled in ALP, formal school, and so forth) of girls that we would be interviewing, re-contact procedures, what to do in case a replacement girl was needed, research ethics, child protection, and how to use the paper tracking sheet system.[281] Enumerators were trained on:

- The program, its activities, and the overall evaluation methodology
- Child protection and research ethics
- Learning assessment administration
- Household survey administration
- Re-contact procedures

[281] All enumerators already have experience using ODK Collect for mobile data collection and interviewing. The SOMGEP-T evaluation also relies on a backup system of paper tracking sheets, however, which requires training on its and emphasis on its importance.

- Specific scenarios they might face during re-contact and during the learning assessment

They also spent significant time practicing the household survey and learning assessments in one-on-one sessions.

This year, enumerators were required to take an assessment to evaluate their understanding of the tools. The assessment was reviewed by Consilient's technical team to identify struggling enumerators and gaps in their training. The training team used Poll Everywhere, a classroom polling software, to evaluate enumerators' understanding of the tools in real time during presentations and discussions. The software helped poll enumerators to both establish a grading consensus for the learning assessments as well as for the training team to understand weak points in enumerator understandings. A final evaluation was conducted using ODK Collect.

To aid in learning assessment training, Consilient pre-recorded example scenarios. Enumerators were asked to judge the scenarios, how well the girl performed on the subtask/assessment, as well as how the enumerator administered it. The recorded scenarios included examples of enumerators not properly administering assessments as well as examples of common issues seen in previous rounds. One example included an enumerator prompting a girl with answers, in another an enumerator did not let the girl finish reading.

Child protection training and research ethics were incorporated into the curriculum. Child protection was covered during a dedicated session with all evaluation team members – enumerators, fieldwork team leaders, the quality assurance officer, fieldwork manager, and team leader – CARE staff, and the MOE staff that were tasked with accompanying some of the field teams during their work. The training was led by CARE's specialist in child protection, and included theoretical precepts of child protection as well as interactive discussion, group work, and a discussion of specific scenarios. It also included a discussion of the process for reporting child protection issues that may arise during fieldwork or which researchers may become aware of during data collection. Part of this discussion emphasized the multiple available pathways for reporting child protection issues, including reporting directly to the Research Officer overseeing the project, or referring cases to Consilient Research's Human Resources focal point; additionally, CARE staff could be approached directly – the main purpose of having multiple channels for reporting is to facilitate reporting for all individuals. All researchers signed both Consilient's internal child protection policy and CARE's more specific child protection protocol.

Team leaders also participated in several group-wide sessions, but spent considerable time in separate training. These breakout sessions focused on administration of the attendance headcount, classroom observation, head teacher survey, and qualitative tools. A participatory approach to training was emphasized. For example, team leaders conducted a mock classroom observation while observing one of Consilient's staff members training enumerators on the household survey, assessing the extent to which he encouraged participation, asked open-ended questions, explained concepts in a different way, etc. Newer team leaders led mock focus group discussions in which other team leaders played the role of mothers or teachers in a focus group, confronting the moderator with problems they had personally experienced on past evaluations.

Consilient and CARE organized the pilot with the help of the MOE. The pilot was conducted on 4 November 2021. Schools were closed on that day but the MOE organized for on school to be open so that the Enumerators could practice administering the surveys and assessments. The pilot, a small sample (n=40), allowed enumerators to practice data collection as they would in the field, as well as help us work out the

problems in the survey. The enumerators after the pilot then returned to the training venue to debrief, ask questions, and plan for deployment to their respective regions.

Sample Design and Sampling

A full description of the sample design for SOMGEP-T is provided in the main body of the report and in this annex, combined. In general, issues related to overall sample design (selection of schools), new samples of cohort girls (i.e. ABE girls) or discussion of re-contact and attrition are discussed in previous of the report. In contrast, sampling for tools other than the household survey and learning assessments – including qualitative tools, classroom headcounts, classroom observations, and head teachers surveys – are described in the following two sections.

Qualitative Sampling

Qualitative research in the endline shifts focus from prior rounds. Previous rounds' qualitative interviews focused on understanding the barriers to girls' education, different axes of marginalization, and other details that could guide program implementation and also help the evaluators understand dynamics within SOMGEP-T communities. In contrast, qualitative research in the endline focuses more on understanding whether changes have occurred in programme outcomes over time, and why those changes have occurred. To over-simplify somewhat, the goal of qualitative analysis in the endline is to assess the programme's impact, study whether impact varies across relevant subgroups, and understand the mechanisms of change, including whether changes observed can be attributed to the programme's interventions. This was an explicit methodological recommendation made in the ML2 reporting round – one which has been actively incorporated into the tools designed by CARE's technical team and reviewed and revised by the evaluation team. The qualitative analysis also set out to provide insights on girls' own perspectives and interpretations of their experiences of change and use of skills acquired through project interventions.

The process for selecting participants in the varied qualitative interviews was designed to meet two disparate goals: first, ensure representation of a wide range of viewpoints by, for instance, targeting a variety of geographic locations and using random assignment of schools to participation in qualitative interviews where appropriate; second, ensure the quality of data by selecting schools and participants that were of particular relevance to the research goals (e.g., selecting schools with active Girls Empowerment Forums (GEFs) for participatory exercises, as GEFs are arguably the most important vector through which SOMGEP-T hopes to impact girls' life goals and self-confidence).

As was the case in ML2, we have moved away from using a geographically clustered approach to sampling for qualitative interviews. Instead of selecting 10 target locations and conducting FGDs with mothers, teachers, and CEC members in each of those 10 locations, we have instead assigned FGDs to communities independent of one another. The reason is to ensure the widest geographic coverage of qualitative interviews possible. As was the case in prior rounds, random assignment of interviews to communities (i.e. random selection of communities) are stratified by geographic region proportional to the overall sample. This proportional allocation according to region applies to FGDs with mothers, teachers, and CEC members.

The other qualitative tools cannot be conducted using random selection. FGDs or participatory exercises with girls – risk mapping, story-telling or vignettes, and girls' network mapping – were conducted exclusively within intervention communities, with preference for communities with an active Girls' Empowerment Forum (GEF). In ML2, we identified schools with an active GEF on the basis of how many cohort girls reported awareness of and participation in the GEF in their school. The threshold we used to identify "active

GEFs" was necessarily arbitrary, but a useful indicator for identifying these schools. We followed a similar approach in this round.[282] Within schools that have an active GEF, we targeted a mix of girls for inclusion in the FGDs/participatory exercises: girls who are members of the GEF and girls with disabilities (GWDs). The choice to target girls who are participants in the GEF ensures we interview girls who have benefitted directly from this particular programme activity; the inclusion of GWDs – even if they are not participants in the GEF – was done with the intent to provide insight into how benefits from the GEF activity have or have not filtered outward to other girls.

Finally, KIIs conducted with MOE officials were also not randomly assigned. Participants in these interviews were selected purposively. CARE provided the names and contact information for MOE officials in each region, with a total of ten officials being interviewed.

The table below presents the achieved qualitative sample, relative to the targets established prior to the start of fieldwork.

TABLE 81: ACHIEVED QUALITATIVE SAMPLE

Type of Interview	Target	Completed
FGD with CEC Members	10	10
FGD with Mothers	10	10
FGD with Teachers	10	10
Risk Mapping FGD with girls	10	10
Vignette FGD with girls	10	10
Network Map FGD with girls	10	10
KII with Regional Education Officers	10	10
Total	70	70

Sampling of Schools and Centers

In line with guidance from the FM, the evaluation of SOMGEP-T takes a joint sampling approach, in which the girls selected at the baseline constitute a cohort to be tracked over time for the purposes of assessing *both* learning and transition outcomes.

During the baseline, the evaluation team sampled girls at their households, rather than at the project's schools. This sampling approach avoided the significant bias that would occur if girls were selected at schools, since

[282] Note that, due to the expansion of participatory exercises (at ML2, we conducted 8 total participatory exercises, while the endline will include 30), we have opted to combine the risk mapping and girls' network tools into a single FGD. This allows us to target fewer schools (20 in total) that have active GEFs, allowing us to target more active GEFs, overall. While this strategy results in a slightly narrower sample of GEF girls taking part in participatory exercises, it is necessary, because there are just 37 intervention schools in the sample, and not all have active GEFs – identifying 30 intervention schools with active GEFs has not been possible during qualitative sampling. Moreover, the girls' network tool is relatively short and can be reasonably combined with the risk mapping exercise. In total, the 20 participatory exercises (10 vignette; 10 risk mapping and girls' network combined) were completed in unique schools.

only in-school girls would have been included in the latter case.[283] Because girls were selected at their households, the sample included a considerable number of out-of-school girls; moreover, the approach allowed the evaluation team to collect data on the household characteristics of girls in the cohort.

While the sampling design of ML1 was slightly different – reducing the sample to only in-school girls from the baseline and reducing the sample of schools slightly – we do not discuss this in detail, because these changes were temporary and specific to ML1. In other words, out-of-school girls were added back into the sample in ML2, with the field teams re-contacting OOS girls from baseline at ML2.

At ML2, the sample consisted of the following schools and centers:

- 69 primary schools
- 35 ABE centers
- 32 ALP centers

This sample represented all the same ALP centers originally recruited into the sample at ML1. It also included most of the 76 original primary schools from baseline – five were excluded following the baseline as outliers, while two additional schools were excluded following the baseline for accessibility reasons. The endline school sample matches the ML2 sample exactly. The same schools, ALP centers, and ABE centers will be visited. The five outlier schools from baseline remain outside the sample; the two schools removed due to inaccessibility remain outside the sample for security reasons (Xingod) and because only lower-grade classes are available (Dhumay).

The table below describes the target sample size for each of the non-cohort tools employed in the evaluation, and the sample size actually achieved. The next table describes the geographic breakdown of the sample of each tool. Note that the overall school sample for this evaluation is now 69 formal schools, reduced from 76 schools at baseline. Five schools were removed after the baseline due to their exceptionally high learning scores during that round; two additional schools were dropped for accessibility reasons.

TABLE 82 TARGET AND ACHIEVED SAMPLE SIZES FOR EACH TOOL

Tool	Target Sample Size	Achieved Sample Size
Head Teacher Survey	**69**	**68**
Classroom Observations	**138**	**127**
Attendance Headcounts	**N/A**	**499**
Comparison Learning Assessments	**50-60**	**48**

Within schools, a single head teacher survey was completed in each school. No second-stage sampling will occur for headcounts, either. As at baseline and ML1, a census of classrooms, covering grades 1 to 8, was completed, with headcounts completed in each. In past rounds, our field teams have completed 7.02

[283] In fact, the bias is even greater than simply the exclusion of out-of-school (OOS) girls. Sampling at schools also ensures that girls who *attend* school most often have a higher probability of selection, resulting in a sample that is heavily biased toward both enrolled girls and the girls who attend school most frequently.

headcounts, on average, per school. We achieved a total sample size of approximately 499 headcounts as a result.

The only tool requiring second-stage sampling is the classroom observation tool. Within each school, two teachers coached or trained by the project were randomly selected for participation in a classroom observation. Teachers or classrooms were randomly selected from among those teaching grades 3-8 only; moreover, observations were only completed during lessons focused on Somali, English, or mathematics. Team leaders listed all teachers/classrooms with the assistance of the head teacher. Using a random number generator, they select a number between one and the number of teachers listed, selecting the teacher corresponding to the random number.

The only other second-stage sampling occurs for the selection of boys to complete learning assessments. Boys between the ages of 14 and 24 were targeted to complete a learning assessment, if they were located in the household of a girl being interviewed. That is, households with a boy aged 14-24 will completed a boys' learning assessment, which consisted of the same five learning assessments as the girls. However, if a household has multiple eligible boys, a kish grid programmed into the phones will be used to randomly select the boy to complete the assessment.

Data Quality Assurance

Data collection took place during a single wave, without staggering for different tools or modes of data collection. Qualitative and quantitative data collection occurred simultaneously for logistical reasons. Given the remote areas in which SOMGEP-T schools are located, visiting communities twice, in separate waves, was not considered feasible. Team leaders were responsible for most qualitative data collection.

Prior to the start of data collection, CARE's Monitoring & Evaluation team and Consilient's Fieldwork Manager both reviewed the list of target schools for accessibility and security concerns. Consilient's Fieldwork Manager and CARE staff were in contact with teams throughout fieldwork, to discuss security concerns and contingency plans if necessary. CARE staff or members of the MOE also accompanied most of the teams during fieldwork.

During data collection, Consilient's Fieldwork Manager and the technical evaluation team provided extensive quality control of the incoming data. However, quality assurance began in earnest prior to fieldwork, as the survey scripts were reviewed for mistakes, incorrect or unclear translations, errors in skip or filter logic, and general areas of concern. The quantitative tools remained mostly the same from the previous round, with adjustments for the sampling and re-contact process and addition of a few new questions. During the inception phases of each data collection round, the evaluation team sought to script in purposeful checks of data quality. For instance, each girl's information – and that of her household – is scripted into the survey, such that it is shown to the enumerator during the re-contact process. We also added confirmatory steps, asking the enumerator to confirm the outcome of the re-contact procedures and what steps they followed after failing to locate a girl. As an example, if a girl was replaced and the replacement girl's grade level did not match the replacement procedures established during training, enumerators were required to provide a reason why they were interviewing a replacement who did not meet the formal requirements. Finally, we imposed constraints on questions where appropriate, and generally sought to minimize measurement error through careful survey design.

Once data collection began, the evaluation team conducted quality control testing using a pre-designed Stata .do file, which was designed specifically for this project. The process included checking for known error "hotspots":

- Any learning assessment question that involved counting the number of letters known or words read
- Physical headcounts and enrolment numbers from school records
- Later subtasks in learning assessments, where enumerators can skip questions if the girl was completely unable to answer earlier subtasks

In addition to these areas of complexity where we wished to monitor the data for errors, we also monitored re-contact and attrition rates, as well as the number of households in which an eligible boy was located. We also monitored GPS coordinates and survey duration to guard against data fabrication, and checked for "enumerator effects" in learning assessment scores to check whether any enumerators appeared to be administering the assessment in fundamentally different ways.

When errors were discovered through the quality control process, the information was relayed to the Fieldwork Manager and Team Leader, so that corrective action could take place. Where corrections to the data were needed, the evaluation team made these changes immediately once feedback was received from the team leader in the field. All data cleaning that occurred concurrently with fieldwork was recorded in a Stata .do file, with justifications for all cleaning decisions, to ensure replicability and transparency.

Qualitative data was checked in a similar, though obviously less structured, fashion. Team leaders submitted audio recordings of their qualitative interviews to Consilient's office-based researchers and the Fieldwork Manager. The first audio files submitted by each team leader was reviewed by the Fieldwork Manager for quality of the interview, and they provided feedback to the team leader to improve their interviewing techniques.

With regard to data transmission and storage, survey responses were transmitted as soon as network coverage allowed from enumerator's phones to secure servers hosted by Ona, which runs a server for ODK Collect submissions. The data is encrypted end-to-end from the phone to Ona's servers, preventing data loss and ensuring confidentiality of the data. Once receipt of the data was confirmed by the Fieldwork Manager and technical team, surveys were deleted from enumerator phones as, again, a method of ensuring anonymity. Qualitative interviewers were audio recorded. Audio recordings were transmitted to Consilient Research's Google Drive folder dedicated to the project and with access restricted to the technical evaluation team, and then deleted from enumerator phones and audio devices. All data storage took place on Consilient Research's Google Drive account. When data was transmitted to CARE's Monitoring & Evaluation team, data was password-protected in a .zip archive and transmitted via email; the password was provided separately to CARE's focal point via Skype.

Post-Data Collection

During and after the completion of fieldwork, the technical evaluation team performed additional checks of the data for consistency and cleaned the data. Our strategy was to flag interviews or observations which had some form of inconsistency, and follow up with the team leader in question for clarification or additional information. In some cases, the Fieldwork Manager called girls' households or head teachers to clarify information provided.

We emphasise that these cases were not dropped from the sample or cleaned automatically. Rather, they were reviewed in light of additional information elsewhere in the survey, from prior rounds, and after discussions with the team leader or enumerator who conducted the interview. Much of this quality control occurred during fieldwork, to maximize the likelihood that enumerators would recall the interview in question.

As noted above, qualitative data was also subject to quality control. In addition, we conducted post-fieldwork debriefing sessions with several of the team leaders, with the goal of improving data collection for future rounds, and to understand subtleties of the qualitative data – or issues that were revealed outside of qualitative interviews. At times, these post-fieldwork interviews with team leaders are cited in this report.

Qualitative data was transcribed verbatim in Somali and then translated into English, resulting in complete English-language transcripts of all FGDs, KIIs, and participatory exercises. Qualitative analysis was the responsibility of all evaluation team members, rather than assigning a dedicated qualitative analyst. The reason for this decision was to maximize the insights gained from the qualitative data, as multiple individuals reading the interviews notice different aspects of participants' answers. This approach also ensured that there was not a sharp disjuncture between quantitative and qualitative analysis, as the goal was for the two sources of data to "speak to each other".

In terms of process, the evaluation team prioritised reading the complete qualitative interviews before turning to in-depth quantitative analysis. This first reading produced a number of specific findings – many of which were not explicitly addressed in the quantitative tools – which were noted for incorporation into a given section of the report. This process ensured that themes could emerge organically from the qualitative data, rather than being tied too tightly to explaining or providing nuance to quantitative results. Next, the team performed the core quantitative analysis; again, the benefit of fully reading the qualitative transcripts in advance was that insights from the qualitative interviews often suggested adjustments that should be made to the quantitative analysis. Following the quantitative analysis, team members went back to the qualitative data in a more purposeful way, seeking to understand specific findings from, provide nuance to, and find evidence that either confirmed or contradicted, the quantitative findings. While our approach to qualitative analysis was not as systematic – in the sense of establishing a formal and rigorous coding scheme – as some forms of qualitative data analysis, our preferred approach ensured that we maximized insights from the qualitative data and that findings from the one data source influenced analysis from the other in productive and symbiotic ways. We employed a similar approach, where appropriate, to the incorporation of multiple quantitative data sources, attempting to triangulate across multiple data sources in a manner that maximized the quality of our inferences and the value of the analysis.

The EL evaluation constitutes the fourth round of contact between the evaluation team and its fieldwork researchers, on one hand, and the SOMGEP-T intervention and comparison communities, on the other. In actual practice, some of the communities were part of the SOMGEP phase I evaluations, which were completed by many of the same fieldwork researchers. For the purposes of facilitating effective fieldwork, we made an effort during the EL to dispatch field teams back to sites they had visited previously. This approach enhances their ability to work productively with the school administrators, teachers, and the community; it

also increases the likelihood that they will successfully re-contact girls, because they are familiar with the community and may actually remember some of the girls in question.

Data analysis was conducted primarily in Stata, using .do files for all analysis to enhance replicability both internally and externally. The evaluation lead reviewed the cleaning .do files that had been constructed during fieldwork to correct day-to-day quality control issues; they then completed additional cleaning and merged the data with the baseline, ML1, and ML2 datasets, and combined datasets where needed for the analysis (e.g., combining the head teacher survey, which provides information about school characteristics, with the household survey). The cleaning and data compilation process was documented via .do files as well.

Qualitative data was analysed in Excel. The English transcripts were completed in Excel, and the design of the transcript form allowed the evaluation team to filter questions by topical area and respondent type, in order to read many related responses at once. Rather than code qualitative interviews in a formal manner, we made an effort to read transcript relevant to a given section verbatim, as critical information is often hidden in responses to questions that one would not expect. The goal of qualitative analysis was to allow natural themes to emerge, which we accomplished by reading the qualitative transcripts, as much as possible, prior to starting the quantitative analysis. Following the quantitative analysis, we revisited the qualitative data with new questions and themes in mind, seeking to understand the trends and anomalies in the quantitative data.

Annex 2. Learning Test Design and Calibration

Equivalence of learning assessments across rounds

Learning constitutes one of the core outcomes for SOMGEP-T, and learning scores are measured through adapted versions of the EGMA and EGRA tools, as noted in our previous discussion of the evaluation design. In that discussion, we noted that CARE developed new versions of the learning assessments as part of this midline evaluation, in line with revisions made during the previous midline, in late 2018, and ML2, in 2019. Revisions to the data collection tools beg the question of whether the learning assessments are of comparable difficulty from round to round, as changes that made the assessments more or less difficult would influence the learning scores achieved by students.

At the outset, it is important to note that changes in the difficulty of the learning assessment do not, in any way, threaten the inferences we draw regarding program impact on learning. Even if the assessments became markedly more or less difficult from previous years, these changes would not influence or render invalid our conclusions regarding program impact. Thanks to the difference-in-differences design of the evaluation, any change in difficulty over time is applied to both intervention and comparison girls equally, so that the difference does not influence our estimates of program impact.

Our interest in this section is on the impact of changes in assessment difficulty. Changes in assessment difficulty apply equally to all girls– the easier or more difficult assessments are taken by all girls at midline, so any change in the gap between intervention and comparison group learning scores cannot be a function of the change in assessment difficulty, but because of the program itself.[284] Because the change in assessment difficulty applies to all girls equally, the change does not threaten causal inferences we draw in this report.

Although overall conclusions regarding program impact are unaffected by the equivalence of learning assessments from round to round, there are aspects of the analysis where such equivalence is necessary. For instance, our analysis of changes in ALP girls' learning scores does not have the benefit of a comparison group or difference-in-differences design, so small changes in assessment difficulty could produce bias in our analysis of ALP girls.

In order to investigate the equivalence of learning assessments over multiple rounds, the endline round of data collection included data collection using the previous ML2 assessment. Specifically, a sample of 47 cohort girls, who completed the standard learning assessments administered in this round, were recruited into a comparison sample. Following their completion of the standard learning assessments used in this round, they completed the Somali literacy, English literacy, and numeracy assessments from the second midline round (ML2, from 2019). By completing both assessments with the same group of girls, we can compare their scores on the two assessments to judge their equivalence; collecting data from the same girls has the advantage of eliminating sampling variation from the analysis, strengthening conclusions regarding equivalence.

The table below reports the mean scores on each assessment among this sample of 47 respondents; the right column reports the difference in means. Across all three tests, scores are very similar from ML2 to EL, with the largest gap occurring in numeracy.

[284] Or some other exogenous shock, aside from changes in learning assessment difficulty that differentially impacts one of the two groups.

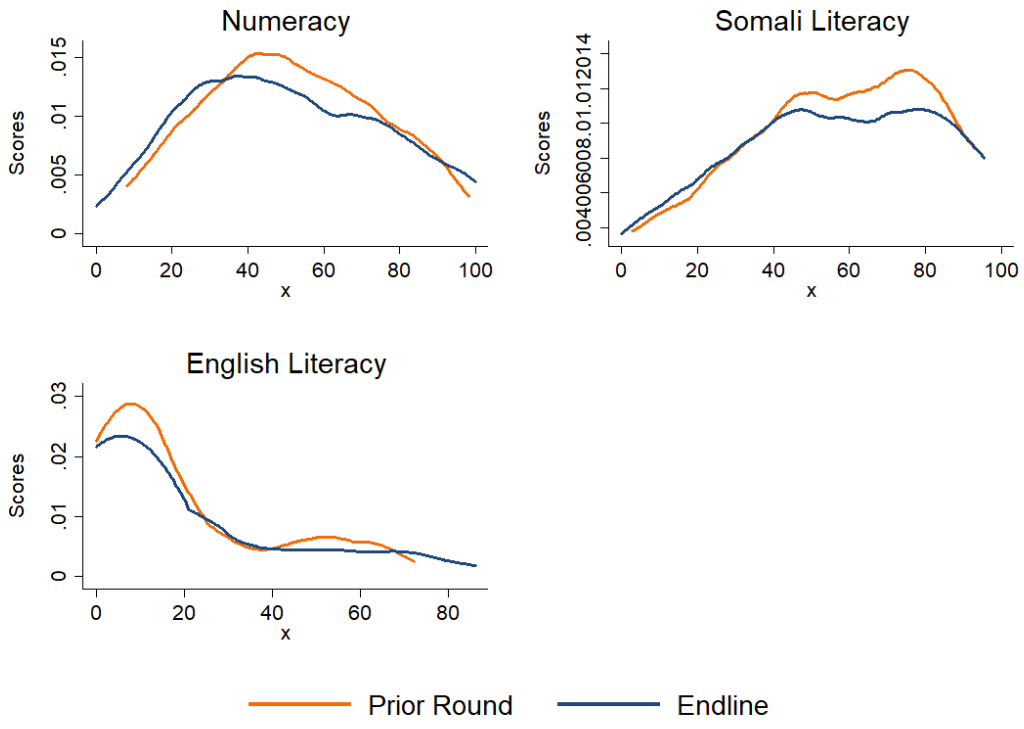
TABLE 83: PRIOR ROUND AND ENDLINE ASSESSMENT SCORES, COMPARISON SAMPLE

Assessment Subject	ML2 Round	Endline Round	Difference in Scores
English Literacy	18.8	18.2	-0.6
Numeracy	51.2	50.4	-0.8
Somali Literacy	55.2	54.8	0.3

Given the careful and deliberate approach to learning assessment revisions made between evaluation rounds, the similarities are not surprising. In this round (which compares EL to ML2), differences between average scores are even less than previously observed (comparing ML1 and ML2), though by small margins. In changing learning assessments, the subtasks maintained their fundamental objective and the skills that they tested. In most cases, the subtasks were altered in small ways, such as small changes in the numbers employed in addition or subtraction problems, adjustments sufficiently small that there is no theoretical reason to expect them to be more or less difficult as a result.

To further illustrate the similarities we plotted the three scores – including Somali literacy, English literacy, and numeracy, for the sample of 47 girls from both ML2 and EL assessments. Distributions between rounds remain similar, with slight variations. Across all three assessments, we see a slightly flatter distribution whereby there is moderately lighter clustering.

FIGURE 54: PRIOR ROUND AND ENDLINE ASSESSMENT SCORES, COMPARISON SAMPLE (N = 47)



As noted above, differences in assessment difficulty will not affect our conclusions regarding overall program impact on cohort girls, but may influence other findings we report. We suggest caution when it comes to analysing changes in subtask-specific outcomes, as individual subtasks are not necessarily equivalent between evaluation rounds. It is not clear whether this is due to bias from a relatively small sample, but performance in the comparison sample of 47 girls yielded statistically significant differences on two of the subtasks: numeracy subtask 10 and English literacy subtask 6. In practice, these subtask-specific differences balanced each other across the entire assessment; however, it is worth noting that individual subtasks may show differences in difficulty across rounds, even if the assessments overall do not.

TABLE 84: SUBTASK-LEVEL DIFFERENCES BETWEEN ENDLINE AND PRIOR ROUND ASSESSMENTS, COMPARISON SAMPLE (N = 47)

Subtask Number	ML2 Round	Endline Round	P-Value for Difference
Numeracy			
1	50.9	49.6	.31
2	90.4	88.9	.18
3	77.9	77.4	.66
4	65.1	66	.74
5	52.3	52.3	1.0
6	79.8	77.1	.13
7	51.5	52.8	.61
8	19.1	20.9	.21
9	31.5	34.5	.16
10	11.1	15.7	.05*
11	24.5	27.7	.37
Somali literacy			
1	81.8	80.6	.15
2	69.1	69.1	1.0
3	72.9	71.8	.53
4	51.9	52.1	.75
5	47.3	48.9	.44
6	48.9	49.6	.8
7	52.5	51.1	.49
8	41.8	44.7	.21

9	27.2	28.5	.44
English literacy			
1	46.7	46	.27
2	31.1	31.7	.43
3	23.4	22.9	.74
4	21.9	22.9	.1
5	11.2	11.2	1.0
6	10.1	12.8	.06*
7	9	11.2	.21
8	5.3	3.2	.42
9	5.3	7.4	.49

Annex 3. Supplemental Learning Tables

In Section 4, we occasionally referred to additional tables provided in the annexes, typically in cases in which tables had been provided in previous rounds but were of less relevance to the endline evaluation. Below, we report learning outcomes at endline among a variety of subgroups, in line with reporting from previous rounds. Note that the results below are focused on cross-sectional differences in outcomes between subgroups, aggregating across both intervention and comparison groups. They are not concerned with estimating programme impact; for estimates of programme impact within particular subgroups, see Section 4.2 of the main report.

This annex also provides additional tables documenting performance of girls on foundational skills. In Section 4.7 of the report, we analysed foundational skills – and gaps in those skills – among cohort girls in intervention schools only. The tables were intended to facilitate comparisons between intervention girls, as a whole, and those who were enrolled in school at the endline. In the tables below, we provide the same analysis, but allowing contrasts to be drawn between intervention and comparison girls (for the first three tables). A second set of three tables has been added to report foundational skill gaps among only comparison girls, comparing those enrolled in school to the overall sample of comparison girls (the second set of three foundational skills gap tables).

Subgroup Learning Outcomes

The tables below document learning outcomes at endline among relevant subgroups of girls, where subgroups are defined by household, girl, and school characteristics, among others. These tables mirror the subgroup analyses conducted in previous rounds – especially the baseline – when attempting to understand patterns or and barriers to learning.

TABLE 85: LEARNING OUTCOMES, AT ENDLINE, AMONG VARIOUS SUBGROUPS

Subgroup	Numeracy	Somali Literacy	English Literacy	Financial Lit.	Obs. (n)
Overall	51.5%	54.5%	21.7%	22.0%	1194
Household Demographics					
Female-headed household	55.7%	59.6%	25.3%	25.2%	557
Part orphan	51.0%	53.5%	20.9%	20.3%	188
Full orphan	53.0%	65.1%	20.5%	24.0%	11
Lives without parents	48.2%	50.7%	19.2%	23.2%	281
HoH has no education	47.9%	49.4%	17.8%	19.2%	430
HoH has no formal education	51.5%	53.3%	21.5%	21.6%	767
CG has no education	47.2%	49.4%	17.6%	18.8%	484
CG has no formal education	50.5%	52.8%	20.9%	21.0%	879
Household Economic Status and Wealth					

HoH pastoralists	41.9%	42.7%	12.2%	15.2%	120
HoH no occupation	53.4%	56.6%	21.9%	22.5%	524
Seasonal migration	39.1%	41.0%	12.7%	17.5%	100
Owns camels	42.1%	43.0%	13.4%	19.1%	160
Owns medium livestock	49.8%	52.4%	19.8%	20.3%	767
Owns a mobile phone	51.6%	54.6%	21.8%	21.9%	1141
Has regular access to water	51.5%	55.1%	22.1%	21.8%	607
Owns land alone	52.4%	55.4%	22.2%	21.9%	747
Owns land	52.2%	54.9%	22.1%	21.8%	808
Poor quality roof	41.6%	42.5%	12.4%	15.2%	263
Gone without food many days	51.7%	54.7%	22.5%	26.3%	90
Gone without food most days	46.8%	51.4%	20.1%	23.6%	20
Gone without water many days	48.6%	51.9%	19.2%	20.4%	290
Gone without water most days	47.2%	48.6%	17.5%	21.4%	77
Gone without medicines many days	48.9%	50.8%	18.8%	19.8%	373
Done without medicines most days	47.2%	48.2%	19.9%	20.6%	160
Gone without cash income many days	50.4%	53.8%	20.4%	20.1%	462
Gone without cash income most days	48.7%	49.5%	19.4%	18.7%	151
School Infrastructure and Environment					
Caregiver believes travel to school is unsafe for girls	37.5%	37.5%	7.9%	16.7%	12
Difficult to move around school	56.4%	61.7%	24.7%	25.2%	133
Doesn't use drinking water facilities	48.7%	53.6%	20.7%	17.0%	74
Doesn't use toilet at school	53.9%	59.6%	26.0%	20.2%	115
Doesn't use areas where children play/socialise	59.5%	64.9%	31.2%	25.9%	404

Teachers treat boys and girls differently in the classroom	54.1%	61.0%	26.3%	24.0%	208
Agrees teachers often absent from class	54.5%	59.7%	24.6%	25.1%	157
Afraid of teacher	57.5%	62.2%	25.4%	25.0%	538
Teacher punishes wrong answers	57.2%	62.9%	26.2%	23.1%	416
Teacher uses physical punishment	53.1%	62.3%	27.9%	28.3%	104
No computers at school	58.8%	63.3%	27.9%	24.5%	667
Cannot use learning materials at school	58.3%	59.2%	25.7%	25.3%	104
Not enough seats for all students	63.4%	67.8%	31.5%	31.9%	59
Girl says teacher asks girls/boys more questions (not equal)	53.2%	59.3%	26.5%	25.5%	61
Girl says teacher asks girls/boys harder questions (not equal)	51.2%	58.0%	25.3%	21.6%	100
Girl and Household Decision-Making					
High chore burden (whole day spent on chores)	36.6%	35.0%	9.9%	15.8%	281
Girl attends school most days	59.5%	65.1%	29.0%	25.8%	681
PCG and family member not involved in CEC	61.2%	63.7%	28.4%	26.6%	93
Girl feels no choice whether to attend or stay in school	51.3%	54.4%	21.7%	21.0%	949

Subgroup Programme Impact on Learning

The tables in this section expand on the analysis of heterogeneous programme impacts, in which we are interested in how the programme affected learning outcomes among particular subgroups. This differs from the previous section, which simply considers learning outcomes in the endline cross-section, i.e. learning outcomes at the endline, without reference to whether the programme was associated with learning outcomes. The tables below report difference-in-differences estimates of how the programme impacted *specific* subgroups of girls; in other words, the sample used for each analysis is limited to girls who fall into particular subgroups. This is a straightforward way to ask whether the programme had differential impact among, e.g., pastoralist households, compared to the programme's aggregate impact across the entire sample.

We report two sets of tables. The first, in the subsection immediately below, is a "naïve" difference-in-difference estimate, which simply reports the mean learning score for intervention and comparison groups at both baseline and endline, limiting the sample to girls with a particular characteristic (e.g., girls from pastoralist households). We calculate the difference-in-differences from these simple means, without controlling for additional factors that might influence learning scores. The advantage of these tables is that they make clear the trends over time in both intervention and comparison groups; they also are a very good first approximation of the programme's subgroup impacts. The first set of tables include, separately, results for the in-school girl and out-of-school girl cohorts.

The second set of tables reports difference-in-differences estimates from a linear regression model, assessing impact from baseline to endline (first two columns) and from ML2 to endline (right-most two columns). The tables address learning in numeracy, Somali literacy, and English literacy, respectively. In each table, "Impact" refers to the difference-in-differences estimate, or regression coefficient from a difference-in-differences model estimated in a regression framework.

Simple Subgroup Impacts

TABLE 86: NUMERACY SCORES, BY ROUND AND INTERVENTION STATUS, FOR THE IN-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	41.6	59.5	36.9	57.3	1222	-2.6
Pastoralist head of household	33.4	49.7	35.2	59.1	182	-7.6
Head of household has no occupation	43.3	59.8	39.7	58.6	752	-2.4
Poor quality roof	34.9	51.7	30.4	51.9	444	-4.8
Gone without enough food many/most days, last year	50.6	58.9	36.2	58.2	166	-13.7
Gone without enough clean water most days, last year	42.7	57.7	36.3	65	162	-13.7
Female-headed household	42.7	60.9	37.7	58.1	792	-2.2

Lives without parents	49	58.5	38.7	57.2	288	-9.2
HoH has no education	42.5	59.4	36.3	56.4	886	-3.1
HoH has no formal education	42.4	60.1	36.9	56.9	978	-2.3
CG has no education	40.8	58.9	36.6	55.6	986	-0.9
CG has no formal education	42.2	59.5	36.8	56.4	1116	-2.3
Neither HoH or caregiver has no education	41.7	58.5	36.1	54.8	764	-1.9
High chore burden (whole day spent on chores)	56.2	66	39.6	60.2	146	-10.9
Girl feels she has some influence over schooling decisions	42.4	61.7	35.2	57.9	898	-3.3
Schooling decisions made entirely by adults, per caregiver	42	59.7	38.2	58.3	956	-2.4
Girl attends school most days	41.7	59.7	37.7	57.4	1086	-1.6
No family member is involved in CEC	41.2	60.1	31	52.8	150	-2.9
Girl has mental health disability	38.5	59	31.9	51.8	116	0.6
Girl has disability other than mental health (physical, cognitive, communicative)	38.5	58	44	51.3	76	12.2
Girl has any disability	37.1	57.5	36.5	52.3	178	4.6

TABLE 87: SOMALI LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, FOR THE IN-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	46.8	69.7	43	67.3	1222	-1.4

Pastoralist head of household	41.8	55.8	39.8	67.3	182	-13.5
Head of household has no occupation	46.7	69.9	44.9	70	752	-1.9
Poor quality roof	36	61.1	32.6	61.7	444	-4.1
Gone without enough food many/most days, last year	45.5	66.6	41.5	69.3	166	-6.7
Gone without enough clean water most days, last year	44.9	66.8	42	70.1	162	-6.2
Female-headed household	49	70.4	45.9	68.6	792	-1.2
Lives without parents	51.4	70.1	50.6	69.6	288	-0.3
HoH has no education	46.6	68.4	40.6	65.7	886	-3.3
HoH has no formal education	46.7	69.5	41.7	65.5	978	-1
CG has no education	43.7	68.6	40.1	65.5	986	-0.5
CG has no formal education	46.4	69.4	42.1	66.1	1116	-1
Neither HoH or caregiver has no education	44	67.8	38.9	64.3	764	-1.6
High chore burden (whole day spent on chores)	68.8	72.6	62.2	68.7	146	-2.6
Girl feels she has some influence over schooling decisions	49	72.2	41.3	67.8	898	-3.3
Schooling decisions made entirely by adults, per caregiver	47.7	69.2	44.6	68.2	956	-2.1
Girl attends school most days	46.2	69.5	41.3	68	1086	-3.4
No family member is involved in CEC	48.2	73.1	43.2	64.3	150	3.8
Girl has mental health disability	34.7	64.7	41.7	63	116	8.6
Girl has disability other than mental health	39.7	69.7	54.5	69.6	76	14.8

(physical, cognitive, communicative)						
Girl has any disability	36.7	66.4	46.3	65.9	178	10.2

TABLE 88: ENGLISH LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, FOR THE IN-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	13.3	29.6	7.6	23.2	1222	0.8
Pastoralist head of household	8.7	18.5	7.3	19.1	182	-2
Head of household has no occupation	13.3	30.7	8.9	22.8	752	3.5
Poor quality roof	9.4	21	5.2	15.4	444	1.3
Gone without enough food many/most days, last year	15	33.2	10.2	25.5	166	2.8
Gone without enough clean water most days, last year	14.8	28.6	12.9	25	162	1.7
Female-headed household	13.7	31.6	8.5	25.1	792	1.2
Lives without parents	12.4	30.3	8.1	23	288	3
HoH has no education	14.5	29.5	7.5	22.6	886	-0.1
HoH has no formal education	13.9	30	7.7	22.6	978	1.3
CG has no education	12.5	28.4	7.3	21.5	986	1.8
CG has no formal education	13.2	29.5	7.6	22.1	1116	1.9
Neither HoH or caregiver has no education	13.3	27.8	7.3	22.4	764	-0.6
High chore burden (whole day spent on chores)	24.8	33.2	14.5	27.8	146	-4.9
Girl feels she has some influence over schooling decisions	14.3	31.6	7.4	21.8	898	2.9

Schooling decisions made entirely by adults, per caregiver	13.7	30.2	8.6	23.9	956	1.2
Girl attends school most days	13.3	29.8	7.4	23.6	1086	0.4
No family member is involved in CEC	13.4	31.7	7.5	20	150	5.8
Girl has mental health disability	11.8	28.2	6.4	23.3	116	-0.5
Girl has disability other than mental health (physical, cognitive, communicative)	8.6	26.5	9.9	21.5	76	6.3
Girl has any disability	10.8	28.7	7.6	23.5	178	2.1

TABLE 89: FINANCIAL LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, FOR THE IN-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	5.3	25.6	5.3	24.6	1222	1
Pastoralist head of household	6.5	16.9	6.4	19.1	182	-2.3
Head of household has no occupation	3.7	25.7	5.6	24.7	752	3
Poor quality roof	3.3	18.5	3.9	21.3	444	-2.2
Gone without enough food many/most days, last year	5.5	30.3	2.8	25.6	166	2
Gone without enough clean water most days, last year	5.5	27.8	5.6	21.7	162	6.2
Female-headed household	5.9	28.2	6.4	25.5	792	3.3
Lives without parents	7.9	27	5.9	27.3	288	-2.3
HoH has no education	5.8	25.2	5.3	23.2	886	1.5

HoH has no formal education	5.6	26.1	5.5	23.7	978	2.2
CG has no education	6	25	5.4	24.1	986	0.4
CG has no formal education	5.9	25.6	5.6	24.3	1116	1
Neither HoH or caregiver has no education	6.2	24.4	5	22.7	764	0.6
High chore burden (whole day spent on chores)	7.9	36.4	9.1	28.1	146	9.5
Girl feels she has some influence over schooling decisions	5.5	27.9	5.2	24.2	898	3.4
Schooling decisions made entirely by adults, per caregiver	5.7	26.4	5.7	26	956	0.5
Girl attends school most days	5.2	25.7	5.2	24.8	1086	0.8
No family member is involved in CEC	3.7	27.3	0.9	27.9	150	-3.4
Girl has mental health disability	2.9	28.7	1.3	18.6	116	8.5
Girl has disability other than mental health (physical, cognitive, communicative)	4	25.3	6.6	28.1	76	-0.3
Girl has any disability	2.9	27.3	3.3	21.8	178	5.9

TABLE 90: NUMERACY SCORES, BY ROUND AND INTERVENTION STATUS, AMONG OUT-OF-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	14.6	36.5	10.1	31.4	708	0.7
Pastoralist head of household	12.6	33.8	6.7	23.9	156	3.9

Head of household has no occupation	15.4	38.2	11.6	32.9	446	1.5
Poor quality roof	5.8	28.8	6.6	22.7	332	6.9
Gone without enough food many/most days, last year	9	31.3	11.5	30.6	152	3.2
Gone without enough clean water most days, last year	12.9	38.2	7.4	24.5	164	8.1
Female-headed household	16.4	37	9.3	35.9	434	-6.1
Lives without parents	16.6	36.6	11.7	33.7	266	-2
HoH has no education	13	36	9.4	29.7	586	2.7
HoH has no formal education	14	36.4	9.3	31	618	0.7
CG has no education	11.6	34.6	9	30.1	612	1.9
CG has no formal education	13.8	35.8	9.1	30.9	658	0.2
Neither HoH or caregiver has no education	11.4	35.1	9.4	29.6	534	3.5
High chore burden (whole day spent on chores)	15.9	35.2	7.2	28.7	330	-2.1
Girl feels she has some influence over schooling decisions	11.1	34.9	10.1	33.7	518	0.2
Schooling decisions made entirely by adults, per caregiver	16.8	39.4	11.9	32	496	2.5
Girl has mental health disability	13.4	37.2	8.6	33.6	78	-1.2
Girl has disability other than mental health (physical, cognitive, communicative)	6.6	25.2	11.6	26.1	40	4.1
Girl has any disability	10.6	34.5	10.1	32.3	106	1.7

TABLE 91: SOMALI LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, AMONG OUT-OF-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	14.4	36.1	11.9	29.4	708	4.2
Pastoralist head of household	14.5	39	5.3	19	156	10.8
Head of household has no occupation	14.7	37.1	13.7	32.9	446	3.2
Poor quality roof	3.9	24.6	6.4	18	332	9.1
Gone without enough food many/most days, last year	12.3	35.6	14.6	35.6	152	2.4
Gone without enough clean water most days, last year	12.2	33.5	9.2	21.3	164	9.1
Female-headed household	14.6	34.8	12	34.4	434	-2.1
Lives without parents	14.9	35	13.6	29.6	266	4.1
HoH has no education	13.5	34.9	11.6	28.7	586	4.3
HoH has no formal education	14	35.4	11.2	29.2	618	3.5
CG has no education	11.4	32.7	10.6	28	612	3.9
CG has no formal education	12.7	34.4	10.6	28.3	658	3.9
Neither HoH or caregiver has no education	11.7	32.4	11.4	28.3	534	3.8
High chore burden (whole day spent on chores)	16.3	36.7	7.8	26.7	330	1.5
Girl feels she has some influence over schooling decisions	11.5	33.9	11.6	32	518	2.1
Schooling decisions made entirely by adults, per caregiver	18.3	40.5	13.6	29.6	496	6.1

Girl has mental health disability	18.6	39.5	7.3	30.8	78	-2.6
Girl has disability other than mental health (physical, cognitive, communicative)	11.8	15	14.9	27.1	40	-8.9
Girl has any disability	16	31.9	11.1	32.1	106	-5.2

TABLE 92: ENGLISH LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, AMONG OUT-OF-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	6.6	10.9	4.4	6.7	708	2
Pastoralist head of household	6.5	7.7	4.5	3.4	156	2.2
Head of household has no occupation	6.9	11.6	4.4	7.7	446	1.4
Poor quality roof	6.6	3.5	3.2	2.8	332	-2.7
Gone without enough food many/most days, last year	6.4	10	4.6	10.7	152	-2.5
Gone without enough clean water most days, last year	6.4	13.5	4.4	5.8	164	5.6
Female-headed household	6.1	11.9	5.4	8.7	434	2.5
Lives without parents	5.5	10.3	3.3	5.7	266	2.4
HoH has no education	6.5	10	4.7	6.4	586	1.7
HoH has no formal education	6.5	10.6	4.5	6.9	618	1.7
CG has no education	6.9	9.4	4.5	6.1	612	0.9
CG has no formal education	6.5	10.1	4.4	6.8	658	1.2
Neither HoH or caregiver has no education	6.5	9.3	4.8	5.9	534	1.6

High chore burden (whole day spent on chores)	5.7	8.2	3.5	5.3	330	0.8
Girl feels she has some influence over schooling decisions	6.8	9.5	4.7	7.9	518	-0.5
Schooling decisions made entirely by adults, per caregiver	7.3	14.2	4.6	6.4	496	5
Girl has mental health disability	6.2	13.5	4	6.4	78	5
Girl has disability other than mental health (physical, cognitive, communicative)	5.8	5	7.4	2.7	40	3.9
Girl has any disability	6.5	10.9	5.5	5.4	106	4.5

TABLE 93: FINANCIAL LITERACY SCORES, BY ROUND AND INTERVENTION STATUS, AMONG OUT-OF-SCHOOL GIRL COHORT

Subgroup	Intervention		Comparison		Obs. (n)	Diff-in-Diff
	Baseline	Endline	Baseline	Endline		
Full Panel Sample	3.1	15.3	2.2	14.2	708	0.2
Pastoralist head of household	3.9	16.1	1.9	9.5	156	4.5
Head of household has no occupation	3.2	14.7	2.2	14.7	446	-1.1
Poor quality roof	1.2	11.4	0.6	8.9	332	1.8
Gone without enough food many/most days, last year	1.4	15.6	3.8	16.5	152	1.4
Gone without enough clean water most days, last year	4.7	16.8	2.5	12.1	164	2.6
Female-headed household	4.2	15.2	3.4	15.8	434	-1.5
Lives without parents	3.9	20	4.3	16.8	266	3.6

HoH has no education	3.1	13.6	1.9	12.6	586	-0.2
HoH has no formal education	3.2	14	1.9	13.3	618	-0.6
CG has no education	2	13.6	1.8	12.9	612	0.5
CG has no formal education	3	14.2	1.7	13.6	658	-0.6
Neither HoH or caregiver has no education	2.1	13.1	2	12.4	534	0.5
High chore burden (whole day spent on chores)	5	16.8	2.2	13.8	330	0.2
Girl feels she has some influence over schooling decisions	1.3	14.2	2	15.3	518	-0.3
Schooling decisions made entirely by adults, per caregiver	4	15.6	2.7	15	496	-0.7
Girl has mental health disability	0	14.1	1.3	16.5	78	-1
Girl has disability other than mental health (physical, cognitive, communicative)	0	14.8	3	12.9	40	4.9
Girl has any disability	0	14.4	2.2	16.3	106	0.3

Regression-Based Subgroup Impacts

TABLE 94: SUBGROUP PROGRAMME IMPACTS ON NUMERACY SCORES

Subgroup	Baseline-to-Endline			ML2-to-Endline		
	Obs.	Impact	P-Value	Obs.	Impact	P-Value
Full Panel Sample	965	-2.1	0.4	975	-2.4	0.1
Pastoralist head of household	169	-0.6	0.9	159	-1.8	0.6
Head of household has no occupation	599	-2.1	0.5	633	-2.7	0.2
Poor quality roof	388	-0.4	0.9	349	-0.7	0.7
Gone without enough food many/most days, last year	159	-5.4	0.3	169	-4.4	0.2
Gone without enough clean water most days, last year	163	-3.3	0.5	150	-3.8	0.3
Female-headed household	613	-4	0.1	631	-2.3	0.2
Lives without parents	277	-7.1	0	247	-0.8	0.8
HoH has no education	736	-1.1	0.7	581	-4.1	0
HoH has no formal education	798	-1.5	0.5	803	-3.3	0
CG has no education	799	-0.8	0.7	661	-2.8	0.1
CG has no formal education	887	-2.1	0.4	881	-2.5	0.1
Neither HoH or caregiver has no education	649	-0.5	0.9	512	-3	0.1
High chore burden (whole day spent on chores)	238	-4.3	0.3	148	0	1
Girl feels she has some influence over schooling decisions	708	-2.5	0.3	493	-4.9	0
Schooling decisions made entirely by adults, per caregiver	726	-1.5	0.5	727	-2.6	0.1
Girl has mental health disability	97	3.3	0.5	104	-3.4	0.4

Girl has disability other than mental health (physical, cognitive, communicative)	58	9.4	0.2	67	-3.1	0.6
Girl has any disability	142	5.7	0.3	154	-1.7	0.7

TABLE 95: SUBGROUP PROGRAMME IMPACTS ON SOMALI LITERACY SCORE

Subgroup	Baseline-to-Endline			ML2-to-Endline		
	Obs.	Impact	P-Value	Obs.	Impact	P-Value
Full Panel Sample	965	0.2	0.9	975	-0.9	0.6
Pastoralist head of household	169	-0.4	1	159	3.7	0.3
Head of household has no occupation	599	-1.2	0.7	633	-1.8	0.4
Poor quality roof	388	2.3	0.5	349	3.5	0.2
Gone without enough food many/most days, last year	159	-1.8	0.7	169	-3.2	0.5
Gone without enough clean water most days, last year	163	1.9	0.7	150	-3.8	0.5
Female-headed household	613	-2	0.5	631	-1.1	0.6
Lives without parents	277	-0.2	1	247	-1	0.8
HoH has no education	736	-0.1	1	581	-1.9	0.4
HoH has no formal education	798	0.7	0.8	803	-1.4	0.5
CG has no education	799	0.8	0.7	661	-2	0.4
CG has no formal education	887	0.5	0.8	881	-1.2	0.5
Neither HoH or caregiver has no education	649	0.7	0.8	512	-3.4	0.2
High chore burden (whole day spent on chores)	238	-1.3	0.8	148	0.4	0.9
Girl feels she has some influence over schooling decisions	708	-1.4	0.6	493	-2.7	0.3
Schooling decisions made entirely by adults, per caregiver	726	0	1	727	-0.1	1

Girl has mental health disability	97	6	0.3	104	-3.1	0.4
Girl has disability other than mental health (physical, cognitive, communicative)	58	14.2	0.2	67	-1.7	0.8
Girl has any disability	142	7.8	0.1	154	-1.4	0.7

TABLE 96: SUBGROUP PROGRAMME IMPACTS ON ENGLISH LITERACY SCORES

Subgroup	Baseline-to-Endline			ML2-to-Endline		
	Obs.	Impact	P-Value	Obs.	Impact	P-Value
Full Panel Sample	965	1.9	0.5	975	1	0.6
Pastoralist head of household	169	1.8	0.5	159	3.9	0.3
Head of household has no occupation	599	3.1	0.3	633	2.7	0.3
Poor quality roof	388	1.4	0.6	349	3.7	0.1
Gone without enough food many/most days, last year	159	1.6	0.8	169	2.1	0.7
Gone without enough clean water most days, last year	163	5.2	0.2	150	2.7	0.6
Female-headed household	613	1.7	0.6	631	2.2	0.4
Lives without parents	277	2.2	0.5	247	-0.1	1
HoH has no education	736	1.7	0.5	581	2.3	0.2
HoH has no formal education	798	2.3	0.4	803	1.7	0.4
CG has no education	799	2.1	0.4	661	1.3	0.5
CG has no formal education	887	2.3	0.4	881	1.2	0.5
Neither HoH or caregiver has no education	649	1.3	0.6	512	2	0.3
High chore burden (whole day spent on chores)	238	-1.2	0.7	148	1	0.8
Girl feels she has some influence over schooling decisions	708	2.6	0.4	493	1	0.6

Schooling decisions made entirely by adults, per caregiver	726	2.9	0.3	727	0.8	0.7
Girl has mental health disability	97	3.6	0.5	104	2.3	0.6
Girl has disability other than mental health (physical, cognitive, communicative)	58	8.7	0.2	67	3.2	0.5
Girl has any disability	142	5.7	0.2	154	4.5	0.1

Foundational Skill Gaps

Additional tables highlighting gaps in numeracy and literacy skills, among varying samples of girls, are provided beginning on the next page.

TABLE 97: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION AND COMPARISON GIRLS, NUMERACY

Subtask	1	2	3	4	5	6	7	8	9	10	11
Skill Assessed	Missing number	Addition (Level 1)	Subtraction (Level 1)	Addition (Level 2)	Subtraction (Level 2)	Word problems (add/sub)	Multiplic. (Level 1)	Multiplic. (Level 2)	Division (Level 1)	Division (Level 2)	World problems (mult/div)
Intervention Girls											
Non-Learner (0%)	3.6%	5.3%	7.8%	16.8%	25.7%	12%	26.3%	65.9%	41.4%	77.4%	52.8%
Emergent Learner (1-40%)	34.3%	2%	4.4%	13.9%	12%	5.1%	11.1%	12.5%	21.3%	10%	0%
Established Learner (41-80%)	50.8%	15.4%	11.2%	31.6%	32.2%	28.8%	33.2%	10.6%	17.6%	7.8%	19.9%
Proficient Learner (81-100%)	11.4%	77.3%	76.6%	37.7%	30.1%	54%	29.4%	11.1%	19.6%	4.8%	27.3%
Full Cohort (intervention and Comparison Girls)											
Non-Learner (0%)	4.6%	6.1%	9.5%	18.3%	28.7%	13.3%	28.5%	69.2%	43.7%	79.7%	54.7%
Emergent Learner (1-40%)	36.3%	3.3%	4.7%	14.7%	11.8%	4.6%	12%	10.6%	22.5%	9.1%	0%
Established Learner (41-80%)	48.9%	15.7%	10.9%	31.2%	31.4%	29.5%	32.5%	11.1%	16.6%	6.9%	18.4%
Proficient Learner (81-100%)	10.2%	74.9%	74.9%	35.8%	28.1%	52.6%	27.1%	9.2%	17.2%	4.3%	26.9%

TABLE 98: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION AND COMPARISON GIRLS, SOMALI LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Reading Words	Reading Comp (easy)	Reading Comp (medium)	Reading Fluency	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)	Sentence completion
Intervention Girls									
Non-Learner (0%)	8.7%	16.8%	18.4%	13.7%	27.6%	31.3%	37.4%	44.9%	59.3%
Emergent Learner (1-40%)	4.5%	7.5%	3.3%	16%	16.8%	13.4%	3.6%	3.4%	17.4%
Established Learner (41-80%)	12.1%	36.6%	28.5%	25.1%	43%	23.5%	5.9%	6.9%	7.9%
Proficient Learner (81-100%)	74.6%	39.1%	49.8%	45.2%	12.6%	31.8%	53.1%	44.9%	15.3%
Full Cohort (intervention and Comparison Girls)									
Non-Learner (0%)	10.7%	19.1%	21.1%	15.7%	29.7%	33.9%	39.9%	47.9%	60%
Emergent Learner (1-40%)	4.7%	7.2%	3.7%	17.3%	17.4%	12.6%	3.2%	3%	17.5%
Established Learner (41-80%)	12.7%	37.2%	28.3%	27.9%	40.7%	22.4%	6.1%	5.8%	9%
Proficient Learner (81-100%)	71.9%	36.5%	46.9%	39.1%	12.1%	31.2%	50.8%	43.3%	13.6%

TABLE 99: FOUNDATIONAL SKILL GAPS AMONG INTERVENTION AND COMPARISON GIRLS, ENGLISH LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Letter Identification	Word Recognition	Reading Comp (easy)	Reading Fluency (medium)	Reading Comp (medium)	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)
Intervention Schools									
Non-Learner (0%)	28.7%	39.4%	59.5%	53.7%	75.7%	78%	76.9%	86%	91.7%
Emergent Learner (1-40%)	13.7%	14.8%	8.1%	11.7%	7.6%	6.4%	9%	0%	0%
Established Learner (41-80%)	15.9%	21.8%	16.8%	15.4%	14.3%	12.9%	11.1%	5.8%	2.3%
Proficient Learner (81-100%)	41.7%	24%	15.6%	19.2%	2.3%	2.6%	3%	8.3%	5.9%
Full Cohort (intervention and Comparison Girls)									
Non-Learner (0%)	31.2%	41.8%	62.9%	56%	79%	81.9%	80.7%	88.4%	93.5%
Emergent Learner (1-40%)	14.8%	16.5%	8.5%	13.1%	7.9%	5.1%	7.2%	0%	0%
Established Learner (41-80%)	16.4%	21.5%	16.1%	15.1%	11.6%	10.9%	9%	4.4%	1.9%
Proficient Learner (81-100%)	37.6%	20.2%	12.6%	15.7%	1.6%	2.1%	3.2%	7.3%	4.6%

TABLE 100: FOUNDATIONAL SKILL GAPS AMONG COMPARISON GIRLS, NUMERACY

Subtask	1	2	3	4	5	6	7	8	9	10	11
Skill Assessed	Missing number	Addition (Level 1)	Subtraction (Level 1)	Addition (Level 2)	Subtraction (Level 2)	Word problems (add/sub)	Multiplic. (Level 1)	Multiplic. (Level 2)	Division (Level 1)	Division (Level 2)	World problems (mult/div)
Comparison Girls											
Non-Learner (0%)	5.8	7.1	11.6	19.9	32.2	14.9	31	73	46.4	82.4	56.9
Emergent Learner (1-40%)	38.6	4.7	5.1	15.8	11.6	4	13	8.3	23.9	8.2	0
Established Learner (41-80%)	46.7	16.1	10.5	30.8	30.4	30.3	31.7	11.6	15.4	5.8	16.7
Proficient Learner (81-100%)	8.9	72.1	72.8	33.5	25.7	50.9	24.3	7.1	14.3	3.6	26.4
Comparison Girls Enrolled in School											
Non-Learner (0%)	0.9	1.4	4.9	10.1	23.5	6.7	15.7	63.5	32.5	76.5	46.4
Emergent Learner (1-40%)	33.6	3.8	4.3	15.9	10.4	4.9	12.8	9.9	29.3	9.9	0
Established Learner (41-80%)	54.8	15.9	8.1	32.8	32.5	31.9	37.1	16.8	18	8.4	20
Proficient Learner (81-100%)	10.7	78.8	82.6	41.2	33.6	56.5	34.5	9.9	20.3	5.2	33.6

TABLE 101: FOUNDATIONAL SKILL GAPS AMONG COMPARISON GIRLS, SOMALI LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Reading Words	Reading Comp (easy)	Reading Comp (medium)	Reading Fluency	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)	Sentence completion
Comparison Girls									
Non-Learner (0%)	13	21.7	24.3	17.9	32.2	37	42.8	51.4	60.7
Emergent Learner (1-40%)	4.9	6.9	4.2	18.8	18.1	11.6	2.7	2.5	17.6
Established Learner (41-80%)	13.4	37.9	28.1	31.2	38	21	6.3	4.5	10.1
Proficient Learner (81-100%)	68.7	33.5	43.5	32.1	11.6	30.4	48.2	41.5	11.6
Comparison Girls Enrolled in School									
Non-Learner (0%)	2.6	9.3	11.3	5.8	20	22.9	28.4	40	49.6
Emergent Learner (1-40%)	3.8	8.1	3.2	18.6	19.1	12.2	2.9	3.2	23.5
Established Learner (41-80%)	12.2	38.3	31	35.1	45.2	25.8	7.2	6.1	12.8
Proficient Learner (81-100%)	81.4	44.3	54.5	40.6	15.7	39.1	61.4	50.7	14.2

TABLE 102: FOUNDATIONAL SKILL GAPS AMONG COMPARISON GIRLS, ENGLISH LITERACY

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Letter Identification	Word Recognition	Reading Comp (easy)	Reading Fluency (medium)	Reading Comp (medium)	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)
Comparison Girls									
Non-Learner (0%)	34.1	44.6	66.8	58.7	82.8	86.4	85	91.1	95.5
Emergent Learner (1-40%)	16.1	18.5	8.9	14.9	8.2	3.6	5.1	0	0
Established Learner (41-80%)	17	21.2	15.2	14.7	8.3	8.5	6.5	2.7	1.4
Proficient Learner (81-100%)	32.8	15.8	9.1	11.8	0.7	1.4	3.4	6.2	3.1
Comparison Girls Enrolled in School									
Non-Learner (0%)	18.3	28.1	55.1	46.7	78	81.7	79.4	87.5	93.6
Emergent Learner (1-40%)	17.4	22.3	13	18	10.1	4.3	7	0	0
Established Learner (41-80%)	21.4	28.4	19.1	18.3	10.7	11.9	8.7	4.1	2
Proficient Learner (81-100%)	42.9	21.2	12.8	17.1	1.2	2	4.9	8.4	4.3

Annex 4. Intervention Roll-out Dates

TABLE 103: INTERVENTION ACTIVITIES AND IMPLEMENTATION TIMELINES

Activities	Start	End
Output 1: Improving access to post-primary options		
Meetings with MoEs, specialists and other stakeholders to develop ALP model	October, 2017	December, 2017
MoE subject specialist workshop to develop ALP modules	January, 2018	February, 2018
Validation, translation, production and distribution of ALP modules	July, 2018	September, 2018
Roll out of ALP classes (ALP implementation)	September, 2018	December 2021
Develop girls' life skills in upper primary through ALP, including leadership skills, financial literacy and business selection and management of income generation activities; participation in Girls' Empowerment For a	July, 2018	December 2021
Expand ALP to enrol 2,345 OOSG within an additional 34 villages, increasing ALP coverage from 76 villages to 110 villages	June,2019	December 2021
Provide 2 years of Alternative Basic Education (ABE) classes for 2,029 marginalized girls (older girls who have never attended school, particularly girls from displaced families) and link them with existing schools to join formal education depending on learning achievement	August 2019	December 2021
Training CECs across 199 villages in identification of	April 2019	December 2021

different type of disabilities and support to girls/boys with disabilities		
Work with CECs to liaise with parents of displaced OOS girls and girls with disabilities, provide targeted social support and track their attendance	April 2019	December 2021
Assessment of girls with disabilities for placement in regular schools or referrals to special needs facilities	February 2020	December 2021
Training of CECs to improve retention and transition (33 additional secondary schools)	February, 2018	July, 2018
CEC Coaching on improving retention and transition (199 schools)	March, 2018	December 2021
Provide partial grants to girls from poor families	November, 2017	March 2021
Equip and enrol girls into boarding schools	June 2019	December 2021
Output 2: Supportive school practices and conditions for marginalised girls		
Train teachers on improved delivery of literacy and English language, supported by digital content in all 148 primary and 55 secondary schools	February, 2019	Ongoing – end date May, 2019
Recruitment of consultant to develop manual and train teachers on improved delivery of numeracy	October, 2017	December, 2017
Refresher and advance numeracy TOT training	January, 2018	March, 2018
Train teachers on improved delivery of numeracy in all 148 primary and 55 secondary schools (cluster training)	July, 2018	September, 2018
Train teachers to provide structured remedial support to	July 2018	December 2021

students at primary and secondary level		
Train and coach teachers to deliver the ALP curriculum	July 2018	December 2021
Recruit additional ALP teachers to mitigate the effect of teacher's attrition and retraining of new ALP teachers	June,2019	December, 2020
Train and coach teachers to deliver the ABE curriculum	August 2019	December 2021
Train and coach teachers to deliver ABE in villages with large populations of displaced pastoralists	August 2019	December 2021
Training of teachers and MoE staff to identify and support girls with disabilities	October 2019	November 2019
Train teachers on inclusive and special needs education. The training will include basic special education; identification and basic assessment of girls with disabilities; building inclusive classroom environments; guidance and counselling. The training will include residential training and follow up on-site sessions.	October 2019	November 2019
Work with CECs and teachers to address corporal punishment, particularly against overage/ displaced adolescents and those who are struggling to learn, and promote community-managed self-monitoring of community efforts in addressing corporal punishment	August 2019	December 2021
Strengthening Student and Teachers Attendance tracking and monitoring	March 2019	December 2021

Increase reading time by establishing and supporting CEC managed reading clubs associated to GEFs/BEFs and promote the use of culturally appropriate local learning materials	October 2019	December 2021
Construct additional classrooms in remote primary schools; build water facilities in new secondary schools; and provide solar chargers for mobile devices/tablets and sanitary pads to schools	April, 2018	March, 2019
Incorporate life skills and financial literacy training into GEFs and BEFs	April 2019	December 2021
Provide career guidance in schools	November 2018	December 2021
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Engage community-level stakeholders including religious leaders, women's groups, men and boys	February, 2017	September, 2018
Expand and strengthen GEFs and create BEFs to develop leadership and mentorship skills	September, 2018	December 2021
Provide adult literacy and financial literacy classes for mothers	May, 2018	March, 2019
Support the financial empowerment of mothers through savings groups (VSLA), business selection, and business coaching and mentoring	February, 2018	December 2021
Incorporate sessions on identification and support for Girls with Disabilities in stakeholder forums and others NFE and VSL groups.	Jan 2020	December 2021

Provide support to VSLA groups to start business upon completion of the VSLA cycle through competitive selection of most viable business ideas	Oct 2019	March 2020
Encourage girls and boys from pastoralists families to participate in empowerment forums to enhance their confidence and address negative stereotypes associated with their itinerant lifestyle	April 2020	December 2021
Social mobilization campaigns to encourage pastoralists to bring their children to school and actively participate in their education.	Jan 2020	December 2021
Output 4: Enhanced MoEs' capacity to deliver quality and relevant formal and informal education		
Strengthen Gender Departments' capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points especially in rural areas	December, 2017	December 2021
Support quality assurance and standards (QAS) functions at all MoE levels	September, 2018	December 2021.
Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition	January, 2018	December 2021
Work closely with MoE on NFE for mothers and entrepreneurships skills for girls	April, 2018	December, 2018
Development of project IEC materials in conjunction with	July, 2018	March, 2021

MoE for use at stakeholder advocacy and promotion events		
Strengthen the MoEs' Special Education Unit through tailored trainings on advocacy and support for special and inclusive education	October 2019	March, 2021
Training of education officials at national, regional, district levels on basic special education concepts, support required by children with disabilities and including special needs education data in EMIS	October 2019	March, 2021
Introduce components of inclusive education/special education in pre-service teacher training curriculum for both primary and secondary school level.	October 2019	December, 2020

Annex 4: Characteristics and Barriers

Table 5.1: Girls' characteristics

	Intervention				Comparison				Source – variable name[285]
Sample breakdown (Girls)									
	Baseline	Midline1[286]	Midline 2	Endline	Baseline	Midline1 [287]	Midline2	Endline	
Single Orphans	11.1%	13.7%	11.7%	4%	9.8%	13.1%	12.8%	4.1%	PCG_11 g & PCG_13 g
Double orphans	0.3%	0%	0.3%	0.3%	0%	0%	0.5%	1.6%	PCG_11 g & PCG_13 g
Living without both parents (%)	6.1%	9.3%	10.1%	25.2%	5.9%	8.2%	13.3%	33%	E_4f_1 & E_4f_2
Living in female headed household	42.6%	45.5%	51.9%	46.1%	44.5%	44.9%	53.5%	42.4%	HH_8
Married (%)[288]	0.9%	1.2%	4.4%	16.7%	1%	2%	6.6%	16.1%	PCG_22 g
Motherhood									
Under 18	0.3%	0.3%	0.7%	2%	0%	1.3%	1%	4.2%	PCG_23 gi
Under 16	0.3%	0%	0.1%	0.1%	0%	1%	0%	6.3%	PCG_23 gi
Poor households (%)									

[285] Where not mentioned, the figures provided were copied from the final evaluation report.
[286] Midline 1 data is NOT comparable to any other evaluation round, as only in-school girls were assessed.
[287] Midline 1 data is NOT comparable to any other evaluation round, as only in-school girls were assessed.
[288] Married/ ever been married

Difficult to afford for girl to go to school	21.1%	12.9%	7.4%	2.6%[289]	18.1%	7.6%	19.5%	2.35%	PCG_7enr
Household doesn't own land for themselves	22.9%		31%	32%	19.6%		26.4%	32.2%	PCG_11econ
Poor quality roof material	32.4%	36.7%	25.5%	21.6%	37.7%	40%	33.7%	23.4%	PCG_2econ
Household unable to meet basic needs	35.1%		29%	18%	35.0%		30.8%	19.2%	PCG_5econb
Gone to sleep hungry for many days/most days in past year	7.3%	9.1%	12.8%	7.2%	8.6%	12.9%	10.9%	8%	PCG_7econ
Language difficulties:									
LoI different from mother tongue (%)	12.3%		9.8%	31.6%[290]	10.4%		8.6%	27.4%	PCG_2enr[291]
Girl doesn't speak LoI (%)	1.6%		3%	3.3%	4.2%		2.7%	7.2%	PCG_3enr
Parental education									
HoH has no education (%)	67%	49.1%	47.8%	66%	64%	51.2%	44.8%	34.6%	HH_13
Primary caregiver has no education (%)	67.6%	51.3%	53.7%	73.2%	75.1%	59.7%	53.4%	39.7%	PCG_6

[289] Including all girls (in school and out of school). 4% of the caregivers of enrolled girls affirmed that it is difficult to afford girls' education.
[290] Based on CARE's secondary analysis of the data.
[291] Validated when contrasted to PCG_1enr

Table 5.2: Potential barriers to learning and transition

	Intervention				Comparison				Source-Variable name[292]
Sample breakdown (Girls)									
	Baseline	Midline1[293]	Midline 2	Endline	Baseline	Midline1[294]	Midline2	Endline	
Home/ Community									
Safety									
Fairly or very unsafe travel to schools in the area (%)	4.8%		2.5%	3%	5.0%		1.2%	3%	
Doesn't feel safe travelling to/from school (%)	3.2%	3.7%	1.4%	2.4%	9%	4.1%	0.7%	0.8%	
Girl travels more than 30 minutes to school	7.7%	5.2%	8.2%	1.8%	4.1%	6.5%	7.2%	2.7%	PCG_7
Parental/caregiver support:									
Sufficient time to study: High chore burden	11.1%		14.9%	23.8%	11.3%		18.2%	27.1%	
Doesn't get support to stay in school and do well (%)	6.1%	1.1%	3.7%	3.7%	1.1%	2.9%	5.2%	5.9%	
Girl has no choice in whether to attend school	86.3%	86.1%	85.4%	81.4%	85.6%	93.5%	84.2%	77.1%	HHG_6
Family decides for girl whether she will attend school	21.8%		28.5%	24.3%	20.4%		31.1%	31.3%	Is22
Family not a member of CEC	58.2%		52%[295]	55.7%[296]	46.9%		72.4%[297]	83%[298]	SM_5h
Parent has never visited school	10.5%		15.6%[299]	18.9%[300]	13.5%		16.9%[301]	20.7%[302]	PCG_TQC1
School Level									

[292] Where not mentioned, the figures provided were copied from the final evaluation report.
[293] Midline 1 data is not comparable to any other evaluation round as only in-school girls were assessed.
[294] Midline 1 data is not comparable to any other evaluation round as only in-school girls were assessed.
[295] Calculated based on all caregivers of enrolled girls, regardless if the school has a CEC or not.
[296] Calculated by the project using the datasets provided by the external evaluator. Value differs from the figure provided by the EE.
[297] Calculated based on all caregivers of enrolled girls, regardless if the school has a CEC or not.
[298] Calculated by the project using the datasets provided by the external evaluator. Value differs from the figure provided by the EE.
[299] Enrolled girls only
[300] Enrolled girls only.
[301] Enrolled girls only
[302] Enrolled girls only

Attendance									
Attends school half the time (%)	1.2%	1.8%	2.8%	1.4%	2%	1.4%	1.6%	2.9%	PCG_6enr
Attends school less than half time (%)	0.9%	0.3%	0.7%	1.9%	0.7%	2.9%	1.4%	0.06%	PCG_6enr
Doesn't feel safe at school (%)	4.9%	2.2%	0.7%	1.5%	6.4%	1.4%	1.1%	1.2%	
School Facilities									
No seats for all students (%)	20.9%	13.3%	2%	5.9%	30.5%	15.8%	2.9%	5.5%	
Difficult to move around school (%)	16.4%		14%	17.5%	21.9%		14.4%	18.1%	
Doesn't use drinking water facilities	19.7%	10.5%	11.2%	10.3%	32%	23.7%	15.5%	6.7%	
Doesn't use toilet at school	22.6%	17.9%	26.3%	13%	30.1%	28.8%	24.7%	15.7%	
No computer in class	89.6%	95.7%	87.6%	84%	95.1%	93.2%	92.6%	83.5%	CSG_2s
Cannot use books or other learning materials at school(%)	19.9%	20.1%		10.6%	25.1%	28.3%		13.9%	
Doesn't use areas where children play/ socialise	21.2%		27.6%	53.6%	21.7%		33.6%	49.8%	
Teachers									
Disagrees teachers make them feel welcome	5.1%	4%	3%	2.9%	7.5%	5.8%	4.3%	2.4%	
Agrees teachers treat boys and girls differently in	41.1%	43%	32.6%	28.7%	41.1%	46.5%	33.2%	25%	
Agrees teachers often absent from class	35%	17.2%	23.9%	21.2%	40.1%	29.2%	23.2%	20.4%	
Afraid of teacher	57.7%	75.9%	66.7%	67.6%	58.3%	82.7%	70.1%	72.4%	HHG_7f_n
Uncomfortable asking teachers question	5.5%	1.5%	1.9%	2.9%	4.9%	2.5%	2.4%	1.2%	HHG_7c_n
Teacher punishes/disciplines when students gets lesson wrong	76.8%	80.2%	66.6%	54%	79%	84.9%	66.3%	52%	TQ_6s
Physical punishment witnessed last week	63%	54.5%	48.3%	53.9%	49.6%	51.2%	54.4%	52.2%	TQ_8s
Caregiver rates principal or head teacher lowly	3.3%	1.6%	1.8%	2.3%	4.8%	2.9%	1.4%	2.6%	SM_3h
Caregiver rates quality of teaching as poor	2.8%		1.4%	0.6%	4.6%		1.4%	0.8%	TQ_3h
Teacher asks questions unequally of girls and	7.3%		7.4%	8%	10.6%		6.2%	6.3%	TQ_1s
Teacher asks harder questions unequally of girls and boys	9.7%		9.9%	10.9%	10.8%		8.7%	11.3%	TQ_2s

Annex 5. Logframe

The project's logframe is attached as a separate file.

Annex 6. Outcomes Spreadsheet

The filled Outcomes Spreadsheet is attached as a separate file.

Annex 7: Project Design and Intervention

The description of SOMGEP-T's design and intervention is provided below, as reported by the project.

TABLE 104: PROJECT DESIGN AND INTERVENTION

Intervention types	What is the intervention?	What output will the intervention contribute to?	What Intermediate Outcome will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition and sustainability outcomes?
Access	Developing and implementing Alternative Learning Program for Out of School Girls	Output 1	IO- 1	By offering an alternative pathway for girls who may have otherwise dropped out, transition rates will improve. Girls will have increased exposure to higher learning, which will boost learning outcomes. ALP's particular focus on developing life skills will ensure this intervention produces sustainable outcomes, or outcomes that are relevant to the individual and community.
	Provision of partial grants to girls from poor families	Output 1	IO- 1	Increased attendance and retention is expected to improve transition rates and learning outcomes, as girls who are in school and are properly equipped are more likely to succeed. Girls from poor families who may not have otherwise had access to education will be better equipped to participate in decision-making and economic activities.
	Equip and enrol girls in 2 boarding schools	Output 1	IO- 1	
School Governance/management	Capacity building of CEC's to improve retention and transition	Output 1	IO- 2	A focus on retention and transition is expected to have a direct impact on transition rates and learning outcomes, as girls will have better access to higher education levels. The focus on the community level will ensure buy-in and contribute to the project's sustainability at the community level.
Teachers capacity building	Train teachers on improved delivery of literacy and	Output 2	IO- 3	Improved teaching quality contributes to enhance learning and transition outcomes, as children are equipped with the literacy skills in Somali, English, numeracy and life skills

	English language supported by digital content			necessary to progress to higher levels of education. Interventions focused on improving teaching quality are expected to boost transition rates and learning outcomes in a sustainable way, by equipping children with the skills they need to succeed not only in school, but outside school as well
	Train teachers on improved delivery of numeracy	Output 2	IO- 3	
	Train teachers to provide structured remedial support to students at primary and secondary levels	Output 2	IO- 3	
	Train and coach teachers to deliver ALP curriculum	Output 2	IO- 3	
	Train and coach teachers on career guidance	Output 2	IO- 3 & IO- 4	Encouraging girls to think about their futures and how to achieve their aspirations will impress on them the importance of knowledge and education. It will also give them a clear pathway to achieving their goals.
Community-based attitudes and behaviour change	Engage community – level stakeholders including religious leaders, women's groups, men and boys	Output 3	IO- 1	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Shifts in gender and social norms are expected to have a long-term, sustainable impact on the communities in which SOMGEP-T will operate.
	Provide adult literacy and financial classes for mothers	Output 3	IO- 1	

Economic empowerment	Support the financial empowerment of mothers through savings groups(VSLA) , business selection and business coaching and mentoring	Output 3	IO- 1	
	Work closely with MoE on NFE for mothers and entrepreneurship skills for girls	Output 4	IO- 1	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
Life Skills	Develop girls life skills in upper primary through ALP	Output 1	IO- 4	The project's learning outcomes are focused on literacy, numeracy, and financial literacy. This intervention is designed to boost these specific learning outcomes, as well as increasing the likelihood of transition into ALP or secondary education. Additionally, the focus on leadership skills and other skills relevant to the job market contributes to the sustainability of SOMGEP-T.
	Incorporate life skills and financial literacy training into GEFs and BEFs	Output 1	IO- 4	Financial literacy training is one of the specific learning outcomes SOMGEP-T is expecting to influence. Financial literacy and life skills training will increase the likelihood of girls succeeding in higher levels of education, and will also equip them to contribute to the local economy through income-generating activities. These skills are expected to increase the relevance of education for students and families. Life skills – specifically leadership skills – are expected to boost students' voice and self-confidence, enhancing classroom participation among girls.

Girls self Esteem	Expand and strengthen GEF's and create BEFs to develop leadership and mentorship skills	Output 3	IO- 4	Girls who receive leadership and mentorship skills through life skills development will be better equipped to participate in class, breaking traditional norms that restrict girls' voice; to engage in the local economy; and to contribute to their communities in the future. Additionally, the capacity of GEFs and BEFs to track attendance and retention rates will contribute to improvements in learning and transition outcomes, and will encourage community-based organizations to think about how their actions have a direct effect on important student outcomes.
MoE Capacity building	Strengthen Gender Departments' capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points especially in rural areas	Output 4	All IO	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
	Provide support to Regional Education Officers (REOs) and District Education Officers	Output 4	All IO	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.

	(DEOs) to mainstream improved teaching practices and address retention/ transition			
Construction	Construct additional classrooms in remote primary schools; building water facilities in new secondary schools and [provide solar chargers for mobile devices/tablets and sanitary pads to schools	Output 2	IO- 2	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Infrastructure development will benefit not just the current cohort of students with which SOMGEP-T is engaged, but will also benefit future students

Annex 7: Beneficiaries tables

Table 7.1: Direct beneficiaries

Beneficiary type	Total project number	Total number of girls targeted for learning outcomes that the project has reached by Endline	Comments
Direct learning beneficiaries (girls) – 26,290 girls – learning beneficiaries 1,814 girls with disabilities	32,862	18,778	Overall reach is calculated based on: • 16115 girls enrolled in primary school • 3,577 girls enrolled in secondary schools • 2,138 girls enrolled in ALP • 1,643 girls enrolled in ABE

Table 7.2: Other beneficiaries (Total over lifetime of the project)

Beneficiary type	Number	Comments
Learning beneficiaries (boys) – as above, but specifically counting boys who will get the same exposure and therefore be expected to also achieve learning gains, if applicable.	19,428	Considering 80% of the boys in school as learning beneficiaries.
Broader student beneficiaries (boys) – boys who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal	24,286	Considering all boys in primary school (18547), secondary school (4577), 1104 ABE and 58 ALP.

change, etc. but not necessarily achieve improvements in learning outcomes.		
Broader student beneficiaries (girls) – girls who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.	8762	Defined as in-school girls not benefitting from improved learning outcomes (estimated as 30% of 23473/ total) and out of school girls not benefiting from improved learning (1720). This calculation is based on the number of OOS girls who did not graduate.
Teacher beneficiaries – number of teachers who benefit from training or related interventions. If possible /applicable, please disaggregate by gender and type of training, with the comments box used to describe the type of training provided.	595 teachers coached on literacy, numeracy, English and structured remedial classes 346 teachers trained on PFA	
Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.	6,595 mothers receiving NFE training 2,356 community members participating in VSLA	

Table 7.3: Target groups - by school

School Age	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Lower primary	Yes - Grade 1-4	19,989	272
Upper primary	Yes - Grade 5-8	5,820	233 (+93 in benchmark)
Lower secondary	Yes - Form 1-2	1,912	12 (benchmark only)
Upper secondary			

Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 7.4: Target groups - by age

Age Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Aged 6-8 (% aged 6-8)		9,120	This group will benefit from teacher training, improved school management and conditions at the household. However, the sample tracks only girls age 10-19.
Aged 9-11 (% aged 9-11)	√	6,885	241
Aged 12-13 (% aged 12-13)	√	6,885	268
Aged 14-15 (% aged 14-15)	√	5,461	192
Aged 16-17 (%aged 16-17)	√	3,086	109
Aged 18-19 (%aged 18-19)	√	1,425	62
Aged 20+ (% aged 20 and over)			
Total:		**32,862**	[This number should be the same across Tables 3, 4, 5 & 6]

Table 7.5: Target groups - by sub group

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Disabled girls (please disaggregate by disability type)	√	1,814	60
Vision impairment	√	197	5
Hearing impairment	√	230	6
Mobility impairment	√	230	6
Cognitive impairment	√	263	7
Self-care impairment	√	230	6
Communication impairment	√	329	9
Mental health impairment	√	1709	45
Orphaned girls	√	3,615	96
Pastoralist girls	√	3,943	105
Child labourers	√	87[303]	13
Poor girls	√	32,862	872

[303] This number does not include girls who support the family business or do unpaid work at home.

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Other (please describe)			
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 7.6 : Target groups - by school status

Educational sub-groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Out-of-school girls: have never attended school	√	1285	92
Out-of-school girls: have attended school, but dropped out	√	3855	275
Girls in-school	√	27722	505
Total:		32,862	[This number should be the same across Tables 3, 4, 5 & 6]

Table 7.7 : Beneficiaries matrix

Outcomes	Direct beneficiaries			Indirect beneficiaries				
	In-school girls (6-10 grade)	OSG (6-9 years)	OSG (18-25)	In-school boys	HT/Teachers	Parents	SMC/ PTA	Local government
Learning	✔			✔	✔	✔		
Transition	✔	✔	✔	✔	✔	✔		
Sustainability	✔	✔	✔		✔	✔	✔	
IO 1: Attendance					✔	✔		
IO 2: Self-esteem and empowerment	✔	✔	✔					
IO3: Parental engagement	✔	✔	✔			✔		
IO4: Quality of teaching	✔				✔	✔	✔	✔
IO5: School management and governance	✔				✔	✔	✔	✔

Annex 8: Key Findings on Output Indicators

A description of key findings on output indicators has been provided by the project, below.

TABLE 105: OUTPUT INDICATORS

Logframe Output Indicator	Means of verification/sources	Collection frequency
Number and Indicator wording	List all sources used.	E.g. monthly, quarterly, annually. NB: For indicators without data collection to date, please indicate when data collection will take place.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	ALP monitoring tool	Monthly
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	ALP monitoring tool	Monthly
Output 1.3: Percentage of girls receiving partial grants who remain in school	Partial Grants Fidelity of Implementation	Termly
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	Midline Evaluation, GEF/BEF Fidelity of Implementation	Yearly
Output 2.2: Percentage of teachers using the digital learning platform	Classroom Observations	Monthly
Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	GEF Fidelity of Implementation	Bi annual
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Output 3:1 Number of women mentors providing support to marginalised girls	GEF Fidelity of Implementation , Monitoring Visits	Monthly

Output 3:2 Number and percentage of mothers completing literacy courses	NFE Completion records	Annual
Output 3:3 Percentage of active village savings groups in project areas	VSLA Fidelity of Implementation [FOI]	Monthly
Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		
Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	Gender Units Reports	Monthly
Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	Joint Monitoring Reports	Quarterly

TABLE 106: MIDLINE STATUS OF OUTPUT INDICATORS

Logframe Output Indicator	Midline status/midline values Relevance of the indicator for the project ToC	Midline status/midline values
Number and Indicator wording	What is the contribution of this indicator for the project ToC, IOs, and Outcomes? What does the midline value/status mean for your activities? Is the indicator measuring the right things? Should a revision be considered? Provide short narrative.	What is the midline value/status of this indicator? Provide short narrative.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	The ALP sites established by the project offer out of school girls post primary opportunity. Originally the project planned to establish 76 ALP programs by end of Year 2 and an additional 34 by end of Y3. Girls enrolled in ALP are expected to have increased learning outcomes as well as developing essential life skills which will enable them to be productive members of the society. Constant measurement of the coverage of ALP is vital. The indicator is still relevant; no modifications are required.	Midline Wave 1 Status = 96.05% functional ALP sites. At the time of the assessment 73 ALP sites were functional out of the targeted 78 ALP sites. Midline Wave 2 Status = 97.3% functional ALP sites. At the time of the assessment 107 ALP sites were functional out of the targeted 110 ALP sites. Three ALP centres are not functioning due to prolonged insecurity at the borderline. The project will identify new ALP villages to replace those in conflict areas. Endline Status = 100% at the end of December 2021, 114 ALP were functional, these include ALP sites in comparison schools
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	Girls are learning relevant life skills that will not only boost their learning outcomes and attendance, but will also enable them to contribute to the local economy once they leave school. This intervention boost learning outcomes, as well as increasing the likelihood of girls transiting into formal schools. Life skills remain a key	Midline Wave Status 1 =96.05% Midline Wave 2 Status = 97.3% A total of 3410 girls were provided with training on life skills as we as training on basic financial literacy meant to inculcate the culture of savings among the girls. ` Endline Status = 84.4%.

	component of the program it is vital to constantly monitor the delivery of life skills training. The indicator is still valid; modification is not required.	A total of 3693 girls joined ABE and ALP. Project target of 4374 was not met as a result of multitude of problems; the main challenge include; migration of pastoralists, various economic challenges. A separate FOI survey was done on the children who dropped out of school. When asked about the main reasons for dropping out, the most common responses for both boys and girls was migration (32% of girls and 43% of boys). The other most popular reasons for girls were affordability (14%) obligations at home i.e. house work (12%) and getting married (11%). For boys, other popular reasons included the family deciding that school was not useful / relevant (10%), having obligations at home (9%) and being displaced (8%).
Output 1.3: Percentage of girls receiving partial grants who remain in school	The provision of partial grants to girls enabled girls who are at risk out of school to continue with their education.	The intervention was completed in Year 2
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	Addressing corporal punishment will improve conditions for learning this enable girls to attend schools regularly and improve their learning outcomes. The prevalence of corporal punishment should continue to be monitored. The indicator is still valid.	Midline Status Wave 1 =76.8% Midline Status Wave 2 = 45.3% Endline Status = 54% FOI results show that over half the classrooms observed, boys and girls are physically disciplined, with these proportions higher for girls (58%) compared to boys (50%). Despite this, in 88% of the classrooms, teachers interacted with students in a respectful and positive manner overall
Output 2.2: Percentage of teachers using the digital learning platform	The digital learning platform is expected to improve the quality of teaching, this will increase student performance and motivation is likely to have a positive effect on attendance and learning.	Midline Wave 1 Status = 0 Midline Wave 2 Status = 0 Endline Status = 100% The project provided 133 tablets in Year 4 to formal schools to deliver online content during the COVID-19 pandemic. Based on the analysis of tablet usage, the teachers who received tablets early in the year used the tablet for roughly 1.8 months of the 2020/2021 term 2 academic year, which is significantly high because schools were closed due to COVID-19. Relevance and attitudes towards learning were also measured. Approximately 92% of the 38 teachers surveyed indicated that their way of teaching has changed since they started using tablets. Nearly all the teachers indicated that the lesson content

		uploaded in the tablet was easier to use because it follows the same structure with the conventional standard lesson. Teachers usually teach for five hours a day, whereas the MTR FOI on the uptake of tablet content show that teachers have 1.9 hours' access to technology content per day
Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	Life skills – The girls or boys led activities boost their voice and self-confidence, enhancing classroom participation and improved learning outcomes. The indicator is still relevant	Midline Wave 1 GEF Status = 43.75% Midline Wave 2 GEF Status = 68.57% Endline GEF Status = 80.7% This indicator assesses the activities led by girls or boys in their school or communities; these activities are designed to build girl or boys' confidence and participation in the classroom. The data for this indicator was collected through the GEF/BEF fidelity of implementation checklist. The fidelity checklist asked a series of questions to understand the various activities implemented by GEF/BEF; the seven activities assessed include facilitation, debating sessions, competitions, fundraising, sanitation campaigns, community sensitization on girls' education, following up on girls who dropped out of school and participation in other community-related activities. GEF/BEF's who implemented at least 4 out of 7 of the activities, were considered to have met the fidelity of implementation minimum standards. At the end of Year 3, at least 80.7% of GEF met the fidelity of implementation minimum standards as compared to 68.6% at midline 1/FOI wave 1. An analysis of the activities led by GEF show marked improvements in the proportion of GEF's leading various activities at school and community level.
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Output 3:1 Number of women mentors providing support to marginalised girls	Girls who receive mentorship skills from women mentors, will be better equipped to participate in class, breaking traditional norms that restrict girls' voice; to engage in the local economy; and to contribute to their communities in the future. This is expected to contribute to improvements in learning and transition outcomes. Indicator still valid.	Midline Wave 1 Status = 66% of GEFs have active mentors Midline Wave 2 Status = 65.7% of GEFs have active mentors[121 mentors from a total of 184 GEFs provided support to marginalised girls] Endline Status = 81.8% of GEFs have active mentors The results of the fidelity of implementation show that 81.8% of the women mentors provided support to the GEFs to undertake various activities which include; facilitating; debating sessions, competitions, fund raising sanitation campaigns, community sensitization on girls' education, following up on girls who dropped out of school and participation in other community-related activities.

Output 3:2 Number and percentage of mothers completing literacy courses	Mothers in NFE classes acquire essential literacy skills that enable them to support their girls with homework. This will ultimately improve the girl's learning outcome. Skills learnt from the NFE classes will enable them to venture into business, improving their financial capacity to meet the basic education necessities. Girls with adequate basic education necessities are likely to attend school regularly, learn and improve their learning outcomes. Indicator is still valid.	Completed in 2020
Output 3:3 Percentage of active village savings groups in project areas	Increased financial capacity of vulnerable households, are expected to contribute to improvements in attendance transition and learning outcomes. Indicator is still valid.	Midline Wave 1 Status = 100% Midline Wave 2 Status = 98 % [134 out 137 VSLA are functional] active village savings established] Endline Status = 88.56% VSLA are active [124 out of 141 VSLA are functional A total of 124 VSLA groups are functional, 17 are no longer functional. The non-functional groups collapsed because economic challenges caused by COVID-induced lockdowns
Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		
Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	Enhancing the capacity of MoEs to develop plans, administer trainings, and provide incentives will contribute to all four intermediate outcomes by sending a strong, positive message about the importance of girls' education from the government, and by giving the government clear and actionable ways to contribute to positive changes in girls' education outcomes. Indicator is still valid.	Midline Wave 1 Status = 3 Midline Wave 2 Status = 3 Endline Status = 3 The project continued to provide Incentives to 13 all-female Gender Focal Person's (GFPs) in the project zones (6 each in SL and PL, 1 in GM). The GFPs worked closely with the project officers to conduct activities earmarked to promote girls' transition. They have been engaged in coaching of Gender Empowerment Forum (GEF), promoting of inclusive education approaches, coordinate the provision of sanitary wear.

Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	Regular joint field monitoring visits will improve the quality of project delivery more importantly the quality of teaching. This is expected to lead to improvements in attendance, transition and learning outcomes. Project will take timely adaptations to ensure identified gaps in programming there by improving the quality interventions ultimately contributing to project outputs, intermediate and outcomes. Indicator is still valid.	Midline Wave 1 Status = 60.1 % [89/148 primary schools] Midline Wave 2 Status = 47.7% [69/148 primary schools] – The quality assurance visits were adversely by the COVID Pandemic Endline Status =74.5%[Quality assurances visited 110 /148 primary schools]

TABLE 107: OUTPUT INDICATOR ISSUES

Logframe Output Indicator	Issues with the means of verification/sources and the collection frequency, or the indicator in general?	Changes/additions
Number and Indicator wording	E.g. inappropriate wording, irrelevant sources, or wrong assumptions etc. Was data collection too frequent or too far between? Or no issues?	E.g. change wording, add or remove sources, increase/decrease frequency of data collection; or leave as is.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	None	None
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	None	None
Output 1.3: Percentage of girls receiving partial grants who remain in school	None	None
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	None	None
Output 2.2: Percentage of teachers using the digital learning platform	None	None
Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	None	None
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Output 3:1 Number of women mentors providing support to marginalised girls	None	None

Output 3:2 Number and percentage of mothers completing literacy courses	None	None
Output 3:3 Percentage of active village savings groups in project areas	None	None
Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		
Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	None	None
Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	None	None

Annex 9: MEL Framework

The project's Monitoring, Evaluation, and Learning (MEL) Framework is attached as a separate file.

Annex 10. Inception Report

The inception report for the endline evaluation, submitted prior to the start of data collection, is attached as a separate file.

Annex 11. Data Collection Tools

The quantitative and qualitative data collection tools are attached separately, given their length. Five quantitative tools are provided:

- Household survey, which encompasses the survey module completed by girls, the learning assessments, and the module and assessments completed by boys
- Classroom observation tool
- Headcount tool
- Head teacher (school) survey
- Brief phone-based survey script used for collecting transition data from girls who could not be interviewed in-person.

The qualitative tools include:

- FGD Mothers
- FGD Teachers
- FGD CEC Members
- KII MOE Officials
- Risk Mapping FGD with girls
- Vignette Exercises with girls
- Girls Networking Exercise