Project Evaluation Report

Report title:	Educate Girls, End Poverty – Transition (EGEP-T) in Somaliland, Puntland, Galmadug, Hirshabelle and Banadir Midline Report
Evaluator:	Forcier
GEC Project:	Educate Girls, End Poverty (EGEP)
Country	Somalia
GEC window	GEC-Transition
Evaluation point:	Midline
Report date:	January 2020

Notes:

Some annexes listed in the contents page of this document have not been included because of challenges with capturing them as an A4 PDF document or because they are documents intended for programme purposes only. If you would like access to any of these annexes, please enquire about their availability by emailing uk_girls_education_challenge@pwc.com.

Girls' Education Challenge

UKaid from the British people

Educate Girls, End Poverty – Transition (EGEP-T) in Somaliland, Puntland, Galmudug, Hirshabelle and Banadir

Midline Evaluation

Final

Brenton Peterson, Jonathan Forney, and Sam Ha

JANUARY 27, 2020

Abbreviations

ADRA	Adventist Development and Relief Agency
CEC	Community Education Committee
CISP	Comitato Internazionale per lo Sviluppo dei Popoli
DFID	Department for International Development
EGEP	Educate Girls, End Poverty Project
EGEP-T	Educate Girls, End Poverty – Transition Project
FGD	Focus Group Discussion
FM	Fund Manager
FRS	Federal Republic of Somalia
GEC	Girls Education Challenge
ISG	In-School Girls
KII	Key Informant Interview
MELF	Monitoring, Evaluation and Learning Framework
MoE	Ministry of Education
MOECHE	Ministry of Education, Culture, and Higher Education
OOS	Out-of-School
PWC	PriceWaterhouseCoopers
RI	Relief International

TABLE OF CONTENTS

List of Figures

List of Tables

Executive Summary

This report constitutes the midline evaluation of the Educate Girls, End Poverty – Transition (EGEP-T) project undertaken in Banadir, Galmudug, Hirshabelle, Puntland and Somaliland by Relief International (RI) and its implementing partners. The project is funded by UK Aid, as part of its Girls Education Challenge (GEC) Fund. This report analyses the chosen approach of EGEP-T to improving girls' educational outcomes in Somalia, draws comparisons between baseline benchmarks and midline levels of key indicators in order to draw conclusions about possible intervention effects, and makes recommendations for future evaluations and project implementation.

As at baseline, education in Somalia continues to be carried out in a context of rampant insecurity. In 2018 alone, 64 schools were attacked and 21 forced to close.[1] The political climate in the country broadly has improved somewhat in recent years, but is still marked by tentative and geographically-limited gains in stability. While al-Shabaab was driven out of Mogadishu in 2011, the conflict has worsened in recent months, reaching a point at which midline data collection in the capital city was cancelled due to security concerns. Meanwhile, Somaliland has experienced a different conflict trajectory. Fighting in Tukaraq occurred sporadically throughout 2018, resulting in occasional school closures. In general, however, Somaliland became more settled during much of 2018, and conflict has become less widespread since the baseline. In contrast to Somaliland, Puntland and Galmudug are both heavily impacted by the presence of al-Shabaab, in addition to Puntland's border conflict with Somaliland. In 2015 and 2016, conflicts in Galmudug prompted the closure of schools, though conflict has been reduced somewhat since that time. Certainly, the sample of schools evaluated at the midline is less affected by conflict than they were at baseline, partially as a result of the easing tensions in Somaliland and partially as a result of the exclusion of schools in Mogadishu. In short, the importance of conflict to the evaluation has been reduced due to both omission from the sample and thanks to a reduction in conflict.

It is in this difficult and varied context that Relief International (RI), the Adventist Development and Relief Agency (ADRA) and Comitato Internazionale per lo Sviluppo dei Popoli (CISP) are implementing the Educate Girls, End Poverty – Transition (EGEP-T) project. The complexity of each local environment makes both implementation and evaluation more difficult. Implementation is more difficult because access to schools and coordination with officials must be negotiated with different actors in different locations, and the poorly-institutionalized nature of government ministries overseeing education means that a standard approach applied to all schools is unlikely to be successful. Evaluation of the project's impact is also more difficult, and the results more tenuous, because there are often idiosyncratic local conditions that can drive major shifts in project outcomes, especially enrolment or transition rates, and changes in response to temporal shocks like drought or localized conflict are common.

This midline evaluation takes a mixed-methods approach, employing both qualitative and quantitative methods. The evaluation does not include a control group of schools or students; as such, conclusions regarding project impact are drawn on the basis of comparisons of student performance over time vis-à-vis benchmarks established at baseline where appropriate. Respondents in the primary evaluation sample were in Grade 6 through Form 2 at the time of the baseline, and findings are presented with reference to baseline grades, with midline grade levels (that girls transitioned into)

[1] https://reliefweb.int/sites/reliefweb.int/files/resources/20190118_humanitarian_response_plan.pdf

presented in parentheses. The remainder of this executive summary will provide a brief overview of key results for each outcome area, in order of the outline of the report.

Theory of Change

The project's Theory of Change identifies six barriers to girls' educational achievement, which project activities have been designed to address:

> Barrier 1: Household-Level Economics: Households often view their finances as an obstacle to educating their children, and particularly their girls. Schooling imposes a financial burden on households, and to the extent that education is seen as investment, the financial cost of enrolment is often too high for households to justify educating girls who will leave the labor force and become part of another family after marriage. To ease the financial burden of schooling, EGEP-T provides payment for school fees, school uniforms, examination fees, solar lamps, and sanitary kits.
>
> Barrier 2: Lack of Confidence, Life Skills and Psychosocial Support: Girls often lack confidence in their abilities, which reduces their willingness to participate in class and their desire for schooling. Previous evaluations for EGEP and EGEP-T have found that girls with greater psychosocial wellbeing had higher attendance rates and that girls feel lower levels of agency than do boys. In order to target this barrier, EGEP-T plans for girls' clubs and boys' clubs to promote leadership, confidence and provide children with valuable training on life skills such as their reproductive rights and social skills.
>
> Barrier 3: Poor Learning Environment: Student who attend class regularly may nevertheless face poor-quality teaching, an unwelcoming environment, and a lock of teaching and learning resources. To improve teaching quality, EGEP-T trains teachers in subject content in math, subject-specific pedagogical techniques, remedial teaching practices, English proficiency, and gender-responsive techniques.
>
> Barrier 4: Weak Government Outreach and Engagement: Schools require monitoring to ensure that they are keeping accurate records, that teachers are consistently showing up for work, that teaching quality is good, and that child protection systems are developed and actually in use. However, there are system-level deficiencies in monitoring, promotion of education, and other key activities of government industries. In order to engage government officials, EGEP-T will develop monitoring tools that will allow ministry official to more accurately and efficiently monitor school, teacher, and student performance.
>
> Barrier 5: Weak School Governance: In the absence of effective and continued ministry oversight, schools in Somalia have become reliant on community involvement for management and support such as community education committees. CECs can promote accountability, can influence community opinion regarding education, can leverage and aggregate community resources to improve infrastructure or pay teachers, and can promote efficiency at the school level. EGEP-T is providing cash grants to schools as part of their drought response and CECs are participating in grant management and CECs are receiving more general capacity-building efforts.
>
> Barrier 6: Lack of Complete Community Support for Girls' Completion of the Full Education Cycle: While the first phase of EGEP found that there was a high level of community support for girls' schooling, the support did not necessarily extend to completion of secondary school. In order to change community attitudes and behaviors to promote girls' completion of education, EGEP-T is engaging in a broad campaign targeting the community attitudes by

promoting community dialogues, workshops with men and boys, community mobilisation events, and outreach efforts through radio, loudspeakers, banners and t-shirts.

Learning

To measure learning outcomes, a panel of 996 girls were tracked from baseline to midline. Learning assessments for numeracy and Somali literacy were adjusted since the time of the baseline in order to compensate for ceiling effects at baseline. As a result, not all subtasks included in those assessments were comparable between the two rounds of data collection. Progress since baseline was calculated using comparable scores for the pooled sample of cohort and bursary girls (including only subtasks that are comparable between baseline and midline), using the panel-only sample, and using a truncated sample (removing all girls who scored above 85 percent on a given assessment at baseline) in order to minimize ceiling effects. Baseline and midline scores for the pooled, truncated, panel-only sample are included here to provide a basis for understanding progress since baseline. The comparable baseline numeracy score for cohort and bursary girls in the truncated sample was 48.4 percent, as compared with 56.3 percent at midline. For Somali literacy, the baseline score was 59.1 percent, as compared with 63.7 percent at midline. For English literacy, the baseline score was 31.5 percent at baseline, as compared with 45.9 percent at midline. For numeracy and English literacy, the midline target was reached. In the case of numeracy, the target was 4.99 and the achieved difference over and above benchmark was 5.25, exceeding the midline target by 5 percent. For English literacy, the midline target was 5.40 and the achieved difference over and above benchmark was 7.59, exceeding the target by 41 percent. These increases in score from baseline to midline are also statistically significant. In contrast, the midline target was not achieved for Somali literacy: the target was 4.49 and the difference over and above benchmark was 1.43. Over all, when changes in score from baseline to midline are compared with benchmark expectations (using an arithmetic difference in differences approach), the findings suggest that there have been substantial intervention effects on girls' numeracy and English literacy learning since baseline. When testing for differential intervention effects between cohort girls and bursary girls, we find that bursary girls are keeping pace with, cohort girls and may be closing the gap with cohort girls in numeracy and Somali literacy. When comparing bursary girls and boys, we find that cohort girls are keeping pace with boys, but not closing the substantial gap that has existed since baseline.

In terms of marginalization status, girls' scores are significantly lower than boys' scores across all assessments. This gendered learning gap has been present since baseline, and there is no evidence yet that project interventions are reducing this gap. Among girls, the strongest indicators of marginal status signalled by lower than average learning outcomes are girls who members of poor households and who live in rural communities. In addition, girls' whose caregivers or heads of household have never received any formal education, score much lower than average across all learning assessments, as do girls who have a disability, and girls who are old for their grade.

Transition

Relative to the baseline benchmark sample, transition rates at midline have increased by 0.7 percentage points, from 88.3 percent in the comparable portion of the baseline benchmark, to 89.0 percent among cohort girls at midline. This difference is not statistically significant at any standard level. The observed improvements in transition come from an overall reduction in dropout rates.

EGEP-T only considers in-school progression or re-enrolment to constitute successful transition, and all cohort girls were enrolled at the baseline. As a result, the breadth of potential variation in transition

pathways is fairly limited, and only three possible outcomes are observed in the transition sample analysed in this report: dropping out, being held back a grade, and progressing to the next grade. At midline, dropout rates had declined from 6.8 percent to 3.6 percent, a gap that is statistically significant at the 5 percent level. A decline in dropout rates is meaningful, even if much of the shift has been toward girls repeating a grade, because it means girls remain enrolled, continue learning, and may continue to progress in school during the following year.

Subgroup analysis did not reveal strong predictors of successful transition, in general. Substantial individual-level barriers to education, such as motherhood and early marriage, were associated with lower transition rates, as were some forms of physical impairment, though the latter were based on exceedingly small sample sizes. Despite ample qualitative evidence that the financial burden of schooling on families poses a barrier to continued enrolment, girls from economically marginalized households do not have systematically lower transition rates.

Sustainability

The sustainability of the project has improved somewhat over time, though improvements are not broad-based across different metrics of sustainability. In general, improvements have occurred at the level of schools, while community-level indicators of sustainability have stagnated at their baseline values. The project defined sustainability in schools based largely on material support from CECs; CECs are more likely than they were at baseline to provide bursaries to girls in their communities, which should encourage higher enrolment rates and improved transition outcomes even after bursary support from EGEP-T ends. However, increased support in one realm appears to have been offset by decreased financial support from CECs for school improvements -- CECs are less likely to provide this support at midline, possibly because financial resources have been shifted to the provision of bursaries.

Other school- and system-level sustainability indicators include the extent to which schools actively participate in the project's teacher mentoring program and have implemented mechanisms for child protection. At baseline, the project's teacher mentoring program had not begun. By comparison, the vast majority (82.5 percent) of schools now have at least one teacher actively participating in the program. Participating teachers are enthusiastic about the program, reporting that they use the skills they have learned, though this has not yet been reflected in observable teaching quality.

Perhaps the most significant gains in sustainability have been in the implementation of child mechanism procedures. Schools are much more likely at midline to have appointed a focal point within their school for child protection issues, are more likely to maintain a record of such cases, and are more likely to have an established procedure for handling, escalating, and following up on reports. These improvements are reflected in the comfort level of girls, more of whom feel they could report mistreatment by a teacher or bullying by a male classmate now than at midline.

Where sustainability has not improved over time is in community attitudes toward girls' education. Awareness-raising activities within communities, meant to encourage enrolment and galvanize public support for girls' education, have not increased since the baseline, though the baseline likely captured a concerted increase in awareness-raising at the start of the school year. Unfortunately, male support for girls' education – in the form of tangible actions taken by men and boys in the community to support their sisters and daughters to complete school – has not changed since baseline. Caregiver perceptions of male support for girls' education has not improved overall, and has regressed on some

indicators. Teachers report that fewer – 29.2 percent, versus 32.8 percent at baseline – fathers would likely attend a parent-teacher meeting regarding their daughter's education. Girls' perceptions are similar, but mixed: the share who feel male peers encourage girls to answer questions in class has risen from 79.9 to 84.1 percent at midline. But girls feel less support from their brothers at home, with fewer believing their brother would help them with their chores to allow them time to prepare for an examination.

Net improvements in sustainability have been driven primarily by changes in school-level outcomes, as noted above. In general, sustainability indicators based on participation in project interventions – teacher mentoring and the establishment of child protection procedures – have shown the greatest gains, while indicators based on changes in community sentiment or tangible actions by community members have shown no progress.

Gender Analysis

EGEP-T interventions have the potential to be gender-transformative. In particular, interventions at the output level including provision of bursaries and cash grants, provision of solar lamps, recruitment of female mentors, and teacher training on learner-centred pedagogy all have the potential to make gender-transformative contributions. By most measures, the project has yet to have a gender-transformative effect, but this is understandable given the severity of pre-existing gender inequalities and the fact that school-based interventions including teacher training have not yet had sufficient time to take effect.

In learning, a significant gap between girls and boys was found at baseline and persists at midline. In transition, boys' outcomes are not being tracked, and this is problematic in terms of the potential to assess gender transformative effects on transition within EGEP-T. Analysis of community attitudes (more on this below) suggests that, at best, there have been modest improvements in attitudes toward girls' education, and qualitative evidence suggests that girls still have more challenges than boys with regard to feeling safe on the way to, and at, school. As girls age, they also face more pressure than boys to miss school to perform household duties, and to drop out of school to marry. Qualitative evidence suggests that social barriers, such as early marriage, are being actively contested, and thus more progress may be evident at the next evaluation point.

Attendance

The average rate of attendance in the midline among panel cohort girls has risen modestly to 94.6 percent from 93.2 percent in the baseline evaluation. While this increase in attendance rates was not statistically significant, panel girls in the midline were significantly less likely than panel girls in the baseline to have attendance rates lower than 85 percent, 80 percent, and 75 percent, suggesting that anecdotal qualitative evidence of improved teacher efforts to track and monitor attendance may be accurate and these efforts may indeed be helping to limit cases of extremely low attendance.

Girls in Puntland attended school at significantly lower rates than girls from other zones, while girls in Galmudug had significantly higher attendance rates. Neither drought nor urbanicity were strongly correlated with attendance rates, but living in a conflict-affected area was a significant predictor of lower attendance rates. Though girls in both Galmudug and Puntland live in areas with conflict, girls in Puntland may have experienced far more sustained conflict than girls in other zones because girls in Puntland were coded as being in conflict-affected schools in both the baseline and the midline (while none in Galmudug were coded in this way).

Few barriers were found to be strong correlates of school attendance; however, girls who have 30 minutes or more of travel to get to school or who feel unsafe on the way to school were significantly more likely to have lower attendance rates than peers.

Targets for girls' overall attendance in the attendance record data of cohort girls and in the headcount data were met, but the boys' headcount attendance target was not achieved. Lastly, we find that girls in urban, IDP, and conflict-affected areas achieved headcount attendance targets, however targets were not met in rural, non-IDP, and non-conflict areas.

Self-Esteem and Empowerment

Unlike boys, girls reported high levels of confidence in their abilities to lead their peers and friends to do an activity. While girls' leadership skills did not change to a statistically significant degree from the baseline to midline, girls' overall agency and self-esteem (as measured through the combined score of agency and self-esteem indicators) did increase significantly from baseline to midline, even after controlling for girls' geographic, demographic and school-level characteristics. Girls tend to attribute these positive changes in their levels of self-confidence to their participation in the EGEP-T project activities.

Teaching Quality

In keeping with the baseline, at midline the primary indicator of teaching quality is an index score of teaching practices observed during classroom observations. Fifteen different behaviours were observed for teaching approaches during the classroom observations, including the practice of referring to previous lessons, and helping students when they did not understand. In the aggregate, the average teaching quality score across all midline schools was 63.9, while the baseline average was 67.3. Thus, the teaching quality score decreased by a total of 3.4 percentage points from baseline to midline, but this difference is ultimately not statistically significant.

At the same time, a more targeted analysis of teachers who participated in the project's Continuing Professional Development (CPD) training intervention showed some evidence of improved teaching practices from baseline to midline, both in general and in comparison to teachers who did not complete CPD training. The improvements documented are substantively large, but based on a small sample of teachers who were successfully tracked over time, but provide suggestive evidence that project interventions are having an impact on those participating, without necessarily improving teaching practices among the entire teaching cadre. Triangulating the teaching quality score with other ancillary indicators of teaching quality suggests that teacher motivation has likely increased since baseline, as reflected in reduced teacher absenteeism and increased teacher preparation, there is broadly no evidence of a measurable intervention effect on teaching quality since the baseline.

School Management and Institutional Governance

School management was measured using a multidimensional scorecard approach. In the aggregate, schools improved only slightly from baseline to midline, improving 1 percentage points on a standardized 100-point scale. Improvements were concentrated in Puntland; in fact, Puntland is the only geographic zone that showed improvement in overall school management.

The score for school management captures the quality and activity level of CEC management, the level of tangible CEC support for the schools and girls in them, the enactment of formal school-level policies key to school functioning and improvement, and the quality of school record-keeping. The biggest gains from baseline to midline were seen in the activity levels of CECs: at baseline, 65.0 percent of head teachers reported their school had a CEC and it met at least monthly; at midline, that share had risen to 71.8 percent. CECs are also more likely at midline to provide bursary support to girls in the schools they manage. At the same time, CECs are somewhat less likely to make financial contributions to manage or improve the school, which may suggest there is a substitution effect, in which CECs can increase their bursary support, but only at the expense of other monetary needs.

Community Attitudes

The project's logframe indicator related to community attitudes focuses on caregivers' aspirations for their daughters. There were concerns prior to the midline about the comparability of this measure from baseline to midline, as the baseline sample covered caregivers from a random household sample, while at midline it would consist exclusively of caregivers of cohort girls. The share of caregivers who expressed high aspirations for their daughters' education was already high at baseline, with 89.5 percent of caregivers in the sample indicating they hoped their girls would go to college or university. At midline, this share had increased to 90.7 percent, though this difference was not sufficiently large to be distinguishable from a null effect. Gains on this indicator were concentrated exclusively in Puntland, where the share of caregivers who aspired to send their daughters to college rose from 88.3 percent to 92.3 percent.

Findings related to potential improvements in community attitudes were mixed. Caregivers had only slightly higher aspirations for their girls in terms of educational attainment, but they were more willing to make tangible sacrifices to support their girls' education (when prompted to in a hypothetical scenario) than they were at baseline. Girls themselves appear to be more inclined to stay in school and recognize its importance. When asked whether going to school was important for what they want to do in the future, the share of girls who responded that school was important rose from 94.7 percent at baseline to 99.3 percent at midline. However, while their views on education seem to have improved, their control over educational decisions that affect them has not increased, at least according to their own self-reports: slightly fewer girls feel they get support from their families to stay in school and do well, and slightly fewer feel they can choose whether to stay in school or not.

Conclusions

The foregoing findings were compared against targets set in the baseline, and with few exceptions, the targets were not met.

> *Learning*: Arithmetic difference in differences analysis suggests that there have been no measurable intervention effects on cohort learning since baseline. The same findings also hold for bursary girls. When testing for differential intervention effects between cohort girls and bursary girls, we find that bursary girls are keeping pace with cohort girls, and may be closing the gap with cohort girls in numeracy and Somali literacy. When comparing bursary girls and boys, we find that cohort girls are keeping pace with boys, but not closing the substantial gap that has existed since baseline. Finally, we find that girls in baseline grade 8 (midline F1) have markedly attenuated learning vis-à-vis their peers in other grade-levels as well as when

compared with benchmark expectations or when compared with boys at the same grade-level.

Transition: Relative to the baseline benchmark sample, transition rates at midline have increased by 1.2 percentage points at midline as compared with the baseline benchmark. This difference is not statistically significant at any standard level. The transition target of 7% improvement of transition rate above the benchmark group in the baseline group was not met with only a 0.7% increase in the rate..

Sustainability: Improvements in sustainability have occurred at the level of schools, while community-level indicators of sustainability have stagnated at their baseline values. The project defined sustainability in schools based largely on material support from CECs; CECs are more likely than they were at baseline to provide bursaries to girls in their communities, which should encourage higher enrolment rates and improved transition outcomes even after bursary support from EGEP-T ends. However, increased support in one realm appears to have been offset by decreased financial support from CECs for school improvements -- CECs are less likely to provide this support at midline, possibly because financial resources have been shifted to the provision of bursaries. Perhaps the most significant gains in sustainability have been in the implementation of child mechanism procedures. Schools are much more likely at midline to have appointed a focal point within their school for child protection issues, are more likely to maintain a record of such cases, and are more likely to have an established procedure for handling, escalating, and following up on reports. These improvements are reflected in the comfort level of girls, more of whom feel they could report mistreatment by a teacher or bullying by a male classmate now than at midline.

Attendance: The average rate of attendance among panel cohort girls has increased moderately from baseline to midline. Girls in the midline are significantly less likely than panel girls in the baseline to have attendance rates lower than 85 percent, 80 percent, and 75 percent. Girls in Puntland attended school at significantly lower rates than girls from other zones, while girls in Galmudug have significantly higher attendance rates. This difference in attendance by geographic area appears to be linked to Puntland's sustained conflict across baseline and midline evaluation points.

Self-esteem: Girls indicated significantly higher levels of self-esteem and agency when compared to the baseline data. The changes in girls' levels of self-esteem were still statistically significant even after controlling for girls' school level factors as well as demographic and zonal characteristics including age and grade.

Teaching quality: There has not been a measurable positive intervention-effect on teaching quality since baseline. In the aggregate, the average teaching quality score across all midline schools was 63.9, while the baseline average was 67.3. Thus, the teaching quality score decreased by a total of 3.4 percentage points from baseline to midline, but this difference is ultimately not statistically significant.

School management: In the aggregate, schools improved only slightly from baseline to midline, improving 2.8 percentage points on a standardized 100-point scale. Improvements were concentrated in Puntland; in fact, Puntland is the only geographic zone that showed improvement in overall school management.

Community attitudes: The share of caregivers who expressed high aspirations for their daughters' education increased modestly from baseline to midline, but this difference was not

sufficiently large to be distinguishable from a null effect. Gains on this indicator were concentrated exclusively in Puntland, where the share of caregivers who aspired to send their daughters to college rose from 88.3 percent to 92.3 percent.

Despite the lack of targets that were met, and though there is a substantial gap between the projects' achievements at this evaluation point and the next evaluation point, the indicators used to measure progress remain relevant for subsequent rounds. The indicators remain important proxies for concepts in the theory of change. The theory of change suggests that learning scores and transition rates ought to be linked to the intermediate outcomes such as attendance, self-esteem, teaching quality, school management, and community attitudes. Our analysis found that only self-esteem, teaching quality, and school management were positively correlated with learning scores and

Evaluation Recommendations

- Following difficulty in this evaluation round creating learning assessments that are comparable in difficulty between baseline and midline, we strongly recommend two rounds of piloting for the learning assessment at endline.
- In the endline evaluation, we recommend timing the evaluation to avoid school closures or exam preparations and thereby maximize the data that can be collected from the various survey tools of the evaluation.
- In order to understand the extent to which students' skill gaps are a reflection of weaknesses in their teachers' understanding, we recommend matching data from teacher assessments during teacher training with the learning assessment scores of their students.
- Grade 8 students who cannot be located in a follow-up round of data collection should be replaced with girls in a different grade (e.g. Grade 7 or Form 1) to avoid repeated high attrition rates in each subsequent evaluation round.
- Prior to the endline, revisions should be made to the quantitative and qualitative data collection tools in consultation between the evaluation team and RI staff to more fully capture data on sustainability, making questions more specific to project interventions where possible.

Programming Recommendations

- To improve the attendance and learning, the project should consider creating a transport program that targets girls with long commutes in conflict-affected or unsafe areas. This could be done through busing and/or drivers who have been vetted. Alternatively, school CECs could coordinate caretakers to escort groups of girls to school when they are traveling from nearby remote areas.
- CECs should be empowered to raise money from communities to invest in school improvements that will provide basic health and sanitation facilities, particularly in rural and under-resourced schools.
- CECs should be trained to target their financial support to the areas of greatest need in their particular schools and to help them identify alternative revenue sources. Increased and more consistent training for CECs should also promote engagement among the schools with defunct CECs at midline.
- We encourage schools to allow girls who are finishing household work in the morning to still attend even if they are tardy in order to allow girls, who usually take on household chores in the morning such as making breakfast for their family.

- In order to improve community attitudes towards girls' education, efforts to encourage enrolment should expand beyond only community leaders to regular men and women in the community.
- Recognizing that teaching quality has not improved by most measures and that teachers are not thinking about teaching methods in the classroom or using student performance as a means of evaluating teaching methods being used, we recommend increasing teachers' attention to issues of pedagogy. This could be accomplished through mechanisms such as teachers observing one another's classes, in-service days focused on pedagogy, self-evaluation, and reflection on teaching methods.
- Given the uneven rollout of project interventions, ensuring widespread uptake of project interventions in a timely fashion would contribute to improved outcomes, especially in those outcomes that tend to improve slowly such as transition, community attitudes, and self-esteem.

1. Introduction

1.1 Project Context

For the last 13 years, Somalia has been ranked among the top three most fragile states in the world.[2] Somalia's long slide into lawlessness and insecurity is typically timed to have begun in 1991, with the ouster of President Siad Barre's military regime, which prompted a descent into civil war between multiple armed groups, and a generalized breakdown in public services and security. This process has naturally impeded the Somali people's access to education, as the collapse of the state included the collapse of formal education systems, leaving funding and governance gaps that regional administrations, international non-governmental organizations (NGO), national NGOs, religious groups, the private sector, and various other groups have since struggled to fill. Despite their best efforts, the involvement of so many actors has led to a lack of uniformity that presents its own set of challenges for the education sector today.

Adding to a lack of uniformity in education is the broader fact that the context in which education occurs in Somalia varies dramatically across geographic space. For instance, Somalia's education system was derived partially from different education systems introduced by the British and Italian colonialists in different regions. In short, the context in which education occurs in Somaliland is very different from that of Banadir, driven by differences in recent history, differences in the nature and severity of past and ongoing conflicts, variation in climate and bouts of drought, as well as differences in the institutional structure of education and the broader political setting in which those institutions operate.[3] The school system has also been influenced by the Islamic nature of Somali society. Quranic schools have an active role in the current system – official statistics in 1981 show that there were approximately 5,480 Quranic schools operating in the country, with girls aged 4 to 14 years old constituting the majority of students enrolled in such schools.

Somalia continues to be affected by a series of conflicts, ranging from short-term inter-clan disputes to the long-term battle against al-Shabaab. Since 1991, three separate international peacekeeping missions (UNOSOM I, UNITAF, and UNISOM II) have been deployed to Somalia, and a regional peacekeeping mission (the African Union Mission in Somalia, or AMISOM) is currently active. These peacekeeping efforts have focused on the broader conflicts, most recently between the federal government and al-Shabaab. During the civil war itself, between 1977 and 1991, 90 percent of the country's school buildings were destroyed, instructional materials were lost, and teachers and students scattered and abandoned the educational system. The fighting between al-Shabaab and the government – and the former's occupation of large areas of the country at times – has furthered this destruction, as schools have been directly targeted by armed groups, either through attacks, occupation, or use as a source of forced recruitment and indoctrination. In 2018 alone, 64 schools were attacked and 21 forced to close.[4]

The political climate in the country broadly has improved somewhat in recent years, but is still marked by tentative and geographically-limited gains in stability. In 2011, al-Shabaab was driven out of

[2] https://fragilestatesindex.org/wp-content/uploads/2019/03/9511904-fragilestatesindex.pdf

[3] For instance, even the descent into civil conflict has varied across the country. Somaliland's greatest experience of conflict was, arguably, prior to the 1991 fall of Siad Barre, while other regions saw their greatest insecurity and violence in the years that followed the collapse of his government.

[4] https://reliefweb.int/sites/reliefweb.int/files/resources/20190118_humanitarian_response_plan.pdf

Mogadishu. Two limited, indirect presidential elections have taken place in recent years—one in 2012, when the Transitional Federal Government (TFG) was replaced by the Federal Government of Somalia (FGS), and another in 2017, during which incumbent President Hassan Sheikh Mohamud conceded defeat to former Prime Minister Mohamed Abdullahi Mohamed. Although these strides have enabled a more effective humanitarian response, al-Shabaab still controls many rural areas in southern Somalia, violence remains high, and there are continued concerns over ineffective governance.[5] In this context, the humanitarian situation has also deteriorated, as recurrent droughts have had a severe impact on vulnerable communities.

The evolution of security over the past 2-3 years underscores the geographic variation in conflict and the fact that the situation in any given area can change quickly. While al-Shabaab was driven out of Mogadishu in 2011, the conflict has worsened in recent months, reaching a point at which midline data collection in the capital city was cancelled due to security concerns.

Meanwhile, Somaliland has experienced a different conflict trajectory. In 1991, Somaliland unilaterally declared independence from the central government, and has governed itself as a de facto state, unrecognized by the international community, since that time. Somaliland, in particular, has not been seriously affected by al-Shabaab, while al-Shabaab has successfully orchestrated attacks in Puntland. Baseline data collection for EGEP-T took place in late 2017, immediately following the Somaliland elections, which prompted clan conflicts centred on disputed election results. Underlying tensions between clans in Ceel-Afwayn erupted, after the election, into open fighting.[6] And the border dispute between Puntland and Somaliland – which has been simmering for many years – devolved into widespread fighting in Tukaraq town in early 2018, just 2-3 months after the election. Fighting in this area occurred sporadically throughout 2018, resulting in occasional school closures. In general, however, Somaliland became more settled during much of 2018, and conflict has become less widespread since the baseline. It is important to note, however, that many conflicts in Somaliland go underreported and are unnoticed by the international community, because they tend to be localized and short-lived, driven by clan differences and historical tensions that flare up temporarily; this does not mean, however, that they are necessarily less destructive to stability and investment in education.

In contrast to Somaliland, Puntland and Galmudug are both heavily impacted by the presence of al-Shabaab, in addition to Puntland's border conflict with Somaliland. In 2015 and 2016, conflicts in Galmudug prompted the closure of schools, though conflict has been reduced somewhat since that time. Certainly, the sample of schools evaluated at the midline is less affected by conflict than they were at baseline, partially as a result of the easing tensions in Somaliland and partially as a result of the exclusion of schools in Mogadishu. In short, the importance of conflict to the evaluation has been reduced due to both omission from the sample and thanks to a reduction in conflict.

The effects of conflict are exacerbated by, and arguably operate in a feedback loop with, poor climatic conditions and drought. The most recent 2016/2017 drought led to an increase in food insecurity and cash shortages and resulted in widespread loss of livelihoods, as many lost all of their livestock. From November 2016 to May 2017, there were an estimated 739,000 drought displacements.[7] Due to sustained humanitarian assistance and above-average *GU* rains (April – June 2018), some

[5] https://fragilestatesindex.org/wp-content/uploads/2019/03/9511904-fragilestatesindex.pdf
[6] Interpeace. September 11, 2018. "Somaliland: 'Declaration of Peace' Halfs Protracted Conflict n Ceel Afwayn." Available at: https://reliefweb.int/report/somalia/somaliland-declaration-peace-halts-protracted-conflict-ceel-afweyn.
[7] https://reliefweb.int/disaster/dr-2015-000134-som

improvements were seen in food security from 2017 to 2018. However, malnutrition rates remain high and food security improvements have recently reversed with the below-average *Dyer* rains in the last quarter of 2018. As a result, the *Jalal* dry season (January – March 2019) has been particularly harsh, with some of the worst-affected areas experiencing faster depletion of pasture and water resources, severe water shortages, and earlier than normal water trucking at high prices.[8]The full effects of conflict and drought on the education system are difficult to estimate, as a census has not been conducted since 1986/87. However, there have been attempts made by different development groups to fill the data gap. A 2013/2014 Population Estimation Survey (PESS) suggests that education indicators vary widely by region, level of urbanization, and wealth status. On the whole, however, the survey found that the gross enrolment ratio (GER) in 2014 was just 30% for primary school, 27% for secondary school, and 9% for tertiary school. In early 2019, it was estimated that a total of 2.6 million people had been internally displaced due to the combined effects of conflict, insecurity, and/or drought, and three million children were out of school across Somalia.[9]

A more recent primary school GER estimate of 32% in 2017 suggests that the situation remains critical and that Somalia still has one of the lowest primary school enrolment rates in the world. Notably, enrolment rates are particularly low among rural households (18%) and in nomadic communities (<10%).[10] Also notable, attendance rates vary by displacement status—there is a significant disparity in attendance rates between non-displaced (45%) and internally displaced (28%) school-aged children.[11]

Even when parents enrol their children in school at a young age, an estimated 50% of children drop out of school by grade 5. In addition to conflict and displacement, research suggests that lack of finances is one of the greatest barriers to education. In one study, 65% of IDP and 49% of non-displaced households who do not have all their school-aged children in school reported that school fees are the obstacle. The opportunity cost of sending children to school is particularly pronounced in some regions and households, as children are often needed to help with household chores or supplement the household income.

This chore burden is especially heavy for girls in Somalia, with girls between the ages of 5 and 14 years old spending an average of 26 hours a week on household chores. This is the highest chore burden of any country and one of the most disproportionate for girls compared to boys.[12] In general, the chore requirements for girls increases with age, meaning that the older girls targeted by EGEP-T programming are likely to experience a particularly onerous burden of household tasks.[13] Girls also face a number of other additional barriers to education, such as low availability of sanitation facilities, lack of female teachers, safety concerns, and social norms that favour sending boys to school.[14]

For girls and boys that remain in school, learning outcomes are adversely affected by other shortcomings in the education system. Somalia's teacher-student ratio is estimated to be at 1:32, a

[8] https://reliefweb.int/report/somalia/somalia-climate-update-february-2019-monthly-rainfall-and-vegetation-cover-issued
[9] https://reliefweb.int/sites/reliefweb.int/files/resources/20190118_humanitarian_response_plan.pdf
[10] https://www.unicef.org/somalia/SOM_resources_annualreport2017.pdf
[11] https://reliefweb.int/sites/reliefweb.int/files/resources/reach_som_final_report_joint_multi_cluster_needs_assessment_2018_1.pdf
[12] https://www.thejournal.ie/un-chores-3014246-Oct2016/
[13] This is consistent with data from the EGEP-T evaluation, where girls are more likely to cite household duties as a barrier to reading and studying as they get older.
[14] https://www.unicef.org/somalia/SOM_resources_annualreport2017.pdf

figure that is believed to "vary significantly" from one zone to another (UNICEF, 2012).[15] According to a World Bank report on non-state education providers, in general, 12 percent of teachers in Somalia do not have post-secondary education.[16]

The educational context in which EGEP-T is being implemented is also extremely challenging. Somalia's education and school systems are in need of much improvement. The 2016 Puntland and Somaliland Education Sector Analysis indicates that the average revenue budget allocated nationally for education is 7 percent, with more than 90 percent of which is paid to support personnel and on-going operations.[17] As a result, teachers are insufficiently trained and unqualified. Many teachers do not have any training in education and get employment as teachers because they could not find employment elsewhere. In 2011, the Somaliland government made their primary schools free for all students in Somaliland. This caused a large influx of students into the schools, often leading to over-crowding, and at the same time, caused teachers to move into private schools or find other occupations because of the worsening working conditions. In addition, the federal government in Mogadishu only began developing a national curriculum in 2017. There is little standardization of curriculum, teaching quality, or other aspects of education in schools around Somalia. Unless education is made a higher priority by the national government, Somalia's education system will remain dependent on non-state education funding.

Somalia's federal structure, and disputes over sovereignty and independence of individual states, renders the educational context both more difficult and widely divergent from place to place. For instance, as we discuss in greater detail in Section 4.1, until recently Somalia had lacked a unified curriculum for either primary or secondary school The former situation, in which individual schools had imported curricula from neighbouring countries, or implement multiple curricula in a piecemeal fashion, has been characterised as "curriculum chaos" by at least one expert on Somali education policy.[18] Emerging from this, Somaliland, Puntland, and FMS have now developed a unified curriculum that is in different stages of rollout. By the time of the next evaluation point, this curriculum may be sufficiently uniform in its implementation that it can serve as the basis for analysis of grade-level competencies. For the purpose of this evaluation, we adhere to the standards established at the baseline prior to unification of curriculum.

The institutional framework within which education takes place is superficially similar across jurisdictions, but there are important differences in policy and implementation on-the-ground. The Federal Republic of Somalia (FRS), Somaliland, and Puntland, for instance, all organise education under their respective Ministries of Education and Higher Education, and on-the-ground efforts are run – at least in de jure terms – by regional and district offices of the same. Similarly, students across the three areas enter school at the same age and primary and secondary education are organised on a 4-4-4 system, with four years each of lower primary, followed by upper primary, and then secondary school..

But aspects of policy, and the reality on-the-ground, can vary markedly. Somaliland and the FRS both offer early childhood education to children aged 3-5 years, prior to enrolment in primary school, but

[15] UNICEF Annual Report 2012 for Somalia, ESARO http://www.unicef.org/about/annualreport/files/SomaliaCOAR_2012.pdf

[16] World Bank (2018) Study on Understanding the Role of Non-State Education Providers in Somalia. Available at: http://documents.worldbank.org/curated/en/487221555358250787/pdf/Study-on-Understanding-the-Role-of-Non-state-Education-Providers-in-Somalia.pdf

[17] Puntland Education Sector Analysis (draft), September 2016, p27, Somaliland Education Sector Analysis (draft), September 2016, p31.

[18] Educational Challenges in Post-Transitional Somalia, 2015, Heritage Institute for Policy Studies.

Puntland does not appear to have a specific policy for early childhood education. Somaliland, meanwhile, is the only project location that has a formal policy of free primary education, though implementation is uneven: as we found at baseline, a majority of primary school head teachers in Somaliland reported that their school charges fees to student, despite the formal policy of free primary education. Similarly, the general trend across locations is for instruction to take place in Somali during primary school and English during secondary school. However, this general trend masks important variation that might influence learning outcomes, especially – all of the schools sampled in Galmudug reported English as the language of instruction, while Somali was almost universally used in Somaliland, even in secondary schools. At an even more fundamental level, ownership and control over schools tends to be different across project locations, with predominantly publicly-owned schools in Somaliland, private schools in Banadir and Galmudug, and community-based schools in Puntland.

Even within the same geographic zones, institutional incoherence presents a significant barrier. Shattered in the wake of the civil war, the Somali educational system was rebuilt largely by private umbrella associations, over a dozen of which are actively operating today. These umbrella associations are typically run by former teachers, financed by some blend of diaspora Somalis, international donors, and school fees paid by students, and serve as the overseers of the schools within their associations, including the maintenance of teaching standards and – to some extent – development of a specific curriculum.[19] The largest umbrella organizations manage several hundred schools, and managing far more than those under the purview of the federal Ministry of Education, Culture, and Higher Education (MOECHE) itself. The role of umbrella associations and other non-state education providers can also vary across geographic areas.[20]

Teacher training is also not uniformly run by official state actors: the primary vehicle for teacher training in the country – Teacher Training Institutes (TTIs) – can also be run by private umbrella organizations, undermining uniformity in training standards.[21] To provide a brief overview of the diversity of TTIs and conditions under which they operate: In 2005-6, when the Programme for the Strengthening of the Capacity of Teacher Training (SCOTT) was designed based on a Teachers' Needs Assessment, Somaliland opted for Private TTIs, while Puntland established GTEC as the public TTI with the Puntland administration exercising clear control over the governance of GTEC and its strategic focus and management. Teacher training in remaining locations was distorted due to insecurity and significant control of umbrellas. Continuous Professional Development programs had not previously existed in Somalia. This is a new intervention based on recently designed INEE, TICC framework.

[19] World Bank (2018) Study on Understanding the Role of Non-State Education Providers in Somalia. Umbrella associations essentially take on all aspects of governance and management of their member schools, leading teacher training sessions, developing curricula, monitoring standards, and so forth.

[20] For example, in south-central Somalia, the majority of schools are owned by communities. In contrast, in Puntland, authorities and communities tend to jointly own schools, whereas private individuals own a large portion of schools in the south and in Somaliland.

[21] World Bank (2018) Study on Understanding the Role of Non-State Education Providers in Somalia. While privatized teacher training is fairly common in other contexts, including those of developed countries, this fact has important consequences for programmatic interventions aimed at influencing teaching quality. To the extent that EGEP-T seeks to influence teacher training curricula, the varied institutional framework of training complicates this effort and makes it less efficient, by necessitating coordination across and interaction with numerous additional organizations and stakeholders. Moreover, very different curricula across training institutes and teacher colleges means developing a unified approach to EGEP-T teacher training is more difficult, as some institutes may provide excellent training on, for instance, gender-sensitive teaching, while others neglect the topic altogether. Such divergence introduces unnecessary inefficiency into the development of EGEP-T interventions.

Somaliland recently brought teacher training under State control/public TTI under MoE&S. The federal government of Somalia has declared Somalia National University as the main Public TTI for Somalia and will collaborate with GTEC to replicate teacher training established in Puntland.

The nature of school ownership also appears to be related to the typical qualifications of the teachers they employ, according to at least one study. Independent schools tend to have the most qualified teachers, with 75 percent of their teachers holding a bachelor's degree; in schools managed by an umbrella association, 68 percent of teachers meet this standard, while qualifications are less prevalent in government and other types of schools.

Coupled with the fact that each state has their own ministry overseeing education – with varied capacities, funding, and policy goals – it becomes clear that education in Somalia takes place within a fractured institutional environment, with overlapping and competing organizations, as well as gaps in coverage and monitoring that occur naturally as a result. To illustrate the idea of institutional competition, there is an ongoing dispute between the Puntland MOE and the Federal MOECHE over the administration of examinations, with the Puntland MOE electing not to implement the Unified National Secondary School Examination off and on for the last several years.[22] These facts make project implementation more difficult in myriad ways. Increasing the number of stakeholders requires more consultation and more effort to satisfy diverse stakeholders. The varied institutional context produces schools with wildly different needs and capacities – these differences inhibit the development of a programmatic approach or unified interventions, because interventions need to be matched to schools of all different kinds and backgrounds. Finally, coordination with official actors becomes less efficient, which can prevent access to schools or slow efforts to influence formal policy. Varied federal- and state-level frameworks, combined with public versus private distinctions, also interact with the roles of Community Education Committees (CECs), head teachers, and other local-level actors who manage specific schools, complicating issues further and in ways that are *a priori* unclear.

The complex institutional environment, with a plethora of actors engaged in education policymaking and oversight, should not be taken as an indication that the *capacity* of actors is sufficient to effectively make policy, monitor implementation, or otherwise manage schools and the education system writ large. By any metric, the Somali state itself – whether considered at the federal or some other level – is almost entirely lacking in bureaucratic or administrative capacity. Government officials often have restricted access in conflict areas, and either conflict or limited budgets can prevent efforts to engage in monitoring or policy rollout in rural areas. The technical competence of the bureaucracy is often limited by the number of educated and experienced administrators, exacerbated by principal-agent problems, budget shortfalls, and other issues common to less-developed contexts. To illustrate, a recent report found that, due to ongoing teacher shortages and lack of funding needed for enforcement, the Somali government has not regulated the qualifications needed to be employed as a teacher, and the state lacked the resources necessary to run its own pre-service and in-service training programs.[23] The state fragility mentioned at the outset of this contextual review means that,

[22] See discussion in local Puntland media, such as the Puntland Post (online April 17, 2019) and Garowe Online (online April 20, 2019). This dispute is illustrative of the problems inherent to competing jurisdictions over schools. It is also worth noting that such conflict could influence transition rates for girls moving from primary to secondary school in Puntland, though analysis of this kind is beyond the scope of this report.

[23] World Bank (2018) Study on Understanding the Role of Non-State Education Providers in Somalia. Available at: http://documents.worldbank.org/curated/en/487221555358250787/pdf/Study-on-Understanding-the-Role-of-Non-state-Education-Providers-in-Somalia.pdf

even when there is broad political agreement on goals, implementation is often marginal and uneven at best.

It is in this difficult and varied context that Relief International (RI), the Adventist Development and Relief Agency (ADRA) and Comitato Internazionale per lo Sviluppo dei Popoli (CISP) are implementing the Educate Girls, End Poverty – Transition (EGEP-T) project. The complexity of each local environment makes both implementation and evaluation more difficult. Implementation is more difficult because access to schools and coordination with officials must be negotiated with different actors in different locations, and the poorly-institutionalized nature of government ministries overseeing education means that a standard approach applied to all schools is unlikely to be successful. Evaluation of the project's impact is also more difficult, and the results more tenuous, because there are often idiosyncratic local conditions that can drive major shifts in project outcomes, especially enrolment or transition rates, and changes in response to temporal shocks like drought or localized conflict are common.

1.2 Project Theory of Change

Relief International (RI) recognizes that the challenges associated with girl's education are multi-dimensional with no single "magic bullet" solution. As such, EGEP-T will take a multi-level and holistic approach similar to EGEP (Phase 1), building on the lessons learned and the work accomplished by EGEP. The project aims to assist hard-to-reach marginalised girls residing in urban and rural areas, as well as those located in IDP camps in Somalia and Somaliland. The project aims to effect sustainable change; the long-term goal is to enhance girls' knowledge by mitigating environmental, economic, and academic obstacles at critical decision points as well as equipping girls for adult life by facilitating their transition through primary and secondary education. Ultimately, girls will be able to enjoy an improved quality of life and communities will benefit from better skilled and educated young women.

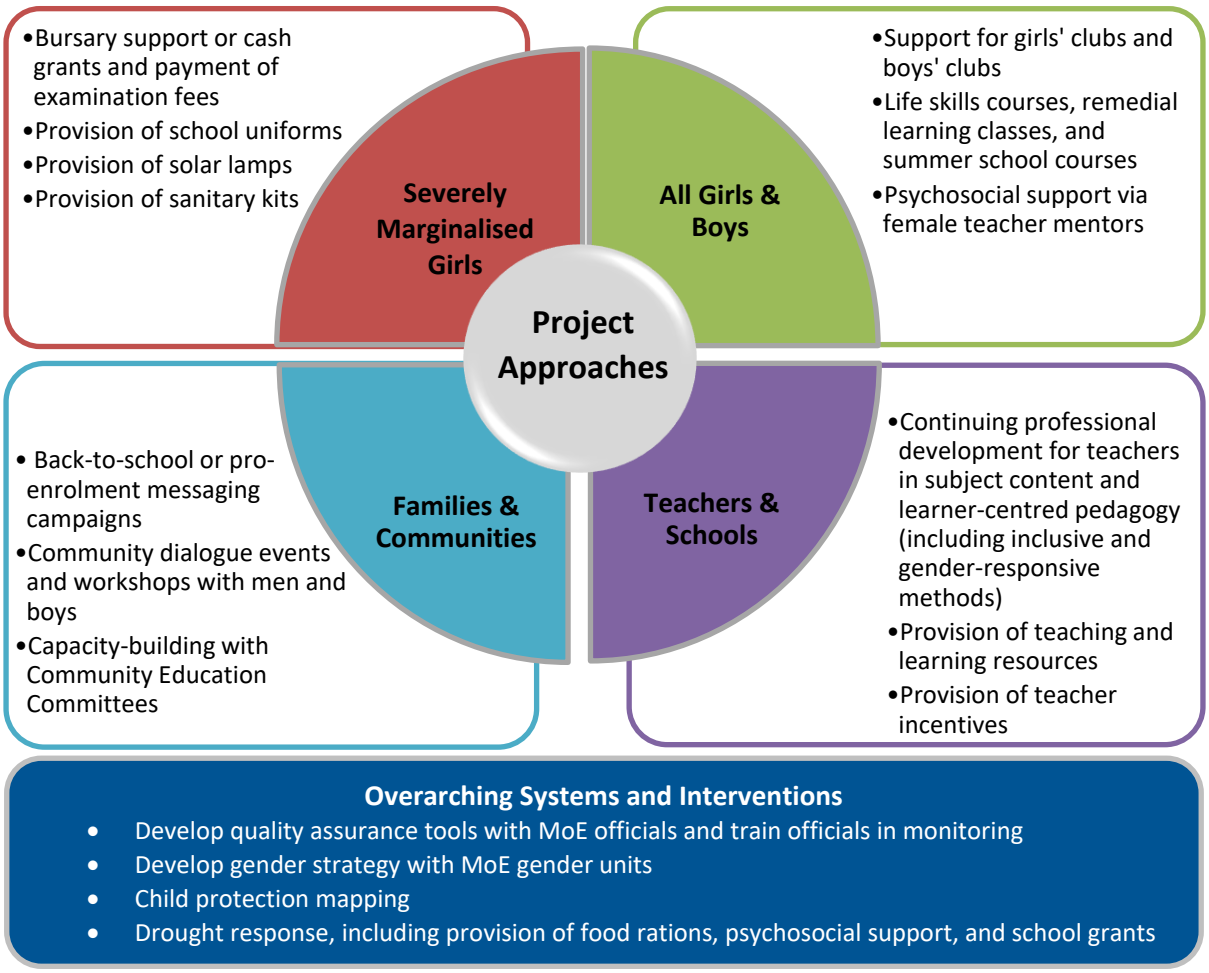
The project's Theory of Change identifies six barriers to girls' educational achievement, which project activities have been designed to address. The first barrier is a lack of economic resources at the household level – girls' education often takes a backseat to other household needs, especially because school fees impose a burden on household finances and girls' schooling prevents girls from either entering the workforce or engaging in housework. The second barrier is girls' lack of confidence and a dearth of psychosocial support for girls to continue their education. The third barrier is poor quality teaching in many schools, which reduces the ability of enrolled girls to learn and may undermine the rationale for continuing to enrol and attend school. The fourth barrier is the often-minimal engagement of relevant government ministries in school governance and management, which produces weak systems of child protection and quality monitoring. The failure of government agencies to monitor schools undermines teaching quality, record-keeping, and the incentives for teachers and schools to provide high-quality education. The fifth barrier is the often-weak management and governance of schools at the community level. Often neglected by the MoE, schools rely heavily on the CEC for their management and on community members for their upkeep, funding, and key aspects of governance. Where schools lack effective CECs and other governance mechanisms, school performance suffers. The sixth, and final, barrier identified by the project is a lack of community support for girls' completion of the full education cycle, especially in the face of social and cultural obstacles to educating girls.

EGEP-T has also been designed to adapt to the specific barriers faced in different contexts, and barriers which might arise or shift during project implementation. Owing to the potential for drought to influence project implementation and outcomes, drought response activities have been integrated

into EGEP-T programming. The barriers identified and targeted by EGEP-T do not exist in a vacuum. Often, they overlap and reinforce one another in complex ways, necessitating an integrated approach that seeks to address multiple barriers simultaneously.

In the sections that follow, we discuss these specific barriers and the interventions designed by EGEP-T staff to target each barrier to girls' educational achievement. The overall set of project activities, organised by the level at which they target girls, families, teachers, schools, and communities, are summarised in Figure 1 below.

FIGURE 1: LEVELS AND TYPES OF EGEP-T PROJECT INTERVENTIONS



Barrier 1: Household-Level Economics

Household finances represent a consistent obstacle to educating all children in Somalia, especially girls. Underlying both EGEP and EGEP-T interventions is an assumption – well-founded and based on significant primary research in Somalia and elsewhere – that schooling imposes a financial burden on households.[24] In Somalia, where unemployment rates are universally high, the incentives to complete secondary-level education are dampened. This is especially true for girls, as the default expectation for most girls is that they will get married and leave the labour force shortly after exiting school. To

[24] See, for instance: Deininger, Klaus. 2003. "Does Cost of Schooling affect Enrollment by the Poor? Universal Primary Education in Uganda." *Economics of Education Review* 22 (3): 291-305.

the extent that education is seen as an investment, considered in strict cost-benefit terms, the financial cost of enrolment is often too high to justify.

To ease the financial burden of enrolment on families, EGEP-T has planned a number of project activities. The most direct activity involves the full payment of school fees for families for severely marginalised girls.[25] This same group of girls will also be provided with school uniforms and the payment of their Grade 8 and Form 4 examination fees. For severely marginalised girls in IDP schools and rural areas, the project will also provide solar lamps and sanitary kits. These activities target the direct financial cost of enrolment (school fees), as well as incidental expenses (uniforms; examination fees). They also target indirect, non-financial costs of girls' enrolment: the reduced ability of girls to perform housework while enrolled. By providing school lamps, the project makes it possible for girls to complete chores during daylight hours and study after sunset. When girls cannot afford sanitary kits, it can result in them staying home when menstruating. The project is therefore distributing sanitary kits to improve girls' attendance rates. Moreover, sanitary kits may have knock-on benefits for girls' self-confidence and their sense of feeling supported in their decision to continue their education.

Barrier 2: Lack of Confidence, Life Skills and Psychosocial Support

Beyond financial barriers, girls in Somalia face additional challenges that reduce their enrolment, attendance, and learning outcomes. Girls often lack confidence in their abilities, which can reduce their willingness to participate in class, and even their desire to continue schooling. Girls also often lack educated female role models. In schools with few or no female teachers, it may be difficult for girls to see how education can improve their future lives, especially those girls from homes without an educated adult female. EGEP's endline evaluation, completed in late 2016, documented the role of psychosocial wellbeing on educational outcomes, finding, for instance, that girls with greater psychosocial wellbeing had higher attendance rates. As this baseline documents, girls indicate that they have significantly lower levels of agency over their actions and key decisions – such as whether to attend school and when to get married – than do boys.[26]

To target this individual-level barrier to girls' achievement, EGEP-T has planned a number of activities to be implemented in schools, including the formation and promotion of girls' clubs and boys' clubs that will engage in leadership and confidence-building activities, and provide children with valuable training on life skills, such as their reproductive rights, and social skills. Remedial courses will also be provided in mathematics and literacy, in an attempt to prevent girls and boys who have fallen behind – who may be demoralised as a result – from dropping out.

Barrier 3: Poor Learning Environment

Promoting enrolment and attendance is not enough to substantially transform girls' educational outcomes. Even if students attend class regularly, poor-quality teaching and a learning environment that is not conducive to learning will inhibit their performance. In addition, poor-quality teaching and an unwelcoming environment may discourage attendance and enrolment in the first place.

---

[25] In the case of Somaliland, where school fees have been formally abolished, cash grants will be provided to families, to deal with informal fees and incidental expenses.

[26] While girls express lower levels of agency, their reported self-confidence and empowerment is not dramatically lower than boys, overall. However, it is important to note that boys were not asked the full battery of self-confidence and empowerment questions; moreover, because these are self-reports, comparability across genders may not accurately reflect the relative confidence levels of each group.

Teachers in Somalia face their own significant barriers, which contribute to poor-quality teaching in the aggregate. For instance, teachers are often not provided the resources they need to teach effectively; their salaries are often delayed; and they work in difficult environments, with overcrowded classrooms, students with inconsistent attendance, and in communities that may not value their work. Moreover, the relatively low pay of teachers and lack of trained, qualified teachers means that many students are taught by teachers who are only marginally qualified.[27]

Schools, teachers, and students alike also face challenges in the form of a lack of teaching and learning resources. A substantial share of teachers lack the supplies they feel are necessary to teach their classes, and many report spending their own money to purchase supplies.[28] Many classrooms are devoid of relevant learning materials posted on the wall – while informational posters may not be the most critical resource for student learning, their absence suggests a broader dearth of resources.[29] Resource constraints are found at the student level as well, of course. Many students share textbooks, for instance, which may prevent them from studying outside of school.

To improve teaching quality and a more positive learning environment, EGEP-T plans a number of activities focused on training teachers and providing teachers with additional resources and additional incentives to promote high-quality teaching. EGEP-T will train teachers in subject content in mathematics, subject-specific pedagogical techniques, remedial teaching practices, and English proficiency. They will also train teachers in gender-responsive techniques. The teacher training programme will take the form of a Continuous Professional Development approach through the use of coaches.

The project's efforts regarding teacher training are necessarily focused on intervention schools and improvements that can be made among the existing teacher cohort. It is beyond the scope of the EGEP-T intervention to increase the net supply of trained teachers more broadly, or to promote higher-level educational qualifications for teachers in their schools. Teachers and school administrators routinely request this type of training, as documented in qualitative interviews conducted at the baseline, suggesting that they will take the training seriously and actively implement changes in their classrooms and in their teaching styles.

Barrier 4: Weak Government Outreach and Engagement

While schools may be capable of promoting positive educational outcomes, the system in which they operate is often not designed to do so. The fourth barrier identified by EGEP-T's Theory of Change concerns system-level deficiencies in monitoring, promotion of education, and other key activities of government ministries. Schools require monitoring to ensure that they are keeping accurate records, that teachers are consistently showing up for work, that teaching quality is good, and that child protection systems are developed and actually in use. Many schools can operate effectively without such monitoring, thanks to the efforts of head teachers, CECs, and other actors that hold them

[27] To illustrate, 6.5 percent of teachers surveyed at the baseline had not completed secondary school, and a further 21.9 percent had completed secondary school but had not received any additional training. In some cases, teachers who had only completed a secondary education were, themselves, teaching secondary school students.

[28] Among teachers surveyed in the baseline evaluation, 32.8 percent indicate that they do not have the basic supplies necessary to teach. In addition, 75.7 percent report that they have spent their own money on supplies at some point during their careers, suggesting that resource gaps would be worse if teachers relied exclusively on materials provided by their schools.

[29] Enumerators observed classrooms at EGEP-T schools and noted whether relevant learning materials were posted on the walls – just 36.7 percent of classrooms were observed to have materials posted.

accountable. However, these school-level and local mechanisms are not always sufficient; it is in these cases that external monitoring by MoE officials and other government representatives is particularly critical.

EGEP-T plans to engage with a number of government officials, at all levels, to promote greater engagement with and oversight of their intervention schools and other schools. For the purposes of monitoring, EGEP-T will – in conjunction with government officials – develop new monitoring tools that will allow ministry officials to more accurately and efficiently monitor school, teacher, and student performance. EGEP-T will also map and help strengthen existing child protection systems at the school level.

Barrier 5: Weak School Governance

The fifth barrier targeted by EGEP-T is related to the fourth, but focuses locally on school-level management. In the absence of effective and continual ministry oversight, schools in Somalia have become increasingly reliant on community mechanisms for management and support, such as CECs staffed by community members and head teachers. Where these institutions are underdeveloped or ineffective, many core functions of the school administration, such as monitoring student and teacher attendance, suffer. CECs can promote accountability, can influence community opinion regarding education, can leverage and aggregate community resources to improve infrastructure or pay teachers, and can promote efficiency at the school level.

EGEP-T will engage in a range of capacity-building exercises with CECs to improve school management. Where schools and communities are particularly strained financially due to drought, EGEP-T is providing cash grants to schools as part of their drought response efforts. CECs are participating in grant management training, as well as more general capacity-building efforts.

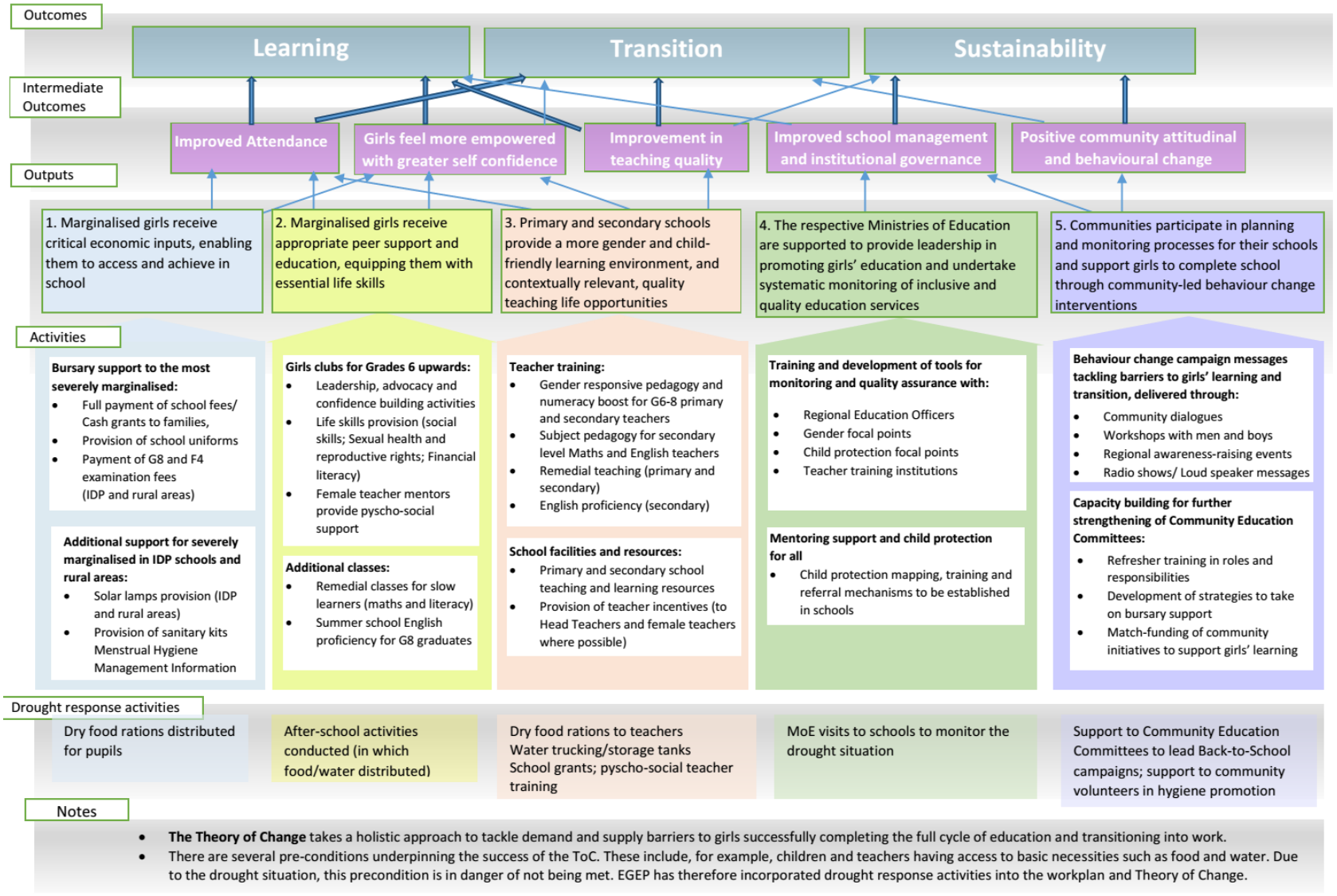
Barrier 6: Lack of Complete Community Support for Girls' Completion of the Full Education Cycle

The sixth barrier targeted by EGEP-T activities is a lack of community support for girls' completion of the full education cycle. In the first phase of the EGEP, the project found a high level of community support for girls' education in general. However, this did not always translate into behaviours to actually facilitate girls' schooling, and critically, the support did not necessarily extend to completion of secondary school.[30] Early marriage and the low likelihood of adult women participating in the labour force reduce the perceived economic value of girls' education, and there is a view – though it is unclear how widespread it is – in some communities that girls who are educated are less valued as wives, mothers, and caregivers.[31] Somali households face inherent trade-offs when deciding how to

[30] This finding from the first phase of EGEP is related to findings that we document in greater detail below regarding support for girls' education in the abstract versus in practical terms. Importantly, while both the EGEP endline evaluation and this baseline evaluation document strong support for girls' education in the abstract (e.g., when respondents are asked whether they think girls should go to school), they express much lower levels of support when faced with less abstract questions (e.g., when asked to choose between household consumption goods and girls' education). We discuss this point in greater detail in Section 5.5 of this report.

[31] According to several interviewees, some members of their community undervalue girls' education because girls often do not use their schooling in the workforce, and their future roles are primarily as wives and mothers (see, e.g.: KII with female teacher, Galmudug; FGD with mothers, Somaliland; FGD with boys, Galmudug). However, interviewees were generally describing the attitudes of other community members, so it is unclear the extent to which these attitudes are actually widespread.

FIGURE 2: EGEP-T THEORY OF CHANGE



allocate their scarce resources; if girls' education is systematically undervalued, either within their household or within their community more broadly, it is less likely that parents will invest in educating their girls. Further, if the community does not value girls' completion of the full cycle of education, girls themselves may feel it is less important; the community may also erect actual barriers to girls' progression through school, either consciously or unconsciously.

In an effort to change community attitudes and promote behaviour in favour of girls' completion of education, EGEP-T will engage in a broad campaign targeting community attitudes by promoting community dialogues, workshops with men and boys in which the value of girls' education is promoted, community mobilisation events, outreach efforts through radio and loudspeakers, the promotion of messages through banners, t-shirts, etc., amongst other approaches. EGEP-T will also conduct capacity-building efforts with CECs, to increase their ability to fund bursaries for girls, and to promote girls' education among CEC members themselves.

External Factors: Drought

At the time of the baseline evaluation, Somalia was in the midst of a widespread drought. Between November 2016 and May 2017, an estimated 548,000 people were displaced as a result of drought, which affected Somaliland and Puntland – among areas where EGEP-T is being implemented – particularly severely but had widespread impacts across the country.[32] Since that time, the drought has ended in some areas, been temporarily relieved in other, and continued unabated in still others. As noted with regard to the project's context, drought is regional and even local, and pockets of severe drought remain, even if the scale of the emergency has subsided.

Given the potential importance of drought in shaping project outcomes, EGEP-T took stock of the situation at baseline and integrated drought response activities into its programming. At the time of baseline data collection, 30 of the 140 sample schools were rated as severely affected by drought, according to RI's Monitoring & Evaluation Team, while a further 7 schools were affected by major drought. As a result, EGEP-T has incorporated drought response activities into its programming – at the time of baseline data collection, this included the provision of food rations to both teachers and children, school grants, teaching and learning materials, psychosocial support, and other interventions designed to ameliorate the impact of the drought.

1.3 Project Beneficiaries

Beneficiary Counts

At the time of the baseline, EGEP-T was slated for implementation in 172 primary and 56 secondary schools over the next three years.[33] RI has estimated the number of EGEP-T beneficiaries based on school enrolment data collected from project schools. In total, the project expects to reach 30,100 direct beneficiaries overall, and a total of 89,212 indirect beneficiaries, with indirect beneficiaries defined as students enrolled in EGEP-T project schools, but not receiving direct support in the form of bursaries, solar lamps, etc. The breakdown of beneficiaries by gender and direct/indirect status is provided in the table below.

[32] US Department of State, Humanitarian Information Unit. May 5, 2017. "Somalia: Drought and Displacement." Available at: https://reliefweb.int/map/somalia/somalia-drought-and-displacement-5-may-2017.
[33] 17 targeted schools combine primary and secondary school, and are therefore counted in both categories.

TABLE 1: ESTIMATED BENEFICIARIES OF EGEP-T PROJECT

Group	Expected Beneficiaries
Direct Beneficiaries – Girls	
Primary School Girls	19,178
Secondary School Girls	10,922
Indirect Beneficiaries – Girls	
Primary School Girls	25,586
Secondary School Girls	0
Indirect Beneficiaries - Boys	
Primary School Boys	46,357
Secondary School Boys	17,269

Beneficiary Types and Interventions

EGEP-T includes a range of interventions tailored to different beneficiary groups. Table 2, below, describes the structure of EGEP-T in relation to each target group. The rightmost column in the table lists the types of support that beneficiaries in each category will receive. Note that Table 2 includes direct and indirect beneficiaries as defined and targeted by EGEP-T. However, it does not describe a broader group of potential indirect beneficiaries – not included in the project's own calculation of beneficiary numbers – who benefit from, for instance, community-level attitudinal and behavioural change campaigns but who either do not attend school or attend a non-project school in the same community.

Disaggregating the expected beneficiary totals presented in the previous section, RI expects to impact 2,628 severely marginalised girls in primary school, and 1,310 severely marginalised girls in secondary school. Both sets of beneficiaries are included in the total reported figure of 30,100 direct beneficiaries.

Table 2 describes specific beneficiary groups, according to their enrolment status, grade, and the urbanicity of their residence. The rightmost column indicates what types of support for which each group is eligible.

TABLE 2: TYPES OF TARGETED SUPPORT, BY BENEFICIARY GROUP AND MARGINALISATION

Enrolment Status	Location	Grade	Support
Severely Marginalised Girls			
In-school	Urban & Rural	G2-G5	• Bursary support • School uniforms • Payment of school fees or cash grants • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
In-school	Urban	G6-G8	• Bursary support • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes

			• Teaching and learning materials • Teacher training
In-school	Rural	G6-G8	• Bursary support • Solar lamps • Sanitary kits • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
In-school	IDP Camp	G6-G8	• Solar lamps • Sanitary kits • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
Out-of-school	Urban	F1-F4	• Bursary support • Community back-to-school campaigns • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
In-school	Urban	F1-F4	• Bursary support • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
In-school	Rural	F1-F4	• Bursary support • Solar lamps • Sanitary kits • Teacher training in inclusive education • Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
		Marginalised Girls	
In-school	Urban & Rural	G6-Form 4	• Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher training
		Marginalised Boys	
In-school	Urban & Rural	G6-Form 4	• Participation in boys' clubs • Remedial classes

	• Teaching and learning materials • Teacher training

Table 2 includes a classification of target groups by their relative level of marginalisation. Within the beneficiary population, individuals fall into one of two categories: marginalised and severely marginalised. RI considers all girls above Grade 6 in EGEP-T target schools to be marginalised. Per the Monitoring, Evaluation, and Learning (MEL) Framework, they note significant pressures placed on girls to engage in unpaid domestic work or paid work in lieu of schooling, the uniformly poor quality of teaching, the often-realized threat of violence within and around project locations, and a fractured and weak educational system. Given Somalia's current conflict situation, as well as the ongoing drought and the structural challenges that face pre-adolescent and adolescent girls in the region, considering all girls in this group marginalised is justified.

Beneficiaries in the Midline Evaluation

Relative to the plan established prior to the baseline evaluation, no major changes have occurred to the nature or number of beneficiaries that EGEP-T seeks to reach. Although the structure of the cohort and school sample have changed significantly from the baseline, this issue is addressed in more detail in the methodological discussion below. For the purposes of targeting programmatic interventions, no meaningful changes have occurred.

Throughout this report, we refer to cohort girls, bursary girls, and cohort boys by the grade in which they were enrolled at baseline. Given that no girls who were out-of-school at baseline are being systematically tracked as part of the learning cohort, this approach is straightforward, because every girl in the learning cohort has a known baseline grade. The table below relates baseline grade levels and ages to expected grade levels and age ranges of girls and boys in this midline report. Note that data collection for the midline evaluation took place around 18 months after the baseline, so age ranges are given rather than precise ages, as some children may have had two birthdays since the baseline. Such uncertainty does not apply to grade levels, because the baseline took place shortly after the 2017-2018 school year began, and the midline took place near the end of the 2018-2019 school year, so most students should have progressed a single grade, though there are many exceptions in practice. At baseline, all cohort girls, bursary girls, and cohort boys were in grades 6 through Form 2, and ranged in age from 11 to 18 years.

TABLE 3: GRADES AND AGES OF COHORT AT BASELINE, AND EXPECTED VALUES AT MIDLINE

Characteristic	Value at Baseline	Expected Value at Midline
Grade	6	7
Grade	7	8
Grade	8	Form 1
Grade	Form 1	Form 2
Grade	Form 2	Form 3
Age	11	12-13
Age	12	13-14
Age	13	14-15
Age	14	15-16
Age	15	16-17
Age	16	17-18
Age	17	18-19
Age	18	19-20

2. Evaluation Methodology

The overall goal of the EGEP-T midline evaluation was to assess the project's impact from baseline to midline. However, the evaluation also has a number of ancillary goals: preparing for the most rigorous possible endline evaluation by making adjustments to the evaluation methodology at midline; testing the Theory of Change outlined by the project; and understanding which of the project's interventions are working and which are not, to help shape programming going forward. In this section of the report, we outline the evaluation questions the midline seeks to answer in the next section. Building on that overview, we discuss the evaluation's overall design, the sampling approach taken at baseline, and adjustments made to the methodology from the previous round, and describe the data collection tools employed. Next, we report on the replacement process, and the actual sample composition achieved at midline for the various population groups targeted. Finally, we discuss methodological limitations that present problems for the analysis, and present the results of analysis meant to shed light on these limitations and the best way forward.

In order to improve the readability of this section, methodological details on sample weighting, data collection tools, and fieldwork-related limitations have been moved to the methodological annex (Annex 18).

2.1 Key Evaluation Questions

The EGEP-T evaluation, as with other GEC-T evaluations, seeks to answer an overarching set of research questions. These questions are asked of every GEC-T project:

- Was the project successfully designed and implemented? Was the project good Value for Money?
- What impact did the project have on the transition of marginalised girls through education stages and their learning?
- What works to facilitate transition of marginalised girls Somalia/Somaliland through education stages and increase their learning?
- How sustainable were the activities funded by the project and was the project successful in leveraging additional interest and investment?

In addition to these broad questions of interest to the GEC-T program as a whole, the evaluation seeks to answer a more specific set of questions targeted to EGEP-T programming and the Somali educational context. These complementary questions are organized below, mapped to the myriad outcomes targeted by the project. Note that these evaluation questions – and the logframe indicators associated with them – have changed slightly from the baseline. Specifically, at midline, an additional system-level sustainability indicator was added; other adjustments impact measurement strategies only, and do not affect the overall research questions.

LEARNING

To what extent has the project impacted on numeracy and literacy levels of girls at upper primary and secondary level?

- What has been the difference in impact across the different marginalised groups?*
- To what extent has the project impacted on learning levels of girls in comparison to boys? And to what extent has the project impacted on boys in their own right?

TRANSITION

To what extent has the project impacted on transition rates of girls from upper primary to lower secondary school, and from lower to upper secondary school?

- What has been the difference in impact across the different marginalised groups?
- What are the factors that lead to 'successful' or 'unsuccessful' transition of girls?

SUSTAINABILITY – COMMUNITY LEVEL

What progress towards sustainability of impact and interventions has been achieved at the community level?

- To what extent has the project impacted on the opinions of community members in relation to girls' school completion? Amongst which groups of community members (mothers, fathers, teachers, girls, and boys) has the greatest attitudinal change been seen?
- To what extent has the project led to men and boys taking action to support girls in attending and completing school?
- To what extent are community leaders leading campaigns and advocacy events?

SUSTAINABILITY – SCHOOL LEVEL

What progress towards sustainability of impact and interventions has been achieved at the school level?

- To what extent are Community Education Committees providing bursary support to marginalised girls in EGEP target schools?
- To what extent are CECs in EGEP target schools providing match funding (financial or in-kind) for school improvement initiatives?
- To what extent is the teacher peer-mentoring programme operational in EGEP target schools?

SUSTAINABILITY – SYSTEM LEVEL

What progress towards sustainability of impact and interventions has been achieved at the institutional level?

- Do Federal level and Galmudug Ministries of Education respectively have Gender Strategies in place?
- To what extent are MoE officials conducting monitoring visits, and follow-up monitoring visits in EGEP target schools?
- Have TTIs (Teacher Training Institutes) integrated components of the CPD approach into their curriculum for teacher training?
- Do EGEP schools have functional child protection mechanisms?

INTERMEDIATE OUTCOME – ATTENDANCE

- To what extent has attendance improved in EGEP target schools over the life of the project?
- Amongst which groups of marginalised girls has attendance been most impacted?
- To what extent are improvements in attendance contributing to improvements in learning levels and transition rates?

INTERMEDIATE OUTCOME – EMPOWERMENT AND SELF-CONFIDENCE

- To what extent has the project positively impacted on girls' self-esteem?
- Has there been an increase in girls willing to take up leadership positions?
- To what extent are increases in self-esteem and empowerment of girls contributing to improvements in learning levels and transition rates?

INTERMEDIATE OUTCOME – TEACHING QUALITY

- To what extent are trained teachers demonstrating improved teaching practices, including gender-friendly teacher approaches?
- To what extent are improvements in teaching quality contributing to improvements in learning levels and transition rates?

INTERMEDIATE OUTCOME – SCHOOL MANAGEMENT AND GOVERNANCE

- To what extent are Community Education Committees contributing to effective management of EGEP target schools?
- To what extent are improvements in school management and institutional governance contributing to improvements in learning levels amongst girls in EGEP target schools?

INTERMEDIATE OUTCOME – COMMUNITY ATTITUDES AND BEHAVIOURAL CHANGE

- To what extent is the project positively impacting on parents' aspirations for level of schooling they hope their girls will reach?
- To what extent is the project positively impacting on community attitudinal and behavioural change in relation to girls' completion of school?

In the case of each outcome – especially learning, transition, and each intermediate outcome – the evaluation will attempt to consider the differential impact of the project on different subgroups of girls, boys, households, schools, and so forth. For instance, results specific to girls, such as learning, transition, and attendance, will be disaggregated to assess the impact of the project on girls who are drought-affected, those from more marginalized households, and those from female-headed households. Similarly, analysis at other levels – where the unit of analysis may be the teacher or the school – will be disaggregated by, for example, school characteristics. The goals of this analysis will be to determine whether the project is only benefitting those most or least vulnerable, and to look for barriers that might be hindering project impact among specific subgroups or in specific contexts.

2.2 Evaluation Design

This section provides an overview of the research design employed by EGEP-T for its evaluation. The first section describes the research design in broad terms, followed by a description of the different data collection tools, and the sampling approaches utilized for each.

Overall Research Design

Phase II of the Girls Education Challenge aims to build on lessons learned from Phase I in terms of both program implementation and evaluation. This evaluation was designed while incorporating general feedback from GEC Phase I, which applies to most contexts, and feedback specific to RI and Forcier's experience conducting GEC evaluations in Somalia.

The EGEP-T evaluation design deviates in important ways from the standard GEC-T design established and recommended by the Fund Manager (FM), a consortium led by PriceWatershouseCooper (PWC). The first major point of difference is the lack of a set of control or comparison schools in the evaluation. Due to security concerns, RI and its partners elected against including a set of comparison schools in the evaluation. The standard GEC-T evaluation design employs comparison schools and a difference-in-differences design; the lack of comparison schools means that the EGEP-T evaluation utilizes a strictly pre-post design – as opposed to difference-in-differences – in which changes are tracked in the treatment or intervention schools exclusively. This design choice has significant implications for the

types of analysis that can be expected at the midline and endline evaluations, as well the strength of any conclusions that can be drawn.

Second, the design diverges from the standard sampling approach recommended by the FM. The approach recommended by the FM has been labelled a “joint sampling” approach, in which the same students who complete learning assessments at schools are also included in the household survey sample. That is, after completing learning assessments at a sample school, enumerators follow up at the same children’s residences, recruiting their households into the household survey sample, including a survey module that applies specifically to children. Contrast this with the "split-sample approach"; using this sampling methodology, the "learning sample" and "household sample" are comprised of different respondents, with children sampled randomly from project schools and households sampled separately using a random walk strategy in communities that include program schools.

The EGEP-T evaluation employs a hybridized version of these two approaches. At baseline, a split-sample approach was taken, with separate samples of cohort children – recruited from within schools – and random households in the community, recruited via a random walk. The midline, however, reverted to something akin to the joint-sample approach: the same children were tracked from baseline, and now their heads of household and caregivers were contacted for the purposes of completing the household survey. In the interest of clarity, we describe the sampling approach employed for children at length here, because these design details impact the nature of the analysis that will be conducted at midline. Later in this report we discuss the sampling approach for specific data collection tools.

In the baseline evaluation, transition and learning outcomes were studied using two distinct samples. The main learning cohort, at baseline, was a stratified random sample of girls, drawn from the population of enrolled girls. That is, *every* girl selected into the main learning cohort was enrolled in and present at school at the time of the baseline evaluation. We refer here to the main learning cohort, but the baseline evaluation captured three distinct samples of children at schools:

- Cohort girls – grades 6 through Form 2
- Bursary girls – grades 6 through Form 2, receiving a bursary through EGEP-T
- Cohort boys – grades 6 through Form 2

The critical aspect of the sampling approach for the learning cohort is that children were sampled *at schools*. As a consequence, nearly every child in the three samples listed above had successfully transitioned from time *t-1* to time *t*, where time *t* represents the baseline.[34] Using these samples to establish baseline transition rates would be misleading, in that the samples consist entirely of children who have remained in school, rather than the overall population of school-age children.

To establish baseline transition rates for later benchmarking, the baseline evaluation also performed a random sample of households in the communities surrounding project schools. This clustered random sample targeted households with a female member aged 11-18 years. In total, the sample of 559 households captured 863 girls in the target age range, for the purposes of establishing pre-intervention transition rates in the community at large.

[34] Exceptions are children who were enrolled in school but had not properly progressed from one grade to another over the last year. These exceptions do not obviate the broader point, however, which is that transition rates among the learning cohorts described here were artificially high.

Naturally, the midline sampling strategy is different, and was developed to explicitly account for the approach taken at baseline and to draw conclusions regarding learning and transition outcomes that are as valid as possible under the circumstances. At midline, data collection will track cohorts of children from baseline, using a sample approach that is common to most or all GEC-T evaluations. At this stage, the cohorts constitute both a learning sample and a transition sample, which are slightly distinct from one another. The learning sample includes the following three groups of children from the baseline:

- Cohort girls
- Bursary girls
- Cohort boys

All three groups were sampled at school during the baseline. For the midline, attempts were made to re-contact every child in these groups. Children who had dropped out of school since the baseline or who could not be located were replaced by drawing a random replacement child from the same grade and child "type" at the school.[35] The outlined approach ensures a learning sample that is stable over time: children who remain in-school will form a true panel sample; children who have dropped out of school will be replaced by in-school children, such that the baseline and midline *learning* samples both consist entirely of children who were enrolled – and therefore exposed to the program's interventions – at the time of data collection.

The transition sample takes a different approach, naturally, because dropping out of school constitutes a part of the transition outcome, and replacing children who have dropped out with in-school children is not sensible when measuring transition outcomes. Transition outcomes were tracked exclusively among cohort girls and bursary girls. For the purposes of measuring transition from baseline to midline, all cohort and bursary girls from the baseline were re-contacted and household surveys were conducted with them, even if they had dropped out of school since baseline. Transition outcomes at the midline are defined among the group of girls who were in-school at baseline and were successfully located and re-contacted at midline.

As noted in the case of the learning sample, girls who were no longer in-school were replaced – for the purposes of assessing learning – with girls drawn from within the school. These girls also constitute a "top-up" sample for measuring transition rates from midline to endline.

2.3 Replacements and Achieved Sample

Given the pre-post research design described above, the evaluation's rigor depends heavily on the ability to maintain comparability of the samples between data collection rounds. To draw valid inferences regarding project impact requires assumptions about the nature of change in project outcomes, such as learning and transition rates, in the absence of intervention. But it also requires that the baseline and midline samples be strictly comparable.

Two factors combine to influence comparability from baseline to midline. The first is the sample of schools included in each round of data collection. At midline, we sought to revisit every school included in the baseline sample. Owing to the exclusion of schools in Banadir, this was not possible; however, we sought to revisit every other school (n = 119) from the baseline evaluation, across four

[35] By child "type", we mean cohort girls, cohort boys, and bursary girls (e.g., a replacement for a bursary girl must be comparable to the replaced girl in terms of her grade *and* bursary recipient status).

zones. Regarding the school sample, schools may not be eligible for the midline evaluation because the project ceased working with the school in question, and schools may be inaccessible for security or other reasons.

The second factor is the sample of children successfully re-contacted. Within each school, we sought to re-contact the same cohort girls, bursary girls, cohort boys, and out-of-school girls. However, some children were no longer eligible for inclusion (e.g., learning sample cohort girls who had dropped out at the time of the baseline) and had to be replaced by new girls; other children were not available for interviewing, could not be located, or refused to participate. To the extent that children are replaced, the comparability of the sample from baseline to midline is affected; when entire schools are replaced, it means that children are replaced en masse, which can have particularly large effects on comparability.

When schools required replacement, we selected a replacement school in consultation with RI's monitoring and evaluation team. The goal, as with replacement of schools during baseline data collection, was to maintain the sample composition by replacing schools with schools that were similar on observable characteristics. Specifically, replacement schools were selected within the same geographic zone (Somaliland, Puntland, and Galmudug/Hirshabelle, respectively). Replacement schools were from the same school level, primary or secondary. Beyond these two considerations, we attempted to select schools in the same region and district as the school being replaced, while simultaneously ensuring the school had a sufficient number of students to serve as a viable replacement (i.e. to have 12 cohort girls, multiple bursary girls, etc.). Given these varied constraints, our focus in selecting replacement schools was on selecting schools that maintain geographic and grade-level balance.

In total, two baseline schools were replaced at the midline. The schools, their replacements, and the basic characteristics of each, are described in the table below.

TABLE 4: REPLACEMENT SCHOOLS IN MIDLINE SAMPLE

Replaced School				Replacement School			
School	**Zone**	**District**	**Setting**	**School**	**Zone**	**District**	**Setting**
Geerisa Primary	Somaliland	Awdal	Rural	Agbar Primary	Somaliland	Maroodi Jeeh	Rural
Omar Samatar Secondary	Puntland	North Mudug	Urban	Bareeda Secondary	Puntland	Gardafu	Urban

The first problematic school encountered was Geerisa Primary School in Awdal, Somaliland. When our teams arrived early in April, the school was closed due to a widespread illness (fever) in the community and among the students. The team moved forward with their fieldwork schedule, planning to return to the school nearer the end of fieldwork. When our team leader contacted the head teacher approximately two weeks later, they were told that the school was still closed. While we briefly discussed conducting learning assessments and other surveys strictly in students' homes, this posed an unnecessary risk to the field team and to the community, as visiting multiple households could serve to spread the illness between families. As a result, a replacement school was chosen.

The second school requiring replacement was Omar Samatar Secondary School in Puntland. Fieldwork for this school ran into examination periods that occur in late April and early May. When the team arrived at the school, students were in the middle of several days of examination review, and were then scheduled to complete several days of examinations, at which point they would not return to

school. Waiting for the completion of the examinations would have required a delay of over one week for the team, and data collection would have been sub-optimal due to the lack of school-level data collection from teacher surveys, classroom observations, and headcounts. A replacement school was chosen.

In both cases, the zonal balance of the sample was maintained from baseline to midline. In addition, the replacement schools matched the school level and urbanity of the schools being replaced. However, the replacements were both drawn from different regions than the schools they were replacing, owing to the small number of available replacement schools.[36]

Beyond school-level replacement, a number of cohort girls, bursary girls, and cohort boys were replaced during the midline. Note that, in cases where an out-of-school girl from baseline could not be located, no replacement out-of-school girl was selected. In total, the target sample of cohort girls at midline was 1,370, which excludes all cohort girls from Banadir. Overall, 1,128 cohort girls were successfully re-contacted, a re-contact rate of 82.3 percent. Importantly, this includes cohort girls who were re-contacted but had dropped out of school, and were tracked exclusively for the transition sample but replaced in the learning sample. Just 996 cohort girls were successfully re-contacted *and* remained enrolled, meaning that the "true panel" of learning girls – those who completed learning assessments at both baseline and midline – was just 996 girls of the original 1,370 targeted (72.7 percent re-contact rate). The table below reports the re-contact rate for different cohorts of children. In Section 2.5, we provide extensive analysis of replacement rates and the predictors of replacement, to investigate the methodological and inferential consequences of replacement.

TABLE 5: RE-CONTACT AND REPLACEMENT RATES, BY COHORT OR RESPONDENT TYPE

Type of Child	Target Sample	Re-contact Rate	Replacement Rate Learning Sample
Cohort girls	1,370	82.3%	24.3%
Bursary girls	365	71.0%	19.2%
Cohort boys	337	58.8%	27.1%

The target sample sizes reported in the table above exclude Banadir schools, which accounts for the smaller sample sizes targeted at midline than analysed at baseline. Of the targeted respondents, the re-contact rate represents the share re-contacted in some form, even if they were replaced in the learning sample. For instance, consider a cohort girl who was located, but had dropped out of school prior to the midline. Such a girl would be interviewed for the purposes of analysing transition rates, and would count as re-contacted in the table above. However, she would not complete a learning assessment and would not be included in the learning sample, so she would count toward the replacements for the learning sample, in the final column.

Note that re-contact rates are lowest among cohort boys due to the nature of the tools/surveys applied to them. Cohort boys were not tracked for transition; therefore, if they were not located at the school and it was confirmed that they had dropped out of school, they were not located or re-contacted at all. This accounts for the much lower re-contact rate among cohort boys, even though

[36] The baseline sample included approximately 80 percent of all schools in the sample frame, limiting the number of possible replacements. This issue was exacerbated by the need for replacements at baseline, as several schools were replaced during baseline data collection.

their replacement rate within the learning sample is not dramatically different from that of cohort girls.

The table below reports the target and achieved midline sample sizes, excluding schools in Banadir, for each child respondent group – cohort girls, bursary girls, cohort boys, and out-of-school girls. The target sample size at midline was exactly equal to the achieved baseline sample size, excluding Banadir, because the midline goal was to re-contact every child from the baseline. The right-most two columns report the size of the "true panel" at midline only and the size of the cross-sectional sample at midline. To be clear, the "true panel" of cohort girls includes 996 girls who were contacted at both baseline and midline, meaning the total sample size for analysis is double that number – 1,992. The cross-sectional sample includes only girls who were either successfully re-contacted ("true panel") or replaced at midline. The only girls excluded from the cross-sectional sample are those for whom a replacement was not chosen, and those whose replacement was removed from the sample during cleaning.

TABLE 6: MIDLINE TARGET AND ACHIEVED SAMPLE SIZES

	Midline Target Sample Size	Midline Achieved Sample Size	Midline Panel Sample Size	Midline Cross-Sectional Sample Size
Cohort girls	1,370	1,346	996	1,316
Bursary girls	365	348	274	339
Cohort boys	337	329	234	318
Out-of-school girls	114	64	64	64

The table below reports the baseline and midline achieved sample sizes for the remaining data collection tools. We do not include Banadir in the baseline counts of observations. Note that, in most cases, no explicit targets were set for the midline sample, beyond sampling instructions. For instance, teams were required to complete one headcount per grade level, but no explicit target number of headcounts in the total sample was set, as it is not known, *a priori*, how many grade levels exist in each school. In the right two columns, we report the baseline and midline "comparable sample" sizes – the number of observations completed in schools that were included in both baseline and midline. Excluded from the comparable sample are the two replaced schools described above (surveyed only at baseline), and the two replacement schools chosen at midline (surveyed only at midline). The "comparable sample" forms the backbone of our analysis for each of these data collection tools, to ensure that our inferences are not influenced by sampling variation at the school level.

TABLE 7: SAMPLE SIZES FOR ADDITIONAL DATA COLLECTION TOOLS, BY ROUND

Data Collection Tool	Baseline Sample	Midline Sample	Baseline – Comparable Sample	Midline – Comparable Sample
Head Teacher Surveys	119	112	110	110
Teacher Surveys	422	516	399	483
Mentor Surveys	N/A	171	N/A	N/A
Headcounts	781	755	722	714
Classroom Observations	223	215	205	205

2.4 Replacement Analysis

A key methodological consideration in all GEC-T evaluations is the tracking of children -- i.e. cohort girls, as well as other cohorts of respondents – over time from one evaluation round to another. Accurate tracking is essential for the validity of GEC-T evaluations, because most of the evaluations rely on panel analysis to draw valid inferences about project impact. EGEP-T is no exception, and significant time and energy was invested at the baseline to record contact information necessary to re-contact girls at the midline; likewise, significant time and effort was dedicated at midline to locating cohort girls, bursary girls, and cohort boys from the baseline, in an effort to reduce panel attrition. As the baseline evaluation report made clear, panel attrition is one of the most important threats to the validity of the EGEP-T evaluation overall.

In this section, we build on the discussion in Section 2.3, above, which described the target and achieved sample sizes for different respondent cohorts. As we emphasized in that discussion, successful re-contact rates varied across cohort types. To complicate matters further, it is important to note that replacement rates also vary across learning and transition samples. We unpack this distinction below, and provide analysis of panel attrition. The goal of this analysis is to investigate whether attrition is "as-if random" or if attrition versus successful re-contact is correlated with a respondent's characteristics. To the extent that attrition is non-random, it has the potential to undermine the quality of our findings.

In the sections of that follow, we investigate the effects of several distinct forms of panel attrition or replacement:

- Attrition due to large-scale sampling changes, i.e. the removal of Banadir from the sample
- Attrition due to replacing entire schools, in cases where a school is inaccessible
- Individual-level attrition, i.e. cases in which a school is followed up at midline but an individual cannot be located or is replaced

We do not address attrition due to school replacement in detail, because only two schools were replaced at midline. While this constitutes a non-trivial share of all girls replaced at midline, we do not believe attrition at the school level was related to underlying school or girl characteristics.[37] The last category is particularly important, and can be decomposed into two primary forms of attrition: cases in which a respondent cannot be located, and cases in which a respondent was located but has dropped out of school and is no longer eligible for inclusion in the learning analysis. We discuss and investigate each of these issues separately below.

Attrition due to changes in sample structure

The most important sampling change that occurred from baseline to midline was the removal of all schools in Banadir (Mogadishu) from the set of schools to be visited at midline. We discuss this issue in greater detail in Section 3, where we show how the removal of Banadir schools influenced the composition of the sample from baseline to midline.

[37] As discussed in the previous section, one school was replaced due to a long-term and ongoing outbreak of illness in the community; the second school was replaced due to the timing of fieldwork vis-à-vis the examination schedule.

The table below briefly illustrates the impact of the removal of Banadir on sample composition. Our goal here is simply to emphasize that cohort girls in Banadir were not removed "as-if randomly"; in short, they are systematically different in many ways from the typical cohort girl in the remainder of the sample. For instance, no cohort girls in Banadir are defined as attending rural schools, compared to 27.6 percent of cohort girls in the overall baseline sample. The final column of the table shows the share of cohort girls who are considered rural in the *remainder* of the sample, indicating that the relative urbanicity of the sample has changed.

TABLE 8: DIFFERENCES IN SAMPLE COMPOSITION AS A RESULT OF BANADIR EXCLUSION

Subgroup	Banadir	Baseline Sample	Midline Sample
Rural	0.0%	27.6%	32.4%
Age	13.9 years	14.6 years	14.7 years
Female-headed households	52.9%	43.7%	42.1%
Head of household has no education	8.3%	24.3%	27.1%
Head of household has only religious education	32.2%	19.8%	17.7%
Head of household completed at least primary school	47.4%	38.9%	37.2%
Somali literacy score	82.9	77.3	76.4
English literacy score	58.6	41.4	38.4
Numeracy score	77.5	71.6	70.6
Transition rate, benchmark sample	76.2%	73.6%	73.2%

The biggest shifts that take place as a result of excluding Banadir concern household characteristics and baseline learning scores. Girls in Banadir were less likely to come from households where the head of household had not completed any formal, non-religious, education. Even more importantly, cohort girls in Banadir outperformed their peers in terms of learning on all three subjects evaluated at baseline, despite the fact that they were younger than the sample average in the other zones. Benchmark transition rates were also higher in Banadir than the sample average. Both of these findings have consequences for the comparison of baseline to midline outcomes, which necessitates the removal of Banadir from the baseline samples that we use in this report.[38]

Predictors of successful individual re-contact

We now turn to analysis of individual-level attrition. Note that our analysis of attrition, re-contact and replacement also excludes Banadir. Our interest is in determining whether individual-level attrition is

[38] The exclusion of Banadir has obvious consequences for the ability of the project to meet targets, established at baseline, for the midline evaluation. Excluding Banadir from the midline may make some midline targets impossible to meet, while making others easier to meet, depending on whether Banadir over- or under-performed the remaining geographic regions. To ensure the project is assessed against fair targets at midline, we have re-calculated the baseline values for each outcome and intermediate outcome, excluding Banadir from the sample; we then re-calculated midline targets on the basis of those revised baseline levels. We did not make adjustments to the *magnitude of change* expected from baseline to midline, applying the exact same expected change, simply adjusting the baseline levels.

related to girl- and, to a lesser extent, community-level characteristics. In contrast, decisions made at the project level (the exclusion of Banadir), and decisions made in light of on-the-ground accessibility (the replacement of two schools at midline) are not a function of respondent characteristics.

This section focuses on successful re-contact as an outcome. We define children as eligible for this analysis if we sought to re-contact them at midline and visited the schools in which they were located at baseline. We include cohort girls, bursary girls, and cohort boys in our analysis, though cohort boys are excluded from some of the multivariate analyses we perform, because the baseline did not capture the level of detail about them as it did about cohort girls and bursary girls.[39] Children who were contacted at baseline but removed from the data during the baseline evaluation are not included. Children in schools that were replaced wholesale or excluded at midline are also not included.

We define successful re-contact as the ability to locate and collect data from the respondent. Our definition includes children who were re-contacted and remained in the learning sample (i.e. children who were re-contacted and remain enrolled in a project school at midline); it also includes children who were re-contacted exclusively as part of the transition sample (i.e. children who were no longer enrolled in a project school but who were successfully re-contacted and surveyed to track transition rates).[40] Unsuccessful re-contacts occur for any number of reasons, but include children whose location is unknown altogether, and children who have moved out of the community. Note that unsuccessful re-contacts may or may not have been replaced in the learning sample or the forward-looking transition sample; for the purpose of this analysis, replacement is not an important criterion.[41]

Overall re-contact as an outcome is directly related to maintenance of the transition sample. While girls in the learning sample were replaced even in cases where they were located (if they had dropped out of school), this is not true for the transition analysis. Our analysis of overall re-contact is *identical* to a focus on attrition from the transition sample between baseline and midline and thus has implications for transition findings throughout this report.

The first finding to arise concerns differential attrition across cohort groups. Cohort boys are much less likely to be re-contacted (70.7 percent) than either cohort girls (83.8 percent) or bursary girls (84.9 percent). Cohort boys are less likely to be re-contacted successfully for two straightforward reasons: first, less information was collected about cohort boys at baseline, as they were given a shorter survey at that time. Second, and more importantly, cohort boys were not being tracked for transition outcomes. For other respondents who were not found at the project school, their transition rates were being tracked, which necessitated follow-up at their household. Cohort boys, on the other hand, were relevant only to the learning sample, and were not tracked to their households in the event they could not be located at the school. Due to this dramatic difference between cohort boys and other respondent types, the remainder of our analysis focuses on cohort girls and bursary girls exclusively.

Beyond respondent type, geographic location, age, and grade level at baseline are important factors in shaping re-contact rates. In terms of zone, re-contact rates were lowest in Puntland, where just

[39] The total sample analyzed for re-contact includes 2,035 respondents: 1,346 cohort girls, 358 bursary girls, and 331 cohort boys.

[40] Importantly, this definition is distinct from the maintenance of the learning sample, which we analyze in the next section. Many children were successfully re-contacted but removed from the learning sample due to the fact that they were no longer enrolled in a project school. The outcome in this section is logically and methodologically prior to the maintenance of the panel learning sample, because it involves the more basic task of locating and contacting children from the baseline evaluation at midline.

[41] While all cohort girls, bursary girls, and cohort boys were replaced in the learning sample if they were no longer enrolled in school, eligible replacements were not always available.

81.9 percent of cohort and bursary girls were re-contacted, compared to 86.4 and 85.1 percent in Somaliland and Galmudug, respectively. Age is also a key predictor of re-contact rates: among girls aged 11-12 years at baseline, 90.5 percent were successfully re-contacted at midline. Meanwhile, among girls aged 17-18 years at baseline, re-contact rates declined to 78.1 percent.

Finally, arguably the most important predictor of successful re-contact, here or in the multivariate analysis that follows, is grade level at baseline. Girls in secondary school – Forms 1 or 2 – at baseline are slightly less likely to be re-contacted than girls in grades 6 or 7, at 85.2 percent versus 89.4 percent. Grade 8 girls represent a special case, though – re-contact rates among this group are a dismal 71.6 percent. How can we explain this finding? Simply put, many girls enrolled in grade 8 at baseline will drop out rather than transfer to a secondary school, at which point the primary school they were attending may lose track of their location; in addition – and likely more important as an explanation – grade 8 girls who move into secondary school may do so in a new location. That is, girls who move into secondary school may move to live with extended family to attend school elsewhere, or attend a boarding school. Once girls leave primary school, teachers and school staff may lose track of their location – especially in urban areas, where community social cohesion is lower – and our ability to locate them is reduced as a result.[42]

Thus far, our focus has been on broad characteristics that are related to unsuccessful re-contact. To expand on this discussion, we performed a more exhaustive multivariate analysis of re-contact outcomes. We estimated a linear (OLS) regression model, predicting successful re-contact among the aforementioned sample of cohort and bursary girls. We experimented with alternative specifications, but our main model incorporates binary variables capturing geographic zone, and grade level, in line with the findings reported above. We also incorporate variables capturing respondent type (bursary girl versus cohort girl) and rural location. There are good theoretical reasons to expect both drought and conflict to influence re-contact outcomes, because they should influence enrolment rates and can cause individuals to migrate away from their previous communities in search of economic opportunities. We code communities that were affected by either drought or conflict at midline based on data provided by Relief International.

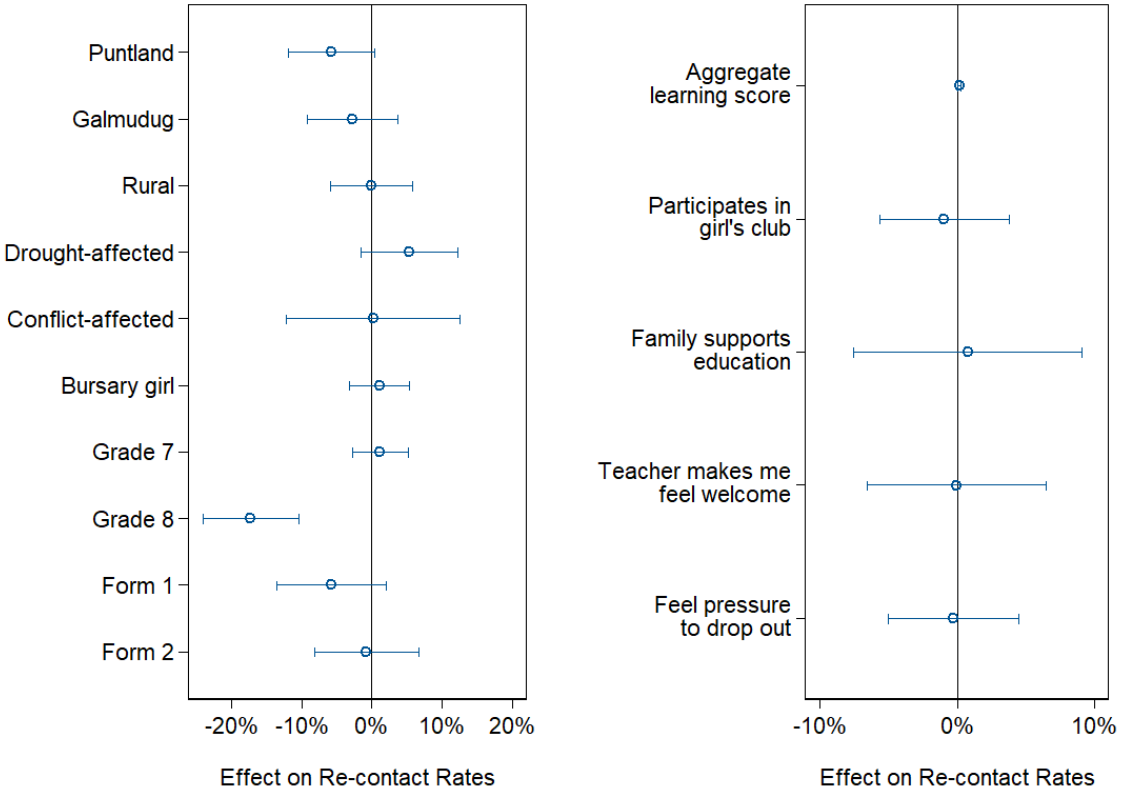
The results of this model are reported in the left panel of the figure below. The association between each variable and successful re-contact is plotted with a hollow dot representing the regression coefficient and the horizontal bars signifying the 95 percent confidence interval around each coefficient.[43] To interpret these results, variables whose 95 percent confidence does not overlap with the vertical line at "0%" are statistically significant at the 5 percent level. Notably, Puntland is associated with a 5.7 point decline in re-contact rates, even after controlling for other factors, and this relationship is marginally significant ($p = 0.07$). As the left panel shows, grade 8 girls are still extreme

[42] Of course, contact information was collected for all girls at baseline. However, because the baseline evaluation did not include household surveys with cohort girls, the precise location of their households vis-à-vis the school was not known prior to midline data collection. In practice, based on reports from team leaders, head teachers, teachers, and girls in the school were the most important sources of information on a girl's current whereabouts, but these sources are less useful if a girl has changed schools or moved away.

[43] A regression coefficient in this context represents the change in the likelihood of successful re-contact (on a 0 to 100 scale) associated with a 1-unit change in the variable written along the y-axis, accounting for the influence of all other variables in the model. For instance, the coefficient for Puntland represents the change in re-contact rates predicted for a girl in Puntland versus Somaliland, *holding all else equal*. Note that the comparison of Puntland to Somaliland is implicit in the model specification, because sets of dummy variables always require one category to be omitted from the model. We selected Somaliland as the reference category for zone, such that the effect of living in Puntland or Galmudug on re-contact rates is reported *relative to re-contact rates in Somaliland.*

outliers, with re-contact rates 17.2 points below those of grade 6 girls, and significantly lower than any other grade level studied.[44]

FIGURE 3: BASELINE CHARACTERISTICS AND SUCCESSFUL RE-CONTACT IN MIDLINE TRANSITION SAMPLE



The right panel expands our analysis by reporting the results of additional variables, incorporated into a more saturated model of re-contact rates. The model includes all of the variables reported in the left panel, meaning that these results control for zone, grade, respondent type, and other girl- and community-level characteristics. We expected girls who reported participation in a girl's club at baseline, and who felt welcomed by their teacher – to cite two examples – might be more likely to remain enrolled and able to be re-contacted at midline. However, this does not appear to be the case, after controlling for other factors that can influence re-contact outcomes.

The exception, though difficult to see in this figure, is learning: girls who achieved higher learning scores at baseline are more likely to be re-contacted at midline. We constructed an aggregate learning score for this analysis, which averages numeracy, Somali literacy, and English literacy scores on a 100-point scale. A 10-point change in mean learning is associated with a 1.7 percentage point change in

[44] Alert readers may notice that the 95 percent confidence interval for grade 8 girls overlaps that of Form 1 girls. However, the gap between these two groups is, nonetheless, statistically significant: 95 percent confidence intervals are calculated relative to a null hypothesis that a given variable is equal to zero; when we compare two coefficients in the same model using an F-test, we take advantage of their pooled sample size, which reduces the width of confidence intervals. In such an F-test, the disparity between grade 8 and Form 1 girls is significant with $p = 0.02$.

re-contact rates. While this shift in re-contact rates is not extreme, it could signal that girls who fall out of the panel are not random – they tend to be those who underperform in school.[45]

Non-random panel attrition can have far-reaching consequences. If girls who underperformed at baseline are more likely to fall out of the panel at midline, this has obvious implications for the analysis of learning scores over time. We turn to a discussion of the learning sample in the next section. But the impact is also felt in the case of the transition sample: if successful re-contact is non-random, it implies that girls who dropped out of school but are successfully tracked for transition may be systematically different from girls who dropped out but cannot be located.

Unfortunately, while we can study learning outcomes at baseline according to the re-contact status of girls at midline, this is not possible for transition, because transition as an outcome was not captured for cohort girls at baseline.[46] The fact that grade 8 girls and girls in Puntland have structurally different re-contact rates, coupled with the observed correlation between learning scores and successful re-contact, suggests that girls who are successfully re-contacted may differ from those who fall out of the transition sample.

We employ two additional sources of data to shed light on this possibility. First, we look at transition-adjacent baseline characteristics of girls as a predictor successful re-contact. The table below reports the successful re-contact rate among girls as a function of three baseline characteristics we would expect to predict attrition. Girls who were not enrolled *in the year prior to the baseline* are more likely to fall out of the panel at midline, as are girls who worked outside their home in the past. Finally, girls who experienced a serious illness in the year prior to the baseline are slightly less likely to be successfully re-contacted at midline.

TABLE 9: BASELINE CHARACTERISTICS AND SUCCESSFUL RE-CONTACT IN MIDLINE TRANSITION SAMPLE

Subgroup of baseline girls	Successful Re-contact rate
Overall cohort and bursary girl sample	84.0%
Was not enrolled in school last year	72.4%
Has worked outside home in past	71.4%
Had serious illness in last year	81.3%

The takeaway from the table is that attrition is non-random. Girls who drop out of the sample at midline are those who were more tenuously enrolled in the past. They may, therefore, be girls who are more likely to drop out in the future – if attrition is systematically correlated with dropping out or failing to progress into the next grade, it will produce bias in our estimates of midline transition rates.

The second additional data source is drawn from school reports completed by team leaders during midline data collection. Field teams who were unable to re-contact a girl from the baseline were asked to gather as much information as possible about the girl's current enrolment status, perhaps from former neighbours or from family members. The data collected should be interpreted with extreme caution – it is not systematic and it may not be reliable. Nonetheless, the results are telling: among girls who were successfully located but who are not enrolled in the project school, enrolment rates

[45] The correlation between learning scores and re-contact rates, though it is not clear from the figure, is significant with $p < 0.01$.

[46] All cohort girls were selected from within schools at baseline, implying transition rates of, effectively, 100 percent.

are 82.9 percent. In contrast, among girls who could not be located, reported enrolment rates are just 41.7 percent.

Of course, this data is extremely incomplete – for instance, we cannot reliably ascertain successful transition without reliable information about grade level, and there are many girls who could not be located *and* for whom we could not collect enrolment information from those who know them. Despite these caveats, the results confirm our concern regarding differential attrition from the transition sample – girls who could not be re-contacted appear to be less likely to transition successfully than even those girls who dropped out but who could be located.

Predictors of re-contact and replacement, learning sample

The analysis above focused on re-contact itself – even if a child was replaced in the learning sample – as an outcome. In this section we re-focus our approach, drilling down to study predictors of re-contact and replacement in the learning sample. Here, at the risk of generating confusion, we alter the outcome of interest slightly. We are now interested not in successful re-contact, but in replacement in the learning sample. Re-contact is a prerequisite for a girl to remain in the learning sample at midline and, as we documented above, re-contact appears to be related to learning performance at baseline.

In this analysis, we are interested exclusively in the learning sample. Our starting point is every cohort and bursary girl from the baseline who was sought for interviewing during the midline. The goal at midline was to maintain the integrity of the *panel*, i.e. the set of girls who are included in both rounds of data collection. Girls who could not be re-contacted were replaced, as were re-contacted girls who had dropped out between baseline and midline. Girls who remain in the learning panel are those who were both successfully re-contacted and remained enrolled at midline. We study predictors of replacement in the midline learning sample. If girls fall out of the learning sample non-randomly – i.e. if panel attrition is correlated with learning or outcomes predictive of learning – it raises concerns about the validity of our main learning results.

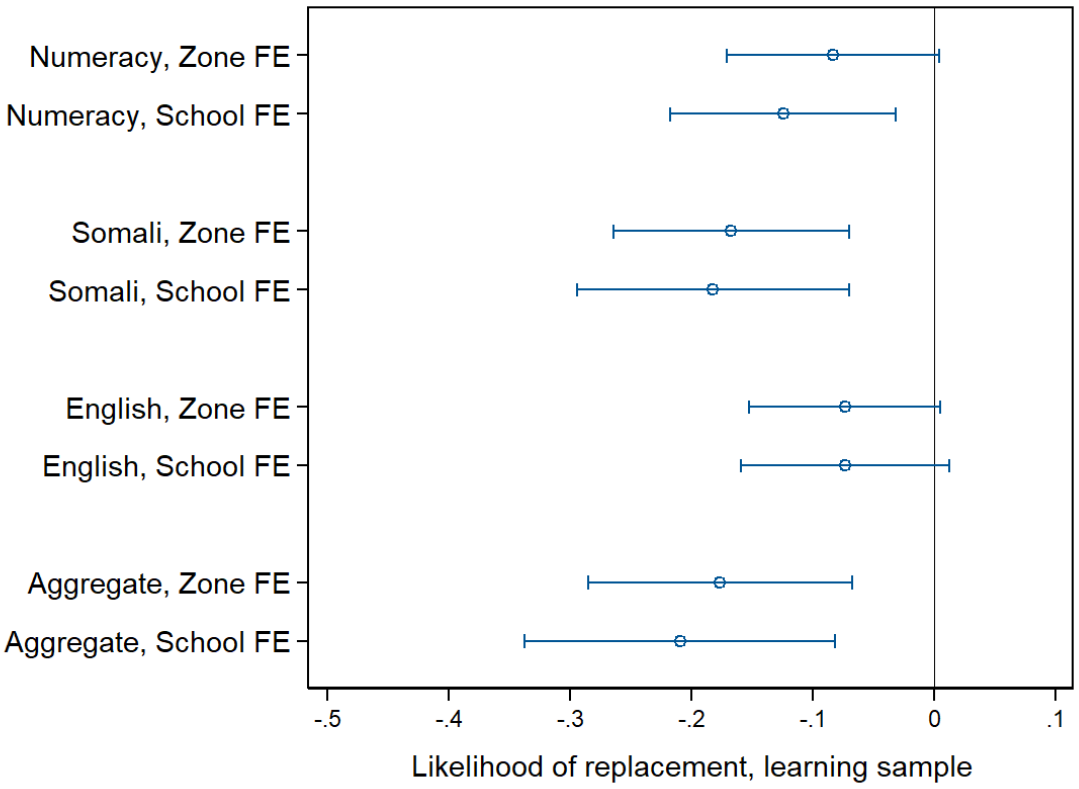
In total, 469 children were replaced in the learning sample at midline, including cohort girls, bursary girls, and cohort boys, representing a replacement rate of 23.1 percent overall. Replacement rates are highest among cohort boys, at 25.4 percent, and lowest among bursary girls, 18.2 percent of whom were replaced in the learning sample. Replacement rates were slightly higher in Puntland than other zones, but not dramatically so. Unsurprisingly, replacement rates are positively correlated with age – as age increases, girls are more likely to drop out and, subsequently, to be replaced. And, as with our analysis of overall re-contact rates relevant to transition, replacement rates are considerably higher (50.0 percent) among girls in grade 8 at the baseline, compared to 13.5 percent among all other grades combined.

Our core interest is in whether replacement is correlated with learning outcomes. If girls with higher learning scores are less likely to be replaced, the replacement process itself will produce upward bias in our estimates of learning gains from baseline to midline. As before, we estimate a series of linear regression models, predicting in this instance replacement as an outcome. Our models control for zone, baseline grade level, and respondent type, because these factors are correlated with pre-existing learning outcomes *and* the likelihood of replacement.

Because our interest is in the role of learning scores on replacement only, we take an aggressive approach to modelling replacement. We employ, alternately, either zone fixed effects or school fixed effects, which capture all observable and unobservable characteristics of a given zone or school. For each learning subject, we estimate its impact on the likelihood of replacement. The results are

reported in the figure below, which plots the regression coefficient and its 95 percent confidence interval for learning scores from each regression. For instance, the first result is the predicted impact of a 1-unit increase in baseline numeracy on the likelihood of replacement at midline, while controlling for zone, grade, and respondent type. The second result is the same predicted impact of a 1-unit increase in baseline numeracy, while controlling for underlying differences in replacement rates and numeracy scores *in each school*. In both cases, increased numeracy scores are associated with a decline in the likelihood of replacement – children who perform better at baseline are more likely to remain in the panel at midline.

FIGURE 4: CORRELATION BETWEEN BASELINE LEARNING SCORES AND REPLACEMENT IN MIDLINE LEARNING SAMPLE



The finding reported for numeracy is consistent in the case of Somali and English literacy, with somewhat stronger effects reported for Somali literacy. The fact that all of our results trend in the same direction is, itself, suggestive: if we were capturing random noise in the case of numeracy, we would expect null effects of Somali and English literacy on replacement rates. Instead, we find a consistent story: girls who perform better are less likely to be replaced. The effects are considerable: for an otherwise identical girl, a 10 point increase in numeracy scores is associated with a 1.2 percentage point reduction in the likelihood of her being replaced at midline. Similar results obtain in the other subjects: a 10 point increase in Somali literacy – scored on a 0-100 scale – is associated with a 1.8 point reduction in the likelihood of being replaced.

Discussion

The results in this section are deeply concerning. In the case of the transition analysis, the sample is potentially biased toward a positive outcome, as girls who are most likely to drop out are simultaneously more likely to fall out of the sample at midline, artificially increasing estimated transition rates. The learning sample is similarly affected: girls with the highest propensity to drop out

and be replaced in the sample are those who underperformed at baseline. Depending on the nature of girls selected as replacements at midline, the outcome could be the widespread replacement of underperforming girls in the sample with higher-performing girls.

Luckily, there is a natural solution to this problem in the case of learning – the use of the "true panel" of girls who appear in both the baseline and midline data. Through much of the learning analysis in Section 4, we rely heavily on this "true panel" because it allows the strongest possible inferences about changes in learning scores from baseline to midline, by eliminating potential bias stemming from the replacement process. For transition, unfortunately, no simple solution exists, and the potential bias due to differential attrition needs to be kept in mind when reviewing the results in this report.

2.5 Challenges and Limitations

This section summarises methodological limitations of the study. For fieldwork-related limitations, please see the methodological annex (Annex 18).

Methodological Limitations

The evaluation reported on here is subject to a number of caveats as a result of methodological issues encountered during the midline. Many of these limitations stem from the specifics of the evaluation design, and were also noted at baseline; in other cases, conducting the midline has exposed additional limitations that must be reckoned with.

LACK OF A COMPARISON OR CONTROL GROUP

The lack of a comparison is arguably the defining limitation of the EGEP-T evaluation, as it shapes both the nature of the evaluation and interacts with other limitations in important ways. The research design employed diverges significantly from the GEC-T standard design, which employs a difference-in-differences approach to impact evaluation, utilizing a comparison group of schools whose progress is tracked over time in parallel to the treatment schools. The alternative design employed in this evaluation was developed in response to security constraints that limit the possibility of a traditional treatment-control panel design.

In the absence of a comparison group, the evaluation employs a pre-post design, which compares outcomes between baseline, midline, and endline, testing for changes over time. These comparisons draw from the same set of schools and respondents, ideally, but cannot account for temporal shifts in outcomes that would have occurred with or without the intervention (in other words, those changes that would have occurred under the counterfactual). The weakness of such comparisons stems from the possibility of positive trends over time in the broader program environment. If community attitudes in Somalia are generally becoming more supportive of girls' school completion, the evaluation may attribute impact to the program that is actually part of a broader trend in Somali society. If the endline evaluation documents a positive change in community attitudes within treated communities, it is not possible to determine whether this change was due to the treatment itself, because community attitudes in control communities were not tracked over the same time period.

In the case of learning, the evaluation seeks to mitigate this concern using benchmarked comparisons. In line with guidance from the FM, the evaluation utilizes cohort girls and other girls surveyed at the baseline to construct a counterfactual based on the observed differences in learning scores between girls in, e.g., grade 6 and grade 7 at baseline. To draw conclusions about changes in learning scores over time for girls who were in grade 6 at baseline, we compare the gains in their scores to those that the difference in scores between grade 6 and grade 7 at baseline; the critical idea is that benchmarking

accounts for maturation or growth effects among girls. The evaluation nests this approach in a modified difference-in-differences calculation.

Benchmark comparisons represent a significant improvement over non-benchmarked comparisons, which are unable to distinguish between expected changes that would occur over time (i.e. natural improvement from grade 6 to grade 7) and program-driven changes over the same time period. However, these comparisons are, nonetheless, susceptible to their own types of bias. The first, and most important, is the same bias that affects over-time comparisons described above: the possibility of broad societal-level changes – either positive or negative – that are not driven by the program itself. Imagine a broad, negative shift in learning outcomes across Somali schools in general, perhaps driven by an increase in conflict, a worsening of drought conditions, or policy changes that damage the educational system overall. "Benchmark girls" (i.e. those in grade 7 at the baseline) will not have been affected by these changes at the time of the benchmark's establishment; in contrast, "program girls" (i.e. those in grade 6 at the baseline) will have been affected from the changes by the time of the endline, when they will be in grade 7 and being compared to the pre-established benchmark.[47] Of course, similar threats to inference stem from broad positive shifts – a widespread increase in learning across Somali schools due to some exogenous shock would bias the results of the evaluation toward finding a positive effect of EGEP-T, even if the project itself had no impact. The lack of a comparison group could result in bias in either direction, depending on the nature of any exogenous shock to educational outcomes during the project's lifecycle.

The lack of a comparison or control group is the most important shortcoming of the evaluation design. Benchmarking and other steps taken to mitigate this shortcoming represent important improvements in the evaluation's design, but are still susceptible to key sources of bias that need to be accounted for and considered in the evaluation.

CHANGES IN LEARNING ASSESSMENTS ACROSS ROUNDS

As discussed at baseline, the evaluation of EGEP-T employs different learning assessments at baseline and midline – and will employ a third, modified, learning assessment, at endline. In response to feedback provided in the baseline evaluation report, Relief International and its implementing partners adjusted the midline learning assessments, increasing their difficulty by adding more complex subtasks. The goal of the redesign was to reduce the threat of ceiling effects in the midline and endline evaluation rounds, as ceiling effects were observed during the baseline.

Regardless of the redesign, a comparable set of subtasks were maintained between baseline and midline evaluations, facilitating comparisons in learning scores over time on a like-for-like basis. Unfortunately, even minor changes to the comparable set of subtasks between rounds – changing the response options or providing a new question with assumed equal difficulty – introduces the potential for differences in assessment difficulty. In a traditional difference-in-differences design, any changes in assessment difficulty will apply equally to treatment and control groups; as a result, it is still possible to obtain unbiased estimates of the project's impact on learning, even in the presence of systematically different examinations.

[47] To be clear, the value of benchmarks lies in controlling for growth or improvement that would occur naturally over time, such as improvement in learning outcomes as a child grows. Over-time comparisons of outcomes that exhibit natural growth patterns, which are not benchmarked, are rendered invalid due to bias stemming from this naturally-occurring growth. However, benchmarking does not account for non-program changes that occur between the baseline and endline, such as changes in societal conditions, or broad-based policy shifts. This type of bias is still present even in benchmarked comparisons, unless a suitable control group is tracked over the same time period.

However, without a comparison group, it is not possible to distinguish precisely between two possible interpretations of changing mean scores. The first interpretation would attribute changes to the project; for instance, if scores are lower at the midline, we might conclude that the project caused the decline in scores. The second interpretation would attribute changes to the test itself – perhaps a more difficult test is responsible for declining scores. In the absence of a comparison group, there is no formal method for adjudicating between these two possible interpretations.

Prior to the midline, Relief International piloted the midline learning assessments alongside those used at the baseline. In all, 50 girls completed each of the numeracy, Somali literacy, and English literacy assessments. Girls were presented with the full set of subtasks to be used at the midline, and the comparable subset of subtasks from the baseline, such that it was possible to assess whether comparable subtasks at the baseline were easier or more difficult than comparable subtasks at the midline. RI provided this pilot data to us prior to the start of training, and we performed a rapid analysis of the results, highlighting a handful of subtasks that seemed to have changed in difficulty across rounds. In some cases, changes in difficulty balanced each other; in others, RI's Monitoring & Evaluation team made small adjustments to the assessment in an attempt to equalize difficulty levels.

CEILING EFFECTS IN LEARNING ANALYSIS

As noted above, the baseline evaluation reported significant ceiling effects – censoring of scores at the top end of the scoring scale – in numeracy and Somali literacy. In response, RI designed more difficult subtasks, which were included in the midline assessments. By increasing difficulty at midline, this approach has reduced the potential for ceiling effects to influence midline-to-endline comparisons.

However, the midline evaluation cannot avail itself of these more difficult subtasks, as they were not included in the baseline assessment. We take a comparable-subtask approach to equalizing difficulty between baseline and midline, but this does not mitigate the issue of ceiling effects for baseline-to-midline comparisons – in fact, it exacerbates the problem slightly. To the extent that ceiling effects are present, it will reduce the estimated gains in learning scores over time, biasing the results toward a null or negative effect.

We have taken a number of steps to address ceiling effects in this round, in consultation with the FM and RI's Monitoring & Evaluation team. First, we disaggregate results by grade level, which should reveal ceiling effects, as ceiling effects are most likely among older girls. Specifically, if we observe large improvements in learning among younger girls, with smaller or non-existent improvements among older girls, it is suggestive that ceiling effects are operative.

We also employ a more drastic approach, limiting our analysis to include only girls who scored below 85 percent at baseline. While ceiling effects should be less pronounced among younger girls, a few primary-age girls did achieve high scores at baseline and may be subject to ceiling effects at midline. In contrast, by specifically removing girls with high baseline scores, we can assess whether learning scores improved among the set of girls among whom ceiling effects are least likely. Although this method is still subject to multiple interpretations – for instance, it is possible that the project has heterogeneous impacts on learning and primarily benefits low-achievers – it should provide greater confidence in findings around learning.

BIASED REPLACEMENT IN LEARNING SAMPLE

The replacement analysis in Section 2.4 revealed that girls who remained in the learning sample at midline and those who fell out of the sample were fundamentally different with respect to baseline learning outcomes. Girls who achieved higher scores on baseline learning were less likely to drop out

of school in the year that followed, and less likely to be replaced in the midline sample. The resulting bias could produce an overestimate of improvement in learning scores, if replacement girls have better pre-existing learning performance than those they replace.

To mitigate this problem, this evaluation focuses primarily on the "true panel" of cohort and bursary girls when studying learning outcomes. The "true panel" consists of girls who appear in both baseline and midline samples, without replacement. Although the available sample size for analysis is reduced, the effect on statistical power is minimal, because girls are removed, but the number of clusters or schools in the analysis remains fixed. This approach yields the most rigorous assessment of learning outcomes possible, under the circumstances.

NON-RANDOM ATTRITION IN THE TRANSITION SAMPLE

Transition rates at midline were calculated among cohort and bursary girls, with the goal of re-contacting every girl from the baseline to track their transition status. In practice, the actual sample employed consists of girls who could be located and re-contacted from baseline, such that panel attrition may influence the calculation of transition rates. As the analysis of re-contact outcomes in Section 2.4 showed, it is likely that girls who drop out of the transition sample altogether – i.e. cannot be re-contacted or tracked for transition – have lower probabilities of successful transition than those who can be re-contacted successfully. The result is an estimate that may overstate midline transition rates, though the severity of this bias is not quantifiable with the available data.

In an attempt to combat attrition of this kind, field teams requested information about girls they were attempting to locate, even if they could not be re-contacted directly, to track their transition status. For instance, if a girl had moved to reside with family members in another city, but the enumerator was able to locate a family member remaining in the community, the enumerator recorded whether the girl was enrolled in her new location. Unfortunately, this approach can only provide suggestive evidence, because the data was inconsistently collected – a minority of re-contact attempts brought teams into contact with family members or others who knew the girls' present enrolment status – and may or may not be reliable. Reducing or eliminating differential attrition in the transition sample from midline to endline should be a key focus of research design in the next evaluation round.

COMPARABILITY GAPS BETWEEN BASELINE AND MIDLINE TRANSITION SAMPLES

At baseline, a benchmark sample was drawn from a random household survey in the communities surrounding EGEP-T project schools. Girls in this baseline transition benchmark sample were selected through a random household survey, while girls in the baseline cohort sample were selected through a random survey of children enrolled in school. The result is that the transition benchmark sample includes a number of out-of-school girls and has a significantly lower overall transition rate than we would expect from a sample comprising exclusively girls who were enrolled at baseline. Enrolled students are much more likely to remain enrolled and successfully transition to the next grade than out-of-school girls are to re-enrol.

The disjuncture between the transition cohort and the transition benchmark sample is illustrated in the baseline report. Among cohort girls at baseline, transition rates were effectively 100 percent, because all girls were enrolled in school. In contrast, the overall rate among the transition benchmark sample was 73.6 percent. This gap stems from the fact that the two sets of girls were selected into the sample in systematically different ways and the benchmark sample included girls who were out-of-school in the year prior to the baseline (time *t-1*). Benchmark girls enrolled at time *t-1* had transition rates at baseline of 86.8 percent; meanwhile, benchmark girls who were not enrolled at time *t-1* had transition rates of just 22.2 percent. Clearly, comparing cohort girls – all of whom were enrolled at baseline – to this mixed benchmark sample is not a fair comparison.

To facilitate a like-for-like comparison between baseline and midline transition rates, we limit the transition benchmark sample to those girls who were enrolled at time *t-1*, ensuring the transition analysis makes valid comparisons. Restricting the sample in this way reduces the available benchmark sample, from a possible 717 benchmark girls to 558.[48] The reduction, while meaningful, does not have significant effects on statistical power, given the clustered nature of the survey design and the fact that girls, rather than clusters, are being removed from the sample. The approach yields a transition benchmark sample for the baseline that is as comparable as possible to the girls being assessed at midline with respect to their underlying propensity for enrolment and transition. To protect against potential bias due to differences in the age distributions of the two samples, we re-weight the benchmark sample by age to match that of the midline cohort transition sample, in line with guidance from the FM.

COMPARABILITY OF HOUSEHOLD SAMPLES

The household sample analysed at baseline consisted of a random sample from EGEP-T project communities, as discussed above. The sample included a number of households in which girls were not enrolled in school, or were enrolled in non-project schools. In contrast, midline household surveys were conducted predominantly with the households of cohort and bursary girls, all of whom were enrolled in an EGEP-T school at the time of the baseline. This represents a qualitative shift in the composition of the sample, toward – for instance – households which may value education more highly. To the extent that a household's support for girls' education is correlated with enrolment, writ large, or enrolment in an EGEP-T project school, specifically, it could result in bias in measures derived from the household survey. For instance, if community attitudes are assessed on the basis of caregiver responses, changing the set of caregivers interviewed from those drawn in a random survey to those of girls enrolled in school may produce pro-education bias at midline, vis-à-vis baseline responses. This issue may affect conclusions drawn regarding the community attitudes intermediate outcome, as well as those sustainability indicators focused on changes in community support for girls' education.

To address this issue, we analyse the comparability of baseline and midline household samples extensively in Section 2.6, below. We provide evidence for systematic but limited differences between baseline and midline household samples. Two solutions are used when we measure community attitudes – the baseline sample is restricted to the set of households whose girls were enrolled at baseline, to ensure comparability with the midline sample, and results are further triangulated across disparate data sources that are not affected by changes in sampling procedures.

It is important to note the scope conditions of this limitation. While the household sample shifted from baseline to midline, it is expected to stay more consistent from midline to endline, because all cohort and bursary girls from the midline will be re-contacted and their households interviewed to track transition rates. Effort at the design stage for the endline can be made to ensure comparability between the two samples. This will facilitate valid comparisons of midline to endline community attitudes, without the limitations noted here in the context of baseline to midline analysis. While the resulting sample will be confined to caregivers – rather than a random sample of households in a community – we view this as a valid indicator of community attitudes writ large, and do not think it should be considered a major problem for the endline report. The limitation we note in this section is about the *change* in the sample composition from baseline to midline; the composition of the midline

[48] Both sample sizes account for the loss of the Banadir sample by removing benchmark girls from Banadir from the analysis. Moreover, girls for whom transition cannot be clearly coded are removed.

sample – and the future endline – may not reflect community attitudes in a perfect sense, but is a reasonable proxy for those attitudes.

SOCIAL DESIRABILITY BIAS IN SURVEY RESPONSES

Several of the outcomes measured in this evaluation are prone to social desirability bias, depending on the precise manner in which they are measured. Attitudinal outcomes, such as community support for girls' school completion, are especially subject to such bias, because respondents may know that expressing support for girls' education is the socially desirable response, i.e. the response that will not be stigmatized socially. This type of bias is especially common when respondents know that the sponsor of a survey supports one view over another.

As at the baseline, a number of steps have been taken to mitigate social desirability bias. First, data has been collected from a variety of sources, where possible, to triangulate outcomes between respondents with different incentives to misrepresent their attitudes. For instance, community attitudes will be assessed via responses from community members, responses from female and male students, and responses from head teachers. As noted above, community members may have an incentive to misrepresent their views on girls' education; however, female students and head teachers do not have obvious incentives to misrepresent the attitudes of other people in their communities.[49]

Second, in many cases, we have incorporated questions which allow respondents distance from the answers that they provide. In the household survey, we ask caregivers direct questions that assess their attitudes toward girls' education. We also ask them to assess the extent to which men *in their household* support girls' education. Because responses to this question may be subject to stigma (i.e. respondents may feel that enumerators or others will judge them or their family members if they respond negatively), we also ask respondents to assess the extent to which men *in their community* support girls' education. By allowing respondents to report the attitudes of community members writ large, we are more likely to receive truthful answers, as respondents may feel more comfortable reporting that "men in this community do not support girls' education" than that "men in this household do not support girls' education" or "I do not support girls' education."

INACCURACY OF SCHOOL ATTENDANCE RECORDS

The primary indicator of attendance rates – one of the project's intermediate outcomes – is derived from school attendance and enrolment records. In principle, school records present the most comprehensive and valid means of measuring school attendance. However, it is known that these records are often inaccurate, sometimes to a severe degree. As the baseline report documented, a substantial minority of girls are missing all or part of their attendance records; another set of girls have records that are obviously inaccurate.

Like the baseline evaluation, we will rely heavily on classroom headcounts performed during fieldwork in the midline due to the incompleteness of the data and the biases inherently present in the attendance record data. In the baseline, the girls' cohort included 1,609 respondents, while we have school records for 1,190 girls. The available sample is smaller when we analyse correlates of attendance, because we are occasionally missing data on other variables. In the midline, the girls'

[49] Head teachers' assessments of community attitudes are subject to their own form of bias, of course, because head teachers are typically well-educated and may be drawn from outside the community, especially in rural areas. Their viewpoint may be biased against communities, if they see education as particularly important and communities as failing to support it adequately. Importantly, biases of this kind will be controlled for directly in over-time comparisons at the midline and endline, because the same respondents will generally be interviewed in later waves.

cohort includes 1,449 respondents and 369 bursary girls. Of the girls' cohort, we have records for 1,046 of the cohort girls and 274 of the bursary girls. Moreover, the sample of girls are sampled from among girls in attendance at the time of data collection; as such, the cohort girls sample is biased toward girls who attend more frequently. Particularly high absenteeism is concentrated among a subset of all students, who are less likely to be selected into the learning cohort sample than those who attend regularly. As such, the individual-level sample has explicitly excluded, to some degree, students who are frequently absent, who likely have systematically different characteristics than those who are consistently present and therefore more likely to be sampled. In contrast, using headcount data includes all grades at a school, and in principle, included all girls who were enrolled at any point during the year as the denominator in our calculations.

Despite the biases in school records data, the data is still useful for analysing the predictors or correlates of attendance at an individual level, though not among a truly random sample of girls. They still serve as an unbiased baseline for attendance rates among this particular sample, because midline and endline evaluations can assess changes in attendance rates among the same subsample of girls as long as all concerned are aware that this sample is not perfectly representative of the population of enrolled girls.

However, this headcount is only an instantaneous measure of attendance (on the day at which the headcount is taken), and is thus not necessarily representative of the overall attendance trend at a given school. The household survey provides a further check on attendance rates, using information reported by caregivers. When reporting on attendance, we make specific effort to triangulate across these distinct data sources, with headcounts and the recollections of caregivers filling gaps left by often-inaccurate school attendance records.

GAPS IN MEASUREMENT OF SUSTAINABILITY INDICATORS

In general, the evaluation captured rich, varied, multi-dimensional data on the intermediate outcomes, especially teaching quality, school management and governance, and community attitudes. At times, however, sustainability indicators were measured in less comprehensive ways, which limited the breadth and quality of the analysis. For instance, one sustainability indicator measured the extent of awareness-raising activities conducted in project communities; while data was collected from caregivers and teachers on this topic, the set of activities that respondents were presented was limited and likely did not capture the full extent of awareness-raising activities that the project sponsored. In at least two cases, sustainability indicators were not directly measured at all.[50]

At the endline, it will be especially important to improve the measurement of sustainability indicators. Measures should be made more robust, allowing for greater triangulation across question types and respondents; in some cases, measures also need to be tied more directly to the project's activities. In contrast to other outcomes, direct and precise comparability between evaluation rounds is slightly less important in the assessment of sustainability, so improvements can be made to the data collection tools to improve this aspect of the endline evaluation.

[50] In one case, shortcomings in the measurement strategy interacted with fieldwork challenges: sustainability indicator 7 – concerning the development of a gender strategy within the federal and Galmudug education ministries – was planned to be a purely qualitative measure, based on just a single interview with a ministry official in Mogadishu. This weak measurement strategy was then impacted by the cancellation of fieldwork in Mogadishu, as the team intended to conduct the interview in question was no longer participating in fieldwork. The evaluation team failed to reallocate the interview to another team, who could have conducted the interview by phone, producing a situation in which *no* data was available to measure the indicator.

2.6 Analysis of Methodological Limitations

Comparability of BL HH and ML HH Samples for Community Attitudes

In addition to sampling complications related to the transition analysis, attitudinal measures are also affected by sampling decisions taken at baseline and midline. To briefly summarize the issue: at baseline, community attitudes were measured on the basis of a random household sample in project locations. A random sample of 559 households (approximately four per school-community visited) was selected, with eligibility determined by the presence of an adolescent girl aged 11-18 years in the household. The households surveyed constitute the transition benchmark sample, but also serves as the sample for measuring underlying community attitudes toward girls' education. In many ways, this sample provides the best possible representation of community attitudes, because it is a random sample of adults in the community, including adults whose children are not enrolled in school.

As noted previously, the midline evaluation did not repeat the random household sample conducted at baseline, instead conducting surveys with the households of cohort girls and bursary girls. This shift from baseline to midline was planned from the inception stage of the baseline, and was necessary to capture demographic and other key household-level information about girls constituting the learning and transition sample (i.e. "cohort girls"). Unfortunately, such a shift means that the population represented has also changed – at baseline, the sample represented community members with daughters aged 11-18 years; at midline, the sample represents community members whose daughters were simultaneously enrolled at baseline, present at the school during fieldwork, and selected into the cohort sample.[51]

Given that the primary goal of the household survey is to collect information about cohort girls and their households, this shift was necessary and has limited consequences. One such consequence, however, is seen in the measure of community attitudes. Theoretically, we would expect the parents of girls enrolled in school to have systematically different attitudes toward girls' education than the parents of a random subset of girls, many of whom are not enrolled. To the extent that this is true, it produces bias in our estimates of changes in attitudes, likely overstating the improvement in attitudes from baseline to midline. The indicators potentially affected by this bias are:

- Intermediate outcome: change in community attitudes toward girls' education
- Sustainability indicator, exposure to awareness-raising activities –percentage of surveyed members of EGEP target communities, who have been exposed to project awareness-raising activities report having changed their opinion positively in relation to the importance of girls' school completion.
- Sustainability indicator, male support for girls' education – percentage of surveyed members of EGEP target communities that report that boys and men are taking action to support girls in attending and completing school
- Sustainability indicator, community leaders' advocacy – percentage of EGEP target communities where community leaders are leading campaigns and advocacy events

In each case, there are strong theoretical reasons to expect differences in outcomes between the populations sampled at baseline and midline. In the case of community attitudes toward girls' education, individuals whose daughters are currently enrolled in school likely hold more pro-

[51] As discussed elsewhere, cohort and bursary girls were selected at baseline via schools, meaning that all selected girls were enrolled in Grade 6 through Form 2 at the time of selection. Further, cohort and bursary girls were selected on the basis of *presence* at schools at the time of baseline data collection, not just enrolment.

education attitudes than a random sample of parents. Similarly, adults whose children are enrolled in school may perceive greater male support for girls' education, as male members of these households may be more supportive of girls' education than the typical male community member.

The remaining indicators are less directly impacted by a change in sample frame, but affected nonetheless. For instance, parents of girls enrolled in EGEP-T project schools are more likely to be familiar with project awareness-raising activities, given their direct affiliation with the project itself. Similarly, parents of enrolled girls may be more likely to be familiar with pro-education activities happening in their communities – such as advocacy campaigns by community leaders – than parents of girls who are not enrolled.[52]

Ultimately, concern about bias due to sampling strategy is an empirical question. The available data do not allow us to test hypotheses regarding differences between community members in the two samples directly, but it does allow us indirect tests that are illuminating. At baseline, the household sample included 557 total households; 75.0 percent of the girls 11-18 years old selected were enrolled in school.[53] The variation in enrolment status allows us to test whether adults in the two different types of households – those in which the girl was enrolled and those in which she was not – have systematically different attitudes toward girls' education.

To preview the results briefly, we find that adults in households with enrolled girls (hereafter ISG households) are significantly more pro-education than their counterparts in households in which the girl is not enrolled (hereafter OOS households), across a number of measures. Our solution to this difference is to subset the baseline household sample to ISG households when analysing changes in community attitudes, producing a like-for-like comparison between ISG households at baseline and households of cohort girls – who were, by definition, enrolled at baseline – at midline. Our approach ignores a more nebulous concern regarding comparability, in that many ISG households at baseline included girls who were enrolled but not specifically in *project schools*. We test for this possibility by comparing ISG households in which the girl was enrolled specifically in a project school versus ISG households in which the girl was enrolled in a non-project school, and discuss these results in more detail later in this section.

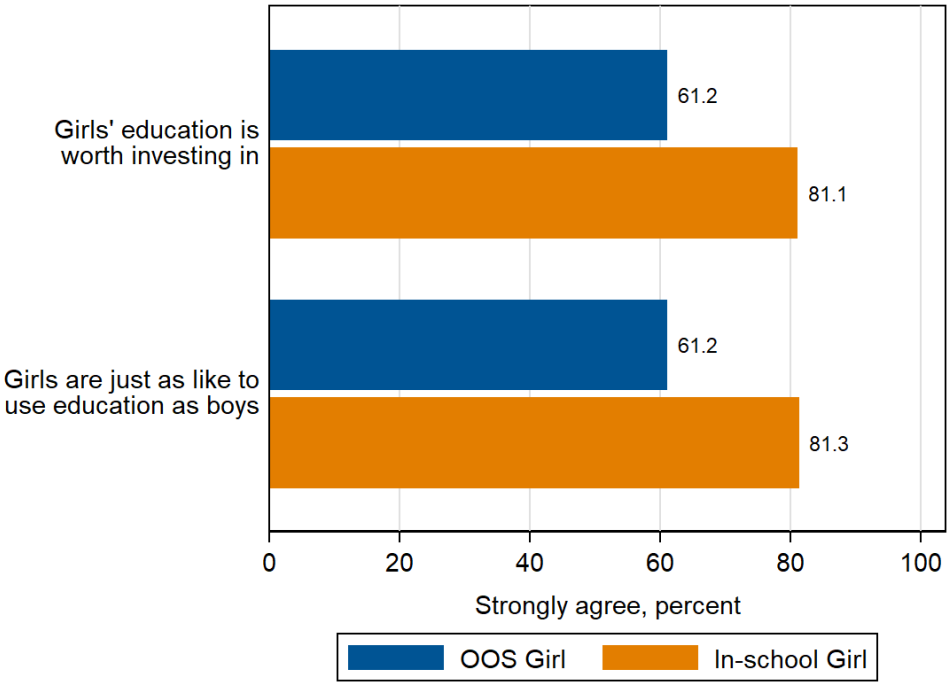
Our first tests concern the difference in community attitudes between ISG and OOS households. The evidence in this case is clear: ISG households exhibit more positive attitudes toward girls' education across a range of indicators. The figure below reports the share of caregivers who strongly agreed with two statements regarding girls' education, disaggregated by the enrolment status of their daughters. In the first case, 81.1 percent of respondents in baseline ISG households strongly agreed with the statement "girls' education is worth investing in, even if funds are limited." In contrast, just 61.2 percent of respondents in baseline OOS households strongly agreed with the same statement. We

[52] On the other hand, the baseline sampling strategy tended to over-represent households nearest schools, because the random walk methodology requires a starting point and teams have a natural tendency to define starting points in proximity to the schools. As a result, it is conceivable that community members surveyed at baseline, who tend to be closer to the schools on average, could be *more* familiar with project activities and leader advocacy campaigns.

[53] In reality, the structure of the sample is slightly more complicated, because some households included multiple girls aged 11-18 years, and their enrolment status often differed. In the analysis that follows, we categorize households according to whether the single, randomly-selected girl was enrolled. However, we also investigated the relationship between adult attitudes and the share of girls aged 11-18 years in the household who were enrolled (i.e. capturing all girls in this age range, rather than a single girl randomly selected via a kish grid) and the results are not affected by this more nuanced analysis.

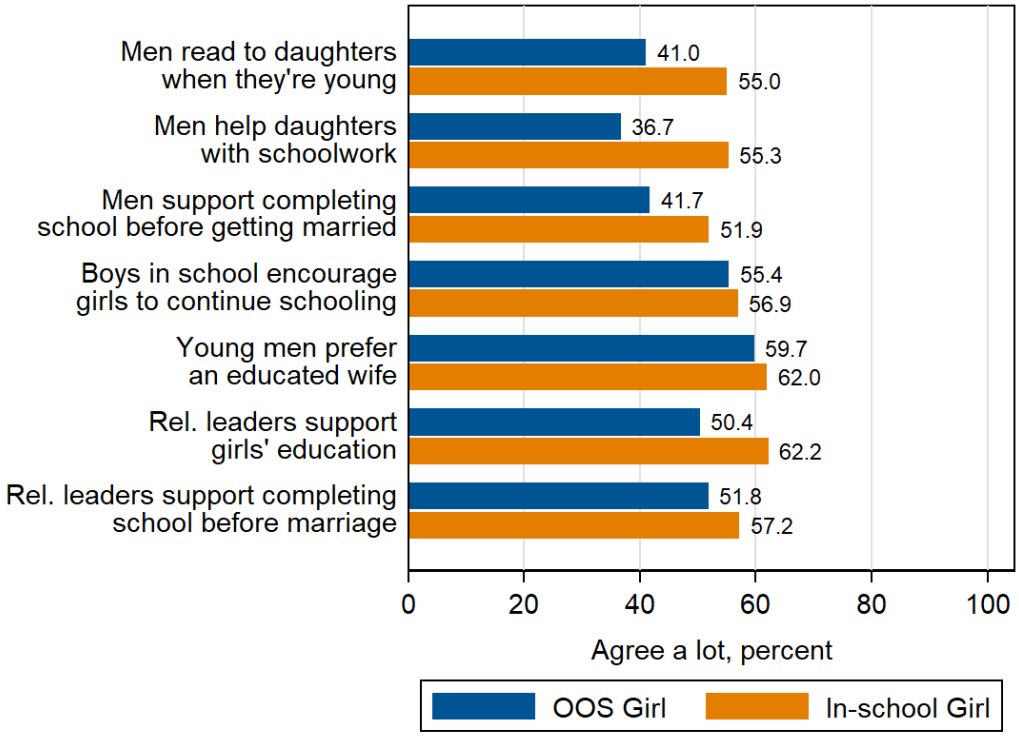
observe nearly identical results when respondents were asked whether they agree that "girls are just as likely to use their education as boys."

FIGURE 5: CAREGIVER ATTITUDES TOWARD GIRLS' EDUCATION, BY ENROLMENT STATUS OF GIRL AT BASELINE



The results above are most problematic for the intermediate outcome regarding community attitudes. However, the large gap between ISG and OOS households is not limited to attitudinal measures. Respondents in ISG households also have different perceptions regarding support for girls' education among male community members and community leaders. As the figure below shows, ISG respondents were more likely to report that men and boys take tangible actions supporting girls' education, and slightly more likely to report that boys in their communities hold positive views regarding the importance of girls' education. This finding is slightly more surprising, because it is not immediately clear why the enrolment status of one's daughter should influence one's perceptions regarding other community members and their attitudes. However, individuals living in the same community have diverse experiences; if girls in more supportive households are more likely to be enrolled in school, it would explain why girls' enrolment status is associated with different perceptions of support among male community members – men in their households and social circles may be more supportive, colouring their perceptions differently than adult members of OOS households.

FIGURE 6: PERCEIVED ATTITUDES OF COMMUNITY MEMBERS BY CAREGIVERS, BY ENROLMENT STATUS OF GIRL



In general, the results regarding male and community leader support for girls' education are less dramatic than the attitudinal differences cited above, in which ISG household members were approximately 20 percentage points more likely to strongly agree with pro-education statements. Most striking are results regarding tangible actions taken by men in their communities: 55.3 percent of ISG household members state that men in their communities help their daughters with schoolwork, compared to just 36.7 percent of OOS household members. Findings regarding boys and, for instance, their preferences for an educated wife are less stark: 62.0 percent of ISG household members agree strongly that young men prefer an educated wife, compared to 59.7 percent of OOS household members. These differences are not statistically significant. However, the gaps are sufficient that they could produce bias when comparing baseline to midline.[54] Perceptions of religious leaders and their support for girls' education represent a middle ground, with significant differences between respondents in ISG and OOS households.

In the sections impacted by this sampling change from baseline to midline, we employ a straightforward solution, restricting the baseline sample caregivers or household members from ISG households. In the case of community attitudes, this means that community attitudes are being compared from baseline to midline with a sample of caregivers whose daughters were uniformly enrolled at baseline, a broadly like-for-like comparison. We employ a similar sample when we utilize household-level data to assess male support for girls' education.[55]

[54] It is also worth noting that the small sample size available for analysis (557 respondents, divided into two groups) means that even substantively meaningful effect sizes may be too small and noisy to attain statistical significance.

[55] In our analysis of community attitudes, male support for girls' education, and many other indicators, we triangulate across multiple data sources. Our concerns about sampling approach are limited to household-level

As briefly mentioned above, our solution obscures a less clear-cut concern regarding the comparability of ISG households in which a girl is enrolled in a project school versus those in which a girl is enrolled in a non-project school. To evaluate comparability between these two groups, we tested the differences in caregiver attitudes toward girls' education in the baseline household sample, comparing caregivers of girls enrolled in project schools (n = 266) to caregivers of girls enrolled in non-project schools (n = 152).

Overall, the results are mixed. Caregivers in project schools are significantly *less* likely to agree that girls are just as likely as boys to use their education, and are slightly less likely to agree that girls' education is worth investing in, even when funds are limited. On the other hands, caregivers from project schools have higher educational aspirations for their daughters on average – this finding is especially consequential, because it is the project's primary logframe indicator of community attitudes.

In the area of project activities, such as back-to-school campaigns and community outreach efforts, caregivers in project schools are, surprisingly, no more likely to be familiar with the activities than caregivers in non-project schools. Caregivers in ISG households were significantly more likely to be familiar with awareness-raising activities of all kinds – e.g., door-to-door visits encouraging enrolment and events highlighting community leaders' support for girls' education – than caregivers in OOS households. This finding underscores the need to subset the baseline household sample by enrolment status of the household's adolescent girl. However, there is *no* significant difference between households with a girl enrolled in a project versus non-project school – for instance, precisely 50.0 percent of both types of caregivers had heard of back-to-school campaigns encouraging re-enrolment of girls. This surprising finding may stem from the fact that all caregivers surveyed, even those with girls enrolled in non-project schools, live in proximity to project schools, and may be familiar with community-oriented project activities.[56]

Based on these results, our preferred approach is to restrict the baseline household sample, when analysing those indicators outlined at the outset of this section, to respondents in ISG households (i.e. those households in which the selected adolescent girl was enrolled at the time of data collection). In the case of community attitudes, we will also perform analysis after limiting the baseline sample to ISG households with girls enrolled specifically in project schools, owing to the occasional differences, noted above, between caregivers from project and non-project schools, respectively. These findings

---

data, and do not influence data collected directly from cohort girls, teachers, head teachers, and so forth, which are often used to assess community attitudes and other outcomes of interest.

[56] We also investigated the question of differences between households with girls enrolled in project versus non-project schools another way, comparing the perceptions of girls themselves. To be clear, we tested whether baseline girls in non-project schools in the household sample viewed their households and its support for girls' education differently than cohort girls when asked the same questions at baseline. The former group represents girls enrolled in non-project schools, while the latter represents a large sample of girls in project schools. We found that respondents in these two groups did have occasionally large differences in how much support they claim to receive from their families for their schooling, and in how likely their parents are to attend meetings with their teachers, a tangible if imperfect measure of parental support for their daughter's education. These results suggest that project and non-project households may provide different levels of support for education and have different underlying attitudes toward girls' education. However, we do not report these results here because they are based on an indirect measure – girls' perceptions of support from their families – and provide a less straightforward assessment of whether caregiver attitudes vary across these two types of households. In general, the findings did not alter our conclusions regarding the best way to approach sample construction and comparability for these indicators.

will only be presented where they are relevant and substantively meaningful, to avoid unnecessary confusion.

3. Context, Educational Marginalization, and Intersection between Barriers and Characteristics

Building on the methodological discussion and brief description of the achieved sample provided in the previous section, this section of the report provides greater detail on the composition of the cohort samples at baseline and midline. The goal of this section is to highlight characteristics of the sample and the distribution of specific barriers to educational attainment within the sample. To the extent that the sample is representative of the project's beneficiaries, the results presented below shed light on the *type* of child impacted by EGEP-T's interventions, and the nature of their particular marginalization. The discussion below focuses on changes in sample composition from baseline to midline, followed by a presentation of sample composition at the midline, broken down by a girl's re-contact versus replacement status, to illustrate differences in cohort composition that stem from the replacement process.

The most important change to the sample from baseline to midline was not a function of replacing schools or individual girls. Rather, the most consequential change took place early in fieldwork, when fieldwork in Mogadishu (Banadir) was cancelled due to security and accessibility concerns. The baseline sample included 21 schools in Mogadishu, out of a total sample of 140 schools. Girls in Mogadishu comprised 14.9 percent of the cohort girl sample, and a similar share of all other sample populations – cohort boys, bursary girls, head teachers, teachers, headcounts, and so forth.

The exclusion of Mogadishu reshaped the sample in critical ways, as demonstrated in the table below. The first column reports the share of the baseline cohort girl sample with a given characteristic (e.g., a female head of household, or attending an IDP school). The second column reports the share of the baseline Mogadishu sample with the same characteristic, facilitating comparison between the full sample and the Mogadishu-specific sample. The final column reports the frequency of a given characteristic in the baseline sample, excluding girls from Mogadishu, to show how the nature of the baseline sample changes (column 2 versus column 4) with the removal of Mogadishu. In the right-most column, an asterisk denotes characteristics whose distribution in the sample is significantly altered by the exclusion of Mogadishu.[57]

As the results show, the exclusion of Mogadishu restructured the sample in important ways, beyond the obvious reduction in sample size. Without girls from Mogadishu, the sample became significantly more rural, and a higher share of girls were considered drought-affected at baseline. In general, the remaining girls are more marginalized than those that were lost in Mogadishu – beyond being more rural and more drought-affected, they are also more likely to reside in a household where the household head lacks education, and more likely to attend a school that has insufficient seats for all students. In addition, Mogadishu schools had lower levels of teacher absenteeism, Mogadishu girls were less likely to have experienced a severe illness in the last year, and less likely to live in pastoralist

[57] To test for changes in the sample composition, we performed a t-test comparing values in the full baseline sample to values in the baseline sample that excludes Mogadishu. Results that are statistically significant at the 5 percent level, in a two-sided test, are marked in the table.

households, though the loss of Mogadishu does not *significantly* alter sample composition in these cases.

Clearly, the nature of the sample has shifted dramatically. Even leaving aside relative marginalization, girls in Mogadishu systematically outperformed girls in other zones in terms of all three learning outcomes at baseline. For instance, the mean English literacy score in Mogadishu was 58.7 percent, compared to 37.3 percent in the remaining zones. Slightly smaller, but still significant, gaps are found for all three learning outcomes. Girls in Mogadishu are also slightly more likely, in the benchmark transition sample, to remain enrolled and progress from grade to grade successfully (i.e. Mogadishu had a higher transition rate at baseline).

We note these shifts in the sample composition because they represent the most important such adjustments at midline. Throughout the remainder of this section, we describe the baseline and midline samples that will be analysed in this report, all of which exclude Mogadishu to ensure comparability across rounds.

TABLE 10: BASELINE SAMPLE COMPOSITION, WITH AND WITHOUT MOGADISHU SCHOOLS

Characteristic	Full Baseline Sample	Mogadishu Baseline Sample	Baseline Sample Excluding Mogadishu
Rural	26.6	0.0	31.2*
Attending IDP school	5.0	0.0	5.9
Drought-affected	53.4	0.0	63.0*
Female HoH	44.4	49.4	43.6
HoH does not have a wage-earning occupation	27.9	25.5	28.3
HoH pastoralist	1.7	0.0	2.0
HoH has no education	26.7	7.0	30.5*
HoH has no education or only religious education	48.9	37.8	50.9
Girl is working outside home	0.6	0.0	0.7
Traveling to school takes 1+ hours	3.8	0.4	4.4
Girl cannot use learning materials at school	29.1	49.8	25.5*
Not enough seats for all students	20.1	2.9	23.1*
Girl will not use water facilities	36	46.4	34.2
Girl will not use toilet	26.1	15.1	28.0
Girl feels unsafe on journey to school	3.2	2.9	3.2
Girl feels unsafe at school	2.0	2.9	1.8
Teacher makes girl feel unwelcome in classroom	3.4	1.7	3.7
Teacher absent frequently	13.1	6.3	14.2
Vision impairment	1.3	0.4	1.5
Hearing impairment	0.1	0.0	0.1
Mobility impairment	0.1	0.4	0.1
Cognitive impairment	0.2	0.0	0.2
Self-care impairment	0.1	0.0	0.1

Communication impairment	0.1	0.4	0.1
Any disability[58]	1.9	1.3	2.0
Serious illness[59]	9.3	1.3	10.6

Outside of the exclusion of Mogadishu, changes in sample composition from baseline to midline are somewhat less systematic. In addition, their provenance is less clear – in some cases, changes from baseline to midline may stem from changes in the characteristics or views of the same girls (e.g., improvement in their home lives); in other cases, changes may stem from panel attrition and the selection of replacements at midline.

The table below describes the characteristics of cohort girls (top panel) and bursary girls (bottom panel) at baseline and midline. The baseline sample reported on here is the full cohort girl sample, excluding Mogadishu. The midline sample is the full cohort girl sample utilized at midline. In theory, the two samples should be very similar, given that approximately four-fifths of girls are shared between the two samples, i.e. they appear in both rounds. However, replacements selected naturally alter the sample composition, and girls' characteristics can change over time, as can the characteristics of their teachers or the schools they attend.[60]

TABLE 11: COHORT SAMPLE CHARACTERISTICS AT BASELINE AND MIDLINE

Characteristic	Baseline	Midline
Cohort Girls		
Rural	31.2	33.1
Attending IDP school	5.9	6.0
Drought-affected	63.0	19.6
Female HoH	43.6	65.4
HoH does not have a wage-earning occupation	28.3	38.6
HoH pastoralist	2.0	3.0
HoH has no education	30.5	50.7
Traveling to school takes 1+ hours	4.4	2.4
Girl cannot use learning materials at school	25.5	16.7
Not enough seats for all students	23.1	11.9
Girl will not use water facilities	34.2	18.6
Girl will not use toilet	28.0	17.2
Girl feels unsafe on journey to school	3.2	2.2
Girl feels unsafe at school	1.8	0.6
Teacher makes girl feel unwelcome in classroom	3.7	6.6

[58] The indicator “any disability” refers to the proportion of girls who meet the criteria of having a vision, hearing, mobility, cognitive, self-care, or communication impairment.

[59] The indicator “serious illness” refers to the proportion of girls who answered affirmatively to the question, “In the last year, have you had any serious illnesses?”

[60] Another potential source of variation in sample composition is change in the way respondents answer, even in the absence of tangible changes in their circumstances. Public opinion researchers are well aware that survey responses are not entirely stable over time – in some well-known cases, respondents may answer questions differently within the same survey, especially if the context of the question is altered in even minor ways. Therefore, changes in characteristics from baseline to midline may arise from this type of instability. A simple example would be a girl who reports at baseline that she has a lot of difficulty walking, while reporting at midline that she has only "some" difficulty walking, resulting in a category change in her disability status.

Teacher absent frequently	14.2	21.6
Vision impairment	1.5	0.5
Hearing impairment	0.1	0.1
Mobility impairment	0.1	0.0
Cognitive impairment	0.2	0.1
Self-care impairment	0.1	0.1
Communication impairment	0.1	0.1
Any disability	2.0	0.6
Serious illness	10.6	16.8

Characteristic	Baseline	Midline
Bursary Girls		
Rural	29.6	30.4
Attending IDP school	5.8	5.6
Drought-affected	67.1	22.1
Female HoH	56.5	69.4
HoH does not have a wage-earning occupation	40.0	45.7
HoH pastoralist	0.8	3.4
HoH has no education	40.0	58.1

No clear pattern emerges from the results presented in the table. Girls are much less likely to attend schools that are drought-affected at midline, but this difference stems from climatic changes – the drought eased between the time of the baseline and the start of the midline. The midline sample has a higher share of girls in female-headed households, and households headed by an individual without any education. On the other hand, the midline sample includes fewer girls that seem uncomfortable with facilities – such as the water source and the toilets – at their schools. During the baseline, 28.0 percent of cohort girls stated that they were unwilling to use the toilet at school, compared to just 17.2 percent at midline. Changes in the bursary girl sample are qualitatively similar – a reduction in the number of drought-affected girls, but an increase in household characteristics that are typically associated with marginalization, such as a female head of household, and an uneducated head of household.[61]

The final table, below, provides insight into the source of changes from baseline to midline. In this table (top panel), we broke down the midline cohort sample into girls who were successfully re-contacted at midline and those that were selected as replacements for girls who could not be re-contacted successfully. The first column, focused on-re-contacted girls, is what we refer to in this report as the set of "true panel" girls from the midline – girls who appear in both baseline and midline samples. However, the table focuses exclusively on these girls at midline, excluding their baseline information.

The second column reports characteristics of the cohort girls selected as replacements. Replacement girls were selected from within schools, so we would expect them to differ somewhat from the girls

[61] The number of characteristics tracked for bursary girls is noticeably smaller than that for cohort girls. While both groups of girls completed an extensive child survey at midline, the baseline survey given to girls was considerably shorter, and no household survey was completed with either cohort or bursary girls at baseline. Critically, bursary girls at baseline received a shorter module focused on household characteristics than did cohort girls, resulting in less demographic data and fewer indicators of bursary girls' views of their school, at least in the baseline dataset.

they replaced, many of whom were no longer enrolled in school at the time of the midline. In total, the combined sample – in the table's right-most column – includes 1,316 cohort girls; the set of re-contacted "true panel" cohort girls comprised 996 girls; the set of replacement girls included 320 additional cohort girls. For bursary girls, reported in the table's lower panel, the combined sample included 339 respondents, 274 of whom were re-contacted "true panel" bursary girls, and 65 of whom were replacements drawn for bursary girls who could not be located, refused to participate at midline, or had dropped out of school and were replaced in the learning sample.

TABLE 12: MIDLINE COHORT SAMPLE CHARACTERISTICS, DISAGGREGATED BY RE-CONTACTED AND REPLACEMENT GIRLS

Characteristic	Re-Contacted Girls	Replacement Girls	Combined Sample
Cohort Girls			
Rural	32.8	34.1	33.1
Attending IDP school	4.9	9.4	6.0
Drought-affected	18.6	22.8	19.6
Traveling to school takes 1+ hours	2.4	2.4	2.4
Girl cannot use learning materials at school	16.9	15.8	16.7
Not enough seats for all students	11.3	13.8	11.9
Girl will not use water facilities	17.3	22.5	18.6
Girl will not use toilet	17.1	17.4	17.2
Girl feels unsafe on journey to school	2.4	1.7	2.2
Girl feels unsafe at school	0.4	1.0	0.6
Teacher makes girl feel unwelcome in classroom	7.5	3.7	6.6
Teacher absent frequently	21.1	23.0	21.6
Vision impairment	0.5	0.3	0.5
Hearing impairment	0.1	0.0	0.1
Mobility impairment	0.0	0.0	0.0
Cognitive impairment	0.1	0.0	0.1
Self-care impairment	0.1	0.0	0.1
Communication impairment	0.1	0.0	0.1
Any disability	0.7	0.3	0.6
Serious illness	17.9	13.4	16.8

Characteristic	Re-Contacted Girls	Replacement Girls	Combined Sample
Bursary Girls			
Rural	28.8	36.9	30.4
Attending IDP school	5.8	4.6	5.6
Drought-affected	21.9	23.1	22.1
Teacher makes girl feel unwelcome in classroom	5.6	7.9	6.0
Vision impairment	1.1	0.0	0.9
Hearing impairment	0.0	0.0	0.0
Mobility impairment	0.4	0.0	0.3
Cognitive impairment	0.7	0.0	0.6

Self-care impairment	0.0	0.0	0.0
Communication impairment	0.0	0.0	0.0
Any disability	1.8	0.0	1.5
Serious illness	16.8	13.8	16.2

The theory of change assumes that subgroup characteristics and barriers intersect, adding complexity to the marginalization experienced by girls targeted by EGEP-T. As shown below in Table 13, both cohort and bursary girls who live in rural areas with poorly resourced schools, as well as girls with uneducated, unemployed parents, tend to have that compounded by going to schools with limited CEC engagement and poor teaching. This confirms that RI is serving a high number and proportion of marginal girls. The intersections of some of those traits comprise between a quarter and a third of the sample.

TABLE 13: INTERSECTION OF CHARACTERISTICS AND BARRIERS OF MIDLINE SAMPLE [62]

Barriers	Female head of household	Head of household has no education	HoH does not have a wage-earning occupation	Rural	Teacher absent frequently	Affected by drought
Cohort Girls						
Household Level Economics						
Gone to sleep hungry many nights	3.1%	3.0%	1.9%	1.5%	0.8%	0.6%
Poor roofing material	7.5%	7.4%	5.5%	8.8%	3.8%	1.1%
Confidence and Psychosocial Support						
Lack familial support	3.2%	2.4%	1.5%	0.4%	1.4%	0.9%
Lack teacher support	3.0%	2.6%	1.8%	1.5%	0.5%	0.4%
Weak School Governance						
No CEC meetings	16.1%	13.2%	11.6%	11.6%	4.9%	8.4%
No CEC in school	20.8%	16.3%	12.1%	16.4%	7.3%	5.1%
Poor learning Environment						
Afraid of teacher	36.0%	27.6%	21.8%	21.4%	15.0%	8.2%
Teacher uses corporal punishment	10.5%	7.7%	7.1%	1.7%	2.9%	2.0%
Barriers	**Female head of household**	**Head of household has no education**	**HoH does not have a wage-earning occupation**	**Rural**	**Teacher absent frequently**	**Affected by drought**
Bursary Girls						
Household Level Economics						
Gone to sleep hungry many nights	5.4%	5.4%	3.8%	0.8%	0.0%	2.7%
Poor roofing material	12.3%	11.5%	6.9%	10.3%	0.0%	3.1%
Confidence and Psychosocial Support						
Lack familial support	2.3%	1.5%	1.1%	0.3%	0.0%	0.3%

[62] Bursary girls were not asked about their teachers' use of corporal punishment, and, as such, that indicator is not included for bursary girls in Table 16.

Lack teacher support	3.2%	3.2%	2.7%	0.8%	0.0%	0.4%
Weak School Governance						
No CEC meetings	22.0%	17.1%	14.2%	12.3%	0.0%	10.7%
No CEC in school	25.8%	20.3%	19.1%	13.3%	0.0%	7.4%
Poor learning Environment						
Afraid of teacher	36.15%	27.69%	25.77%	18.73%	0.00%	8.76%

Most of the characteristics and barriers documented above among the cohort sample are addressed either explicitly or implicitly by the project through its Theory of Change and selected interventions. For instance, teachers are targeted for training that should alter their reliance on corporal punishment and promote an environment more conducive to learning, and households with especially limited financial resources are targeted for the provision of bursaries. In general, the project is oriented toward the girls who face prominent barriers, the existence of which is supported by findings from the baseline and midline evaluations.

Beyond those characteristics documented above, three additional groups could potentially be better or more explicitly targeted by the project's interventions. The first are girls who have married. Among the learning cohort, 2.2 percent of girls at midline were married, though the marriage rate is somewhat higher among girls of cohort age in the broader communities.[63] Girls who are married who remain in school actually have *higher* learning scores than their unmarried peers, which can be explained by the fact that married girls face higher barriers to continued enrolment, and those who remain enrolled are particularly motivated or performing particularly well in school, justifying the continuation of their education. On the other hand, marriage is a significant predictor of dropping out or being held back a grade, which is consistent with past research in and outside of the Somali context. While the project has consistently recognized the importance of early marriage as a barrier to enrolment – by, for instance, engaging with community leaders and parents to change attitudes on this topic – married girls are not explicitly targeted for direct support through bursaries or other outreach. Married girls are indirectly targeted, because the project identifies older girls, who are more likely to be married, as particularly marginalized, but this approach could be more direct, given the importance of early marriage as a predictor of project outcomes.

The second group are girls who are older than the standard age for their grade. The learning analysis in this report shows that girls who are much older than their classmates lag significantly in learning outcomes. We do not think the fact that girls are older causes them to perform poorly; rather, it is likely that girls who perform poorly are held back, thus becoming older than their classmates. Regardless of the mechanism by which girls who are "old for their grade" perform poorly, they are an identifiable group who underperforms in terms of learning outcomes. It is also possible that they face additional barriers, such as embarrassment from having been held back or dropping out and re-enrolling later. The project could arguably target these girls more directly with either additional psychosocial support or inclusion as a criterion of marginalisation for the purpose of allocating bursaries.

The third group are girls who attend schools with relatively limited instructional time. In the context where the project operates, school-days can be short, schools can close early for holiday or summer breaks, and teacher absenteeism is high. The latter issue is addressed at length in this evaluation and

[63] At baseline, the household sample showed that 6.7 percent of girls 11-18 years of age were married. The higher rate among the household sample is expected, because the household sample included girls who were out-of-school, rather than only girls who were enrolled at the time of the baseline.

should be reduced, at least somewhat, through the project's efforts at training teachers and improving school management. However, other factors that reduce instructional time are not explicitly targeted through EGEP-T's training or school management interventions. At the baseline, around one-fifth (19.3 percent) of head teachers stated that their school includes just four hours of instructional time per day, on average. And it is common for schools – especially those in rural areas – to close several days early when a long holiday break is beginning. Given the relationship between instructional time and learning outcomes documented in the learning analysis below, efforts to improve instructional time could be more explicitly incorporated into EGEP-T programming.

Overall, the project has designed interventions to address the characteristics and barriers analysed in this section. For girls in rural areas and IDP camps, the intervention provides bursary support, solar lamps, sanitary kits, teacher training, life skills, remedial classes, English proficiency classes, and learning materials. For girls whose households have substantial economic difficulty, EGEP-T provides full payment of school fees and provides for incidentals such as school uniforms and payment of grade 8 and Form 4 examination fees. EGEP-T interventions also address characteristics that are not discussed specifically above – for instance, although the evaluation does not measure the share of girls who are from minority clans, the project rightfully targets these girls with bursary support.[64] In addition to the standard interventions, in schools affected by drought, EGEP-T provides food and water provisions to teachers and children. To address poor teaching quality for girls with and without disabilities, the EGEP-T project provides teacher training on basic approaches to inclusive education. Nevertheless, there are a number of subgroup characteristics and barriers that are systematic in a way that the project cannot practically address such as the safety of girls on their journey to school, long journeys to school, and the availability of water facilities and gender-separated toilets.

4. Learning Outcome

This section presents findings for learning outcomes related to numeracy, Somali literacy, and English literacy. We begin with an explanation of the assessment design because the design helps to explain the underlying structure of our presentation of findings. To briefly summarize, midline assessments in numeracy and Somali literacy were made more difficult (through the addition of more difficult subtasks) in order to avoid some of the profound ceiling effects encountered at baseline. The result is that midline learning outcomes have a comparable component that will be useful for tracking progress since the baseline as well as a non-comparable component that will be useful for establishing a more sensitive set of measures moving forward.

Thus, we present findings on learning outcomes in two sub-sections: comparable and non-comparable. We begin by presenting findings from the comparable learning components – drawing comparisons between baseline and midline learning using subtasks and scores that are comparable in content and difficulty across the baseline and midline assessments. This longitudinal analysis is in keeping with the standard midline reporting template and involves the presentation of comparable learning scores by grade, analysis of baseline-midline differences by grade (using arithmetic difference in differences), baseline-midline differences in learning gaps, and baseline-midline differences in key subgroups. Separately, we present analysis of the non-comparable midline subtasks and scores in order to lay the foundation for future longitudinal comparison. This midline-only analysis has more in common with baseline report modalities than with midline reporting and will involve descriptive

[64] The evaluation does not include a measure of minority clan membership due to concerns about the sensitivity of the question.

summaries of learning by grade using non-comparable scores, learning gaps, and key differences among subgroups.

4.1 Summary of findings

This section presents a brief and non-technical summary of learning results that are then analysed in greater detail in the sections below. The tables below present a comparison of baseline and midline learning outcomes in terms of aggregate learning scores for cohort girls, bursary girls, and cohort boys, as well as a summary of results by zone. These preliminary results are weighted and based on truncated samples. In order to minimize ceiling effects our analysis of progress against benchmarks uses a truncated sample that excludes individuals who scored over 85 percent on a given assessment at baseline. The analysis in the tables below also uses these truncated scores in order to be consistent in the presentation of these results. In order to maximize comparability, we have used scores that are based exclusively on comparable subtasks which have been matched between baseline and midline (more on this below), as well as the true-panel sample of girls who are matched between baseline and midline.

TABLE 14: LEARNING SCORES BY LEARNER TYPE AND BASELINE-MIDLINE

Learner Type	Assessment	Baseline Score	Midline Score	Difference (Midline - Baseline)	N (Sample size)
Cohort Girls	Numeracy	48.5	56.5	8*	742
	Somali Literacy	59.9	64.3	4.4*	679
	English Literacy	31.3	46.5	15.2*	917
Bursary Girls	Numeracy	47.9	55.7	7.8*	210
	Somali Literacy	56.2	61.3	5.1*	194
	English Literacy	32	43.9	11.9*	255
Boys	Numeracy	56.3	66.8	10.5*	128
	Somali Literacy	61.3	64	2.7	153
	English Literacy	35.3	52.6	17.3*	206

*Note: Asterisk denotes statistically significant differences based on 95% confidence levels.[65]

TABLE 15: COHORT GIRLS' LEARNING SCORES BY ZONE

Assessment Type	Round of Data Collection	Somaliland	Puntland	Galmudug
Numeracy	Baseline	43	57.2	45.5
	Midline	44.1	71.7	67.3
Somali Literacy	Baseline	57.4	63	60.1
	Midline	55.9	74.4	66
English Literacy	Baseline	24.8	36.9	37.7

[65] Statistical significance of each difference is tested in a linear regression that uses baseline/midline as a predictor of a given learning outcome, while adjusting for clustering at the school level. For details of each regression, please see Appendix.

	Midline	35.8	54	65.9

TABLE 16: BOYS' LEARNING SCORES BY ZONE

Assessment Type	Round of Data Collection	Somaliland	Puntland	Galmudug
Numeracy	Baseline	52.3	60.9	56.8
	Midline	57.2	76	77.8
Somali Literacy	Baseline	60.3	61.5	65.9
	Midline	54.3	75.5	66
English Literacy	Baseline	30.4	39.8	38.9
	Midline	40.6	64	59.1

This aggregate analysis does not allow for the use of benchmarking (which is only possible when scores are disaggregated by grade), but it nonetheless provides an overview of major trends in learning. The primary trend that emerges here is that there has been a sizable and statistically significant increase, from baseline to midline, in English literacy for all learner types. Numeracy scores have also increased in the aggregate for all learner types, and this increase is statistically significant for both cohort and bursary girls (but not for boys). Progress in Somali literacy is negative in some cases, and positive in others, but for all learner types the change over time in Somali literacy is indistinguishable from zero.

These results presented in the table above do not speak directly to intervention-effects, but rather address the more fundamental question of whether measurable changes in learning took place since baseline, and whether or not those changes are statistically detectable given baseline ceiling effects and questions of assessment comparability. These findings suggest that learners of all types have made significant gains in terms of their numeracy and English literacy learning since the time of the baseline, and so these are the two places where we may expect to find positive intervention-effects when actual progress is compared against the expected progress based on benchmarking.

Ultimately, the analysis below will find that cohort girls at nearly every grade level have made substantial positive progress in terms of English literacy learning over and above what would be expected based on benchmarking. This result suggests that there may have been a positive intervention-effect on English literacy learning. There are obvious caveats in terms of our ability to make valid causal inferences about intervention-effects in the absence of a control or comparison group, but the finding related to English literacy is the strongest evidence that can be obtained given the design of this study. When considering analogous results for bursary girls and boys, their grade-wise outcomes are mixed (when compared to benchmark expectations) and generally do not indicate progress over and above what would be expected based on benchmarking.

4.2 Assessment Design

EGEP-T focuses on numeracy and literacy as core learning outcomes. In Somalia, the official language of instruction is Somali in primary schools and English in secondary schools. Because the project targets girls in both primary and secondary school, separate assessments were conducted in English and Somali, and all girls took assessments in both languages, irrespective of their current grade-level. By assessing secondary school-age students in Somali (even though they are no longer being

taught to read in Somali), the project is able to track progress in Somali literacy for girls who are in Grade 7 at the baseline and have entered secondary school at the endline. This approach is also in keeping with the testing philosophy of the Early Grade Reading Assessment (EGRA) approach, which is to test students' literacy in their mother tongue. Each subtask comprised a set of individual items, ranging from one to ten per subtask.

For ease of reference, the list of midline subtasks, by assessment type, is as follows:

- Numeracy:
    - Subtask 1: Find missing numbers in list
    - Subtask 2: Subtraction, 2 digits
    - Subtask 3: Word problems (addition and subtraction)
    - Subtask 4: Multiplication, 2 digits
    - Subtask 5: Multiplication, 3 digits and decimals
    - Subtask 6: Division, 2 digits by 1 digit
    - Subtask 7: Division, 3 digits by 2 digits
    - Subtask 8: Word problems (multiplication and division)
    - Subtask 9: Percentages and decimals
    - Subtask 10: Fraction operations
    - Subtask 11: Word problem (equation of 1 variable)
    - Subtask 12: Algebra with 1 variable
- Somali Literacy
    - Subtask 1: Word identification
    - Subtask 2: Reading fluency (difficult)
    - Subtask 3: Reading comprehension (difficult)
    - Subtask 4: Writing
    - Subtask 5: Writing – fill missing words
    - Subtask 6: Writing – fill missing words
    - Subtask 7: Sequence of sentences in story
    - Subtask 8: Reading comprehension
- English Literacy Skill
    - Subtask 1: Letter identification
    - Subtask 2: Word recognition
    - Subtask 3: Reading comp (easy)
    - Subtask 4: Reading fluency
    - Subtask 5: Reading comp (medium)
    - Subtask 6: Reading comp (hard)
    - Subtask 7: Writing (missing words)
    - Subtask 8: Writing (negative form)
    - Subtask 9: Writing (future tense)

The tests were designed by RI under guidance from the GEC FM as well as feedback from the evaluator on the basis of the baseline findings, including extensive analysis of ceiling effects at baseline as well as recommendations for ameliorating these effects moving forward.

RI piloted the learning assessments from March 11 to March 16, in 8 schools, split halfway between primary and secondary schools. The sample size for each assessment during the pilot was 50 (i.e. 50 girls for numeracy, 50 for Somali literacy, and 50 for English literacy), but different girls took different tests in an attempt to minimize the amount of time that any given girl was asked to leave class in order to take the assessment. The pilot included all the new subtasks (i.e. the complete midline test with harder subtasks added to ameliorate baseline ceiling effects). The pilot also included the baseline subtasks that were expected to be equivalent to subtasks included in the midline, allowing the checks for comparability of subtask difficulty presented in the following section.

The scoring methodology ensured that each subtask was weighted equally in the final aggregate score. Specifically, each subtask was scored as the percentage of items correct out of the total number of items (hence ranging from 0 to 100). In keeping with FM guidance, the reading tasks that involved a word-per-minute (WPM) score were censored at a cap of 100 WPM, with individuals who scored above 100 WPM being assigned a score of 100 WPM. The result is that all subtasks were individually standardized to range from 0 to 100. The total score for the numeracy and literacy assessments was then generated by taking the average of the subtask scores for that assessment (with each subtask being given equal weight), presenting the total percentage score based on the averaged subtasks, ranging between 0 and 100. This procedure ensured that each subtask (and the associated skills) made an equal contribution to the final score for a given assessment, and that the final scores for each assessment have a comparable range from 0 to 100.

4.3 Empirical analysis of methodological issues

Equivalence of Assessment Difficulty

The baseline evaluation showed significant potential for ceiling effects in both numeracy and Somali literacy scores. At that time, we recommended supplementing these two assessments with more difficult tasks at midline, to ward off ceiling effects. While this makes the overall assessment incomparable between baseline and midline, it ensures that the midline and endline will be comparable while preventing ceiling effects in these latter two rounds. For baseline-to-midline comparisons, we will employ what the FM calls the "overlapping subtask" approach. In the case of numeracy, where four subtasks from baseline were removed and four new subtasks were added for the midline, this means that we will make comparisons using the seven subtasks that are consistent across the two evaluation rounds.

Note that, when we refer to subtasks as "consistent" across baseline and midline, we do not mean that the subtask questions are identical between the two rounds. Rather, the subtasks have identical natures, they test equivalent skills, and they were designed to be of equal difficulty. Prior to the baseline, we analysed pilot data collected by Relief International which sought to test the equivalence of the baseline and midline assessments. At that time, most subtasks were found to be of roughly equivalent difficulty, though one overlapping numeracy subtask, one overlapping English literacy subtask, and two overlapping Somali literacy subtasks were found to be of significantly different difficulty from baseline to midline. We discuss this issue in greater detail below.

The set of overlapping subtasks are listed in the table below.

TABLE 17: SUBTASK MAPPING BETWEEN BL AND ML FOR NUMERACY AND SOMALI LITERACY ASSESSMENTS

Numeracy Skill	BL Subtask Number	ML Subtask Number
Find missing numbers in list	3	1
Subtraction, 2 digits	5	2
Word problems (addition and subtraction)	6	3
Multiplication, 2 digits	8	4
Multiplication, 3 digits and decimals		5
Division, 2 digits by 1 digit	9	6
Division, 3 digits by 2 digits	10	7
Word problems (multiplication and division)	11	8
Percentages and decimals		9
Fraction operations		10
Word problem (equation of 1 variable)		11
Algebra with 1 variable		12
Somali Literacy Skill	**BL Subtask Number**	**ML Subtask Number**
Word identification	1	1
Reading fluency (difficult)	4	2
Reading comprehension (difficult)	5	3
Writing		4
Writing – fill missing words		5
Writing – fill missing words	6	6
Sequence of sentences in story		7
Reading comprehension		8
English Literacy Skill	**BL Subtask Number**	**ML Subtask Number**
Letter identification	1	1
Word recognition	2	2
Reading comp (easy)	3	3
Reading fluency	4	4
Reading comp (medium)	5	5
Reading comp (hard)	6	6
Writing (missing words)	7	7
Writing (negative form)	8	8
Writing (future tense)	9	9

In the section that follows, we discuss in detail the potential for ceiling effects, finding that ceiling effects will present a significant problem for the midline analysis, especially in the case of numeracy. Nonetheless, we recommend employing the overlapping subtask approach to analysis, for the simple reason that no other method will ameliorate the problem for baseline-to-midline analysis. Score-standardization, in particular, will not resolve the problem in our context, because the evaluation lacks a comparison group. Our best approach is to ensure through direct subtask-wise comparisons that the

matched, consistent subtasks are as similar as possible in terms of their difficulty from baseline to midline.

The following table summarizes the midline learning subtasks, including their correspondence with comparable midline subtasks, as well as a comparison of relative difficulty where subtasks are comparable between the baseline and midline. In order to assess differences in difficulty between baseline and midline, a set of girls in the pilot sample were given both the baseline and midline assessments. The table below presents subtask-level scores for girls who took both assessments, where their scores have been differenced, with baseline score being subtracted from midline score, such that a negative difference indicates that the midline subtask is potentially more difficult than the baseline subtask.

TABLE 18: ANALYSIS OF COMPARABLE BASELINE AND MIDLINE LEARNING SUBTASKS

Numeracy Skill	BL Subtask Number	ML Subtask Number	Difference in Pilot Data (ML – BL)	P-value of difference
Find missing numbers in list	3	1	1.8	0.44
Subtraction, 2 digits	5	2	3.2	0.34
Word problems (addition and subtraction)	6	3	2	0.77
Multiplication, 2 digits	8	4	1.2	0.74
Multiplication, 3 digits and decimals		5		
Division, 2 digits by 1 digit	9	6	1.6	0.67
Division, 3 digits by 2 digits	10	7	4	0.2
Word problems (multiplication and division)	11	8	-15	0.00*
Percentages and decimals		9		
Fraction operations		10		
Word problem (equation of 1 variable)		11		
Algebra with 1 variable		12		
Somali Literacy Skill	**BL Subtask Number**	**ML Subtask Number**	**Difference in Pilot Data**	**P-value of difference**
Word identification	1	1	-7.1	0.02*
Reading fluency (difficult)	4	2	-3.7	0.2
Reading comprehension (difficult)	5	3	-19	0.00*
Writing		4		
Writing – fill missing words		5		
Writing – fill missing words	6	6	-4	0.53
Sequence of sentences in story		7		
Reading comprehension		8		

English Literacy Skill	BL Subtask Number	ML Subtask Number	Difference in Pilot Data	P-value of difference
Letter identification	1	1		
Word recognition	2	2	5	0.12
Reading comp (easy)	3	3		
Reading fluency	4	4	-1.1	0.71
Reading comp (medium)	5	5	22	0.00*
Reading comp (hard)	6	6	-5	0.19
Writing (missing words)	7	7	3.5	0.35
Writing (negative form)	8	8	-8	0.21
Writing (future tense)	9	9	-8	0.04*

The table above shows that, for numeracy and English literacy, there are very few subtasks where there are statistically significant differences between baseline and midline scores, suggesting that the comparable subtasks are essentially similar in their difficulty. In the aggregate, the comparable scores are nearly equivalent between baseline and midline, and the remaining differences between the two are not statistically significant.

For Somali Literacy all of the four comparable subtasks have negative differences in the pilot data, indicating that all four subtasks were potentially more difficult in the midline. Two out of the four comparable subtasks have differences that are statistically significant, with subtask 3 having the largest gap in difficulty as well as the highest level of statistical significance. Ultimately, these problems with Somali score calibration certainly suggest that Somali Literacy, at midline was far more difficult than at baseline. Thus, interpretation of the Somali literacy results below should be cautious and be viewed in light of the calibration issues highlighted here.

Ceiling Effects in the Midline Evaluation

Beyond the issues raised above regarding the equivalence of learning assessments between baseline, midline and endline evaluation rounds, ceiling effects pose an additional threat to valid inferences in the analysis that follows. Ceiling effects occur when the full extent of a student's ability is not captured in the available range of an assessment. For instance, if a student could achieve a high score on a very difficult examination, they are likely to achieve a perfect score on a less difficult test; this process results in truncated scores for high-achieving students – whose scores are capped at 100 percent – and an underestimation of gains in learning outcomes from one round to another.[66] At an individual level, the same process produces a false equivalence between different types of students who achieve perfect scores – those who achieved a perfect score but would have struggled with more difficult tasks, and those who achieved a perfect score and could have performed well on more difficult tasks. Problematically, these two students are observationally equivalent.[67]

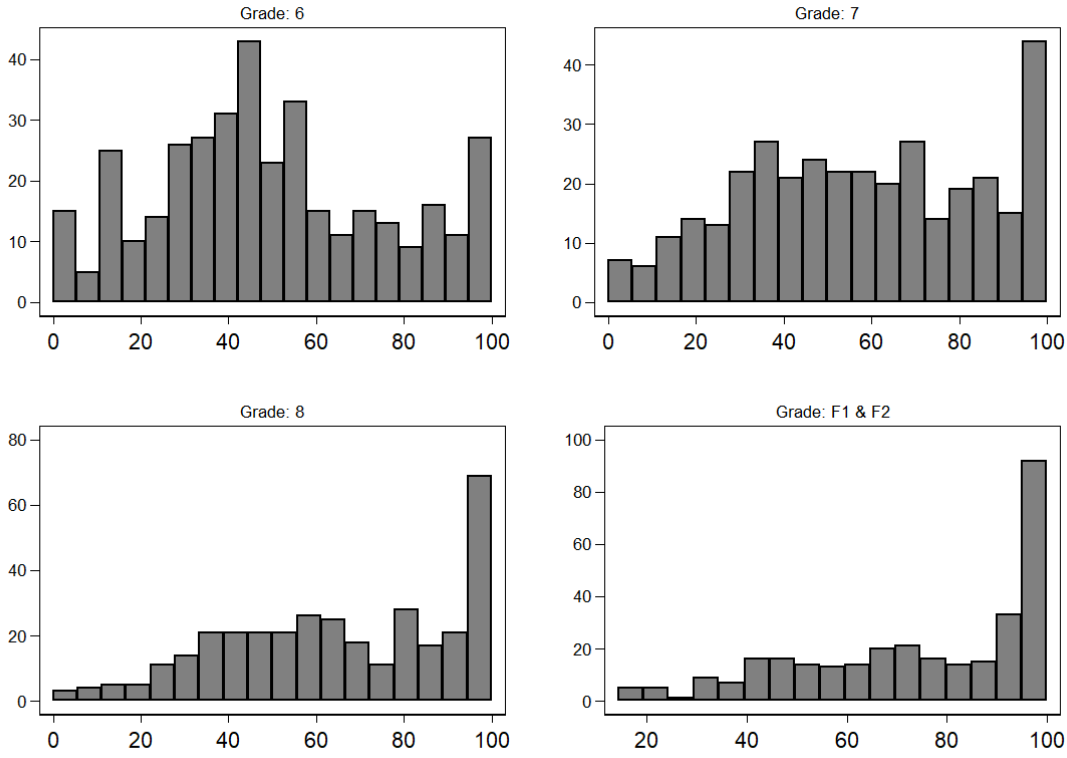
[66] Floor effects can have similar consequences to ceiling effects. However, because floor effects were not observed in the baseline evaluation, we focus the discussion exclusion on ceiling effects here.

[67] Individual-level observational equivalence of this kind is of less concern to us here, because our focus is on the extent to which ceiling effects influence estimates of aggregate outcomes. However, the individual-level issue highlighted does have minor consequences for our study of the Theory of Change, which investigates the relationship between the project's intermediate outcomes and learning scores. To the extent that individual

For our purposes, the role of ceiling effects are of concern because it could result in underestimating project impact or the change in learning scores from baseline to midline. Unsurprisingly, the removal of subtasks from the baseline scoring does little to improve the issue of ceiling effects in either the numeracy or Somali literacy subtasks. Using the full numeracy assessment, just over 20 percent of cohort girls achieved scores of 95 percent or higher; using the more limited set of comparable baseline subtasks described in the table, the share of girls receiving scores of 95 percent or higher is 16.9 percent. This constitutes the set of girls who are most likely to be impacted by ceiling effects during baseline-to-midline comparisons.[68] In the case of Somali literacy, 6.7 percent of girls are at risk of ceiling effects in the comparable assessment.

Baseline score distributions using the comparable set of subtasks for numeracy and Somali literacy, split by the respondent's grade, are provided in the two figures below. As we would expect, the threat of ceiling effects rises monotonically with grade level. To illustrate, consider the case of Somali literacy: 2.4 percent of Grade 6 (at baseline) girls scored 95 percent or higher, compared to 7.0 percent of Grade 8 girls and 14.5 percent of girls in Forms 1 and 2.

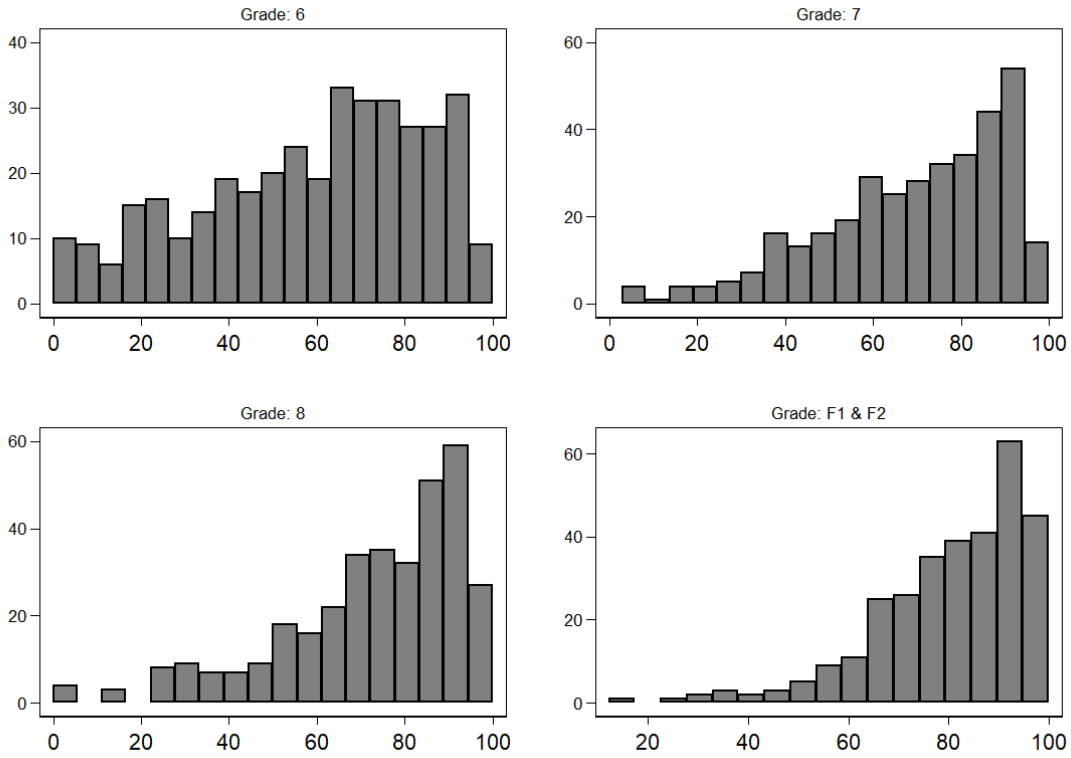
FIGURE 7: BASELINE NUMERACY SCORES USING COMPARABLE SUBTASKS, BY GRADE



scores are truncated, it can distort the relationships between demographic characteristics or intermediate outcomes, on one hand, and learning scores, on the other.

[68] Of course, at endline this issue will be less problematic, because the midline assessment has incorporated several more difficult subtasks. We will assess the potential for floor/ceiling effects during midline-to-endline comparisons in the full midline report.

FIGURE 8: BASELINE SOMALI LITERACY SCORES USING COMPARABLE SUBTASKS, BY GRADE



Unfortunately, even this discussion understates the potential threats to inference from ceiling effects. Unsurprisingly, older students performed better than younger students at the baseline, and therefore exhibited the highest potential for ceiling effects at the baseline. As the figures above clearly demonstrate, students in higher grades are much more likely to be affected by ceiling effects. For instance, consider a case in which we define scores of 95 percent or higher in the baseline as a threshold for potential ceiling effects, though we note that this definition almost certainly underestimates the number of students impacted.[69] Among Grade 7 girls, 11.7 percent fall above this cut-off for numeracy at the baseline; among Form 2 girls, we expect fully 32.9 percent of girls to be affected. To the extent that students in higher grades bump up against the highest achievable scores on the assessment in disproportionate numbers, it will reduce our estimates of project impact among older students and distort our conclusions regarding the distribution of project impact across grades.

The numeracy results show that ceiling effects are particularly pronounced among older girls. Numeracy is not an outlier in this regard: among Grade 7 girls, 13.2 percent achieved scores of 95 percent or higher in Somali literacy at the baseline, while 30.1 percent of Form 2 girls achieved the same distinction. Even in the context of English literacy, where overall ceiling effects and mean scores were lower, as many as 14.4 percent of Form 2 girls will be impacted by ceiling effects (though under 1 percent of Grade 7 girls meet our definition). Ceiling effects in numeracy and Somali literacy are

[69] We based this arbitrary threshold on the idea that a 5-point score increase is likely to be observed in many cases, and students achieving such an increase would be impacted by the score cap. However, given the expectation of growth effects from grade to grade, and the fact that average effects mask widespread individual-level variation, a more plausible threshold for identifying students potentially impacted by the score cap might have been 90 percent instead.

likely to influence our estimates of aggregate learning outcomes, because they are so widespread. But they are likely to have a particularly strong impact among older girls for the reasons outlined above.

To assess the precise impact ceiling effects are likely to have across different grade levels, we performed a Monte Carlo simulation in which we simulated midline learning scores on the basis of baseline learning scores and a set of plausible assumptions regarding the distribution of scores across girls. Specifically, we start from the actual baseline learning data, i.e. the baseline cohort girls included in the sample being analysed in this report. In a first simulation, we assumed a mean score improvement from baseline to midline of 4 percentage points, with score improvements distributed normally with a standard deviation of 8 points.[70] The standard deviation selected implies that 95 percent of girls in the simulation would score between -11.7 and 19.7 points *higher* at baseline than midline. While this is a wide range of potential score improvements, it is actually a narrower distribution than we have observed in other GEC and GEC-T evaluations.

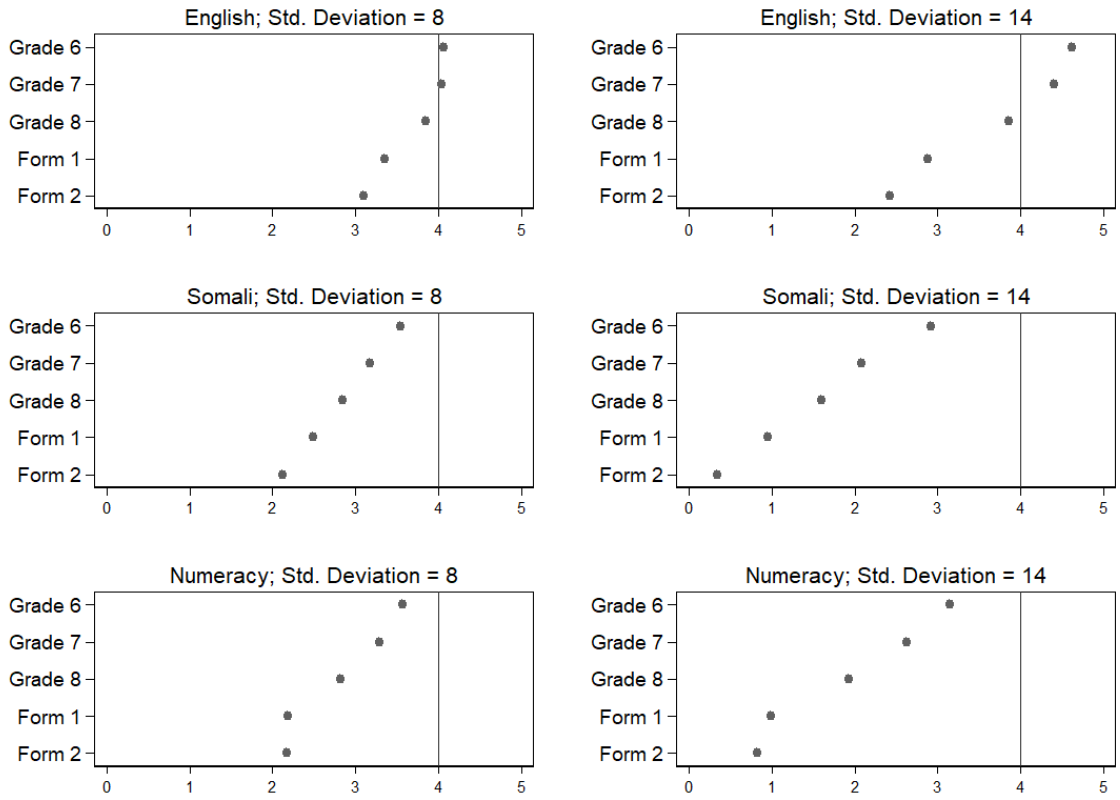
Our simulation consists of randomly assigning girls (n = 1,370) a score improvement, calculating the aggregate improvement from baseline to midline among girls in each grade, and repeating the exercise for 1,000 permutations. The essence of the simulation comes when we calculate aggregate improvements by grade, because we truncate scores to fit on the 0-100 scoring range. To be clear, our assumption regarding score improvements is that the simulated project produces a 4-point gain in *latent* learning, but that we can only observe gains over the 0-100 interval and may subsequently miss the full scope of improvement for some students. The disjuncture between latent and observed learning outcomes is precisely the problem presented by ceiling and floor effects.

The results of the simulation are presented in the left column of the figure below. In the top-left figure, we plot the average *observed* score improvement in English, by grade, across 1,000 permutations. The vertical line represents the average *latent* score improvement; in the complete absence of ceiling effects, our simulation would produce average observed score improvements of 4 points in each grade. Deviations from the 4-point line represent the effect of ceiling and floor effects. As the first results show, girls in Grades 6 and 7 produce score improvements very close to 4 points, which is consistent with the fact that few girls in these grades achieved sufficiently high scores at baseline to be influenced by the score ceiling. However, among older girls, where ceiling effects become more plausible, we observe a decline in estimated score improvements, with Form 2 girls showing a mean *observed* score improvement of just 3.09 points, despite a mean *latent* score improvement of 3.99 points in our simulation.

The results of our first simulation continue in the left-hand column, reporting observed score improvements for Somali literacy and numeracy. As the earlier discussion of aggregate scores would suggest, deviations from the expected 4-point gain are more pronounced in Somali and numeracy scores, because the potential for ceiling effects in these subject assessments is much greater. In the case of numeracy, the simulation produces score improvements of 3.6 points among Grade 6 girls, declining to just 2.2 points among Form 2 girls. In every case, the bias resulting from ceiling effects are more pronounced among older girls.

[70] A 4-point improvement in learning scores is slightly smaller than the target set by the project of 0.25 standard deviations, but is a plausible outcome at midline. For simplicity, we focus on the simplified comparison of girls from baseline to midline, without benchmarking, but the conclusions of the simulation are unlikely to be affected by a more expansive approach.

FIGURE 9: SIMULATION RESULTS – OBSERVED SCORE INCREASES, BY GRADE, IN THE PRESENCE OF CEILING EFFECTS



It is important to note that the results of our simulation stem from assumptions we made regarding the distribution of score improvements across girls (i.e. the mean score improvement and its variation from girl to girl). In the right-hand column of the figure above, we report the results of a second simulation, in which we adjusted a single parameter, increasing the variation in learning score gains from a standard deviation of 8 to a standard deviation of 14.[71] As the results in the right column show, the effects using this assumption are more dramatic: reductions in the observed change to just 1 point or less (compared to the true, latent change of 4 points) are possible in higher grades, and all grades show reductions in observed score changes in Somali literacy and numeracy.[72] Although we have discussed this simulation within the framework of substantive score changes, it is clear that ceiling effects can produce reductions in statistical power, because power is directly related to the target of our simulation – the observed effect size.

The learning results reported later in this section should be viewed In light of these findings. For example, if the learning results produce smaller-than-expected gains vis-à-vis the baseline, it is

[71] Briefly, this results in a fatter distribution, with a wider dispersal of score changes around the same mean of 4 points. Under the latter assumption, in which the standard deviation is 14, 95 percent of girls' score improvements are expected to fall between -23.4 and 31.4 points.

[72] The results in English literacy are complicated by the fact that girls in younger grades were likely to score in the lower range (under 30 points) in English literacy. In that context, and with the potential for large negative score improvements due to the higher standard deviation studied, floor effects may produce *greater* observed score improvements compared to actual, latent improvements. This issue is strictly limited to English literacy among younger girls, and is based on our assumptions regarding standard deviation. In general, our concern is much more focused on ceiling effects, as we described earlier in this section.

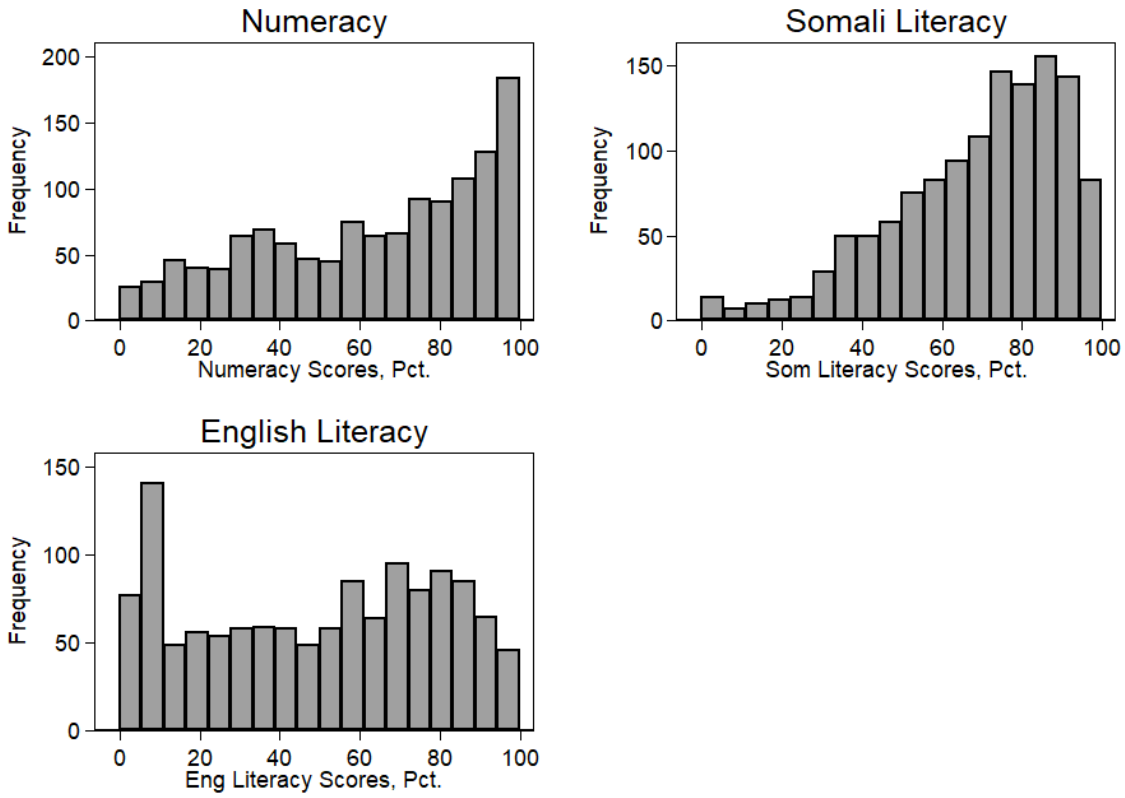
possible that ceiling effects are influencing the aggregate scores. However, this cannot act as a blanket justification for smaller-than-expected improvements. To the extent that ceiling effects are the culprit, we would specifically expect to see sizable gains among younger girls or those girls with lower baseline scores, but smaller gains among older girls and those whose baseline scores approached the score ceiling. Finally, note that we discuss the potential for ceiling effects going forward (i.e. the potential risk of ceiling effects in the endline evaluation) in Section 4.5, in which we report learning scores using the full midline assessments, rather than the set of baseline-to-midline comparable subtasks.

4.4 Comparable aggregate results

This section presents learning outcomes results on the basis of comparable learning scores. In all cases, the true-panel sample has been used (comprising only baseline respondents who were successfully re-contacted at midline). For ease of reference, the sample sizes for the panel sample for relevant learner types are: 996 cohort girls, 274 bursary girls, and 234 boys. We begin with a brief presentation of descriptive statistics so that readers are aware of the distributional properties of each comparable learning score at midline, as well as the grade-wise distribution of scores by learner type. We then present results that allow for an assessment of intervention-effects, including a grade-wise assessment of progress against baseline benchmarks (which we refer to as arithmetic difference in differences) as well as formal difference-in-differences analysis of differential intervention-effects across different learner types. This section concludes with an analysis of comparable subgroups between baseline and midline.

The following panel of graphs summarizes the distribution of numeracy, Somali literacy, and English literacy scores for the pooled sample of cohort and bursary girls belonging to the true panel. These graphs are unweighted in order to describe the true distribution of scores from the non-truncated sample. As noted in the analysis of methodological issues above, the comparable numeracy and Somali literacy scores both demonstrate significant ceiling effects, which were also present at baseline. On the other hand, English literacy scores are right-skewed (and without significant floor or ceiling effects) and can therefore be expected to be the most sensitive indicator in terms of detecting changes in learning since baseline.

FIGURE 10: DISTRIBUTION OF COMPARABLE LEARNING ASSESSMENT SCORES



At baseline, summary analysis of learning scores by grade and by learner type confirmed that scores increase monotonically by grade (a fundamental indicator of assessment validity) and that there was a significant learning gap between boys and girls with boys consistently outperforming girls. Both of these observations are true at midline, as will be established below.

The table below provides a grade-wise summary of midline numeracy scores for each learner type. Scores increase monotonically by grade for cohort girls (where the sample size is large enough to support grade-wise disaggregation), while scores are positively correlated with grade-level for bursary girls and boys, but increases are not monotonic due to lower sample sizes and thus higher variability as a result of sampling error. These results and all results below have been weighted such that sample characteristics are as aligned as possible with population characteristics, and such that each school contributes equally to the total score or result.[73]

TABLE 19: COMPARABLE MIDLINE NUMERACY SCORES BY LEARNER TYPE AND GRADE LEVEL

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	54.6	28.8	58.6	31.2	66.4	24.3
7 (8)	61.5	28.3	56.1	30.5	71.6	24.9
8 (F1)	62.4	25.8	59.4	29.7	74.2	24.3

[73] The weight variable is for analysis of learning panel data is panel_tweight.

Form 1 (F2)	67.5	27.7	81.6	22.1	79.0	21.1
Form 2 (F3)	74.0	26.0	74.3	26.0	87.7	20.0
Overall	**61.3**	**28.4**	**61.6**	**30.5**	**72.9**	**24.4**

The main observable trend in the table above is that boys consistently outperform both cohort and bursary girls at each grade-level and in the aggregate. It is also worth noting that the aggregate numeracy scores of cohort and bursary girls is nearly identical.

The table below provides a grade-wise summary of midline Somali literacy scores for each learner type. As with numeracy scores, Somali literacy scores increase monotonically by grade for cohort girls, while scores are positively correlated with grade-level for bursary girls and boys, but increases are not monotonic.

TABLE 20: COMPARABLE MIDLINE SOMALI LITERACY SCORES BY LEARNER TYPE AND GRADE LEVEL

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	62.3	21.4	61.7	25.3	62.5	25.2
7 (8)	67.5	20.0	64.7	22.6	67.5	22.1
8 (F1)	69.6	20.6	62.5	19.2	69.9	20.4
Form 1 (F2)	73.2	21.1	83.5	17.9	75.2	19.5
Form 2 (F3)	79.4	17.5	84.4	13.3	85.9	14.7
Overall	**68.0**	**21.0**	**66.9**	**23.5**	**69.1**	**22.8**

As with numeracy, boys aggregate scores in Somali literacy are higher than those of both cohort and bursary girls, but the gap in performance is much smaller for Somali literacy than for numeracy. The gender-gap in performance will be explored in greater detail below. In the aggregate, there is only a marginal difference between the Somali literacy scores of cohort girls vis-à-vis bursary girls.

The final table below provides a grade-wise summary of midline English literacy scores for each learner type. Consistent with the foregoing analysis, English literacy scores increase monotonically by grade for cohort girls and are positively correlated with grade-level for bursary girls and boys.

TABLE 21: COMPARABLE MIDLINE ENGLISH LITERACY SCORES BY LEARNER TYPE AND GRADE LEVEL

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	36.4	27.0	38.1	29.4	42.7	27.5
7 (8)	44.0	27.6	41.5	28.9	50.2	28.0
8 (F1)	55.2	26.9	49.2	30.4	61.5	28.8
Form 1 (F2)	67.7	24.2	63.4	23.0	65.2	30.8
Form 2 (F3)	70.1	21.9	59.1	22.4	82.4	22.3
Overall	**48.3**	**29.0**	**45.2**	**29.4**	**54.8**	**30.1**

As above, there is a gender gap in English literacy with boys outperforming both cohort and bursary girls. English literacy scores also exhibit a comparatively large gap between cohort and bursary girls, with cohort girls having higher scores than bursary girls both in the aggregate and by grade.

To summarize, we find that all learning scores increase monotonically by grade for cohort girls, and girls' learning scores lag behind those of boys in the same grade, across all grade levels and across all three assessments. The gendered learning gap here was also found at baseline and has clearly persisted at midline, despite interventions focused on closing this gap. Below, we will examine differences among cohort girls, bursary girls, and boys in greater detail using difference in differences analysis. For now, it will suffice to observe that, since the time of the baseline, bursary girls show some evidence of having closed the gap with cohort girls in both numeracy and Somali literacy, while there is no evidence of reduction in the learning gap between boys and cohort girls.

Progress against Benchmark using Arithmetic Difference in Differences

In the absence of a control or comparison group, we have used baseline grade-wise learning trajectories as a means of establishing benchmark expectations for how much a set of learners at a given grade-level would be expected to learn as a result of advancing to the next grade-level (in the absence of any intervention). This benchmark learning trajectory is then compared with the actual learning trajectory of a given set of learners at a given grade level. We refer to this calculation as *arithmetic difference in differences*.

This calculation has the following form, where *M* and *B* refer to midline and baseline scores, and G refers to the learner's baseline grade:

$$Arithmetic\ DiD = \left(y_{M,G} - y_{B,G}\right) - \left(y_{B,G+1} - y_{B,G}\right)$$

With *G* referring to the grade level at baseline, *G+1* naturally refers to the next grade level up. Thus, the first expression in parentheses is the actual progress since baseline (i.e. the difference between learners' scores at midline and their scores at baseline), while the second expression in parentheses establishes a benchmark expectation for progress in the absence of an intervention (i.e. the difference between learners' scores at baseline and the scores of learners in one grade above them at baseline).

In the tables below, we describe the first two terms $\left(y_{M,G} - y_{B,G}\right)$, e.g. for numeracy, as *Midline Numeracy* and *Baseline Numeracy*. The full, differenced expression is described as *Difference Baseline to Midline.* The third and fourth terms $\left(y_{B,G+1} - y_{B,G}\right)$ are described as *Baseline Benchmark Numeracy* and *Midline Benchmark Numeracy.* Their difference is described as *Benchmark Difference*. The difference in the two differences is in the final column.

Thus, to clarify and to aid in future replication, the following table describes the derivation of what we are referring to as the benchmarks, for each grade-level, below:

Grade	Baseline Benchmark Numeracy (Baseline grade)	Midline Benchmark Numeracy (Baseline grade +1)	Benchmark Difference
6 (7)	Baseline Grade 6 Score	Baseline Grade 7 Score	(Baseline Grade 7 Score – Baseline Grade 6 Score)

7 (8)	Baseline Grade 7 Score	Baseline Grade 8 Score	(Baseline Grade 8 Score – Baseline Grade 7 Score)
8 (F1)	Baseline Grade 8 Score	Baseline Grade F1 Score	(Baseline Grade F1 Score – Baseline Grade 8 Score)
Form 1 (F2)	Baseline Grade F1 Score	Baseline Grade F2 Score	(Baseline Grade F2 Score – Baseline Grade F1 Score)
Form 2 (F3)	Baseline Grade F2 Score	Baseline Benchmark Sample Grade F3 Score	(Benchmark Grade F3 Score – Baseline Grade F2 Score)

The table above makes it clear how this calculation can be carried out for learners at each grade-level, and the sections below present these results separately for cohort and bursary girls, and by each assessment type.

Progress for Cohort and Bursary Girls

This section presents an evaluation of learning for cohort and bursary girls comparing their progress since baseline against a benchmark. The results for cohort and bursary girls have been combined in this analysis at the request of the FM in order to increase the overall sample size for the assessment of intervention-effects. Preliminary analysis of arithmetic DiD results suggested that there were significant ceiling effects that were obscuring progress made vis-à-vis the grade-wise benchmarks established. In light of these concerns, we present arithmetic DiD analysis below with a truncated sample that excludes girls who scored higher than 85 percent on a given assessment at baseline (because these girls were the most likely to have exhibited minimal progress from baseline to midline as a result of ceiling effects). The tables below present the results of this analysis. In all cases, the conclusions about intervention effects are more favourable when using the truncated sample. All results of non-truncated analysis have been retained for reference and placed at the end of Annex 14.

It is important to bear in mind that the analysis below is confounded to a degree by the fact that the arithmetic DiD results are sensitive to biases stemming from differences in baseline versus midline assessment difficulty, as well as problems of ceiling effects for learners in higher grades. The problem of differences in assessment difficulty has been ameliorated (though not obviated) through the use of comparable subtasks only. These caveats should be borne in mind and they may help to explain some of the negative difference in differences found below.

The table below presents the grade-wise, arithmetic difference in differences for the combined sample of cohort and bursary girls in terms of their numeracy scores. Looking first at the single difference between baseline and midline scores among girls at a given grade-level, scores have increased from baseline to midline for girls at all grade levels. It is worth noting that, in the non-truncated scores girls at higher grade-levels showed little to no increase in score from baseline to midline as a result of ceiling effects. When we consider progress since midline vis-à-vis benchmark expectations, girls at each grade also increased their scores over and above the benchmark expectations established for that grade with positive difference in differences across all grades.

TABLE 22: TRUNCATED SAMPLE – COHORT & BURSARY GIRL NUMERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Numeracy	Midline Numeracy	Difference Baseline to Midline	Baseline Benchmark Numeracy (Baseline grade)	Midline Benchmark Numeracy (Baseline grade +1)	Benchmark Difference	Arithmetic Difference in Difference
6 (7)	44.6	52.8	8.2	44.6	48.2	3.6	4.6
7 (8)	48.2	56.5	8.2	48.2	50.0	1.8	6.5
8 (F1)	50.0	56.5	6.5	50.0	54.5	4.6	1.9
Form 1 (F2)	54.5	61.8	7.3	54.5	58.4	3.9	3.4
Form 2 (F3)	58.4	67.2	8.7	58.4	56.2	-2.2	10.9
Total	**48.4**	**56.3**	**7.9**	**48.4**	**50.8**	**2.4**	**5.5**

Note: N = 742 cohort girls and 210 bursary girls in the truncated sample

If we assume that the removal of the highest scoring girls for a given subtask at baseline does not bias the sample in ways that would impede our ability to make inferences about intervention effects on learning, this analysis suggests that there has been a substantial, positive intervention-effect on numeracy learning from baseline to midline. Each of the grade-wise differences between baseline and midline score has been tested for its statistical significance, and the differences at each grade-level are statistically significant at the 0.05 level, and the aggregate difference between baseline and midline numeracy scores is also statistically significant.

TABLE 23: SUMMARY OF NUMERACY RESULTS

Result	Details	Comments
Numeracy Baseline - Midline	Beta[74] = 5.25 Target = 4.99 Performance against target = 105%	Midline target of 4.99 was achieved and exceeded by 5%. This indicates a positive intervention-effect on numeracy learning.

In the aggregate, the difference in differences from the outcomes spreadsheet is 5.25, which exceeds the midline target of 4.99 by 5 percent. Thus, we can conclude that the midline target was achieved in numeracy learning and there was a positive intervention-effect.

The table below presents grade-wise arithmetic difference in differences for Somali literacy. Readers will recall (from the discussion of assessment difficulty above) that Somali literacy subtasks were not well calibrated in terms of their difficulty at midline. It is likely that increases in performance from baseline to midline have been muted due to this calibration issue. Nonetheless, with the exception of girls who were at Form 1 at baseline (F2 at midline), girls at each grade-level increased their average scores from baseline to midline. Difference in differences results are mixed, with girls at some grade

[74] Please note that Beta here is derived from the Outcomes Spreadsheet calculation, whereas the total DiD presented in the table above is derived from arithmetic DiD presented in the table. These two figures differ slightly, and this difference is most likely a result of differences in how values are rounded in the Outcomes Spreadsheet versus in the arithmetic DiD calculation presented in the table above.

levels exhibiting positive improvement vis-à-vis benchmark expectations, whereas girls in baseline grade 6 (7) and 8 (Form 1) showed zero or negative progress vis-à-vis benchmark expectations.

TABLE 24: TRUNCATED SAMPLE – COHORT & BURSARY GIRL SOMALI LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Somali Lit	Midline Somali Lit	Difference Baseline to Midline	Baseline Benchmark Somali Lit (Baseline grade)	Midline Benchmark Somali Lit (Baseline grade +1)	Benchmark Difference	Arithmetic Difference in Difference
6 (7)	54.6	59.6	5.0	54.6	59.6	5.1	0.0
7 (8)	59.6	64.3	4.7	59.6	58.7	-0.9	5.6
8 (F1)	58.7	63.6	5.0	58.7	70.3	11.6	-6.6
Form 1 (F2)	70.3	70.3	0.0	70.3	68.3	-2.0	2.0
Form 2 (F3)	68.3	74.7	6.4	68.3	70.3	2.0	4.4
Total	**59.1**	**63.7**	**4.6**	**59.1**	**62.2**	**3.0**	**1.5**

Note: N = 679 cohort girls and 194 bursary girls in the truncated sample

The findings above suggest that there has been a modest intervention-effect on learning since baseline. At three out of the five grade levels under consideration, the increase in mean score from baseline to midline is positive and statistically significant. The aggregate increase in score (across all grades) from baseline to midline is also positive and statistically significant.

TABLE 25: SUMMARY OF SOMALI LITERACY RESULTS

Result	Details	Comments
Somali Literacy Baseline - Midline	Beta[75] = 1.43 Target = 4.49 Performance against target = 32%	The midline target of 4.49 was not reached. Progress against the target is minimal, at 32%, indicating that there has not been a positive intervention effect on Somali literacy learning at midline.

The aggregate analysis presented in the table above suggests that the aggregate difference in differences based on the outcomes spreadsheet is 1.43, which is well below the midline target of 4.49. The midline target for Somali literacy was not achieved, and while there has been progress since baseline, there is no clear intervention-effect on Somali literacy learning as of this midline study. Due to the calibration issues noted above, we cannot conclude that there was no intervention effect on Somali literacy. It is possible that there were substantial improvements in performance since baseline that were muted by the fact that midline subtasks in Somali literacy were significantly more difficult than at baseline.

The table below presents grade-wise arithmetic difference in differences for English literacy. At all grade levels, there is a substantial increase in average scores from baseline to midline as well as a positive increase over and above benchmark expectations.

[75] As above, Beta here is derived from the Outcomes Spreadsheet.

TABLE 26: TRUNCATED SAMPLE – COHORT & BURSARY GIRL ENGLISH LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline English Lit	Midline English Lit	Difference Baseline to Midline	Baseline Benchmark English Lit (Baseline grade)	Midline Benchmark English Lit (Baseline grade +1)	Benchmark Difference	Arithmetic Difference in Difference
6 (7)	23.5	36.4	12.9	23.5	27.4	4.0	9.0
7 (8)	27.4	43.0	15.5	27.4	35.6	8.1	7.4
8 (F1)	35.6	52.5	17.0	35.6	51.2	15.6	1.3
Form 1 (F2)	51.2	64.2	13.0	51.2	51.1	-0.1	13.0
Form 2 (F3)	51.1	65.0	13.9	51.1	57.1	6.0	7.9
Total	**31.5**	**45.9**	**14.5**	**31.5**	**36.6**	**5.1**	**9.3**

Note: N = 917 cohort girls and 255 bursary girls in the truncated sample

The findings for English literacy provide strong evidence for a substantial intervention-effect on English learning. The learners in the truncated sample demonstrate statistically significant improvements from baseline to midline at each grade level as well as in the aggregate.

TABLE 27: SUMMARY OF ENGLISH LITERACY RESULTS

Result	Details	Comments
English Literacy Baseline - Midline	Beta[76] = 7.59 Target = 5.40 Performance against target = 141%	The midline target of 5.40 was achieved and exceeded by 41%. There is clear evidence of a positive intervention-effect on English literacy.

An aggregate analysis of difference in differences based on the outcomes spreadsheet suggests that there has been an increase of 7.59 percentage points from baseline to midline, which exceeds the midline target of 5.40 by 41 percent. This finding suggests that there has been a substantial intervention-effect from baseline to midline on English literacy learning. **Summary Discussion**

Taken as a whole, the findings above suggest that there have been positive intervention-effects and that midline targets have been achieved in both numeracy and English literacy. The midline target was not reached in Somali literacy in the aggregate, but there is nonetheless moderate evidence of positive progress vis-à-vis baseline expectations at some grade-levels.

There is a substantial amount of qualitative evidence to suggest that math and English have been the subjects emphasized in remedial classes, perhaps to the neglect of Somali, and thus the focus of remedial lessons may help to explain why learning has advanced in both numeracy and English literacy where learning is heavily attenuated in Somali literacy. Girls in Puntland explained: that they find the remedial classes particularly helpful, "because most of the students have difficulties in both Math and English, so they have repeated for us what we miss or don't understand."[77] Similarly, a female teacher of remedial classes in Puntland also explained that: "Many subjects are revised, but the most

[76] As above, Beta here is derived from the Outcomes Spreadsheet.
[77] FGD with girls in Puntland.

important and focused on subjects are English, Math, and sometimes Islamic studies."[78] The fact that Somali literacy is not receiving similar attention in remedial classes may help to explain why learning has attenuated in most Somali literacy skills since baseline. Qualitative findings from Somaliland similarly suggested that math and English have been emphasized while Somali has not been the subject of remedial classes.

Differential Intervention-Effects across Learner Types

This section uses a conventional difference in differences approach to explore the differential effects of intervention types on learning. In this analysis, differences in learner type are treated as fundamental differences in the combination of interventions or "dosage" that a given set of individuals received. In this setting, we assume that the two different learner types were on essentially parallel learning trajectories at baseline and that any significant differences in their learning trajectories at midline are likely to be a result of differences in the types of interventions that they received.[79] It should be noted that this analysis does not make use of a truncated learning sample in order to ensure that sample sizes are as large as possible to achieve adequate statistical power in the statistical DiD tests presented below.

Where this assumption of parallel trajectories holds, our analysis may be able to isolate the effect of key differences in interventions between different learner types. For cohort versus bursary girls, the key difference being isolated for analysis is receiving a bursary (condition for bursary girls) or not receiving a bursary (condition for cohort girls). For cohort girls versus boys, the key difference being isolated is benefitting from individual-level interventions and support as well as school- and community-level interventions (condition for cohort girls) versus not benefitting from individual-level interventions and only benefitting from school- or community-level interventions (condition for boys).

Effect of Bursaries

The effect of bursaries can be assessed by comparing changes in cohort girls learning scores over time vis-à-vis changes in bursary girls' scores over time. At baseline, cohort girls and bursary girls had very similar learning scores, with bursary girls having aggregate scores that were slightly lower than cohort girls, but not to a significant degree. It is also the case that bursary girls were selected on the basis of being particularly marginal based on demographic characteristics and barriers that they faced. Ideally, the granting of bursaries to bursary girls would, at minimum, compensate for their marginal status in a way that would allow them to keep pace with cohort girls in terms of their learning over time. It might also be the case that granting bursaries to bursary girls gives them an added advantage that leads them to surpass cohort girls in terms of their learning. While having bursary girls improve their learning faster than cohort girls is not an intended outcome, it is worth monitoring for this possibility.

The analysis below provides support for the hypothesis that girls who receive bursaries are, for the most part, able to keep pace with cohort girls despite the fact that bursary girls are more marginal in ways that might adversely affect their learning. There is no evidence of bursary girls out-pacing cohort girls in terms of their learning scores.

---

[78] Teacher interview in Puntland.

[79] It should be noted that assignment to different intervention groups was not random; rather, it was by design correlated with marginalization status and also potentially with learning outcomes. Thus, the assumption of parallel learning trajectories is not necessarily valid, and the violation of this assumption may lead to biased results.

The table below presents difference in differences analysis in numeracy scores for cohort girls vis-à-vis bursary girls. Changes in scores for bursary girls are subtracted from changes in scores for cohort girls, such that positive difference in differences indicate that cohort girls' scores have increased more since baseline than bursary girls' scores. The difference in differences by grade are mostly negative, and the aggregate difference in differences is -1.1 and is not statistically significant. This finding indicates that cohort girls and bursary girls are generally keeping pace with one-another in terms of numeracy learning, and it may be the case that bursary girls are gaining very slightly on cohort girls, such that their aggregate scores in numeracy are now nearly identical at midline.

TABLE 28: COHORT GIRLS VIS-À-VIS BURSARY GIRLS – NUMERACY

Grade	Baseline Cohort Numeracy	Midline Cohort Numeracy	Cohort Difference (Baseline-Midline)	Baseline Bursary Numeracy	Midline Bursary Numeracy	Bursary Difference (Baseline-Midline)	Difference in Difference
6 (7)	50.8	54.6	3.7	53.2	58.6	5.4	-1.7
7 (8)	58.0	61.5	3.5	52.9	56.1	3.3	0.2
8 (F1)	62.7	62.4	-0.2	65.7	59.4	-6.3	6.1
Form 1 (F2)	71.2	67.5	-3.7	79.5	81.6	2.1	-5.8
Form 2 (F3)	75.2	74.0	-1.1	65.0	74.3	9.2	-10.4
Aggregate	**59.5**	**61.3**	**1.8**	**58.7**	**61.6**	**2.9**	**-1.1**

The table below presents results of difference in differences for Somali literacy, and these results are nearly identical to the results for numeracy above. The grade-wise differences are predominantly negative, and the aggregate difference in differences is -2.0, which is not statistically significant. As above, this finding indicates that cohort girls and bursary girls are generally keeping pace with one-another in terms of Somali literacy learning. As above, there is some evidence that bursary girls, who scored somewhat lower than cohort girls at baseline, are catching up with cohort girls at the midline, although this change is not statistically significant.

TABLE 29: COHORT GIRLS VIS-À-VIS BURSARY GIRLS – SOMALI LITERACY

Grade	Baseline Cohort Somali Lit	Midline Cohort Somali Lit	Cohort Difference (Baseline-Midline)	Baseline Bursary Somali Lit	Midline Bursary Somali Lit	Bursary Difference (Baseline-Midline)	Difference in Difference
6 (7)	61.6	62.3	0.7	62.5	61.7	-0.8	1.5
7 (8)	69.3	67.5	-1.8	61.3	64.7	3.4	-5.1
8 (F1)	73.2	69.6	-3.7	70.9	62.5	-8.4	4.7
Form 1 (F2)	79.0	73.2	-5.8	82.7	83.5	0.9	-6.6
Form 2 (F3)	81.1	79.4	-1.7	75.3	84.4	9.1	-10.8
Aggregate	**69.6**	**68.0**	**-1.7**	**66.6**	**66.9**	**0.3**	**-2.0**

The table below presents the results of difference in differences for English literacy scores. While the results (and the foregoing analysis) suggest that both groups of girls have made significant progress

since the baseline, bursary girls have made comparatively less progress over time in terms of their English learning scores. Difference in differences are positive for each grade-level and in the aggregate. The aggregate difference in differences of 2.9 percent is not statistically significant.

TABLE 30: COHORT GIRLS VIS-À-VIS BURSARY GIRLS – ENGLISH LITERACY

Grade	Baseline Cohort English Lit	Midline Cohort English Lit	Cohort Difference (Baseline-Midline)	Baseline Bursary English Lit	Midline Bursary English Lit	Bursary Difference (Baseline-Midline)	Difference in Difference
6 (7)	24.4	36.4	12.0	26.5	38.1	11.6	0.4
7 (8)	29.1	44.0	14.9	29.2	41.5	12.3	2.6
8 (F1)	40.9	55.2	14.3	35.9	49.2	13.3	1.0
Form 1 (F2)	56.6	67.7	11.2	62.8	63.4	0.6	10.5
Form 2 (F3)	59.4	70.1	10.7	56.4	59.1	2.7	8.0
Aggregate	**35.2**	**48.3**	**13.1**	**34.9**	**45.2**	**10.2**	**2.9**

Assuming that the primary purpose of bursaries is to allow comparatively disadvantaged girls to keep pace with their peers, most of the evidence above suggests that this goal is being accomplished. In numeracy and Somali literacy, bursary girls are keeping pace with cohort girls, and show some evidence that they are closing the small learning gap that existed at baseline. The exception to this trend is found in English literacy scores, **where** cohort girls' scores have improved somewhat more since baseline than those of bursary girls, and the result is a small, but widening gap between the performance of cohort and bursary girls in terms of English literacy. The analysis above does not provide a clear reason for this gap, but merely allows for the identification of this potential problem. One potential explanation for differential progress in English literacy is that the barriers that bursary girls disproportionately face (because they were selected based on their marginality) may be particularly likely to impact their learning of English. Analysis of subgroups and barriers below will be attentive to factors that disproportionately affect English learning and that may therefore serve as an explanation for the divergence between cohort and bursary girls in terms of their English literacy scores.

Effect of Individual versus Community-level Interventions (and Gender)

This section involves a comparison of changes over time in the learning scores of cohort girls vis-à-vis boys. As with the comparison above, we are comparing a group that is expected to be comparatively marginal (i.e. cohort girls) with a group that is expected to face comparatively fewer barriers to learning (i.e. boys). At baseline, cohort girls scored lower than boys across grade levels and across learning assessments, which provides concrete evidence that cohort girls are at a comparative disadvantage vis-à-vis boys in terms of learning. At the same time, cohort girls will be presumably receiving a higher degree of benefit from planned interventions: cohort girls will receive individual benefits as well as school- and community-level benefits, whereas boys will only benefit from the more diffuse benefits at the school and community levels.

Given how baseline learning gaps and interventions come together with girls versus boys, the ideal outcome would be one in which the added interventions that cohort girls receive will help them to at least keep pace with, and perhaps also gradually overtake, boys in terms of their learning scores. The

evidence from difference in differences analysis below suggests that, at this evaluation point, girls are keeping pace with boys in all three learning subjects under consideration, but not overtaking them.

The table below presents difference in differences analysis of numeracy scores for cohort girls and for boys, with boys' progress over time being subtracted from girls' progress, such that positive difference in differences indicates that girls' scores increased more than boys' scores over time. The differences by grade are mostly positive, and the aggregate difference in differences is 1.0, indicating that on average cohort girls improved their numeracy scores slightly more than boys since the baseline. This difference is not statistically significant, and the main conclusion that can be drawn from these findings is that cohort girls are at least keeping pace with boys in term of numeracy scores, and may be gaining on them slightly.

TABLE 31: COHORT GIRLS VIS-À-VIS BOYS – NUMERACY

Grade	Baseline Cohort Numeracy	Midline Cohort Numeracy	Cohort Difference (Baseline-Midline)	Baseline Boys Numeracy	Midline Boys Numeracy	Boys Difference (Baseline-Midline)	Difference in Difference
6 (7)	50.8	54.6	3.7	58.5	66.4	7.9	-4.1
7 (8)	58.0	61.5	3.5	71.5	71.6	0.1	3.4
8 (F1)	62.7	62.4	-0.2	78.7	74.2	-4.4	4.2
Form 1 (F2)	71.2	67.5	-3.7	86.6	79.0	-7.6	3.9
Form 2 (F3)	75.2	74.0	-1.1	87.3	87.7	0.4	-1.5
Aggregate	**59.5**	**61.3**	**1.8**	**72.1**	**72.9**	**0.8**	**1.0**

The table below presents difference in differences in Somali literacy for cohort girls vis-à-vis boys. The aggregate difference in differences of 0.9 is not statistically significant, indicating that cohort girls are keeping pace with (but not overtaking) boys in terms of Somali literacy.

TABLE 32: COHORT GIRLS VIS-À-VIS BOYS – SOMALI LITERACY

Grade	Baseline Cohort Somali Lit	Midline Cohort Somali Lit	Cohort Difference (Baseline-Midline)	Baseline Boys Somali Lit	Midline Boys Somali Lit	Boys Difference (Baseline-Midline)	Difference in Difference
6 (7)	61.6	62.3	0.7	66.9	62.5	-4.4	5.1
7 (8)	69.3	67.5	-1.8	69.4	67.5	-1.9	0.2
8 (F1)	73.2	69.6	-3.7	69.5	69.9	0.3	-4.0
Form 1 (F2)	79.0	73.2	-5.8	84.0	75.2	-8.9	3.1
Form 2 (F3)	81.1	79.4	-1.7	82.9	85.9	3.0	-4.7
Aggregate	**69.6**	**68.0**	**-1.7**	**71.7**	**69.1**	**-2.5**	**0.9**

The table below presents difference in differences in English literacy for cohort girls vis-à-vis boys, and the conclusions that follow are identical to those found for numeracy and Somali literacy above. The aggregate difference in differences of 0.1 is extremely close to zero and is not statistically significant.

As above, cohort girls appear to be keeping pace with boys in terms of their English literacy scores, but cohort girls are clearly not closing the gap with boys.

TABLE 33: COHORT GIRLS VIS-À-VIS BOYS – ENGLISH LITERACY

Grade	Baseline Cohort English Lit	Midline Cohort English Lit	Cohort Difference (Baseline-Midline)	Baseline Boys English Lit	Midline Boys English Lit	Boys Difference (Baseline-Midline)	Difference in Difference
6 (7)	24.4	36.4	12.0	35.6	42.7	7.1	4.9
7 (8)	29.1	44.0	14.9	32.8	50.2	17.4	-2.5
8 (F1)	40.9	55.2	14.3	42.3	61.5	19.2	-4.9
Form 1 (F2)	56.6	67.7	11.2	64.9	65.2	0.2	10.9
Form 2 (F3)	59.4	70.1	10.7	64.2	82.4	18.2	-7.6
Aggregate	**35.2**	**48.3**	**13.1**	**41.9**	**54.8**	**12.9**	**0.1**

To briefly summarize, the aggregate difference in differences in learning scores between cohort girls and boys are very close to zero, inconsistent in the direction of their implied effects, and none are statistically significant. The overarching conclusion that can be drawn on the basis of these convergent results is that cohort girls are keeping pace with boys in terms of their improvements in learning over time, but there is no evidence that girls are closing the gap with boys thus far.

Comparison of Grade Levels Achieved at Baseline versus Midline

This section uses comparable assessment subtasks to report on girls and boys learning with reference to the development of key skills that are expected to be learned at a given grade level, as noted in national curricula. The coding of grade-level expectations in this section is based on an analysis of the curricula of Somaliland and Puntland that was performed at baseline. To understand the analysis undertaken and the results, it is important to briefly rehearse the context of curriculum development in Somalia, which was also detailed in the baseline report. First, EGEP-T schools fall under the jurisdiction of Ministries of Education in multiple jurisdictions (Somaliland, Puntland, and the Federal Government of Somalia, to name three); as a result, in principle, they adhere to different curricula. Second, at least one of the relevant jurisdictions – the Federal Government of Somalia – is currently in the process of developing a national curriculum, and no current curriculum was available for review at the time of this report.[80] Third, the curricula that *are* available concern mathematics and English literacy only; the targeted learning outcomes for specific grade levels in Somali are either not specified (in the case of Puntland) or are very vaguely defined (in the case of Somaliland).

[80] The reality on-the-ground with regard to educational curriculum is even more complicated and varied than this description implies. As we discuss briefly in Section 1.1 of this report, schools use widely-ranging materials to define their curricula, including curricula borrowed wholesale from other countries (primarily, but not exclusively, Kenya), and curricula adapted piecemeal from multiple disparate sources. To illustrate the extent to which schools vary, consider the aforementioned implementation of language-of-instruction policies: despite official guidance that primary schools should be instructed in Somali and secondary schools in English, a significant minority of primary schools are instructed in English, a significant share of secondary schools are instructed in Somali, and a small set of both types of schools are instructed in Arabic.

Given this context, the evaluation team sought to define grade level achievement according to the curricula that are well-documented – those of English and mathematics from Somaliland and Puntland. For more detail on the process by which skills were matched with subtasks, please see the baseline report. The following table provides a brief summary of the correspondence between subtasks and grade-levels as well as notes on the correspondence between comparable numeracy subtasks from baseline to midline.

TABLE 34: GRADE LEVEL STANDARDS FOR MATHEMATICS AND ENGLISH LITERACY

Grade Level Achieved	Mathematics Skills	English Literacy Skills
2	• Subtraction carrying one number (portion of BL ST 5, ML ST 2) • Addition and subtraction word problems with simple underlying arithmetic (BL ST 6, ML ST3) • Division of 2-digit number by 1-digit number (BL ST 9, ML ST 6)	• Letter identification (subtask 2)
3	• Subtraction carrying two numbers (portion of BL ST 5, ML ST 2) • Multiplication of 2-digit numbers (BL ST 8, ML ST 4) • Word problems with simple multiplication and division (BL ST 11, ML ST 8)	• Identification of basic words, e.g., classroom objects, foods, animals (subtask 1)
4	• Division of 3-digit number by 2-digit number (BL ST 10, ML ST 7)	• 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)
7	N/A	N/A
8	N/A	N/A

The tables below present the results of this analysis for numeracy and for English literacy, using the true panel sample without truncation (and with the application of weights). The tables show the proportion of girls and boys in a given grade whose performance places them at a given achievement level. For example, the first cell in the table below shows that 18.8 percent of sampled girls who were in grade 6 (7) performed at a grade 2 achievement level in numeracy as of the midline. In order to

facilitate comparison with baseline achievement levels, the difference between baseline and midline proportions is displayed in parentheses in each cell below the midline proportion. Thus, the first cell in the table below indicates (in parentheses) that the share of learners in grade 6 (7) who are performing at a grade 2 achievement level has increased by 0.5 percentage points since the baseline.

TABLE 35: GRADE LEVEL ACHIEVED IN MATHEMATICS BY GIRLS AND BOYS, BY GRADE

	Grade 6 (7)	Grade 7 (8)	Grade 8 (F1)	Grade F1 (F2)	Grade F2 (F3)
Cohort and Bursary Girls					
Grade 2 achieved	17.5 (-1)	17.6 (-9.3)	24.2 (-2.8)	29.3 (11.3)	24.5 (-5)
Grade 3 achieved	7.9 (5.1)	7.1 (3.8)	7.9 (3.4)	6.8 (2.5)	9.2 (0.1)
Grade 4 achieved	15.9 (6.8)	18.7 (5.6)	17.1 (-6.4)	25.1 (-10)	33.2 (4.6)

	Grade 6 (7)	Grade 7 (8)	Grade 8 (F1)	Grade F1 (F2)	Grade F2 (F3)
Boys					
Grade 2 achieved	39.2 (6.3)	21.8 (-4.6)	18.5 (-9.9)	27 (-2.6)	22.6 (0.9)
Grade 3 achieved	5.7 (-0.1)	9.3 (5)	12.6 (-0.3)	15.4 (6.3)	3 (-1.3)
Grade 4 achieved	16.9 (5.1)	31.1 (-0.4)	29.4 (-3.8)	35.2 (-16.9)	65.7 (5.6)

The tables above suggest that, while progress is being made and the share of learners at higher achievement levels is increasing, the vast majority of learners still have mathematics achievement levels that are well below their actual grade-level. Even at the level of secondary school, the majority of girls and at least one third of boys are still not performing at a grade 4 achievement level.

For both girls and boys, the share of learners at a grade 4 achievement level has increased since baseline among learners in grade 6 (7), and grade F2 (F3). At the same time, the share of learners at a grade 4 achievement level has decreased since baseline for girls and boys in grade 8 (F1) and grade F1 (F2). The reason for this decrease is unclear, but this decrease is consistent with the grade-wise analysis of scores presented above, where girls and boys in grade 8 (F1) and F1 (F2) showed a decrease in their aggregate numeracy scores between baseline and midline. This trend is explored in greater detail in the analysis of midline-only results below.

TABLE 36: GRADE LEVEL ACHIEVED IN ENGLISH BY GIRLS AND BOYS, BY GRADE

	Grade 6 (7)	Grade 7 (8)	Grade 8 (F1)	Grade F1 (F2)	Grade F2 (F3)
Cohort and Bursary Girls					
Grade 2 achieved	3.9 (3)	6.6 (5.9)	6.9 (3.2)	5 (3)	9.7 (6.1)
Grade 3 achieved	5.6 (4.2)	7.8 (6.1)	4 (1.8)	3.7 (-0.8)	5.9 (2.9)

Grade 4 achieved	3.1 (1.9)	3 (1.3)	10.4 (3.3)	5.7 (-4.4)	5.7 (-0.3)
Grade 5 achieved	2.8 (2.1)	7 (5.7)	8.2 (3.6)	11.2 (1.6)	15 (8.5)
Grade 6 achieved	1.3 (1.3)	1.7 (1.7)	5.1 (3.3)	12.9 (7.1)	11.9 (1.4)

	Grade 6 (7)	Grade 7 (8)	Grade 8 (F1)	Grade F1 (F2)	Grade F2 (F3)
			Boys		
Grade 2 achieved	3.8 (3.8)	2.9 (1.9)	12.5 (9.8)	0 (-3.6)	0 (-3.9)
Grade 3 achieved	10.4 (7.8)	3.8 (3)	5.5 (0.8)	0 (-2.6)	3.6 (-1.3)
Grade 4 achieved	4.1 (3.2)	9.2 (8)	4.1 (-2.1)	0 (-3.4)	0 (-6.5)
Grade 5 achieved	10.3 (10.3)	8.1 (6.6)	13.3 (9.7)	2.3 (2.3)	13.9 (13.9)
Grade 6 achieved	0 (-7.2)	7.7 (6.4)	13.5 (11.9)	19.2 (-5.2)	31.2 (14.2)

As with math, there is evidence of progress in English achievement levels, but the majority of boys and girls are performing below a grade 6 achievement level even though they have advanced beyond grade 6 in their schooling. Even among learners at the secondary school level, less than one fifth of girls had a grade 6 achievement level, and less than one third of boys had the same.

It is worth noting that, in terms of English achievement levels, there have been improvements in achievement for both girls and boys at nearly every grade level, including learners in grade 8 (F1). This finding contrasts with the findings for numeracy achievement levels above, where girls and boys who had just entered secondary school (at midline) had made negative progress in terms of their achievement levels since the baseline.

Comparison of Skill Gaps at Baseline versus Midline

This section examines foundational skill gaps on the basis of comparable midline scores, utilizing the weighted, true-panel sample, that has been truncated (85% threshold), with pooled results from cohort and bursary girls. This analysis provides the opportunity to explore learning gaps within the same sample used for the examination of arithmetic DiD analysis and the assessment of potential intervention effects above. Thus, the primary purpose of this analysis is to identify the skills where significant improvements have been made since baseline, thus providing further explanation for the positive intervention-effects identified in the analysis above. Below, a parallel analysis of learning gaps will be presented for the midline-only scores and sample in order to identify learning gaps at midline that need to be addressed in order to ensure that learners continue to progress in their skill acquisition.

The table below presents learning gaps for numeracy skills, showing the percentage of learners who have attained a given skill level (ranging from non-learner to proficient learner) on a given subtask. The midline percentage is the first number that appears in each cell, and the difference (with baseline percentage subtracted from midline percentage) is presented in parentheses within each cell under

the midline percentage. Thus, positive values in parentheses indicate an increase from baseline to midline in the proportion of learners who belong to a given category of attainment under a given subtask.

The table below suggests that the largest share if improvement has been in the proportion of proficient learners in 2 digit multiplication (Subtask 4). This is still a major learning gap, as will be shown in the midline-only results below, but the proportion of proficient learners has increased by 25 percentage points, within the truncated sample, from baseline to midline. This increase in multiplication proficiency is observed alongside a correspondingly large (23 percentage point) increase in proficiency in word problems that make use of multiplication and division. Together, these improvements in complex multiplication and multiplication-related word problems account for a significant share of the overall progress made since baseline. There have also been moderate increases in the proportion of proficient learners across nearly all other subtasks, which helps to explain why there has been a statistically significant increase in numeracy scores across all grade levels from baseline to midline.

TABLE 37: COMPARABLE GIRLS' FOUNDATIONAL NUMERACY SKILLS GAPS

Categories	Subtask 1 Missing Number	Subtask 2 Subtraction (2 digits)	Subtask 3 Word problems (add/subtract)	Subtask 4 Multiplication (2 digits)	Subtask 6 Division (2 digits x 1)	Subtask 7 Division (3 digits x 2)	Subtask 8 Word problems (multiplication/ division)
Non-learner 0%	0.8 (-2.1)	9.9 (-6.9)	8.9 (1.7)	34.9 (-19.6)	20.9 (-6.5)	49.6 (-12.4)	50.2 (-18)
Emergent learner 1%-40%	12.1 (8.2)	5.6 (-2.7)	5.6 (0.4)	8.9 (-4.4)	6.8 (-0.1)	8 (-7.6)	0 (0)
Established learner 41%-80%	28.6 (22.9)	25.1 (-4)	31.7 (-4.4)	17.5 (-0.9)	19 (-3.8)	16.2 (1.3)	15.5 (-4.7)
Proficient learner 81%-100%	58.5 (-29)	59.5 (13.7)	53.8 (2.3)	38.7 (25)	53.3 (10.4)	26.3 (18.7)	34.2 (22.7)
	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: N = 952 cohort and bursary girls in the truncated sample

The table below presents learning gaps for Somali literacy skills and shows that reading fluency (subtask 2) is the only skill for which substantial improvement has been made since the baseline. The primary learning gap at baseline and at midline is in reading comprehension (subtask 3), where there has been a near-zero increase in the proportion of proficient learners from baseline to midline. This gap will be subject to further discussion in the midline-only analysis below.

TABLE 38: COMPARABLE GIRLS' FOUNDATIONAL SOMALI LITERACY SKILLS GAPS

Categories	Subtask 1 Word identification	Subtask 2 Reading fluency (difficult)	Subtask 3 Reading comp (difficult)	Subtask 6 Writing (fill missing words)
Non-learner 0%	2.9 (0.5)	2.7 (-2.1)	6.1 (-7)	20 (-2.1)

Emergent learner 1%-40%	28.5 (4.7)	1.9 (-6.1)	9.9 (-1)	14.7 (-1.3)
Established learner 41%-80%	57.4 (-12.1)	20.2 (-27.2)	50.1 (6.9)	33.3 (11.6)
Proficient learner 81%-100%	11.2 (7)	75.1 (35.4)	34 (1.1)	32 (-8.2)
	100.0	100.0	100.0	100.0

Note: N = 873 cohort and bursary girls in the truncated sample

The table below presents learning gaps for English literacy skills showing moderate improvements in skills across the board (as in numeracy). The most substantial improvement from baseline to midline is in reading fluency (subtask 4), where the share of proficient learners has increased by 27 percentage points. The other noteworthy finding here is that, paralleling the analysis of Somali literacy above, there is very little improvement from baseline to midline in terms of proficiency in English reading comprehension. It is clear from this analysis that fundamental reading comprehension for younger learners in grade 6(7) is a skill that remains under-developed and that is blocking these learners from acquiring new and more complex skills in both Somali and English.

TABLE 39: COMPARABLE GIRLS' FOUNDATIONAL ENGLISH LITERACY SKILLS GAPS

Categories	English ST1 Letter identification	English ST2 Word recognition	English ST3 Reading Comp (easy)	English ST4 Reading Fluency	English ST5 Reading Comp (medium)	English ST6 Reading Comp (hard)	English ST7 Writing (missing words)	English ST8 Writing (negative form)	English ST9 Writing (future tense)
Non-learner 0%	7.8 (0.8)	18.7 (-2.5)	28 (-13.9)	32.7 (-9.8)	42.9 (-16.3)	53.4 (-18.4)	54.4 (-17.9)	64.5 (-18.6)	71.3 (-15.4)
Emergent learner 1%-40%	6.6 (4.1)	15.3 (-8.6)	5.9 (-3.2)	3.6 (-10)	6.6 (-5)	9.3 (3.5)	7.1 (-1.9)	0 (0)	0 (0)
Established learner 41%-80%	14.2 (7.8)	41.2 (-4.1)	34.5 (13.5)	16.7 (-7.2)	26.6 (9.4)	23.1 (9.7)	20.8 (9.7)	13.1 (4.4)	11 (4.4)
Proficient learner 81%-100%	71.4 (-12.6)	24.8 (15.2)	31.6 (3.6)	47.1 (27.1)	23.9 (11.8)	14.2 (5.2)	17.6 (10.1)	22.4 (14.2)	17.6 (11)
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: N = 1,172 cohort and bursary girls in the truncated sample

4.5 Midline-only aggregate results

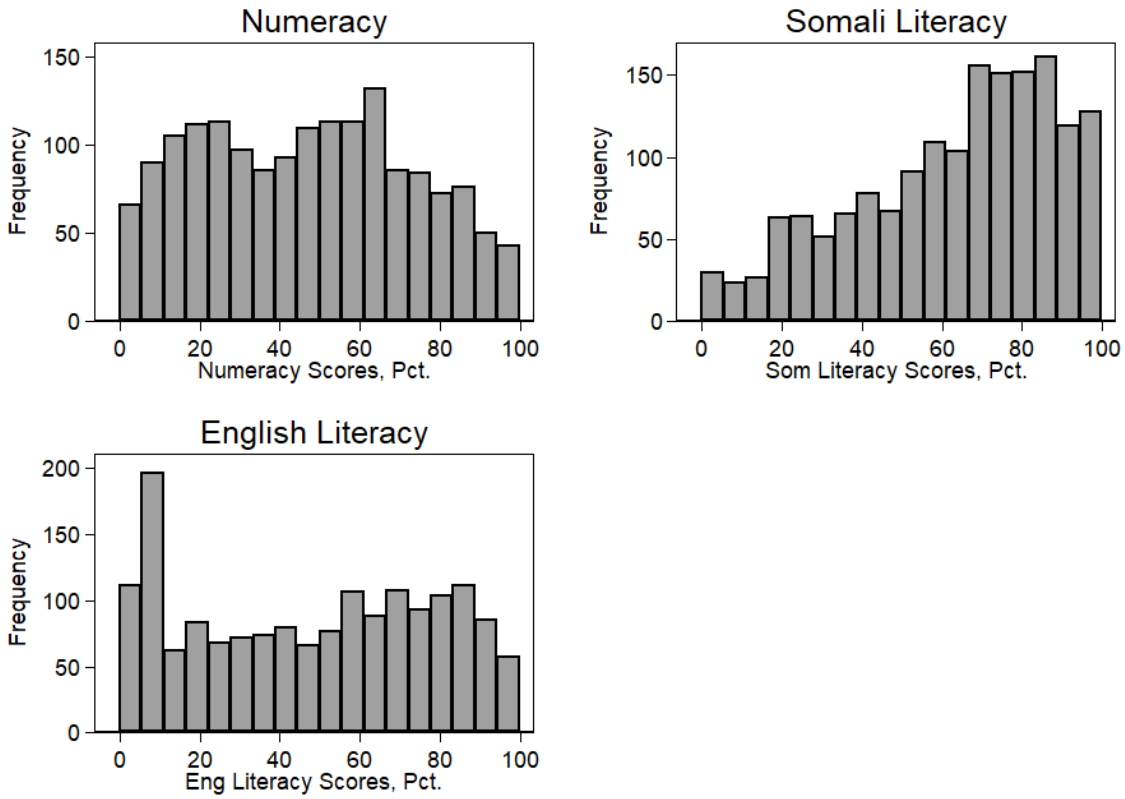
As mentioned in our analysis of methodological issues above, the midline learning assessments in numeracy and Somali literacy were made more difficult at midline in order to ameliorate some of the severe ceiling effects found at baseline. While the foregoing analysis has focused on only comparable subtasks (between baseline and midline), this section summarizes learning scores on the basis of the full array of midline subtasks. This midline-only analysis is necessary in order to establish a basis for comparison for future evaluation points. Thus, the analysis in this section uses complete midline scores (including harder subtasks added at the midline) as well as the cross-sectional sample (which includes replacement girls). For ease of reference, the aggregate learning scores for the full cross-sectional sample are summarized, along with the relevant sample sizes, in the table below:

TABLE 40: AGGREGATE NON-COMPARABLE MIDLINE LEARNING SCORES BY LEARNER TYPE

Learner Type	Assessment	Non-Comparable Midline Score	N (Cross-sectional sample size)
Cohort Girls	Numeracy	47.1	1316
	Somali Literacy	62.8	1316
	English Literacy	49.0	1316
Bursary Girls	Numeracy	46.6	339
	Somali Literacy	63.4	339
	English Literacy	46.0	339
Boys	Numeracy	55.4	318
	Somali Literacy	66.5	318
	English Literacy	56.0	318

The panel of graphs below summarizes the distribution of the midline scores of the pooled sample of cohort and bursary girls, as well as boys, using the full cross-sectional sample (giving a total sample size of N = 1,973.

FIGURE 11: DISTRIBUTION OF (NON-COMPARABLE) MIDLINE LEARNING ASSESSMENT SCORES



The graphs above show that attempts to reduce ceiling effects in numeracy and Somali literacy (by adding more difficult subtasks at midline) have been successful, and there are no evident problems with floor effects. At baseline, 11.8 percent of sampled learners achieved perfect scores in numeracy, whereas only 1.8 percent of learners achieved perfect scores at midline. At baseline, mild ceiling effects were also found in Somali literacy, with 3.2 percent of respondents achieving a perfect score. At midline, only 0.9 percent of learners achieved a perfect score.

The table below presents grade-wise numeracy scores by learner-type for the cross-sectional sample. These results are for the full, non-comparable numeracy score that includes harder subtasks added at midline. As expected, scores increase monotonically by grade for each learner type. In keeping with the analysis of comparable scores above, boys' numeracy scores are higher than girls' scores at each grade level and in the aggregate.

TABLE 41: GRADE-WISE NON-COMPARABLE MIDLINE NUMERACY SCORES BY LEARNER TYPE

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	38.1	22.6	40.8	26.3	44.3	22.3
7 (8)	42.4	24.0	39.4	23.8	50.2	24.0
8 (F1)	45.3	25.6	45.0	26.3	57.0	22.5

Form 1 (F2)	55.8	26.9	59.0	29.1	64.7	25.4
Form 2 (F3)	59.1	27.6	55.0	31.9	72.0	26.0
Overall	**45.0**	**25.6**	**44.6**	**27.2**	**53.9**	**24.9**

The table below presents grade-wise Somali literacy scores by learner-type for the cross-sectional sample. These results are for the full, non-comparable Somali literacy score that includes harder subtasks added at midline. Scores are positively correlated with grade, but do not increase monotonically because both cohort girls and bursary girls in baseline grade 8 (midline F1) have scored lower on average than their peers in baseline grade 7 (8). This trend of attenuated learning among cohort girls in baseline grade 8 (F1) was also observable, albeit much less severe, in the analysis of comparable learning scores above, and was particularly evident when examining arithmetic difference in differences results for comparable scores.

On its own, this grade-wise analysis does not suggest a clear explanation for why girls in baseline grade 8 (F1) are under-performing. One hypothesis is that girls in baseline grade 8 (F1) were at a major transition point from primary to secondary school and that the transition has been somewhat disruptive in terms of their learning. The subtask analysis below will consider the degree to which the lack of mastery of specific skills may have contributed to this result. The other aspect worth noting here is that boys in baseline grade 8 (F1) were also in the same schools and undergoing the same transition as cohort girls but did not experience this same attenuation of learning. It is also possible that girls are marginal vis-à-vis boys in ways that have had a particularly detrimental effect on their learning during the critical transition from primary to secondary school.

TABLE 42: GRADE-WISE NON-COMPARABLE MIDLINE SOMALI LITERACY SCORES BY LEARNER TYPE

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	55.9	25.0	56.8	29.0	57.7	27.0
7 (8)	61.2	23.3	60.7	25.7	61.5	25.1
8 (F1)	61.2	25.2	58.6	24.5	70.0	20.8
Form 1 (F2)	67.7	24.0	75.1	23.7	71.1	23.8
Form 2 (F3)	74.7	21.7	79.1	18.1	79.0	25.0
Overall	**61.7**	**24.7**	**62.2**	**26.6**	**65.3**	**25.3**

The table below presents grade-wise English literacy scores by learner-type for the cross-sectional sample. Unlike the numeracy and Somali literacy scores presented in this section, English literacy scoring makes use of identical subtasks to the comparable score presented earlier. The only difference here is that these results include the full cross-sectional sample (with replacements) while the comparable results presented earlier only included the panel sample. As expected, scores ascend monotonically by grade, and boys' English scores are higher than girls' scores at each grade level and in the aggregate.

TABLE 43: GRADE-WISE MIDLINE ENGLISH LITERACY SCORES BY LEARNER TYPE (CROSS-SECTION)

Grade	Cohort Girls	Standard Deviation for Cohort Girls	Bursary Girls	Standard Deviation for Bursary Girls	Boys	Standard Deviation for Boys
6 (7)	37.4	26.7	38.6	30.5	41.1	26.8
7 (8)	44.5	27.2	41.4	29.0	48.4	29.4
8 (F1)	46.6	29.4	44.6	29.6	60.4	28.2
Form 1 (F2)	66.4	24.3	58.6	26.6	65.8	29.8
Form 2 (F3)	69.9	22.2	65.2	23.4	77.1	26.8
Overall	**47.8**	**29.0**	**45.3**	**30.0**	**54.1**	**30.3**

Summary of Differences in Learning Scores by Learner Type

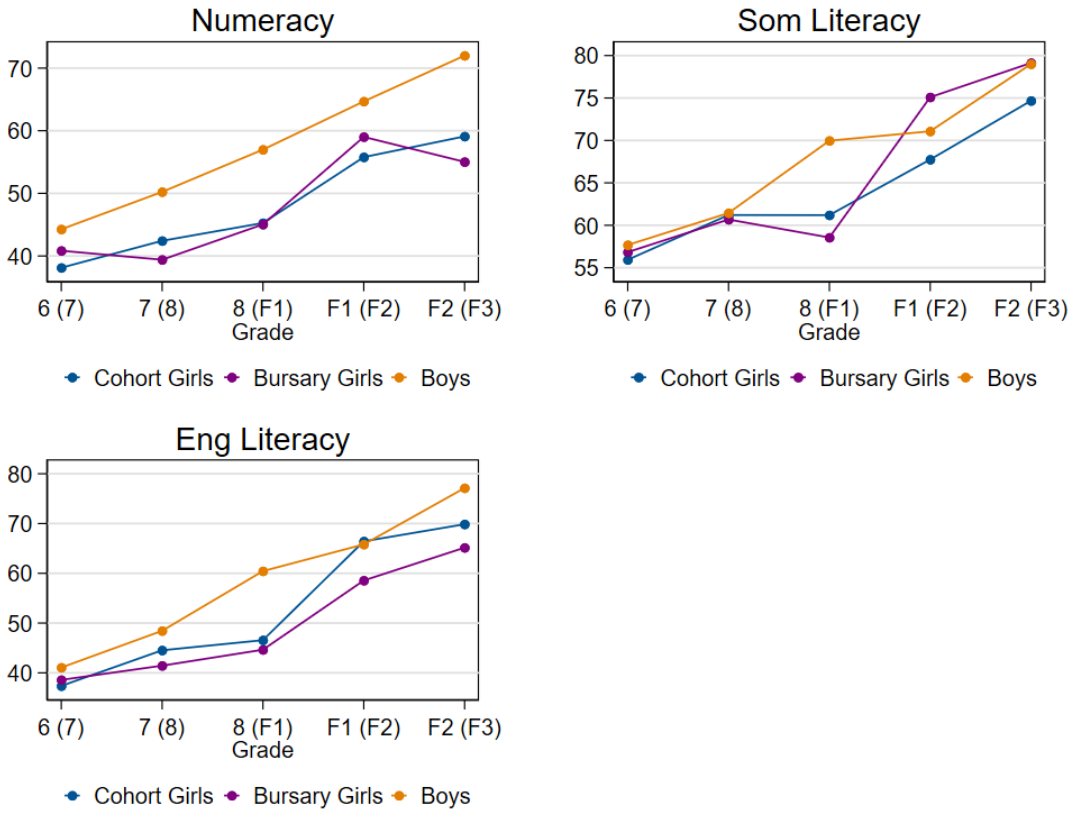
At midline, using the cross-sectional sample and non-comparable scores, boys' learning scores are significantly higher than cohort girls' scores across all three assessments. Numeracy scores exhibit the largest and most consistent gap between girls' and boys' scores. In the aggregate, boys averaged scores were 8.3 percentage points higher than those of cohort girls. This gap was statistically significant at the baseline and remains so at the midline.[81] In Somali literacy, boys' scores also exceed girls' scores in the aggregate, but by a much smaller margin (3.7 percentage points) than that observed for numeracy scores, although this difference is still statistically significant.[82] In English literacy, boys' aggregate scores are 6.7 percentage points higher than cohort girls' aggregate scores, and this difference is also statistically significant.[83]

This general trend of boys outperforming girls suggests that there is still significant room for girls to improve vis-à-vis boys, and this finding is consonant with the difference in differences analysis above that suggested that girls are generally keeping pace with boys in terms of learning scores, but have not been gaining on boys since the baseline. Beneath this consistent aggregate-level trend, there are important differences in grade-wise learning trajectories that deserve further examination.

The graphs below provide a visual overview of the findings presented in the tables above in order to better facilitate comparisons of differences in the grade-wise learning trajectories of different learner types.

[81] For the midline cross-sectional sample, boys' non-comparable numeracy scores are significantly higher than girls' scores at p = 0.00 in a linear regression that is adjusted for clustering at the school level.
[82] For the midline cross-sectional sample, boys' non-comparable Somali literacy scores are significantly higher than girls' scores at p = 0.01 in a linear regression that is adjusted for clustering at the school level.
[83] For the midline cross-sectional sample, boys' English literacy scores are significantly higher than girls' scores at p = 0.00 in a linear regression that is adjusted for clustering at the school level.

FIGURE 12: DISTRIBUTION OF (NON-COMPARABLE) MIDLINE LEARNING ASSESSMENT SCORES



The critical trend that these graphs highlight is the major inflection point at baseline grade 8 (F1) for both cohort and bursary girls in terms of their Somali and English literacy scores. This inflection point, as noted above, involves grade 8 (F1) girls under-performing relative to benchmark expectations as well as under-performing relative to their peers in grade 7 (8) at midline. In contrast, boys do not exhibit this trend at all, showing steady and monotonic increases in learning across grade-levels. Given this marked difference between boys' and girls' learning trajectories, the analysis of foundational skill gaps below will be particularly attentive to gender-based differences in Somali and English literacy skill gaps that might help to explain this finding.

Analysis of Midline Foundational Skill Gaps

This section examines foundational skill gaps at midline, using the full set of non-comparable midline subtasks, with a prospective focus on establishing the main skill gaps that need to be addressed by the project moving forward.[84] In addition, the foregoing analysis has suggested that there is a critical divergence in boys' and girls' learning trajectories at grade 8 (F1). This divergence in performance may be explicable in terms of differences in the subtasks at which girls and boys are experiencing their largest skill gaps, and the analysis below will comment on skill gaps that may explain divergent learning trajectories. This section examines girls' skill gaps in isolation, followed by boys' skill gaps, and then presents a comparative analysis of boys versus girls' skill gaps in the final subsection.

[84] In light of the significant ceiling effects observed at baseline, a longitudinal analysis of learning gaps by subtask is unlikely to yield useful results.

Girls' Skill Gaps

Our analysis of girls' foundational skill gaps uses the cross-sectional sample of girls and pooling the samples of cohort and bursary girls. The total sample size for this analysis is thus, N = 1,655 girls at midline.

The table on the following page presents the percentage of girls performing at a given proficiency level in numeracy, by subtask. The majority of girls are proficient in subtraction and word problems involving simple arithmetic. Proficiency levels decrease moderately as the difficulty of the subtasks increase, up to the point of subtask 4, which is multiplication of medium difficulty.

The primary learning gap that emerges for numeracy is at subtask 5, which involves harder multiplication (with 3-digit numbers and decimals). From subtask 4 to subtask 5, the proportion of proficient learners drops by 22.4 percentage points, and proportion of non-learners increases by 31.4 percentage points, such that the majority of girls fall into the non-learner category for subtask 5. Harder division tasks are also associated with this same skill gap, but the gap is somewhat less severe with reference to harder division problems than for harder multiplication problems. This finding suggests that even girls who are proficient in basic multiplication and division are having difficulty applying their skills to more complex problems with higher numbers of digits and with decimals.

The second major learning gap in numeracy is at subtasks 10 and 12, which deal with fraction operations and with single-variable algebra, respectively. At subtasks 10 and 12, the proportion of proficient learners drops below 10 percent, and nearly three quarters of girls fall into the category of non-learners for these subtasks. These subtasks deal with advanced skills that a large proportion of girls have not learned yet at their grade-level, so these learning gaps are not particularly problematic at this point.

TABLE 44: GIRLS' FOUNDATIONAL NUMERACY SKILLS GAPS

Categories	Subtask 1 Missing Number	Subtask 2 Subtraction (2 digits)	Subtask 3 Word problems (addition/subtraction)	Subtask 4 Multiplication (2 digits)	Subtask 5 Multiplication (3 digits and decimals)	Subtask 6 Division (2 digits x 1)	Subtask 7 Division (3 digits x 2)	Subtask 8 Word problems (multiplication/ division)	Subtask 9 Percentages and decimals	Subtask 10 Fraction operations	Subtask 11 Word problem (equation of 1 variable)	Subtask 12 Algebra with 1 variable
Non-learner 0%	0.6	8.4	8.2	30.7	63.0	18.7	44.0	45.3	57.4	73.7	72.3	76.4
Emergent learner 1%-40%	11.2	5.0	4.9	9.2	0.0	6.2	8.2	0.0	7.3	7.4	0.0	10.2
Established learner 41%-80%	25.0	23.1	31.2	18.0	16.7	18.4	17.0	16.2	17.1	12.3	14.5	7.0
Proficient learner 81%-100%	63.1	63.5	55.7	42.1	20.3	56.6	30.7	38.5	18.2	6.6	13.2	6.4
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

The table below presents the percentage of girls performing at a given proficiency level in Somali literacy, by subtask. The first finding that is worth observing is that word identification at subtask 1 was far more difficult than expected, with only 14 percent of girls being proficient at word identification (although the majority of girls were still at the established learner level with regard to word identification). In contrast proficiency in reading fluency was at nearly 80 percent.

TABLE 45: GIRLS' FOUNDATIONAL SOMALI LITERACY SKILLS GAPS

Categories	Subtask 1 Word identification	Subtask 2 Reading fluency (difficult)	Subtask 3 Reading comp (difficult)	Subtask 4 Writing	Subtask 5 Writing (fill missing words)	Subtask 6 Writing (fill missing words)	Subtask 7 Sequence of sentences in story	Subtask 8 Reading comprehension
Non-learner 0%	2.6	2.3	5.1	21.8	14.6	16.5	33.1	31.5
Emergent learner 1%-40%	24.6	1.8	9.0	18.9	11.9	13.6	13.2	12.6
Established learner 41%-80%	58.7	17.5	46.8	21.6	21.6	34.3	20.8	31.6
Proficient learner 81%-100%	14.1	78.4	39.1	37.6	52.0	35.6	32.9	24.3
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

The main learning gap that emerges in Somali literacy is between reading fluency (subtask 2) on one hand, and reading comprehension and writing (subtasks 3 and 4) on the other hand. While the majority of girls were proficient at difficult reading (involving reading short stories out loud in subtask 2), their comprehension of what they read was much lower, with only 41.5 percent of girls being proficient at reading comprehension. Their writing abilities were even lower than their reading comprehension abilities, with only 39.2 percent of girls being proficient at writing, and with the proportion of non-learners jumping by over 15 percentage points when moving from reading-related subtasks to writing.

The table on the following page presents foundational skill gaps for English literacy. As with Somali literacy, learners have difficulty with word identification or word recognition in English. While nearly three quarters of girls are proficient in the identification of letters, only about one quarter of girls are proficient in recognizing common words. In both Somali and English literacy, the ability of girls to recognize words lags behind their reading fluency, which suggests that there may be some need for review of phonetics and reading fundamentals. The main reason why girls might be able to perform better on reading involving stories than on reading random words is that when girls are reading stories, they are better able to identify words using contextual clues from a given sentence. Lack of mastery of the ability to identify and read unfamiliar words is an important and consistent learning gap (in Somali and English), and remedial work on these skills may be necessary in order for girls to continue to acquire more complex skills involving writing.

TABLE 46: GIRLS' FOUNDATIONAL ENGLISH LITERACY SKILLS GAPS

Categories	English ST1 Letter identification	English ST2 Word recognition	English ST3 Reading Comp (easy)	English ST4 Reading Fluency	English ST5 Reading Comp (medium)	English ST6 Reading Comp (hard)	English ST7 Writing (missing words)	English ST8 Writing (negative form)	English ST9 Writing (future tense)
Non-learner 0%	8.8	16.7	24.5	28.3	38.1	47.8	48.4	59.5	67.8
Emergent learner 1%-40%	6.8	14.0	5.4	3.6	6.4	9.5	8.3	0.0	0.0
Established learner 41%-80%	13.5	42.3	34.8	16.8	29.8	28.5	23.8	15.8	12.8
Proficient learner 81%-100%	70.9	27.0	35.3	51.3	25.8	14.2	19.4	24.6	19.4
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Also in keeping with the findings related to Somali literacy above, only about 38 percent of girls in the sample are proficient in English reading comprehension (subtask 3). We can think of English reading comprehension as also being the primary learning gap in English literacy, with less than half of girls being proficient at easy reading comprehension. As with Somali literacy, reading comprehension lags well behind reading fluency, with a reduction of 14.9 percentage points between the proportion of girls who are proficient in reading fluency (subtask 4) and the proportion of girls who are proficient in answering easy comprehension questions based on the material that they read (subtask 3).

As expected, the proportion of proficient learners continues to decrease in an almost linear fashion as the difficulty of the reading comprehension subtasks increases. This monotonic decrease can be seen across subtasks 3, 5, and 6. This is not a learning gap, per se, but rather a further illustration that reading comprehension skills are pivotal and that girls who have not acquired more fundamental comprehension skills cannot hope to advance onward to more complex comprehension skills.

Boys' Skill Gaps

This section presents an analysis of boys' foundational skill gaps that parallels the analysis of girls' skill gaps above. This analysis allows for the use of boys' learning patterns as a means of contextualizing and better understanding the places where girls are struggling most.

The table on the following page presents the percentage of boys performing at a given proficiency level in numeracy, by subtask. For all numeracy subtasks the share of boys who are proficient learners exceeds the share of girls who are proficient learners. This finding is consistent with findings above related to the large and statistically significant gap between boys' and girls' numeracy scores in the aggregate and at each grade level.

Beginning with the most fundamental subtasks, approximately three quarters of boys are proficient in subtraction and word problems involving simple arithmetic. Proficiency levels decrease moderately as the difficulty of the subtasks increase, up to the point of subtask 4, which is multiplication of medium difficulty.

As with girls, the primary learning gap that emerges for boys in numeracy learning is at subtask 5, which involves harder multiplication (with 3-digit numbers and decimals). From subtask 4 to subtask 5, the proportion of proficient learners decreases by 26.1 percentage points, and the corresponding share of non-learners increases by 37.7 percentage points.

The second major learning gap in numeracy for boys is also parallel to the main learning gap for girls and is at subtasks 10 and 12, which deal with fraction operations and with single-variable algebra, respectively. At subtasks 10 and 12, the proportion of proficient learners drops to approximately 10 percent, and well over half of boys fall into the non-learner category for subtasks 10 and 12.

TABLE 47: BOYS' FOUNDATIONAL NUMERACY SKILLS GAPS

Categories	Subtask 1 Missing Number	Subtask 2 Subtraction (2 digits)	Subtask 3 Word problems (addition/subtraction)	Subtask 4 Multiplication (2 digits)	Subtask 5 Multiplication (3 digits and decimals)	Subtask 6 Division (2 digits x 1)	Subtask 7 Division (3 digits x 2)	Subtask 8 Word problems (multiplication/division)	Subtask 9 Percentages and decimals	Subtask 10 Fraction operations	Subtask 11 Word problem (equation of 1 variable)	Subtask 12 Algebra with 1 variable
Non-learner 0%	0.7	3.3	2.7	17.2	55.8	9.8	30.8	27.4	49.6	70.9	66.0	67.7
Emergent learner 1%-40%	6.9	2.9	2.0	6.8	0.0	4.1	9.8	0.0	7.4	6.8	0.0	10.4
Established learner 41%-80%	20.8	16.9	23.6	24.7	18.9	12.5	18.9	19.4	11.5	12.2	15.4	12.7
Proficient learner 81%-100%	71.6	76.8	71.6	51.2	25.3	73.5	40.5	53.3	31.5	10.0	18.6	9.2
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

The table below presents the percentage of boys performing at a given proficiency level in Somali literacy, by subtask. As with girls, boys found word identification at subtask 1 to be far more difficult than expected, with only 19.5 percent of boys being proficient at word identification while 77 percent were proficient in reading fluency (subtask 2).

TABLE 48: BOYS' FOUNDATIONAL SOMALI LITERACY SKILLS GAPS

Categories	Subtask 1 Word identification	Subtask 2 Reading fluency (difficult)	Subtask 3 Reading comp (difficult)	Subtask 4 Writing	Subtask 5 Writing (fill missing words)	Subtask 6 Writing (fill missing words)	Subtask 7 Sequence of sentences in story	Subtask 8 Reading comprehension
Non-learner 0%	0.6	1.1	6.1	15.8	13.7	14.6	31.8	26.6
Emergent learner 1%-40%	25.4	1.9	5.6	17.1	9.1	15.0	10.8	11.9
Established learner 41%-80%	54.7	21.5	44.3	25.2	15.2	27.9	20.2	29.6
Proficient learner 81%-100%	19.4	75.5	43.9	41.9	62.0	42.5	37.2	32.0
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Again, paralleling the performance of girls, the main learning gap for boys in Somali literacy is between reading fluency (subtask 2) and the two harder subtasks that follow. While 77 percent of boys demonstrated proficiency in reading passages, fewer than half were able to comprehend what they were reading, and boys' writing skills are similar to their reading comprehension skills.

The critical finding here is that despite the fact that boys outperformed girls in Somali literacy, their skill gaps are identical. We will return to this point in the comparative discussion below.

The table on the following page presents boys' foundational skill gaps for English literacy. Paralleling the findings for girls' learning, boys have difficulty with word identification or word recognition in English. While just over three quarters of boys are proficient in the identification of letters, only about one third of boys are proficient in recognizing common words. Similarly, the ability of boys to recognize words lags behind their reading fluency, and this finding is also in keeping with the findings for girls above.

The primary learning gap for boys is in reading comprehension (subtask 3), but this gap is not as pronounced for boys as it was for girls. The other noteworthy difference between boys and girls in terms of their proficiency levels is that boys' proficiency levels in reading comprehension (subtask 3) are nearly identical to their proficiency levels in reading fluency (subtask 4). Readers will recall from above that both boys and girls tended to have lower levels of proficiency in reading comprehension than in reading fluency for Somali literacy, but this same relationship between fluency and comprehension does not hold true for boys when it comes to English literacy.

TABLE 49: BOYS' FOUNDATIONAL ENGLISH LITERACY SKILLS GAPS

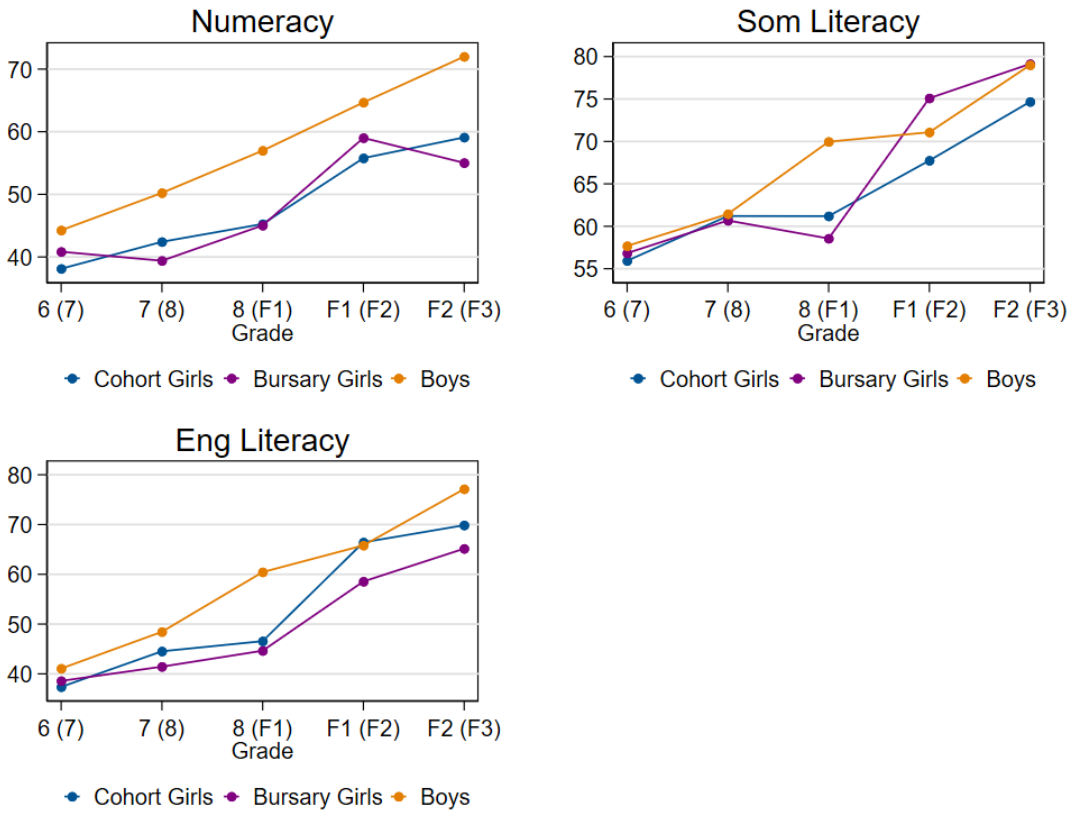
Categories	English ST1 Letter identification	English ST2 Word recognition	English ST3 Reading Comp (easy)	English ST4 Reading Fluency	English ST5 Reading Comp (medium)	English ST6 Reading Comp (hard)	English ST7 Writing (missing words)	English ST8 Writing (negative form)	English ST9 Writing (future tense)
Non-learner 0%	5.2	10.0	21.5	21.8	31.0	42.1	43.4	54.2	63.0
Emergent learner 1%-40%	5.6	16.1	2.1	3.2	6.5	9.5	6.3	0.0	0.0
Established learner 41%-80%	13.8	42.0	27.7	20.4	24.2	23.6	21.1	16.1	9.6
Proficient learner 81%-100%	75.4	32.0	48.7	54.6	38.3	24.8	29.2	29.6	27.4
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Comparative Analysis of Girls' and Boys' Skill Gaps

In numeracy and Somali literacy, girls and boys have parallel skill gaps. In particular, both boys and girls struggled with the identification of words (out of context) in both Somali and English, and also had skill gaps in reading comprehension in both languages. These findings suggest that additional or remedial efforts should be made to target skill-development in the reading of unfamiliar word and in reading comprehension. These skills appear to be insufficiently addressed at the moment (as evidenced by boys and girls having the same foundational gap in reading comprehension).

The only notable divergence between girls' and boys' skill gaps were in terms of the relationship between reading fluency and reading comprehension in the English language. For girls, proficiency in reading fluency was much higher than proficiency in comprehension. For boys, the two proficiency levels were nearly the same. The reasons for this gendered divergence in skill development are not clear, but the identification of this divergence may shed some light on the question of the major grade-wise divergence between girls' and boys' literacy scores at grade 8 (F1). The graph illustrating this divergence is reproduced below for ease of reference.

FIGURE 13: DISTRIBUTION OF (NON-COMPARABLE) MIDLINE LEARNING ASSESSMENT SCORES



The inflection point for grade 8 (F1) girls in Somali and English literacy may be directly related to reading comprehension skills and the fact that boys appear to be developing these skills more dependably than girls, especially in the case of English language learning. The underlying cause of this divergence in learning trajectories between girls and boys is not clear from the quantitative analysis, but the available qualitative evidence does suggest that increased chore burden for older girls may help to account for the dip in their performance around grade 8 (F1).

Broadly, girls face increasing pressure within their households and from society as they age, and this may be contributing to attenuated learning outcomes around grade 8 (F1). Tasks such as caring for infants and young children, watching over the household when parents are away, and assisting with household chores tend to fall on girls – particularly older girls. When asked why girls might miss attending school regularly, one girl explained, "They may get sick, or she may face challenges from her family like that she may have to care for young children in the house. In addition, if the girl is the oldest person in the home she may have to stay at home and handle household chores."[85] Another explained how household expectations can lead to large gaps in education: "…when moms give birth, girls must stay at home around 40 days."[86] Aside from tardiness and absence, one obvious effect of this high chore burden is that girls do not have as much time to do homework or practice what they have learned at home.

The tardiness and absences that result from this household chore burden can also add up to have a significant negative effect on girls' performance in school and motivation to continue schooling, particularly if the school does not make special efforts to accommodate girls: "Similarly, parents assign girls household chores while boys have been prepared since early morning. If the school administration puts pressure on girls for their absence and calls their parents, this will be difficult for them. Same as this – if school lessons become difficult for girls, it leads female students to start being absent…"[87] One girl explains that although there are some who try to continue learning once they have failed subjects, "similarly there may be other disappointed girls who will stay at home,"[88] or "skip learning if she fails an exam."[89] As subject matter becomes more difficult, the likelihood of girls entering this negative cycle of absence/tardiness leading to more absences appears to increase: "There are girls who miss school mostly during secondary school – they don't skip learning during primary school. The reason is that girls may face a difficult subject during secondary school or get married."[90] Thus, the process of girls aging and taking on increased (and heavily gendered) household responsibilities tends to produce a vicious cycle of tardiness, absence, and under-performance among many girls as they age, and this helps to explain the divergent trajectories between girls' and boys' learning as they reach secondary school.

[85] FGD, Girls, Puntland
[86] FGD, Girls, Galmudug
[87] KII, Male Teacher, Somaliland
[88] FGD, Girls, Puntland
[89] FGD, Girls, Puntland
[90] FGD, Girls, Puntland

Subgroup and Barriers Analysis

In this section, we disaggregate learning outcomes to provide insight into particularly disadvantaged subgroups of the target population, and to investigate potential barriers to learning. Due to the split sampling scheme at the baseline, there is a minimum of baseline subgroup data available that can be matched with learning data for the purpose of analysis. For that reason, our analysis in this section concentrates on identifying vulnerable subgroups on the basis of rich midline data, which has the benefit of using learning scores that are not affected by ceiling effects, as well as subgroup data that draws on the girls survey, the household survey, and information from the school head teacher survey which has been merged into the girls learning dataset.

The table below presents analysis of subgroups for a pooled sample of cohort and bursary girls. Girls were only included in this analysis if they had girls survey and household survey data that could be matched with their learning data. This restricts the sample to girls from the true panel (excluding replacements). The total sample size for analysis is 1,190 cohort and bursary girls. Each row in the table below presents a different subgroup and the average learning scores associated with that subgroup. The right-hand column of the table presents the number of individuals who belong to that subgroup, hence the N used for estimating the average score for that subgroup. All subgroup averages were compared with the sample average, and statistically significant differences (at the 95% level or higher) are marked with an asterisk.[91]

Beginning with geography, there are large and statistically significant differences by zone, with girls in Somaliland scoring significantly lower than average on all three assessments and girls in Puntland scoring significantly higher than average on all three assessments, as well as significant differences by urbanicity. Girls in Galmudug also have learning scores that tend to be higher than average (to a statistically significant degree for both numeracy and English literacy). In examining geographic differences, neither drought nor conflict are predictors of lower than average learning outcomes, and thus do not help to explain zone-based differences observed in the table below. In particular, it is worth noting that drought was a significant predictor of lower learning outcomes at baseline but is no longer a significant predictor at midline.[92] On the other hand, living in a rural (as opposed to comparatively urban) community was a significant geographic predictor of lower learning outcomes at both baseline and midline, and differences in urbanicity thus provides an important explanation for broader zone-based differences. Somaliland, which has the lowest average learning scores, also has the highest proportion of rural girls (at 60.4 percent

[91] The statistical significance of each subgroup was tested using a linear regression in which the subgroup in question is included as the independent variable, and a given learning score is the dependent variable. All regressions were adjusted for clustering, and results are marked with an asterisk if they were statistically significant at p = 0.05. The only exception is that individual impairments were not tested for statistical significance because the subgroup sizes were too small to permit meaningful tests. The pooled sample of all girls with any disability (N = 11) was tested for statistical significance.

[92] This change in the salience of drought as a predictor of learning outcomes may be a result of two changes. First, the drought became less severe, in general, from the baseline to the midline, which could explain its reduced predictive power with regard to learning scores. Second, the mix of schools affected by the drought also changed over time – at the midline, more schools in Puntland and fewer schools in Somaliland were affected by the drought than at baseline. It is possible the changing mix of affected schools explains the change in drought's impact.

within Somaliland). By way of comparison, only 14.3 percent of girls in the Puntland sample are from rural communities, and there are no girls from rural communities in the Galmudug sample.

In terms of girls' characteristics, the primary predictors of lower than average learning outcomes are girls having a disability and girls being old for their grade. In general, disabilities are rare among the sampled girls, making it difficult to analyse the effects of specific physical or cognitive disabilities on learning. When we pool together all 11 girls who have disabilities, we can see that their scores are on average much lower than those of their peers on all three assessments (and to a statistically significant degree for English literacy). Girls who are old for their grade have scores that are exceedingly low relative to their peers across all three assessments (and to a statistically significant degree for all three assessments). In light of these findings, it is likely that girls with disabilities and girls who are old for their grade (which may also indicate that they have a learning disability) are facing disadvantages in the classroom or at home that help to explain their attenuated learning. As noted above, the qualitative data makes it clear that older girls tend to have higher levels of chore-burden at home, and this helps to explain why girls who are old for their grade are under-performing vis-à-vis their younger peers. In addition, the qualitative evidence suggests that having extreme mixes of students (in terms of ages and ability levels) within a single grade is a daunting challenge for teachers. This extreme diversity of learners within classrooms applies both to girls who are old for their grade as well as girls with disabilities that affect their learning. As one teacher explained, "First of all, the big challenge is from students because the students are divided into three levels: upper, lower, and middle. If some of them do not understand the lesson, that will be a challenge - you have to repeat again the class or the lesson. So that is the only challenge teachers face; some students understood the lesson and some do not."[93] A number of teachers interviewed openly admitted that they did not feel adequately trained or equipped to address their students' unique educational needs in such diverse classrooms. Lastly, there is some evidence in the qualitative interviews that girls who are old for their grade may not feel comfortable learning with younger students in school, although it is unclear whether this is because they do not understand the content of the lessons, face bullying from other students, or some combination of these two explanations: "After droughts, pastoralists moved to cities, and parents enrolled their children in school. The challenges they face are that when they enrolled, their ages are above the schooling age, and they see themselves as being unable to study with this age and they leave the school and go to private places to improve."[94]

In terms of girls' characteristics, it is also important to note that being orphaned (living without one parent) as well as having a mother tongue that is different from the language of instruction are both *not* associated with lower than average learning outcomes, despite what intuition might suggest.

In terms of household characteristics, the primary predictors of lower than average learning outcomes are the education levels of the head of household and primary caregiver, as well as various indicators of household poverty. Girls whose head of household reported having no formal education have substantially lower learning scores across all three assessments (and to a statistically significant degree across all three assessments). The same holds for girls whose primary caregivers have no formal

[93] KII, Female Teacher, Puntland
[94] KII, MOE, Puntland

education: their scores are significantly lower than those of their peers. In terms of household poverty, five different indicators of household economic deprivation are used, and all five of these tend to be associated with lower than average learning scores. The strongest association between poverty and lower learning outcomes is with households that reported going most days without monetary income. Girls belonging to such households had significantly lower scores across all three assessments.

TABLE 50: LEARNING SCORES AMONG DISADVANTAGED SUBGROUPS FOR COHORT AND BURSARY GIRLS

	Average numeracy score (aggregate)	Average Somali literacy score (aggregate)	Average English literacy score (aggregate)	Number of observations for subgroup
Characteristics:				
All bursary and cohort girls	46.8	62.7	49.5	1190
Somaliland	29.3*	49.2*	35.6*	505
Puntland	60.4*	74.4*	56.7*	551
Galmudug	56.8*	65.6	71.7*	134
Rural	33.8*	53.1*	33.5*	385
Living without one parent	43.5	59.7	51.3	164
Mother tongue different to LOI	61.1*	72.1*	62.9*	187
Disability				
Vision impairment	44.7	63.4	42.6	7
Hearing impairment	85.4	86.0	91.6	1
Mobility impairment	0.0	0.0	11.1	1
Cognitive impairment	20.9	25.2	12.2	3
Self-care impairment	13.8	25.9	22.0	1
Communication impairment	2.5	29.2	11.1	1
Any disability	34.4	49.9	31.4*	11
Serious illness	45.1	62.6	50.1	208
HOH and Carer Characteristics				
HOH female	47.1	63.7	51.4*	789
HOH no education	44.2*	60.7*	47.4*	624
HOH no wage-earning occupation	47.1	62.5	50.0	478

HOH Pastoralist	43.4	61.7	41.7	35
Carer no education	44.6*	60.6*	46.9*	643
Carer non-literate	38.5*	55.1*	42.8*	428
Household Assets				
Owns land	45.5	61.0*	49.4	777
Owns mobile phone	47.0	62.7	49.1	1082
Poverty				
House is informal/temporary structure	34.8*	53.6*	40.7	67
Gone to seep hungry most days	39.9	44.5*	42.6	16
Gone without enough clean water most days	36.1	66.8	43.3	23
Gone without medicines or medical intervention most days	34.9*	59.7	39.1*	68
Gone without cash income most days	38.1*	53.3*	38.9*	78
Migration				
Household migrates seasonally	58.9	70.9	57.4	12
IDP household	41.6	60.4	53.7	60
Other				
Married	53.0	65.8	51.6	26
Mother	49.5	64.3	52.7	25
Old for grade	36.1*	51.1*	36.3*	189

Employing the same strategy used for subgroup analysis above, we analyse barriers to learning in the table below. Importantly, several of the barriers to learning are measured at the school level using data from the school survey that has been merged into the learner-level dataset. This analysis is thus limited to the sample of cohort girls for whom we have matched household information as well as school-level information. The total sample size for analysis of barriers is thus N = 907 cohort girls, although this sample size is reduced to 832 for barriers that were derived from the school survey because we do not have head

teacher survey data from every school (please see the Methodology section for a detailed discussion of why this is the case).

The table below explores the effects of a large number of potential barriers to learning. The majority of the potential barriers do not have consistent or statistically significant effects on learning outcomes, and some of the statistically significant relationships are in directions that run counter to intuition and that are likely to be the product of spurious correlations. Our analysis thus focuses on those clusters of barriers where all effects point in the same, intuitive direction, and where a large number of the relevant relationships are statistically significant.

The most consistent predictors of lower than average learning are limited school infrastructure and school resources, as well as low teaching quality. It is worth noting that, despite what intuition might suggest, the majority of indicators of gender equity in the classroom (e.g. measures including gender of teachers and mentors, as well as the way teachers are reported to treat girls versus boys in the classroom) are not strongly associated with learning outcomes. On the other hand, there are a number of infrastructural issues noted that may indicate gender inequities (e.g. non-use of toilet facilities), as well as household-level indicators of gender unequal treatment of boys and girls that are significant predictors of lower learning outcomes.

TABLE 51: BARRIERS TO LEARNING AND LEARNING OUTCOMES

	Average numeracy score (aggregate)	Average Somali literacy score (aggregate)	Average English literacy score (aggregate)	Number of observations for subgroup
Barriers:				
All cohort girls	46.4	62.5	50.5	907
School Infrastructure				
Difficult to move around school	35.2*	52.8*	41.1*	208
Doesn't use drinking water facilities	39.3*	57.8	47.8	160
Doesn't use toilet at school	33.7*	53.7*	41.9*	161
Doesn't use areas where children play/socialise	38.0*	53.0*	46.7*	264
School Resources				
No computers at school	46.1	62.2	49.9	807
School does not have learning materials	43.9	60.6	51.4	152

Not enough seats for children at school	49.1	64.1	49.3	91
No electricity	32.6*	51.2*	34.6*	189
Only dirt floors	44.1	58.9	45.3	60
Offers remedial classes	47.3	63.2	50.7	557
Offers life skills classes	43.6	61.3	53.4	167
Separate toilets for boys and girls	48.1*	62.9	51.5	778
School feeding program	37.9*	56.5*	39.5*	355
Offers sanitary kits	39.7*	59.5	41.5*	368
Offers solar lamps	41.1*	56.5*	49.5	299
Has girls club	46.0	62.2	50.7	784
Girls club meets at least once per week	35.6*	52.4*	40.1*	109
School has CEC	46.2	62.3	50.1	791
CEC meets frequently	41.7	57.4	45.3	212
Teaching Quality				
Disagrees teachers make them feel welcome	43.2	60.6	50.4	65
Agrees that they are afraid of teacher	41.5*	58.6*	45.1*	490
Teacher punishes students who get things wrong	45.3	62.3	50.8	553
Classes held fewer than 5 hours per day	36.1*	52.6*	40.7*	153
Agrees teacher is often absent from class	47.8	65.6	51.0	190

Teacher uses corporal punishment	51.5*	64.2	56.6*	141
Carer says principal performance is poor	40.7	55.7	33.4*	22
Carer says teaching at school is poor	26.3*	54.9	51.5	5
Gender Equity				
At least one female teacher	43.3*	60.7*	46.5*	684
Female mentor	44.5	61.0	47.8*	697
Male mentor	45.3	62.5	48.4	591
No mentor	46.2	64.2	58.1	77
Teachers treat boys differently from girls	44.5	62.0	47.1*	363
Family prioritizes boys' education	38.3*	52.2*	48.8	94
Other Barriers				
Agrees she has no choice in schooling decisions	45.5	62.6	48.3*	634
Over 30-minute travel time to school	39.5	51.4*	46.4	23
Feels unsafe on way to school	52.2	66.8	61.6	22
Feels unsafe at school	36.0	54.9	36.8*	4
Parents are disengaged	61.7*	68.7	38.3*	42

Beginning at the top of the table, all four variables related to school infrastructure are strong predictors of learning outcomes, with poorer infrastructure (based on girls' reports) being consistently associated with lower learning outcomes. Where girls reported difficulty moving around the school, not using the toilets at school, or not using playground areas, learning scores were lower across all three assessments (and to a statistically significant degree across all three assessments). Learning scores were also lower

across all three assessments where girls reported not using drinking water facilities at the school, and this result was statistically significant in the case of numeracy scores.

Taken as a whole, there is strong evidence that, where girls find school infrastructure to be insufficient or unusable, their learning is attenuated as a result. One possible mechanism explaining this connection between infrastructure and learning is supported by qualitative evidence, which suggests that girls may have to leave schools during the middle of the school day to find adequate places to, e.g. use the toilet or get drinking water, and this detracts from their ability to attend classes consistently and learn. Another possible mechanism is that schools with better infrastructure also tend to be better funded, with better teachers, and thus better able to support student learning in the classroom. While there is less qualitative evidence to support this mechanism, this explanation is consistent with another finding below which is that girls have consistently lower learning outcomes at schools that do not have electricity and at schools that have only dirt floors (and the finding that learning scores are lower in schools without electricity is statistically significant across all three assessments). Poorly funded schools in rural communities are those that have the highest likelihood of having only dirt floors and of not being electrified, and this correlation between rural status and low infrastructure helps to make sense of the finding in the subgroup analysis above that girls living in rural areas have significantly lower learning scores than their peers. All of these findings point in the same direction – namely that poorly resourced schools produce poor learning outcomes.

Moving on to school resources, the key relationships between electrification, infrastructure, and learning have already been noted above. The other aspects of school resources that are worth noting is the fact that none of the school-level interventions provided (e.g. offering of remedial classes, life skills classes, school feeding, or sanitary kits) have had a measurable positive impact on girls' learning. In particular, it is worth noting that the offering of remedial classes, which is the output-level intervention most directly related to learning, is not a significant predictor of higher learning outcomes. This finding does not necessarily call the project's theory of change into question, but it does suggest that future monitoring of outputs should focus on determining why remedial classes have not had a more direct effect on learning. The other school-level, output-related interventions are more directly tied to the outcome of attendance, and therefore we have weaker expectations that these outputs will be strongly correlated with learning. For example, sanitary kits, which are intended as a means of helping girls with menstrual hygiene management, are intended to make it easier for girls to continue attending school during menstruation, but have no direct effect on learning. This finding does not necessarily suggest that these interventions are ineffective. Rather, the only conclusion that can be drawn is that these interventions have not yet had a measurable effect on learning via their effects on attendance. Questions of linkage between intermediate outcomes (such as attendance) and learning will be examined in greater detail in section 7.6 below.

As with school resources, the indicators of teaching quality are many in number, and several of these have consistent, intuitive, and statistically significant relationships with learning outcomes. Similar to the baseline findings, several indicators of lower teaching quality are significantly correlated with lower learning outcomes. Taken from the perspective of cohort girls, low teaching quality (measured as teachers making girls feel unwelcome and girls being afraid of their teachers) are both predictors of lower learning

outcomes across all three assessments (and to a statistically significant degree across all three assessments in the case of girls reporting that they are afraid of their teachers). Similarly, caregiver reports of low-quality principals and low-quality teachers are also consistently correlated with lower learning outcomes (with the correlation being statistically significant between numeracy and poor teaching quality and between English literacy and poor principal performance). Finally, the strongest predictor of lower learning outcomes among the indicators of teaching quality was for schools where classes were held for fewer than 5 hours during the day. Girls in schools with shorter school-days scored substantially lower on all three learning assessments (and to a statistically significant degree across all three assessments). These findings confirm the emphasis that the project's theory of change places on teaching quality.

Barriers that might be associated with gender inequity in the classroom (measured through the gender of mentors and teachers and the reported treatment of boys versus girls in the classroom) were not predictors of lower learning outcomes. The qualitative data from teachers, caretakers, and girls strongly suggests that girls and boys are both equally motivated and engaged in class. There is very little evidence that, within the classroom, girls are significantly disadvantaged vis-à-vis boys. When asked whether there have been any changes in the way girls ask or answer questions in class, perform in school, or lead activities in class, a female remedial teacher answered that "they are better than boys for all these" and that sometimes teachers hold "competitions between boys and girls, and it seems girls are more active than boys."[95] Numerous other respondents also highlighted the improvements in girls' participation and performance, explaining that "girls are better than boys in studying."[96] A particularly confident girl respondent boasted that girls "come first in school compared to boys,"[97] Caretakers have recognized this trend as well, and one mother suggested that, "these days, it seems that girls education is going even better than boys and they are so interested in learning."[98] Taken as a whole, this evidence suggests that gender disparities in learning are less attributable (than in the past) to differences in how girls and boys are treated in the classroom. Rather, these disparities would seem to arise more from broader cultural and structural impediments at the household and community level.

At the household level, there is some evidence from the barriers analysis here that caretakers' beliefs about the relative priority of boys' versus girls' education may have a strong influence on girls' learning outcomes. Caretakers were presented with a hypothetical scenario in which they had the choice to choose how to use limited household resources and could choose between prioritizing boys' education over girls', vis-versa, or prioritizing both equally. Girls belonging to households where caregivers suggested that they would prioritize boys' education over girls (in the hypothetical scenario) scored lower on all three learning assessments (with the correlations for numeracy and Somali literacy being statistically significant).

Finally, there are two other barriers that are worth noting because they were also significant predictors of lower learning outcomes at the baseline: long travel-times to school and girls feeling unsafe while at school. Both of these barriers were correlated with lower than average scores across all assessments (and

[95] KII, Female remedial teacher, Puntland
[96] KII, Male Teacher, Puntland
[97] FGD, Girls, Galmudug
[98] FGD, Mothers, Somaliland

the correlations were statistically significant for long travel times as a predictor of low Somali literacy and for girls feeling unsafe at school as a predictor of English literacy). Long travel times are easy to understand as an impediment to learning because girls who have to walk for 30 minutes or more to reach school may tend to be tardy to class, or may have to miss days of school when the weather is extremely bad.

The reasons why girls reported not feeling safe at school are not clear from the quantitative data, but the qualitative data suggests a number of candidate explanations, including discomfort or even fear because of poor and unsafe school infrastructure, as well as threats of violence and bullying at school, and possibly sexual violence while at school. Boys in a focus group in Puntland gave examples of unsafe school infrastructure: “it will affect that on the students because when a student gets dressed then dirt falls on his/her head from the roof, you can imagine the feeling.” Girls from a focus group in Puntland also talked about school infrastructure, but also linked infrastructure with violence and stone throwing (which seems to conflate two discrete issues), saying: “The most not safe areas are the classes with wood ceilings while the concrete ceiling are better in terms of safety. When we are somewhere under concrete ceiling we don't scare […] while we face stone throwing when we are under the wood ceilings.” The problem of bullying and stone throwing surfaced across a number of qualitative interviews, which suggests that this is one of the more widespread problems leading girls to feel unsafe. A few other girls in Galmudug complained that fighting is relatively common at school, with one girl explaining that, “I dislike violence and fights to happen during school time.” The most direct testimony about sexual violence and rape at schools came from a father in Galmudug, who said that, "This year in the school, girls were feeling scared of the boys who might rape them or kidnap them. But after community co-operation, this is being solved." A few other qualitative interview respondents in Galmudug highlighted problems of sexual violence at or near school, corroborating the preceding quote from a father in Galmudug.

The generalizable lessons here that can be applied to future programming are principally related to the potential utility of empowering CECs to invest in improvements in school infrastructure that will provide basic health and sanitation facilities where they are missing and will enable girls to feel more comfortable using such infrastructure where it exists. In particular, attention should be focused on rural and under-resourced schools because these school-level factors are some of the most consistent barriers to learning according to the analysis above. In keeping with the project theory of change and the focus on the intermediate outcome of increasing the number of active CECs, most of the mechanisms through which limitations of school resources and infrastructure can be addressed are likely to follow from CEC involvement and investment.

In addition, teaching quality is the other principal barrier that needs to be addressed and that was also identified at baseline. The intermediate-outcomes section on teaching quality below will assess the degree to which teaching quality has changed since baseline, but for now it will suffice to observe that low teaching quality and short school days still appear to be contributing to lower learning outcomes among cohort girls.

5. Transition Outcome

Transition represents the second core outcome of EGEP-T, alongside learning and sustainability. Transition encompasses several distinct pathways, including progression through school, re-enrolment in school, and movement into alternative education, or age-appropriate employment, pathways which we describe in more detail below. Before turning the main results, it is important to highlight the methodological issues associated with studying transition in the EGEP-T midline; following this discussion, we report aggregate transition reports against targets established at baseline, and analyse transition rates among various subgroups of cohort girls.

For ease of reference, the table below provides a summary of transition results by zone. The table that follows also reports the project's progress against the established transition outcome targets, and establishes the targeted improvement for the endline evaluation.

TABLE 52: TRANSITION RESULTS BY ZONE

Round of Data Collection	Somaliland	Puntland	Galmudug	Total
Baseline	89.6%	86.6%	94.2%	88.7%
Midline	86.7%	90.2%	93.5%	89.0%

TABLE 53: TRANSITION TARGET PERFORMANCE AT MIDLINE AND TARGET-SETTING FOR ENDLINE

Round	Transition Target (above benchmark group)	Project Progress	Pct. of Target Achieved
Midline	7%	0.3%	4.3%
Endline	7%	TBD	TBD

5.1 Methodology of Transition Analysis

Transition Sample

The sample analysed for the transition outcome is distinct from that employed elsewhere in this report, including in comparison to the learning sample. The EGEP-T cohort was selected at baseline from within schools, meaning that all selected girls were enrolled at the time of the baseline. Although enrolment at the time of the baseline does not guarantee that every girl had successfully transitioned relative to the previous year – girls could be enrolled at baseline but have been held back a grade, rather than progression from the previous year – it does mean the vast majority of cohort girls had successfully transitioned at the time of the baseline.

The cohort girls sampled at baseline constitute half of our transition sample, as they are tracked from baseline to midline. At midline, we attempted to re-contact every cohort girl to obtain their current enrolment status, grade level, work status, and other details, allowing us to determine whether they transitioned successfully from baseline to midline. In contrast to the learning sample analysed in Section 4, cohort girls who dropped out of school between baseline and midline are included in this transition sample, because their outcome (dropping out) falls within the transition calculation. If such girls were not included, it would bias the estimated transition rates upward, toward 100 percent.[99]

At the same time, locating every girl who dropped out of a project school between baseline and midline was not possible, and this may also produce bias toward higher estimated transition rates than their true value. If girls who drop out of school cannot be located, they are not included in our calculation of transition rates; to the extent that these girls have lower transition rates than girls who are successfully located – which is likely – our estimates of transition could be biased upward. We provide analysis of this issue in Section 2.5 of the report, but a summary of our findings is that differential attrition from the transition sample likely results in bias.[100] Finally, note that replacement girls selected for the learning sample are *not* included in the transition sample. Replacement girls were selected randomly from among girls at the project schools; if they were included, it would bias our estimates of the transition upward.

The other half of our transition sample is drawn from the baseline evaluation, when a benchmark sample of girls was drawn for the purposes of benchmarking transition rates in the communities at the start of EGEP-T programming. These girls were selected using a random walk methodology, and constitute a random sample of girls aged 11-18 years in project communities at the time of the baseline. All analysis of transition rates in the baseline report relied on this sample. Importantly, the benchmark group was distinct from the cohort girls recruited at baseline – the former were sampled through households, and a substantial minority were not enrolled in school; the latter were selected through the schools, with 100 percent enrolment rates. The benchmark group of girls constitutes the counterfactual for our midline analysis – the group of girls against which midline cohort girls are compared.

Fundamental differences in the benchmark and cohort samples give rise to methodological difficulties in our analysis of transition. The transition benchmark sample included a number of out-of-school girls and, as a result, had a significantly lower overall transition rate than we would expect from a sample comprising exclusively girls who were enrolled at baseline. Enrolled students are much more likely to remain enrolled and successfully transition to the next grade than out-of-school girls are to re-enrol. To illustrate this issue,

[99] Girls who dropped out between baseline and midline were re-contacted with the help of school officials and various information collected at baseline, including household telephone numbers, names of household members, and written directions to the households.

[100] For girls who could not be located at midline, field teams attempted to collect transition data from knowledgeable individuals (head teachers, family members, neighbors). Among the 130 girls for whom such data was collected, just 63.8 percent were reportedly enrolled in a new, non-EGEP-T, school. If we assume that a small share of those girls enrolled would not have progressed to the next grade level, this suggests transition rates among girls not successfully re-contacted could be below 60 percent. If this is the case, our estimated transition rate could be reduced by as much five percentage points. Unfortunately, more precise estimates of this kind are not possible, as this analysis is speculative, based on limited data and tentative assumptions.

consider two types of girls from the benchmark transition sample: the first girl was out-of-school the year prior to the baseline. For this girl, the probability of her successful transition from time *t-1* (the year prior to baseline) to time *t* (baseline) is 77.8 percent. Now imagine a second girl, who was enrolled in school the year prior to the baseline. For this girl, the probability of her successful transition is 86.8 percent, significantly higher than that of the first girl. Such a large differential is unsurprising: out-of-school girls have self-selected into non-enrolment, and the barriers to re-enrolment are potentially very large. For girls who are already enrolled in school, transition consists of maintaining enrolment and advancing a grade – given that they have self-selected into enrolment in time *t-1*, they may enjoy advantages in terms of familial support or other factors which could make transition more likely in their cases.

A 9-point difference in transition rates is sufficiently large that it could produce significant bias in our analysis of transition rates. Every girl in the cohort sample was enrolled at baseline, meaning they more closely mirror the second girl in our above example. Comparing them to a sample comprised entirely of benchmark girls who were enrolled at time *t-1* is a more appropriate approach; comparing them to a sample that includes a large number of girls who were out-of-school at time *t-1* would reduce the validity of the inferences drawn.

The approach we follow in this report ensures the greatest level of comparability between benchmark and cohort samples. We subset the benchmark sample to include only those girls who were enrolled at time *t-1*, ensuring the transition analysis makes valid comparisons. Restricting the sample in this way reduces the available benchmark sample from 717 girls to 465.[101] This approach reduces sample size for the benefit of additional comparability.[102] For the sake of completeness, the table below reports the number of benchmark girls, by age, who were or were not enrolled at time *t-1*.

TABLE 54: ENROLMENT STATUS OF GIRLS IN BENCHMARK TRANSITION SAMPLE, BY AGE AT TIME *T-1*

Age	OOS at time *t-1* (Excluded from benchmark sample)	In-school at time *t-1* (Included in benchmark sample)
10	25.0%	75.0%
11	20.6%	79.4%
12	20.2%	79.8%
13	16.7%	83.3%
14	18.6%	81.4%

[101] The transition rate in the full sample was 73.2 percent, while the restricted subsample had an unweighted transition rate of 87.7 percent. These values differ from those referenced earlier, because they exclude Banadir from the benchmark sample as well.

[102] We do not expect the reduction in the transition benchmark sample size to have significant consequences for statistical power. The sample size reduction is significant, but is distributed across schools. Due to the clustered nature of the sample, reducing the number of respondents but maintaining a fixed number of clusters has less impact on statistical power than does reducing the number of clusters/schools in the analysis. Prior to analysis, we performed a number of monte carlo simulations aimed at estimating expected statistical power under varied sample constructions. Similar sample size reductions in the context of the learning sample – tested for the sake of guiding our analysis – did not have significant consequences in terms of power.

15	20.3%	79.7%
16	27.9%	72.1%
17	28.4%	71.6%
Total	22.2%	77.8%

In line with guidance from the FM, the benchmark and cohort transition samples are adjusted for discrepancies in their age distribution. The table below documents the age breakdown of each sample. Note that the age referenced in the table – and throughout this section – corresponds to the girl's age at time *t-1*. That is, for cohort girls, age refers to their age at baseline; for benchmark girls, it refers to their age at time *t-1*, which is the year before baseline data collection. Two issues related to age require additional explanation. First, the two samples do not entirely overlap with regard to their age distribution – the benchmark sample includes a number (n = 125) of girls aged 10 at time *t-1*, while the cohort sample includes no equivalently-aged girls. As a result, benchmark girls aged 10 years are excluded from the analysis. On the other end of the spectrum, the cohort sample includes 44 girls aged 18 years or over, with no equivalent girls included in the benchmark sample.[103] The consequence of this mismatch is that girls aged 18 and up are not included in the transition analysis, because they have no valid counterfactual from the benchmark sample.[104]

TABLE 55: AGE DISTRIBUTION OF BASELINE BENCHMARK AND MIDLINE COHORT TRANSITION SAMPLES

Age	Benchmark N	Share of Benchmark Sample	Cohort N	Share of Cohort Sample
11	79	17.2%	41	3.8%
12	71	15.5%	93	8.5%
13	69	15.0%	202	18.5%
14	70	15.3%	261	24.0%
15	62	13.5%	212	19.5%
16	61	13.3%	157	14.4%
17	47	10.2%	123	11.3%
Total	459	100.0%	1,089	100.0%

The second issue also concerns the age distribution of each sample. Leaving aside the mismatch in the "tails" of each distribution, the benchmark sample is more evenly distributed, with many girls aged 11-12,

[103] The exclusion of girls aged 19-21 at baseline was an oversight on the part of the external evaluation team, discussed in more detail in the baseline report.

[104] The age mismatch stems from two issues in baseline data collection. First, the benchmark sample was intended to capture transition data for girls aged 11 to 21 years; however, an error during baseline data collection limited th sample to girls aged 11-18 years. Girls aged 18 years in the benchmark sample were 17 years old at time *t-1*, providing no valid counterfactual for cohort girls who were 18 years at the baseline. Second, the benchmark sample included girls who were aged 11 years at the baseline (10 at time *t-1*); these girls do not provide useful data for comparison to the cohort girl sample, because the cohort girls were all 11 years or older at the baseline.

while the cohort sample tends to clump in the age ranges 13-16 years. The disjuncture in the distributions is not severe, but – because age tends to be a strong correlate of transition rates – it is important to adjust for age imbalances in the transition analysis. When we present aggregate transition results, we follow the FM's guidance and weight each age level in the benchmark sample to match the share of the cohort sample that age level comprises. For instance, we apply a weight to 13-year old girls in the benchmark sample to increase their representation from their unweighted, natural share of 15.0 percent to more closely match the 18.5 percent they represent in the cohort sample. Other age levels are downweighted, as appropriate, to match the age distribution of the cohort sample, explicitly controlling for differential transition across age groups in the main results.[105]

It is important to note that we cannot weight age groups to match across rounds precisely because we also employ post-stratification weights to adjust the sample to more closely match the population distribution of schools in which EGEP-T is being implemented. We discuss this issue in greater detail in the methodological notes elsewhere in this report. Our sampling weights serve these two purposes – adjusting to match the population distribution of school-level characteristics and to ensure balance in the age distribution between baseline and midline samples – but satisfy neither goal completely.[106]

Defining Transition

As the discussion in the previous section implied, the possible transition pathways available to girls in the transition sample are somewhat truncated relative to many other GEC-T projects. EGEP-T does not include support for alternative training or education programmes and focuses particularly on encouraging transition through the school grades, including from primary to secondary school. As per the project's MEL Framework, successful transition will be measured on the basis of yearly progression across all targeted grades.

The table below describes the full range of successful and unsuccessful transition outcomes available to four different groups of girls in EGEP-T communities. Both upper primary and lower secondary school girls' transition outcomes are defined by progression in school from grade to grade. Because the cohort girls tracked for transition ranged from Grade 6 to Form 2 at the time of their recruitment at baseline, they are not expected to progress beyond Form 4 prior to the endline evaluation and the conclusion of the project.

[105] Note that we do not employ weights that adjust for age distribution in the context of the barriers and subgroup analysis. This is because subgroup analysis utilizes subsamples, and the same weights employed to ensure age-equivalence in the aggregate results would not be valid for a given subgroup sample. At the same time, the subgroup and barriers analysis does employ the same post-stratification weights mentioned above, which correct for differences between the sample and population in terms of school-level characteristics.

[106] To illustrate, in the weighted sample, 19.8 percent of benchmark girls are aged 15 years, compared to 19.5 percent in the cohort sample. At the same time, Somaliland respondents comprise 44.5 percent of the weighted sample, while Somaliland schools make up 44.7 percent of all schools in the EGEP-T population. Weighting to ensure balance in the age distribution alone would allow us to match ages precisely; alternatively, post-stratification weighting would allow us to recover the population characteristics precisely; however, combining the two approaches yields a compromise – greater balance and a closer match to population characteristics overall, without precisely matching either.

Therefore, the project has not defined transition pathways that beyond Form 4, such as enrolment in higher education or transition to employment following the completion of secondary school. As noted in the final row, transition is considered successful for out-of-school girls if they re-enrol in school. As the table makes clear, transition into age-appropriate employment and/or alternative education programs are not considered as successful transition pathways for the purposes of this evaluation.

TABLE 56: TRANSITION PATHWAYS

Grade Level	Baseline point	Successful Transition	Unsuccessful Transition
Upper Primary (G6-G8)	Enrolled in Grades 6, 7, 8	• In-school progression • Transitions from G8 to secondary school	• Drops out of school • Remains in same grade
Lower Secondary (F1-F2)	Enrolled in Forms 1, 2	• In-school progression	• Drops out of school • Remains in same grade
Upper Secondary (F3)	Enrolled in Form 3	• In-school progression	• Drops out of school • Remains in same grade
Out of School (age 11-18)	Dropped out or never enrolled	• Enrol or re-enrol in appropriate grade level	• Remains out of school

It is important to note that only girls in the top two rows actually appear in our transition analysis. Out-of-school girls – defined by their status at baseline – are not included in the sample of cohort girls. While the baseline evaluation did capture a small number of out-of-school girls through the household sample, it is unclear how these girls could be included in the transition analysis, as their representation in any combined sample of cohort and OOS girls would not be proportional to their true share of the population in EGEP-T communities.

In practice, girls in this analysis have three possible transition outcomes: dropping out of school, remaining enrolled but being held back a grade (or regressing a grade), and remaining in school while progressing upward in terms of grade level. Throughout this section, we report the transition rate, which represents successful transition; in some cases, we also report the dropout and held-back rate, simply to draw a distinction between those two unsuccessful outcomes, as they represent qualitatively different outcomes, even if both are considered equally unsuccessful.

5.2 Main Transition Results

EGEP-T defines successful transition as progression from grade-to-grade within school or re-enrolment, for out-of-school girls. As noted above, the latter pathway is not relevant to our analysis at midline, because all girls in the cohort transition sample were enrolled at the baseline, and we have restricted the baseline benchmark sample to the set of girls who were enrolled in the year prior to the baseline. Girls have three possible outcomes in this analysis: successful transition, dropping out of school, and continued enrolment but failure to advance a grade level year-on-year. We refer to successful transition as transition and the transition rate as the share of girls successfully transitioning.

Overall, the transition rate in the baseline benchmark sample was 88.7 percent. The table below reports transition rates, by age, for the baseline benchmark sample (left two columns) and the cohort sample at midline (right two columns). Within each age group, transition rates are weighted to match the population of EGEP-T schools. The totals presented along the bottom are weighted further to account for differential age distributions, as discussed in the previous section. In comparison, the cohort sample achieved a transition rate of 89.0 percent, an increase of 0.3 percentage points in transition rates from baseline to midline.

TABLE 57: TRANSITION RATES, BY ROUND AND AGE

Age	Observations Benchmark Sample	Transition Rate Benchmark Sample	Observations Midline Sample	Transition Rate Midline Sample
11	79	90.4%	41	88.1%
12	71	84.4%	93	89.1%
13	69	93.1%	202	90.5%
14	70	88.8%	261	91.6%
15	62	89.1%	212	83.9%
16	61	88.9%	157	91.0%
17	47	80.8%	123	87.4%
Weighted Total	459	88.7%	1089	89.0%

Although the weighting approach employed can be confusing, it does not substantially affect the reported transition rate for the benchmark sample.[107] Nonetheless, it is instructive to compare transition rates within each age level by reading across the table above. For instance, transition rates were 84.4 percent at baseline for 12-year old girls, compared to 89.1 percent for the same age of girls in the midline sample. The table shows significant variation between benchmark and cohort transition rates in individual age categories – benchmark girls outperformed cohort girls in the 11-, 13-, and 15-year old categories, but

[107] A fully unweighted comparison from baseline to midline yields a change from 88.0 to 89.1 percent (+1.1 points); a comparison weighted by age alone yields similar results; neither is statistically or substantively different from the results with age and post-stratification weights described here.

underperformed in the 12-, 14-, 16-, and 17-year old groups. The noise from one age group to another is most likely attributable to the small sample sizes involved, as there is not a consistent relationship in which cohort girls outperformed benchmark girls more often in the younger or older age ranges. We do not report p-values from hypothesis tests for each age level for the same reason; due to the small sample size available within individual age-group samples, such tests will likely capture random noise. Most importantly, the aggregate difference between baseline (benchmark) and midline (cohort) is not statistically significant (p = 0.92), but transition rates are slightly higher at midline than those observed in the benchmark sample.

The small reported gain in transition rates is broadly consistent, geographically, though there are minor differences worth highlighting. Restricting our attention to individual zones, we find that schools in Somaliland experienced a modest 2.9 point decline in transition rates; conversely, schools in Puntland saw a small, but meaningful, 3.6 point improvement over its zone-specific benchmark sample, though neither finding is statistically significant at the conventional levels.[108] In Galmudug, transition rates at midline were 0.7 points below its benchmark at baseline, though this difference was not statistically distinguishable from a null effect. Notably, baseline transition rates were highest in Galmudug, and they remained the highest at midline (93.5 percent, compared to 90.2 percent in Puntland and 86.7 percent in Somaliland). While Galmudug remained the highest performing area in terms of transition in both rounds, the midline results represent a shift in the other two zones studied – at baseline, Puntland had the lowest transition rate, but surpassed Somaliland in the midline.

The results above focus on successful transition. However, there is an important distinction to be drawn between girls who drop out of school and those who remain enrolled but are held back a grade – the latter continue to learn and are more likely to continue their education than the former. In the table below, we disaggregate transition into the three possible pathways described previously. The top panel of the table reports results for the baseline sample, disaggregated by age, and split into dropout, held back, and successful transition outcomes, from left to right. The bottom panel reports the same results for the midline sample. It is important to note the sample size (count) referenced in each row, as some age levels include very small numbers of observations, and their values should be treated with caution.

TABLE 58: DISAGGREGATED TRANSITION OUTCOMES, BY AGE

Age	Dropout, Count	Dropout, Pct.	Held back, Count	Held back, Pct.	Success, Count	Success, Pct.
Baseline Sample (Transition Benchmark)						
11	3	3.5%	5	6.0%	71	90.4%
12	4	5.4%	7	10.2%	60	84.4%
13	4	4.8%	1	2.1%	64	93.1%
14	5	7.6%	3	3.6%	62	88.8%
15	3	4.4%	4	6.5%	55	89.1%
16	5	8.0%	2	3.0%	54	88.9%

[108] For the latter result, focused on Puntland, p = 0.25 after accounting for clustering at the school level.

17	7	15.4%	2	3.7%	38	80.8%
Total		6.8%		4.5%		88.7%
Midline Sample (Cohort Girls)						
11	0	0.0%	5	11.9%	36	88.1%
12	0	0.0%	10	10.9%	83	89.1%
13	6	3.1%	12	6.4%	184	90.5%
14	8	3.1%	14	5.3%	239	91.6%
15	14	6.5%	21	9.7%	177	83.9%
16	4	2.6%	10	6.4%	143	91.0%
17	6	5.2%	9	7.4%	108	87.4%
Total		3.6%		7.5%		89.0%

The table above illustrates a broad difference between baseline and midline transition outcomes. While successful transition is only slightly more common at midline, dropout rates are substantially lower. At baseline, 6.8 percent of girls who were enrolled at time *t-1* dropped out by the time of the baseline; in contrast, in an equivalent sample of cohort girls, the same year-on-year dropout rate was just 3.6 percent, a difference that is statistically significant.[109] At midline, girls were slightly more likely to transition successfully and slightly more likely to be held back a grade, both of which are more positive outcomes than dropping out – the latter because being held back but remaining enrolled implies that a girl can continue learning and may progress into higher grades in the next year.[110] Lower dropout rates are consistent across the age spectrum: while we observe a marked increase in dropout rates among older midline girls – consistent with the baseline benchmark sample – that increase begins from a lower starting point, and dropout rates are lower than at baseline in essentially all age categories.

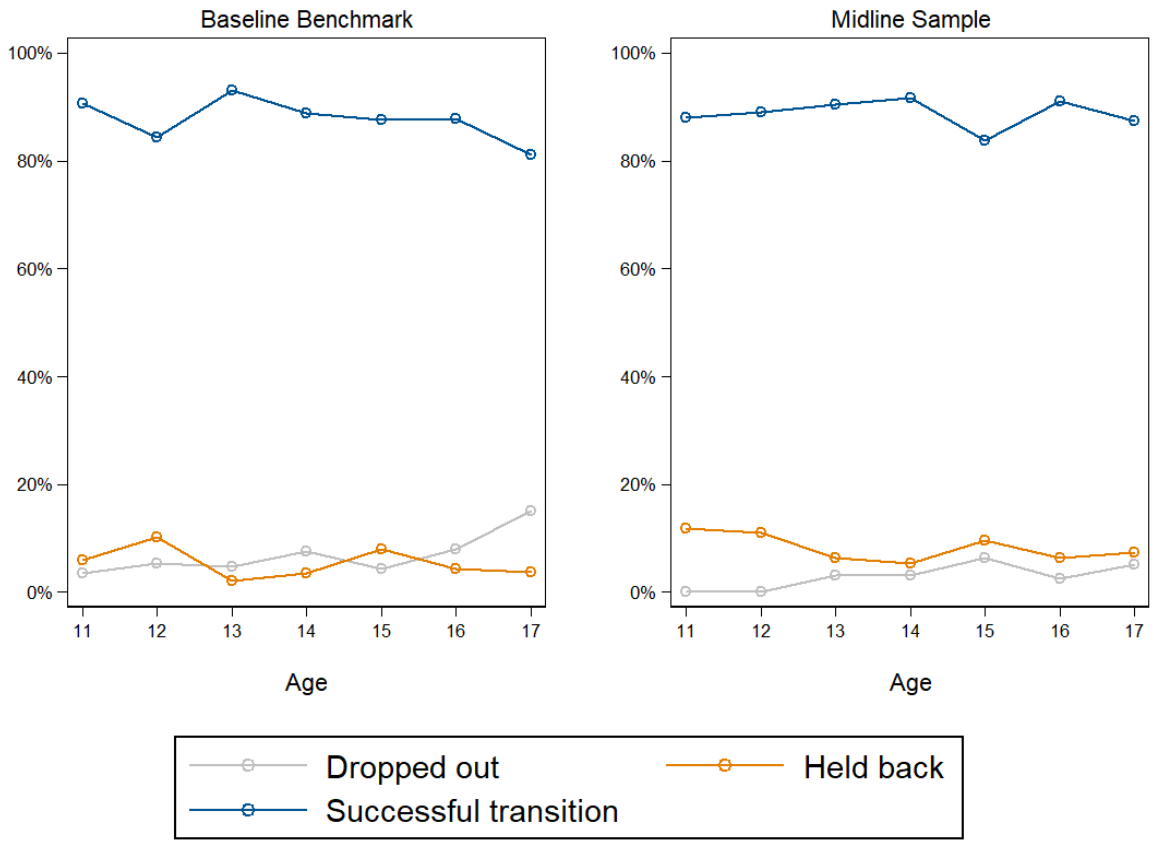
In general, dropout rates tend to be correlated with age. In both samples, girls are more likely to drop out as they get older – aggregating across both rounds of data collection, dropout rates were just 1.1 and 1.6 percent for 11- and 12-year old girls, rising to 4.1 and 7.6 percent for 16- and 17-year old girls, respectively. The figure below emphasizes trends by age group, plotting transition pathways for the baseline and midline samples by age. The left panel reports the share of girls in each transition pathway for the baseline benchmark sample; the right panel reports the same results for the midline cohort sample. In general, transition rates only decline with age in the baseline sample – in the midline sample, successful transition rates are broadly stable as girls get older. Further, the downward shift in transition rates among the baseline group seems to be driven primarily by an increase in dropouts, which represent a more negative

[109] $P < 0.05$, using a two-sided t-test, after accounting for clustering.

[110] The higher rate at which girls have been held back at midline is somewhat surprising, given the age and grade level of the girls being studied. It is generally assumed that girls in higher grades who fail to advance a grade rarely repeat it and drop out instead. In our data, girls in secondary level grades are still likely to be held back but remain enrolled, at least according to their own self-reports and those of their caregivers. It is important to note the source of this information, as girls who failed to advance a grade may be nominally enrolled, and consider themselves so, without actually attending school.

outcome than being held back a grade. We do not see any equivalent increase in dropout rates as girls get older in the midline sample.

FIGURE 14: TRENDS IN TRANSITION PATHWAYS, BY AGE



As part of the EGEP-T baseline evaluation report, targets were established for improvements in transition rates from baseline through endline. At that time, an improvement of 3.5 percentage points – from 72.7 to 76.2 percent – was proposed as a target for the midline.[111] Applying this proposed 3.5 point gain to the revised baseline transition rate utilized above generates a target of 92.2 percent. The project has not met this target, with transition rates rising from 88.7 to 89.0 percent over time.

[111] The baseline transition rate of 72.7 percent was based on the full transition sample – including girls who were out-of-school in time *t-1* and girls in EGEP-T communities in Banadir – accounting for the significant difference between the baseline rate reported at the time of the baseline report and that reported here.

5.3 Transition Analysis – Subgroup Results

Outside of the zonal and age-based disaggregation reported above, the results discussed thus far have centred on aggregate comparisons of baseline and midline transition outcomes. Given that one goal of this evaluation is to assess differences in project impact across distinct subgroups – in an effort to determine whether the project is benefitting the most marginalized and evaluate whether particular interventions or approaches to the project are more effective than others – the aggregate effects presented above need to be supplemented with subgroup analysis. In this section, we discuss transition outcomes among a wide variety of subgroups, investigating the potential role of a wide variety of barriers – ranging from poor school infrastructure to household poverty to exogenous shocks – on successful transition.

Data necessary to establish subgroups come from numerous sources. The most important single source is the household survey, conducted with the head of each household and the primary caregiver of the girl in the sample. Other data is drawn from the child survey, a module completed by girls, which includes questions about their self-esteem and empowerment, as well as information about their school, teachers, and so forth. School-level subgroups are established primarily from data collected from head teachers and from physical observation of schools. Finally, other data was provided directly by Relief International, including information regarding which project schools were drought- or conflict-affected at both baseline and midline.

In the table below, we report on transition outcomes for a range of subgroups. In each case, we report the share of girls in the midline cohort sample who successfully transitioned, the share who were held back and did not progress from grade to grade, the share who dropped out, and the number of observations in the subgroup. We emphasize that the sample for this analysis is limited to the midline cohort sample; while combining this data with the benchmark sample would provide additional data and statistical precision, the benchmark sample has already been extensively analysed during the baseline round.

The best way to approach this table is to consider the successful transition rate for each subgroup, comparing each subgroup's share to the transition rate observed in the top row, among the full sample of cohort girls (88.8 percent).[112] We denote subgroups in which transition rates are significantly different from the overall sample with an asterisk, which represents significance at the 5 percent level in a two-sided t-test. This test adjusts for the sample design by clustering standard errors by school. Note that we do not report statistical significance for the share of girls who are held back or dropped out, because distinguishing between these two outcomes, as opposed to distinguishing between successful and unsuccessful transitions writ large, is of less obvious importance, and statistical significance in one outcome (e.g., successful transition) will be directly correlated with statistical significance in the other

[112] The overall transition rate is slightly different in this subsample because we include girls who fall outside the age range eligible for baseline-to-midline comparisons reported in the overall transition analysis above. Specifically, the subgroup and barriers analysis includes girls who are above 17 years old; the reason for their inclusion is to preserve as large of a sample as possible for subgroup analysis.

two, unsuccessful outcomes. We provide information on these two additional outcomes to allow readers to complete held-back and dropout rates, in case there are interesting patterns – as noted above, being held back is a less negative outcome than dropping out, in our view.

Finally, we encourage readers to pay careful attention to the number of observations in each subgroup, reported in the right-most column. Many of the subgroups analysed include 10-20 respondents, or even fewer, and we discourage drawing conclusions from such small samples.

The subgroups are organized in broad categories, such as disability status and the characteristics of a girl's head of household or caregiver. In the first panel, we report results from groups defined by geography and essential characteristics of girls. Two findings are particularly compelling: first, transition rates in Galmudug are much higher than the sample average or either of the other two zones in the data. This finding is consistent in the benchmark sample, as we discussed in the previous section. Second, girls who face dramatic personal challenges to school enrolment are much less likely to successfully transition. Just 49.1 percent of girls who are married successfully transitioned at the time of the midline; similarly, 61.9 percent of girls who have borne children successfully transitioned.

Beyond these primary findings, girls in rural areas are at a small disadvantage relative to the overall sample. Girls who experience linguistic difficulties, such as those whose households use a language different from the language of instruction at their school, are no more or less likely to remain enrolled and progress in school than other girls.

TABLE 59: MIDLINE TRANSITION RATES AND PATHWAYS, AMONG KEY SUBGROUPS

	Successful Transition	Held Back a Grade	Dropped Out	Observations in Subgroup
Characteristics:				
All cohort girls	88.8	7.4	3.7	1117
Somaliland	86.6	7.7	5.7	468
Puntland	90	8.8	1.2	526
Galmudug	93.5	0	6.5	123
Rural	87.9	6.6	5.5	377
Urban	89.3	7.9	2.8	740
Single orphan	88	9	3	112
Married	49.1*	21	29.9	37
Mother	61.9*	13.6	24.4	28
Old for grade	88.9	6.8	4.3	444

Mother tongue different to LOI	89.1	10.9	0	174
Speaks LOI poorly	90.5	9.5	0	86
Disability				
Vision impairment	82.8	8.4	8.8	11
Hearing impairment	100*	0	0	8
Mobility impairment	81.3	18.7	0	5
Cognitive impairment	59.8	19.6	20.6	5
Self-care impairment	100*	0	0	2
Communication impairment	100*	0	0	2
Any disability	77.5	13.2	9.3	21
Serious illness	84.5	10.4	5.1	165
HOH and Carer Characteristics				
HOH female	88.2	7.6	4.2	696
HOH no education	88.2	7.2	4.6	550
HOH no wage-earning occupation	88	6.9	5.1	425
HOH Pastoralist	78.5	15.2	6.3	32
Carer no education	88.8	7.1	4.1	571
Carer non-literate	89.1	6.5	4.4	383
Household Assets				
Owns land	87.3*	8.4	4.3	708
Owns mobile phone	89.3	7.4	3.3	967
Poverty				
House is informal/temporary structure	94.1	5.9	0	68
Gone to sleep hungry most days	76.6	11.7	11.7	10
Gone without enough clean water most days	93.6	6.4	0	18

Gone without medicines or medical intervention most days	92.7	5.4	1.9	59
Gone without cash income most days	87.2	6.5	6.3	65
Migration and Exogenous Shocks				
Household migrates seasonally	88.8	11.2	0	10
IDP household	77.6	8.4	14	60
Drought-affected at baseline	87.8	6.2	6	375
Drought-affected at midline	93.4	4.2	2.4	216*
Conflict-affected at baseline	94	4.4	1.6	182*
Conflict-affected at midline	94	6	0	65

In the next panel, we report results for girls with a variety of disabilities. As we observed at the baseline, the set of girls with disabilities in the cohort sample is simply too small to allow valid inferences regarding the relationship between disability status and EGEP-T outcomes, including learning and transition. The best approach to analysing the impact of disability is to aggregate across all types of disability studied, which provides a sample size of 21 girls with disabilities, with the majority of these girls reporting vision and hearing impairments (though some girls reported multiple impairments). Among this group, transition rates are 11.3 points lower than the overall sample, though even this stark difference is not large enough, given the small sample size, to be distinguishable from a null effect. Unfortunately, the analysis of specific disability types is hampered by even smaller sample sizes.

The qualitative data suggest that access and discrimination may lead some girls with disabilities to drop out of school. One respondent enumerated two reasons children with disabilities drop out, “The disabled children drop out for two reasons. First, lack of school facilities. For example, if they are learning upstairs, then they cannot access the classroom which causes them to drop out. Second, the long distance to school is another challenge which may cause them to drop out.”[113] In addition, another respondent suggested the lack of separate facilities for girls with disabilities may lead to discrimination against girls with disabilities that will lead them to drop out, “If the children are using the same classroom, then they may discriminate against the disabled, and they may drop out.”[114]

[113] FGD, Fathers, Puntland
[114] FGD, Fathers, Somaliland

Moving to caregiver and household characteristics, there is very little evidence linking any standard measures of household disadvantage and transition rates. Girls whose caregivers lack education have transition rates of 88.8 percent, and similar rates obtain in the case of girls whose head of household lacks any formal education.[115] Girls residing in female-headed households are no more or less likely to remain in school and progress through the grades than other girls, based on the results in the table.

The next two sections of the subgroups analysis focus on household assets and household poverty. We do not find a consistent relationship between typical indicators of household asset ownership or household economic deprivation and transition rates. On one hand, girls whose households own land and those who have gone to bed hungry frequently in the recent past are both less likely to remain school. On the other hand, girls living in temporary or informal structures (defined by use of a lower-quality roofing material) are more likely to transition successfully, as are girls who have experienced deprivation in the form of a lack of water or medicines. Given the importance of hunger in the Somali context, experiences of recent hunger are arguably the strongest indicators of household deprivation, because every household requires food and going without it is an undeniable sign of family struggles. The fact that this variable is associated with lower transition rates is consistent with our expectations, though the results are not statistically significant.

The final panel presents arguably the most surprising set of findings from any of the subgroups. While we expected girls from structurally-disadvantaged backgrounds – either in the form of household poverty or uneducated parents or caregivers – to be more likely to drop out, our strongest expectations were around drought, conflict, and IDP status. There are strong theoretical reasons to expect girls affected by drought to drop out of school in response, either because they are forced to migrate to urban areas or live with family members, or because the household lacks the funds to pay their school fees. Likewise, conflict disrupts "normal life" and should present a particularly onerous burden on girls' enrolment status. However, we find no evidence that conflict affects transition rates. Indeed, in some cases, girls in such households – especially those affected by recent conflict – seem to be more likely to remain enrolled and progressing through the grades. Girls in schools coded as conflict-affected at baseline have transition rates of 94.0. While girls in drought-affected areas are not *more* likely to transition successfully, drought also does not appear to be a meaningful predictor of *lower* transition rates.

These findings are broadly consistent with findings from qualitative interviews with a wide range of participant types. Married girls and mothers, as highlighted above, are much less likely to remain in school than their unmarried and childless peers. Marriage, in particular, came up frequently in the qualitative data. When asked to describe hypothetical reasons why girls drop out of school, one boy said "When girls get married, they can't continue their education regularly."[116] Other boys noted that a marriage proposal is a barrier to schooling.[117] Some boys believed that girls – at least some of them – would continue their schooling after getting married – and one father mentioned that it has become more common in recent

[115] Girls whose caregivers completed only religious education (madrasa education) are also at no obvious disadvantage in terms of transition – 88.5 percent of girls in this group transitioned at midline.
[116] FGD, Boys, Puntland.
[117] FGD, Boys, Puntland; FGD, Girls, Puntland.

years to see married girls continue their education – but most recognized that this was unlikely in practice, and that cultural norms and household demands would prevent it.[118] One female teacher specifically highlighted cultural barriers to enrolment for girls who get married before completing their education, and one girl in Galmudug, participating in a focus group, drew attention to the distinction between boys and girls who get married – noting that boys will continue their schooling after marriage, but girls will not.[119] Another girl, in Puntland, said that such a boy would receive support from others, but an equivalent girl would not be encouraged to continue her schooling.[120]

In at least one case, the qualitative data did not always agree with the quantitative results presented above. We found limited quantitative evidence linking transition rates to household poverty. Transition is slightly less likely in households that own land, and other markers of deprivation are only loosely connected to lower transition rates. In contrast, financial strain was remarked upon repeatedly by interviewees of all kinds.[121] The effect of financial strain is not always strictly about the burden of school fees; in other cases, girls seek out paid work to help support their families.[122]

Moving beyond the subgroups reported above, we now turn to a closely-related analysis of barriers to successful transition. The subgroups considered previously are typically characterized by their stability – girls are either mothers or not, their geographic location does not change dramatically, and their household is either poor or affluent, with relatively little temporal variation. In contrast, the barriers analysed below are characteristics of project schools, including the existence of EGEP-T interventions, and aspects of familial support for education that can be changed in response to advocacy and outreach. As before, we report the share of girls who successfully transition, who are held back, and who drop out, along with the number of respondents in the subgroup in the final, right-most column. Comparisons are best made to the sample average transition rate of 88.9 percent; statistically significant differences in transition rates are denoted with an asterisk.

The first set of barriers we consider are infrastructural. Schools that have poor facilities, lack access to water, or do not provide separate toilets for girls are less hospitable and provide girls with less incentive to continue their education. Girls are often uncomfortable using toilets that are shared with boys, and some girls travel home or elsewhere to use toilets that they feel are more appropriate or cleaner. A lack of separate toilets first presents a barrier to consistent attendance, but may, in extreme cases, prompt girls to drop out altogether, or fall so far behind that they are held back a grade.

In general, there is some limited evidence that school infrastructure shapes transition rates. Schools that entirely lack electricity or water have transition rates somewhat lower than other schools (86.4 percent, in the case of schools that lack electricity, and 83.0 percent in the case of schools that lack access to water altogether). Other infrastructural features do not seem to have any impact: for instance, the availability

[118] FGD, Boys, Puntland; FGD, Fathers, Puntland; FGD, CEC, Puntland; FGD, Girls, Somaliland; FGD, Mothers, Somaliland.
[119] KII, Female Remedial Teacher, Puntland; FGD, Girls, Galmudug.
[120] FGD, Girls, Puntland.
[121] FGD, Girls, Puntland; FGD, CEC, Puntland.
[122] FGD, CEC, Puntland.

of separate toilets, equipped with privacy walls, do not make girls more likely to remain in school. Even the finding regarding electricity, noted above, is likely a spurious correlation – there is not much theoretical reason to expect the presence of electricity to discourage continued enrolment, but electricity is less common in more marginalized, and more rural/remote, communities. Thus, any decline in transition associated with a lack of electricity may simply be capturing lower transition rates in poorer communities.

In the qualitative data, there is evidence from all three zones that CECs and school management prioritize installation of water systems at schools in their communities when they are able to raise money or request support from projects like EGEP-T. When asked what school improvements have been made recently, one CEC member explains that the CEC "built many classes as well as made a water installation system,"[123] and a CEC member from a different zone describes that their school did a lot of planning with Relief International around installing a water tank."[124] In Galmudug, one boy explains that recently, "there have been classrooms, toilets, and a water well-constructed."[125] Additionally, when asked what he dislikes about his school, one boy explains, "There is nothing that I dislike in this school but I would like that feeding and water are added and to complete the school equipment."[126] This may be particularly concerning for students who have to travel a long distance to school and are not easily able to access food or water when attending school.

Additionally, although girls mostly spoke about the importance of electricity when asked specifically about the benefits of solar lamps at the household level, some respondents did describe the general lack of access to electricity in certain communities. For example, in one FGD, a girl explained that solar lamps are "beneficial because during rainy season, the electricity cuts off."[127] When asked about the support their community has received, one respondent requested that Relief International consider the lack of access to electricity: "In addition to that, the electricity is not available in town, so we are asking for help with that too."[128] However, it is unclear whether these quotes are referring to lack of electricity at schools or lack of electricity in rural areas where students live. Regardless, it is also clear from the qualitative data that installation of fans is also prioritized when money becomes available for school improvement fees. Given that schools in some areas also adjust their school years due to the heat, this may be one explanation for why electricity would be considered important at the school level. As one student describes, "Also, the number of fans at the school is limited and the school is hot in summer."[129]

TABLE 60: MIDLINE TRANSITION RATES AND PATHWAYS, AS A FUNCTION OF EXPECTED BARRIERS TO TRANSITION

	Successful Transition	Held Back a Grade	Dropped Out	Observations in Subgroup

[123] FGD – CEC – Somaliland
[124] FGD – CEC – Puntland
[125] FGD – Boys – Galmudug
[126] FGD – Boys – Galmudug
[127] FGD – Girls – Somaliland
[128] FGD – Fathers – Puntland
[129] FGD – Girls – Somaliland

Barriers:				
All cohort girls	88.8	7.4	3.7	1117
School Infrastructure				
No electricity	86.4	7.5	6.1	238
No water at school	83.0	8.8	8.2	129
No clean water at school	87.2	6.6	6.2	238
Only dirt floors	89.3	7.0	3.7	83
Separate toilets for girls and boys	88.5	7.7	3.8	963
Separate toilets for girls, with privacy and available for use	88.0	8.0	4.0	895
School Resources and Support				
School feeding program	88.2	5.6	6.2	449
School has CEC	88.9	7.3	3.8	1005
CEC meets every 2 weeks or more	92.4	3.5	4.1	282
CEC provides bursaries to girls	90.1	6.6	3.3	408
CEC provides at least 10 bursaries to girls	89.9	8.2	1.9	309
Offers remedial classes	91.3*	5.6	3.2	693
Offers life skills classes	90.1	3.6	6.3	200
Offers sanitary kits	88.4	7.2	4.3	469
Offers solar lamps	90.8	4.3	4.9	353
Has girls club	89.5	6.6	3.9	966
Girls club meets at least once per week	90.6	5.8	3.6	137

Gender Equity and Support at School				
At least one female teacher	88.5	6.7	4.9	851
Female mentor	89.0	6.3	4.7	878
Male mentor	87.1*	8.2	4.7	750
No mentor	90.1	9.9	0.0	107
Absentee teaching staff	84.2	9.5	6.3	305
Caregiver Attitudes and Intra-Household Decision-making				
Family would withdraw daughter from school to pay emergency bill	89.7	6.9	3.4	207
Family would not support schooling after marriage for girl	87.9	8.4	3.7	440
Disagree: men in this household support girls' education	87.5	8.1	4.4	121
Caregiver aspires to less than university education for girl	82.2	11.5	6.3	83
Schooling decisions made by adults only	89.6	6.2	4.2	555

The availability of school resources and support from EGEP-T and other sources has a mixed and insubstantial relationship with transition outcomes. For instance, the presence of a CEC does not appear to impact transition rates, though an active CEC – in terms of meeting frequency – is associated with a very small uptick in transition rates. School feeding programs, which we would expect to encourage girls to stay in school, especially in drought-affected or marginalized communities, are not correlated with an increase in transition. Other forms of support are also unrelated to higher transition rates: CECs that provide bursary support to at least one girl are associated with a transition rate of 89.7 percent, almost exactly the same rate as the sample average. Even especially supportive CECs, providing a minimum of

ten local girls with bursary support, based on reports from head teachers, are not related to higher transition rates.

In contrast to the activities of CECs, a few other EGEP-T interventions appear to provide very small, but meaningful, improvements in transition rates. Schools that offer remedial instruction and life skills classes have transition rates of 91.3 and 90.1 percent, respectively, though only the former is statistically different from the overall sample transition rate. Schools that have provided solar lamps within the last year have significantly higher transition rates, at 90.8 percent. And, while schools with girls clubs have only a slightly higher transition rate than other schools, those with particularly *active* girls clubs – meeting once a week on average – show a greater gain in transition rates, relative to the typical school.

Perhaps most surprisingly, the presence of a female mentor does not appear to influence transition rates, nor does the presence of a female teacher. We might expect girls to respond positively to the presence of an effective female role model, in the form of either a female teacher or a female mentor. Neither is associated with an increase in transition rates from the sample average of 88.8 percent. Schools with high levels of teacher absenteeism are associated a marginal decline in transition rates, however. We identified schools at which two or more teachers (out of up to 5 for whom we collected data at each school) had been absent at least one day in the past two weeks, classifying them as schools with problematic teacher attendance. Transition rates are 4.6 percentage points lower at schools with high teacher absenteeism, when compared to schools with no absenteeism problem.

The final set of barriers concerns the attitudes of household members toward girls' education. The barriers in this section are tied fairly directly to the project's Theory of Change, in the sense that increasingly positive community attitudes are expected to generate more consistent enrolment among girls. The evidence presented in the last panel do not appear to support this argument *in general*, however. In the household survey, caregivers were asked what level of schooling they would *like* their daughters to achieve. We analyse this question at length in Section 7.5, on community attitudes. Briefly, the vast majority of caregivers (91.2 percent in the midline cohort transition sample) indicated that they would like their girl to complete college or university.[130] We identified caregivers who set their sights lower, aspiring to secondary completion or less for their girl. Among girls whose caregivers have lower aspirations, transition rates were just 82.2 percent, substantially below the sample average.[131]

We also asked caregivers to imagine a scenario in which their sister needs help paying for an emergency hospital bill. We asked caregivers whether they would sell household goods or livestock to help their

[130] The share of caregivers with high educational aspirations for their daughters is covered at length in Section 7.5, partially due to concerns regarding the quality of this measure and the potential for social desirability bias, in which respondents tend to provide respondents they think interviewers or others around them wish to hear. For the purpose of analyzing transition rates, we still find this to be a useful measure, because it identifies caregivers who are *willing* to express less supportive views of girls' education, though we admit the measure should be viewed with a grain of salt.

[131] This finding is not statistically significant (p = 0.15), despite the substantial reduction in transition rates documented. This is likely due to the small number of caregivers (n = 83) in the subgroup and the clustered nature of data collection.

sister, or withdraw their daughter from school to save on school fees. Based on responses to this question, we classified caregivers who would be willing to withdraw their daughters from school; in our view, these individuals support girls' education more weakly than those who would make a tangible sacrifice – selling goods or an animal – to keep their daughter in school. However, girls whose caregivers expressed less support for their education transitioned successfully 89.7 percent of the time, slightly more than the sample average.

In fact, despite our initial finding regarding familial attitudes, which were based on caregiver aspirations, none of the other measures of the same concept – support for girls' education from male household members and caregivers who say adults are solely responsible for educational decisions affecting their girl, to take two examples – are affiliated with systematically higher transition rates. Even in the case of familial attitudes, the evidence for a connection to transition rates is mixed.

What should we conclude from the subgroup and barriers analysis, in which relatively few variables are associated with meaningful differences in transition rates? The simplest explanation is that transition is a multidimensional outcome, influenced by a wide range of community-, school-, household-, and personal-level factors. Girls from marginalized households may have a harder time paying their school fees, but their parents may be especially supportive of girls' education; girls may prefer better school infrastructure, but remain in school thanks to the support provided by EGEP-T in the form of a solar lamp and a female mentor to whom they can look up to. The multi-causal nature of transition makes it unlikely that bivariate analysis will reveal many relationships of interest.

5.4 Transition and Bursaries

Among a range of interventions that are expected to influence learning and transition outcomes, bursaries provided to severely marginalized girls stand out as particularly focused on influencing transition. As emphasized by numerous qualitative interviewees, the financial strain of education is a major reason girls drop out of school. Bursaries are designed to target this very specific barrier to continued enrolment for marginalized girls.

The midline evaluation provides a useful opportunity to test the differential impact of bursaries on transition rates, isolating their effect from other project outputs. Our goal in this section is to test the difference in transition rates between "regular" cohort girls – who do not receive a bursary but otherwise receive the full package of EGEP-T interventions – and bursary girls, who receive the same set of interventions, in addition to a scholarship to pay for their school fees.[132] At baseline, both bursary and cohort girls were recruited through schools, resulting in a set of both types of girls who were uniformly

[132] Although there is some variation in the package of project interventions applied to different groups of girls (e.g., girls in rural areas are provided with solar lamps; girls in IDP camps are not provided bursaries, because the camp schools do not charge fees), there is also significant similarity in the set of interventions across groups. Our interest in this section is on the impact of bursaries alone, holding constant – to the extent possible – the other interventions girls receive.

enrolled at baseline. Girls in the two samples are similar to each in many ways, including the grade range and schools from which they were selected. This structure allows us to compare dropout, held-back, and transition rates between the two groups, to study whether bursaries have impact above and beyond any effect of other EGEP-T interventions that are shared between the two groups.

Our sample for this analysis consists of cohort and bursary girls who were successfully re-contacted at midline for the purposes of tracking transition. Girls in this transition sample were not replaced, as replacement would have occurred at schools and biased estimates of transition rates upward (given that girls sampled from schools are invariably enrolled in school at the time of data collection). In total, the sample includes 1,117 cohort girls – the same midline cohort sample analysed in the previous section on subgroup transition rates, generally, including girls of all ages – and 298 bursary girls.[133]

Overall, bursary girls are less likely than cohort girls to remain enrolled at midline, and are less likely to progress to the next grade level (i.e. transition successfully). Transition rates are 87.3 percent among bursary girls and 88.9 percent among cohort girls, respectively. Bursary girls are very slightly more likely to drop (4.0 versus 3.7 percent) and more likely to be held back (8.7 versus 7.4 percent) than cohort girls. The overall finding suggests that bursaries are not associated with more successful transition, at least in this bivariate analysis. Moreover, the ancillary finding around dropouts cuts against the idea that bursaries encourage girls to stay in school – if this were the case, we would expect bursary girls to remain enrolled, even if they were prevented from progressing a grade. Instead, we find that bursaries increase both dropout and hold-back rates, with an overall decline in transition rates among bursary girls. Although the decline cited is small – just 1.6 percentage points overall – it does complicate our view of bursaries.

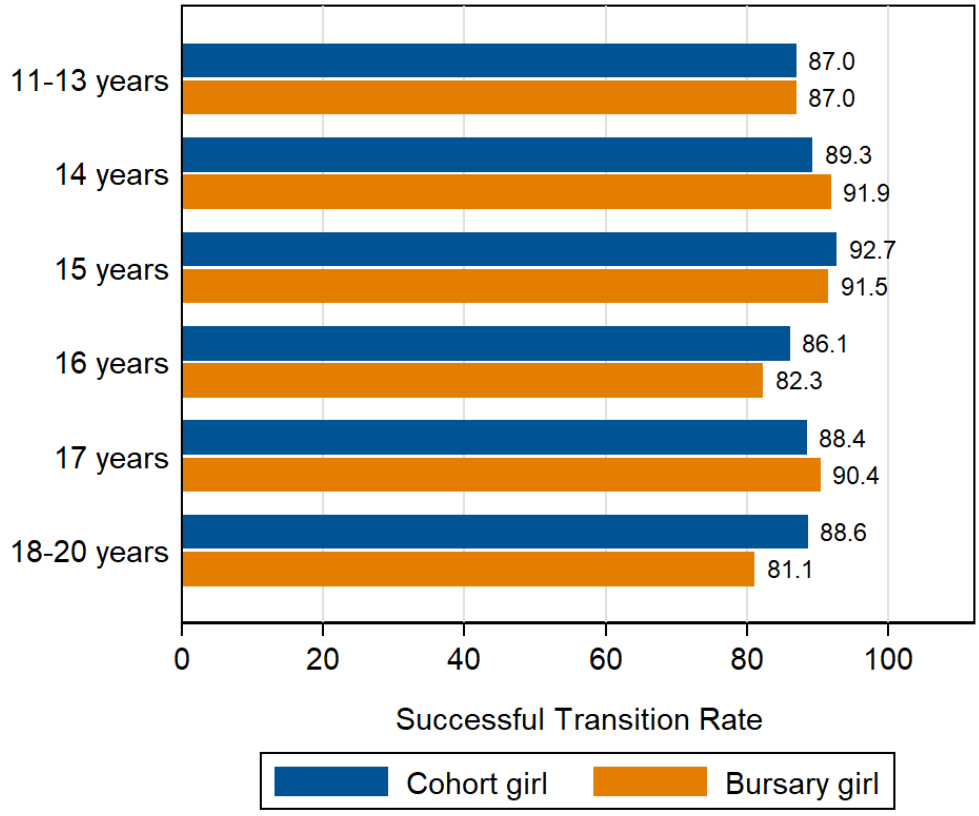
As noted, the gap in transition rates between bursary and cohort girls is small. The difference does not approach statistical significance at any conventional level, with a p-value of 0.44 after accounting for clustering of standard errors by school. The lack of statistical significance suggests a lack of any systematic relationship between bursaries and transition. This idea is buttressed by the noisy results we observe – reported in the figure below – between bursary and cohort girls when we disaggregate by age. The figure below plots transition rates by age group for each type of girl. Bursary girls do not consistently under- or over-perform cohort girls within their respective age groups.[134] The consistency of the differential

[133] In our comparison of cohort and bursary girls, we do not employ post-stratification weights. We also do not employ weights to adjust for the differences in age distribution between bursary and cohort girls, because the regression analysis reported in this section directly controls for age. Post-stratification weights are also unnecessary because our most exhaustive analysis includes school fixed effects, meaning that comparisons between bursary and cohort girls are being made within the same school, and post-stratification weights would not influence within-school comparisons of this kind.

[134] We consolidated the youngest and oldest age groups, as shown in the graph, due to the small number of bursary girls aged 11-12 years, and 19-20 years, respectively. Creating groups that span 11-13 years and 18-20 years provides a cleaner picture of transition rates among bursary girls, untainted by the noise that occurs in extremely small samples (e.g., 3 bursary girls aged 12 years and 2 bursary girls aged 20 years).

suggests that the very small aggregate effect reported above is representative of the true relationship, or lack thereof, between the receipt of a bursary and successful transition.[135]

FIGURE 15: TRANSITION RATES AT MIDLINE AMONG COHORT AND BURSARY GIRLS, BY AGE



In the same vein as controlling implicitly for age, we considered the possibility that the relationship between receipt of a bursary and transition may be obscured by other confounding variables. In other words, if bursary girls are more likely to be attend schools that have naturally higher transition rates, this could produce bias toward a positive effect of bursaries on transition. While each baseline school had an established quota for sampling of 12 cohort girls and 4 bursary girls, variation in the sample size across schools is commonplace. In the case of bursary girls, this can occur because not all schools had four bursary girls available at baseline. In the case of both types of girls, the sample available for analysis is a function of successful re-contact at midline, a process which is not random and may be correlated with other factors that influence transition rates.

[135] While this was unlikely, based on the random sampling approach employed, there are marginal differences in the mean age between the two groups: bursary girls are 0.18 years older than cohort girls in the sample, on average ($p = 0.11$).

A multivariate approach to this question is particularly important, given that bursary girls were *purposively* selected by the project, which targeted the provision of bursaries to girls considered the most marginalized. Unfortunately, non-random assignment of bursaries, as a treatment, makes it difficult to study their causal impact on transition rates, though our analysis can attempt to control for underlying differences in household economic deprivation and individual-level factors thought to be associated with a girls' relative marginalization.

A more rigorous analysis of bursaries' impact would control for a number of variables that influence transition rates, including age, zone, and even idiosyncratic school-specific characteristics. To further assess the impact of bursaries, we estimated a series of linear regressions predicting successful transition.[136] Our key independent variable is respondent type, i.e. whether a girl receives a bursary or not. We added control variables in a stepwise fashion, incorporating dichotomous variables for each age group, followed by adding dichotomous variables to control for underlying differences in transition rates between zones. Each of these dichotomous variables accounts for observed and unobserved characteristics that are associated with a given age group or zone, including differential propensities for successful transition that might otherwise bias a regression model or bivariate analysis.

After controlling for age and zone, the gap between bursary and cohort girls shifted only slightly, from -1.6 percentage points to 1.7 points suggesting that age and zone were not systematic confounders for the relationship between bursary receipt and transition outcomes. Building on this model, we incorporated additional control variables that were predictive of transition rates in our subgroup analysis, such as motherhood, marital status, and urbanity. The logic of this approach is that variables which simultaneously predict successful transition *and* bursary receipt must be included in the regression model to avoid omitted variable bias. Given that bursaries were targeted to marginalized girls – which likely included those most at risk of early marriage or early pregnancy, though this may not have been part of the specific targeting strategy – controlling for these characteristics is especially important. As we showed in Section 5.3, motherhood and marriage are associated with significantly lower transition rates. At the same time, bursary girls are slightly more likely to be married and have children. Therefore, these variables may bias the impact of bursaries toward a null (zero) effect. Including motherhood, marital status, a dummy variable for rural schools, and a dummy variable capturing schools affected by drought at midline resulted in a dramatic increase in the overall predictive value of the regression model. However, they did not substantially change the gap in transition rates between bursary and cohort girls, which remained negative at -0.7 percentage points ($p = 0.70$).

Model development of this kind is prone to a different form of bias – analysts who select variables to include that give them the results they seek. It is possible that our results exist on a knife edge – the

[136] Despite the fact that successful transition is a binary outcome, we employ linear models, which are often referred to in this context as linear probability models. We prefer linear models over logit or probit models in this context, although the latter two functional forms are designed to account for binary dependent variables. Our preference stems from the robustness of linear (OLS) regression to violations of core assumptions required for making valid inferences. In addition, linear models are better able to handle large numbers of fixed effects and high-dimensional regression, as we employ toward the end of this analysis.

inclusion of additional controls or the exclusion of just one control variable could eliminate the finding regarding bursaries' impact. To guard against this possibility, we took an analytically drastic step, employing school fixed effects. Put simply, fixed effects are a single dichotomous/dummy variable for each observed value of a variable. In the case of school fixed effects, we add a dummy variable for *every school in the dataset,* 117 in total. Fixed effects control for latent differences in transition rates across schools. A comparison of bursary and cohort girls that includes school fixed effects is assessing the difference in transition rates between these two types of girls *within the same school*, aggregating these 117 individual school-specific effects to draw conclusions about the effect of bursaries on transition. In most cases, this is the most conservative modelling approach possible, in the sense that it is the least likely to return a false positive result for an independent variable of interest, such as bursaries.

When we add school fixed effects to the model, we observe a small improvement in bursary girl transition rates vis-à-vis cohort girls. However, the gap is still negative, with transition rates among the typical cohort and bursary girl 88.9 and 88.6 percent, respectively. Even the subsequent inclusion of marital status, motherhood, and several indicators of household economic deprivation – such as the frequency of going to bed hungry – did not reveal a meaningful relationship between bursary receipt and successful transition.

The analysis presented in this section is suggestive, but not conclusive. The research design employed does not take into account potential underlying differences in expected transition rates under the counterfactual (i.e. the possibility that transition rates among bursary girls – in the absence of a bursary – would be significantly lower than those among cohort girls). Without data on baseline transition rates among bursary girls *before* they received a bursary, it is difficult to disentangle the impact of bursaries from pre-existing differences in the probability of successful transition.

One specific criticism of the analysis is that it does not account for differences in household economic marginalization between bursary and cohort girls. Because bursary girls were selected on the basis of their relative marginalization, we would expect them to face greater economic barriers to continued enrolment. Indeed, this is the entire premise of EGEP-T's bursary intervention. In additional analysis, we attempted to control for household economic status – especially the experience of hunger and land ownership; inclusion of these control variables did not influence the null results presented thus far.

This approach is not entirely satisfactory, though. The available data on household economic status is limiting. And economic barriers to transition figured prominently in qualitative interviews. A number of interviewees argued that school fees were among the top one or two reasons why girls drop out of school.[137] Household poverty is seen as a significant barrier, regardless of whether schools charge fees, because girls' education has other associated costs beyond school fees – lost income from paid work, lost help on domestic chores, and the cost of learning materials, uniforms, and other sundry items.[138]

[137] FGD, Girls, Puntland; FGD, Boys, Puntland.
[138] FGD, CEC, Somaliland; FGD, CEC, Puntland.

The multivariate analysis provided is the most rigorous possible assessment, given the available data. A more rigorous analysis of bursaries' impact on transition rates would require a more targeted research design, which is unlikely to be possible in most contexts, such as a randomized controlled trial, in which bursaries were randomly assigned to a pool of girls, with follow up 1-2 years later. In the context of non-random assignment, the conservative modelling approach employed here provides reasonably strong evidence for a null relationship between bursaries and transition rates, but it should not be interpreted as entirely conclusive nor the final word on the question.

5.5 Enrolment Counts as a Transition Metric

The results reported in this section have been subject to a number of methodological caveats, which we discussed at length in Sections 2.4, 2.5 and earlier in Section 5. Most importantly, our analysis of transition rates has involved comparing girls from a random household sample at baseline to a sample made up exclusively of cohort girls at midline, which gave rise to concerns about comparability between the two samples.

One method of checking the robustness of our primary transition results – which showed a small but insignificant improvement in transition rates at midline – is to triangulate using an alternative data source. While we do not have transition data for another sample of respondents, classroom headcounts provide an opportunity to check whether gross enrolment rates have changed from baseline to midline. Our goal in this analysis is to study the change in total enrolment in project schools from baseline to midline, using enrolment data collected from individual classes. To the extent that transition rates have improved, we would expect fewer girls to drop out, and enrolment rates to rise. If enrolment numbers have fallen from baseline to midline, it would suggest a decline in successful transition.

It is important to note the shortcomings of this analysis. First, we cannot capture anything approximating true transition, because many girls could remain enrolled in school, but fail to progress to the next grade. This outcome is observationally equivalent to a girl who progresses to the next grade, transitioning successfully. In addition, enrolment rates could change from baseline to midline due to shifts in the population of school-age girls in EGEP-T communities. If the school-age population drops, total enrolment could fall alongside a rise in the *share* of girls who are enrolled.

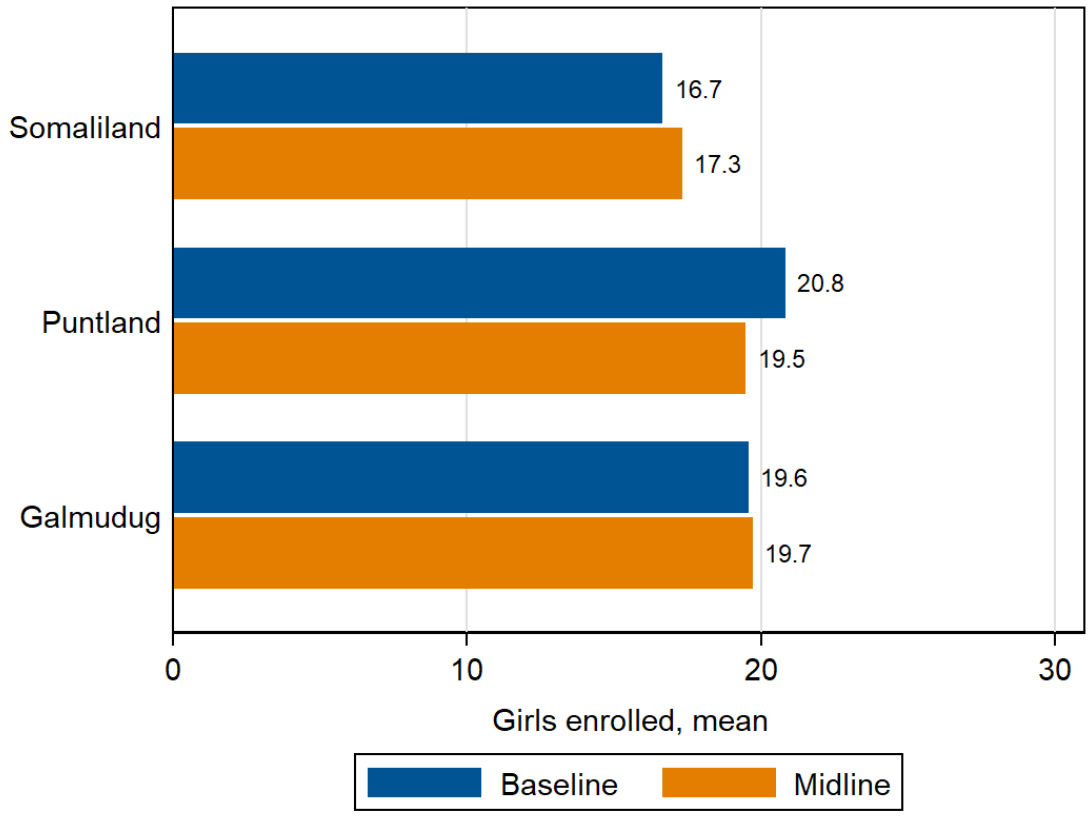
Nonetheless, these caveats do not reduce the utility of the analysis altogether. We have little reason to suspect large-scale population shifts in EGEP-T project communities as a whole. And, while enrolment rates are distinct from transition rates, they are likely to move together. As a broad check on our transition findings, enrolment rates are useful – if we observe very large gains or declines in enrolment rates, it would make us question the earlier finding that transition rates have risen but only slightly.

For this analysis, we employ a comparable sample of 1,298 classroom headcounts conducted at baseline (n = 663) and midline (n = 635). The sample consists exclusively of schools visited in both rounds, and in

which classrooms in both rounds had enrolment numbers available.[139] To the extent possible, the sample was constructed to be comparable over time, to ensure that a few larger or smaller classrooms or schools did not influence the results.

The results of this triangulation exercise are reported in the figure below. Overall, enrolment rates were extremely stable between baseline and midline, declining by 0.34 girls per classroom over this period, from 18.9 girls at baseline to 18.5 girls at midline. Small gains were seen in Somaliland, while a small decline was observed in Puntland, but neither change is statistically significant and likely stems from sampling variation centred on shifts in the classrooms selected.

FIGURE 16: MEAN CLASSROOM ENROLMENT AT GIRLS, BY ROUND AND ZONE



As noted, sampling variation could explain small deviations from baseline to midline. Such variation, especially if it occurs in variables correlated with enrolment rates – such as grade level of classes sampled

[139] A number of classrooms, and entire schools, lacked enrolment numbers for classrooms. In these cases, field teams conducted classroom headcounts but could not calculate attendance as a share of enrolled students. For the purpose of our analysis in this section, a lack of enrolment data prevents us from using the data from that classroom entirely, because our interest is in enrolment rather than attendance.

– could bias the results of our analysis. We tested for this possibility by estimating a series of linear regression models predicting the count of girls enrolled in each classroom, where our key independent variable was the round of data collection.[140] In these models, we controlled for the grade level of the classroom, with a binary indicator variable for each grade level; we also experimented with models employing a fixed effect for every school in the sample. In these models, the core finding is unchanged: enrolment rates declined very slightly from baseline to midline, but the decline is sufficiently small that it is indistinguishable from zero. The results serve as a useful robustness check for our earlier transition findings. If enrolment rates changed dramatically – either going up or down – at midline, it would call into question the idea that transition rates have been relatively stable over the same period. The fact that both transition rates and enrolment rates have been steady since baseline should provide more confidence in our results regarding transition.

6. Sustainability Outcome

EGEP-T's third core outcome, as with other GEC-T projects, is sustainability. EGEP-T seeks to impact learning and transition/enrolment levels for girls and, indirectly, boys in project schools. EGEP-T simultaneously aims to maintain those improvements by promoting sustainable change in schools, communities, and within the educational system in Somalia. Sustainability in the project's Theory of Change is driven by active community support for girls' education, the actions of mobilized CECs, and the development of links between schools and system-level actors – such as Teacher Training Institutes and Ministries of Education – that can promote a range of positive outcomes, including better teaching and improved child protection.

Sustainability outcomes are divided into three categories, the levels at which changes are expected to occur, with 3-4 outcomes per category. The sustainability indicators for EGEP-T are listed below. In the sections that follow, we measure and evaluate each indicator in turn.

COMMUNITY-LEVEL OUTCOMES

- Percentage of surveyed members of EGEP target communities, who have been exposed to project awareness-raising activities report having changed their opinion positively in relation to the importance of girls' school completion
- Percentage of surveyed members of EGEP target communities, that report that boys and men taking action to support girls in attending and completing school
- Percentage of EGEP target communities where community leaders are leading campaigns and advocacy events

[140] We also tested our findings with a poisson model, which is typically better suited to variables in which the dependent variable is a count. We prefer the linear models in this case because their interpretation is straightforward and because the distribution of the dependent variable is not extremely skewed. However, the results are similar using both specifications.

SCHOOL-LEVEL OUTCOMES

- Percentage of CECs providing bursary support to marginalised girls in EGEP target schools
- Percentage of CECs in EGEP target schools providing match funding (financial or in-kind) for school improvement initiatives
- Percentage of schools actively participating in the peer-mentoring programme
- Percentage of schools with functional child protection mechanism

SYSTEM-LEVEL OUTCOMES

- Federal level and Galmudug Gender Units develop gender strategy and Federal level strategy is being implemented within the life of the project
- Percentage of EGEP-target schools receiving follow-up monitoring visits from MoE officials, including Gender Focal Points, District/Regional Education Officers/RES
- TTI (Teacher Training Institutes) integrate components of the CPD approach into their curriculum for teacher training

Sustainability Indicator 1 - Awareness-Raising and Attitudinal Change

Percentage of surveyed members of EGEP target communities, who have been exposed to project awareness-raising activities report having changed their opinion positively in relation to the importance of girls' school completion.

The first indicator of sustainability selected by EGEP-T focuses on awareness-raising efforts around girls' education. The project has implemented a series of campaigns designed to encourage enrolment of out-of-school girls or re-enrolment of girls between school years, and generally to raise awareness of the importance of girls' education and, especially, completion of education through secondary school and beyond. Broad awareness-raising is important for sustainability going forward because shifting community attitudes will ensure a supportive environment for girls' education following the end of EGEP-T interventions.

Our measurement approach to this indicator is more straightforward than that suggested by the indicator itself, which is fairly complex. First, we measure exposure to awareness-raising activities via data collected from caregivers and teachers, triangulating across these two disparate respondent types. In both cases, we asked respondents whether they were aware of any of the following activities or events occurring in their community over the past year:

- Door-to-door visits about the importance of girls' education
- Back-to-school campaigns encouraging the enrolment of girls in school
- Events where community leaders gave speeches encouraging girls' education

It is important to note that this list is unlikely to be exhaustive – Relief International and its implementing partners may have engaged in or promoted other types of outreach that are not captured by these specific categories, in which case our estimates of exposure may be lower than the true value.

Second, our focus in this section is on exposure to these activities, rather than on their effect on individuals' views of the importance of girls' education. We single out exposure to awareness-raising activities because it is not possible to statistically connect awareness-raising activities to more pro-education views, as the indicator implies. For instance, we may find that a caregiver was exposed to awareness-raising efforts and also holds positive attitudes, but we cannot say that exposure changed their views. Attributing causality of this kind is incredibly complex, and cannot be studied effectively with the available data.

It is also the case that community attitudes are addressed at length elsewhere in this report, as they are captured as part of the intermediate outcomes.[141] Moreover, to the extent that awareness-raising activities improve community attitudes, this fact would not be reflected well in our data, because the vast majority of caregivers in our baseline and midline samples are caregivers of girls enrolled in school. Therefore, they are more likely to already hold relatively more positive attitudes toward girls' education, and awareness-raising may have the least effect on them, owing to their higher initial starting values. For these reasons, we have elected to analyse the extent of exposure to awareness-raising efforts, without analysing the next link in the indicator, concerning attitudes.

The overall results of our analysis are reported in the figure below, which draws comparisons between baseline and midline levels of exposure to awareness-raising activities. We asked respondents whether they were aware of any of these activities in the last year in their community; even if they were not directly exposed in the sense of receiving a visit or attending an event, their awareness of the event is sufficient for measuring the frequency of such events. The left panel reports results from a sample of caregivers. As the results make clear – and as we discussed in the baseline evaluation report – awareness-raising activities for EGEP-T had already begun prior to baseline data collection, which explains why so many respondents report exposure at baseline. These activities took place specifically prior to the start of the school year during which baseline data collection occurred.

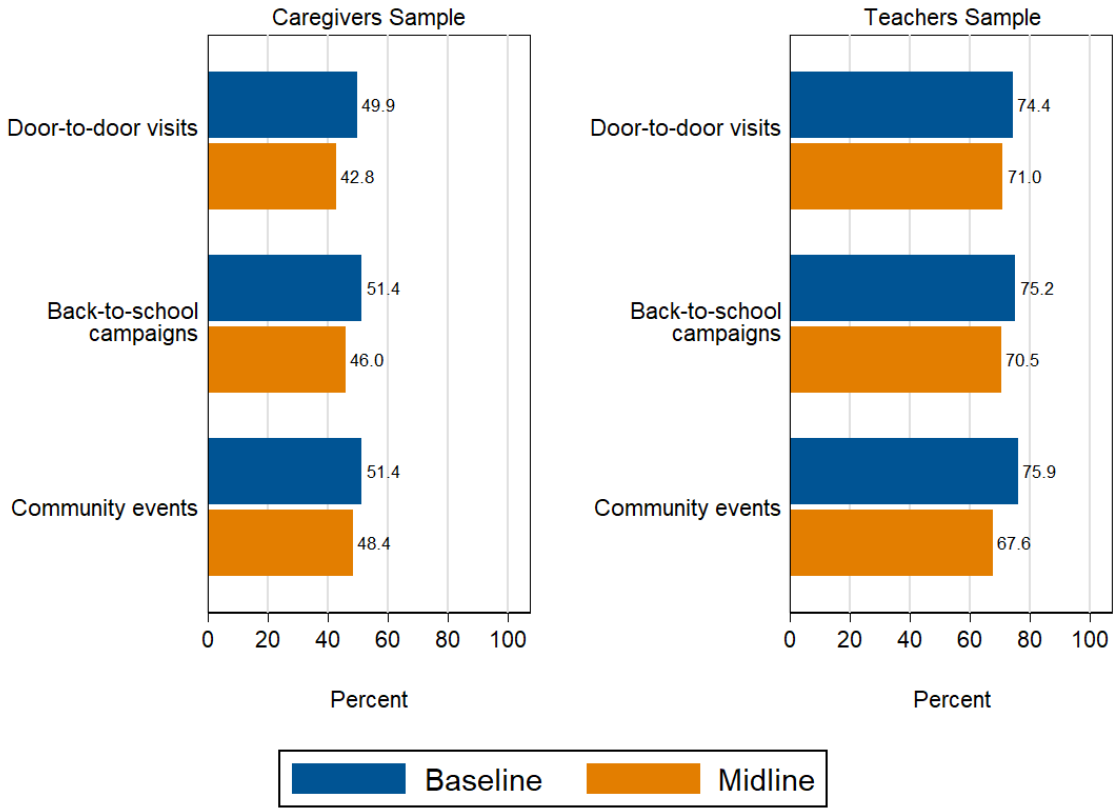
The data suggests that fewer caregivers are familiar with awareness-raising activities at midline than baseline. Back-to-school campaigns are illustrative: at baseline, 51.4 percent of caregivers surveyed were aware of such efforts, compared to 46.0 percent at midline. This finding is especially surprising, given the change in sampling approach for caregivers between baseline and midline. As we have discussed at length elsewhere, baseline caregivers were recruited via a random household sample, and included a significant minority of respondents who were caregivers to out-of-school girls. In contrast, caregivers at midline were exclusively those whose daughters were enrolled at baseline, specifically in EGEP-T project schools. All else equal, we would expect the latter group to be most familiar with awareness-raising activities

[141] To illustrate this idea, we estimated a regression of individual support for girls' education on exposure to awareness-raising activities, using the sample of caregivers surveyed at baseline. Exposure to each activity was measured as a binary (yes/no) variable, and we also controlled for underlying heterogeneity across zones. There was no meaningful relationship between any of the three exposure variables and caregiver aspirations for their daughters' education, either in isolation or jointly. While there are strong theoretical reasons to expect awareness-raising activities to improve community attitudes, such as parental aspirations, it is not possible to make causal attributions of this kind using the type of cross-sectional data available to us from community members.

occurring as part of EGEP-T programming, both because their daughters attend EGEP-T schools and because they are more likely to be aware of education-focused activities occurring in their communities.[142]

On the other hand, the differential timing of baseline and midline may explain a portion of the discrepancy between the two rounds. At baseline, data collection occurred 2-3 months after the start of the school year; at midline, data collection occurred at the very end of the school year, prior to the summer break. Activities like back-to-school campaigns naturally take place prior to the start of the school year and would, therefore, be most fresh in the minds of respondents just 2-3 months later than they would be if they had occurred 8-9 months prior.

FIGURE 17: EXPOSURE TO AWARENESS-RAISING ACTIVITIES AROUND GIRLS' EDUCATION, BY ROUND



Importantly, responses from teachers confirm almost precisely the findings stemming from surveys with caregivers. The right panel of the figure above presents results based on the responses of teachers in

[142] An alternative interpretation is that activities like door-to-door visits are more likely to be targeted to households with out-of-school girls, which would run counter to our initial expectation above. This is certainly possible, but the baseline data do not support it: caregivers of OOS girls were *less* likely to have heard of door-to-door visits or back-to-school campaigns than caregivers whose girls were enrolled in school already.

EGEP-T schools. In both rounds, teachers are more likely to indicate awareness of the activities in question; this is not altogether surprising, given their careers and employment at EGEP-T schools – teachers are more likely to be aware of events regarding education than a typical community member, and teachers in EGEP-T schools should be especially aware of events or outreach campaigns sponsored by EGEP-T.

Although teachers are more likely to indicate exposure in both baseline and midline, we observe the same pattern of decline in exposure rates from baseline to midline. On all three metrics, teachers were less likely to recall an event of the type specified at midline than they were at baseline.[143]

Moving beyond aggregate results, the table below disaggregates exposure to each type of event by zone. The top panel in the table focuses on responses from caregivers, which illustrate a consistent pattern across all three activity types: respondents in Somaliland were least likely to have been exposed to awareness-raising activities, while those in Galmudug were most likely. This pattern is not replicated in the bottom panel, based on responses from teachers: teachers in all three zones were approximately as likely to indicate awareness, with minor variations that are inconsistent across activity types.

TABLE 61: EXPOSURE TO AWARENESS-RAISING ACTIVITIES, AT MIDLINE, BY ZONE

Outcome	Somaliland	Puntland	Galmudug	Overall
Caregivers of girls				
Door-to-door visits about importance of girls' education	27.9	49.6	72.3	42.8
Back-to-school campaigns encouraging enrolment of girls	32.9	52.9	67.6	46.0
Events where community leaders encouraged girls' education	34.1	55.6	73.7	48.4
Teachers				
Door-to-door visits about importance of girls' education	74.1	70.5	61.5	71.1
Back-to-school campaigns encouraging enrolment of girls	75.0	68.1	65.4	70.7
Events where community leaders encouraged girls' education	63.5	71.7	67.3	67.9

Although the table above does not report over-time comparisons disaggregated by zone, there are stark differences here as well. Specifically, the gaps shown in the top panel of the table above – which reflect

[143] The sample of teachers was restricted, as in our analyses of teaching quality and other outcomes that rely on the survey of teachers, to ensure comparability from baseline to midline. The school-level sample was restricted to the subset of schools that appear in both evaluation rounds. While teachers selected in each school varied, this approach ensures school-level differences do not influence the results.

the status quo at midline – are also reflective of baseline-to-midline changes on the same indicators. Among caregivers in Somaliland, exposure to awareness-raising activities declined precipitously from baseline to midline; strictly for illustration, the share of Somaliland caregivers reporting that door-to-door visits occurred in their community in the last year fell from 47.1 percent to 27.9 during this period. Puntland also experienced declines that were relatively consistent across activity type, but of much smaller magnitude than those found in Somaliland. Finally, in sharp contrast, respondents in Galmudug were significant *more* likely to indicate exposure to awareness-raising activities at midline than they were at baseline – on the same metric regarding door-to-door visits, the share of Galmudug caregivers exposed rose from 39.6 percent at baseline to 72.3 percent at midline.

In general in the qualitative data, the positive effects of awareness activities were mentioned with great frequency, but the source of the awareness raising was not always made explicitly clear. This precludes us from providing a nuanced analysis around zonal differences or the prevalence or non-prevalence of certain awareness activities. However, there is evidence in the qualitative interviews that awareness raising activities come from a variety of sources linked to the project. Awareness raising is taking place at the community level through CECs, training with parents, radio advocacy, "door-to-door campaigns to send girls to school,"[144] girls' leadership networks, teachers conducting outreach to parents, and in some limited cases, community leaders.

These activities reportedly have an extremely positive effect in boosting girls' enrolment and attendance, as well as shifting community attitudes toward girls' education. As one CEC member explains, "there is a lot of different advocacy that has been done to support girls' education like Relief Trainings. This training focused on the role of parents in supporting their girls. Girls' enrolment increased because of the awareness training done and women got special considerations for their education."[145] Additionally, "the head teachers and CEC make conferences about girls' educational improvements," and "the awareness on the radio has changed the attitudes of the community." Girls have also taken responsibility for raising awareness in their communities and raising funds for girls' education through leadership networks. One girl described the leadership network as a particularly active advocacy group: "It's a network of girls and we organize activities such as identifying the students who are unable to pay their school fees. We pay the school fees for them and we also buy the school uniform for them."[146]

Awareness raising activities are also having an effect on caregivers' attitudes toward education. In the FGDs with mothers and fathers, parents overwhelmingly express support for girls' education. There is just one notable case where a father said he does not see the value in girls' education and views it as a waste of time, but all other parents interviewed for the midline have positive views toward girls' education, at least when directly asked to share their opinions on the value of girls' education. There is also ample evidence in the qualitative interviews that, as a whole, girls still face heavy discrimination in their households and communities but despite this, there is still notable progress that goes beyond what parents report of their own attitudes toward girls' education – for example, evidence that girls are

[144] FGD, CEC, Puntland
[145] FGD, CEC, Puntland
[146] FGD, Girls, Somaliland

continuing to attend school even after they become married or pregnant and that parents support these developments.

When asked directly to share their views on the value of girls' education, many parents noted that an educated girl translates to an educated community, as girls will pass along their learnings to their children when they become parents and will more closely follow their children's education. Parents also highlight that educated girls will support their families if they receive an education. As one father explains, "Girls' education is very important - it would increase the reputation of the girls, she would help her children, her life, health and overall life. We can say an educated mother has a healthy family. In addition to that, 90% of girls help their parents so uneducated girls could not help her parents."[147] Also noteworthy, some parents specifically highlight the importance of household-level support for girls' education, suggesting that there is an increase in awareness around the influence parents can have on their daughters' motivation to continue schooling: "Firstly, I consider to give equal opportunities to both girls and boys. Also I don't want my daughter to regret why your parents did not teach you. I don't want to discourage my daughters. I like my daughter to be the best girls compared other girls."[148]

Parents do not always attribute these changes directly to project activities, but they do in some cases. For example, when asked what has contributed to changes in girls' education, one respondent explains, "The reason for the change is because the community received a lot of awareness about adolescent girls in school, and even if she gets married to continue the education."[149] In another instance, when asked whether there has been a change in the ways girls are treated in their homes, a respondent explains, "There is a change in how boys and girl been treated. A lot of raising awareness happened about treating boys and girls equally."[150] There are also many cases where respondents simply highlight the changes that have taken place in community attitudes toward girls' education over time: "Yes, because the community itself is changing. In the past time, girls were not used to send to school, and now we are in a level to support the girls in education and health. And the parents understood to educate the girls."[151] However, it should also be noted that the levels of community awareness raising appear to vary widely from location to location. In some locations, caregivers report that community leaders, religious leaders, and government officials have not been active in providing support for girls' education, whereas in other locations, parents report that there has been heavy activity around girls' education. For example, in one community, a parent reports: "The ministry has a strategy, and the religious leaders and parents have done awareness activities. They pay the tuition fee and personal financial source, and the religious leaders tells the community to do actions. A lot of training has been done for the community which Relief participated in. The Ministry of Education provides the salary of some of teachers."[152]

Awareness raising is also taking place in schools, through boys' and girls' groups and other mechanisms established through the project. For example, one girl explains that boys are supporting girls' education:

[147] FGD, Fathers, Galmudug
[148] FGD, Girls, Galmudug
[149] FGD, Fathers, Puntland
[150] FGD, Fathers, Puntland
[151] FGD, Fathers, Galmudug
[152] FGD, Fathers, Puntland

"At school, there is a boys' club that encourages [raises awareness for] girls' education. Teachers also conduct competitions between girls and boys to test their knowledge."[153] CECs, teachers, and community leaders are also active in raising awareness at schools: "She can get encouragement from parents and family too. When we are in school assembly, the religious leaders come to school sometimes for providing awareness about education benefits. Therefore, the student may like more to continue education."[154]

As noted above, it is difficult to always conclude with any certainty from the qualitative data the exact source of awareness-raising activities. It does seem, however, that although respondents from all areas report their communities have been exposed to some sort of awareness raising, certain schools and communities are particularly active in awareness-raising, which could explain why we are not observing consistent increases in prominence of awareness-raising activities in the quantitative data.

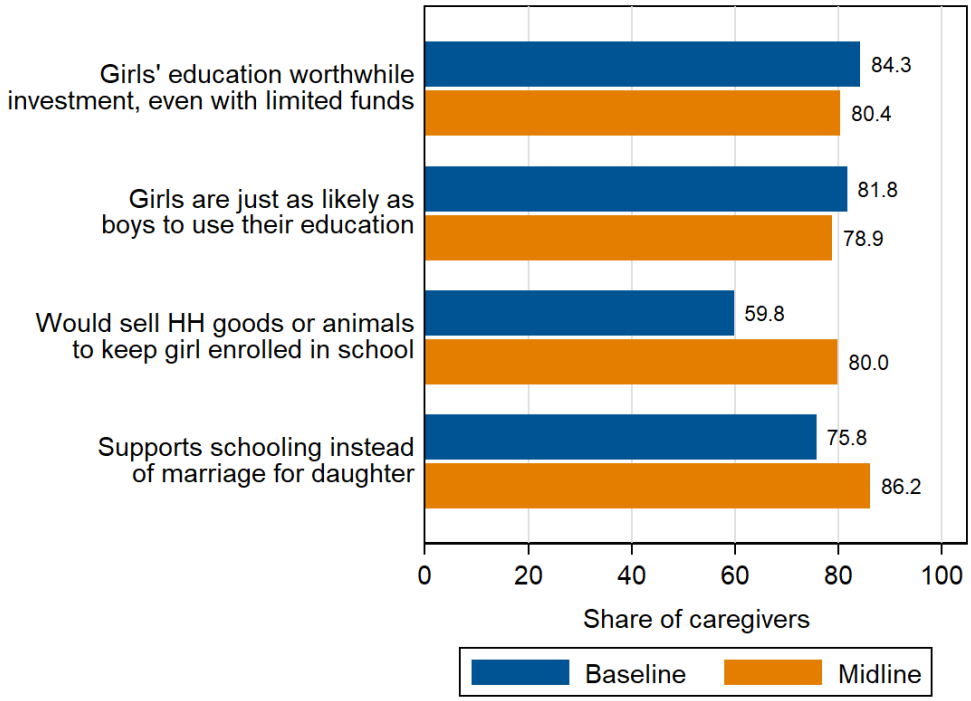
Community Attitudes and Actions

The sustainability of the awareness-raising and attitudinal change activities is dependent on the extent to which those activities have an impact on community attitudes. While section 7.5 goes into greater detail about community attitudes, the following provides an overview of how key indicators of community attitudes has changed from baseline to midline.

Our evaluation team surveyed caregivers regarding their attitudes on girls' education. The surveyed caregivers were asked indicate the extent to which they agreed or disagreed with each statement, and the distribution of their answers are presented in the figure below. The changes in attitudinal support for girls' education are mixed with some positive statements about girls' education having greater agreement among the sampled caregivers and with other positive statements have less agreement.

[153] FGD, Girls, Somaliland
[154] FGD, Boys, Puntland

FIGURE 18: SHARE OF CAREGIVERS EXPRESSING SUPPORT FOR GIRLS' EDUCATION



When caregivers are presented with hypothetical choices regarding girls' education, we find significant positive shifts in the way they would handle the hypothetical situation. We asked respondents to imagine a situation in which their adult sister is sick and needs money to pay for her hospital bill. Respondents were asked to choose whether they would sell some household goods or livestock to help their sister, or withdraw their daughter from school and use the savings from school fees to help their sister. Significantly more caregivers said that they would sell HH goods or animals before withdrawing their daughter from school.

In the second hypothetical presented, we asked caregivers whether they would support the marriage of their hypothetical 15-year-old daughter. We found that significantly more caregivers in all three regions said that they would prefer that their daughter stay in school instead of accepting the marriage proposal.

TABLE 62: CAREGIVER SUPPORT FOR TANGIBLE ACTIONS TO SUPPORT GIRLS' EDUCATION IN HYPOTHETICAL SITUATIONS

Zone	Baseline Percent	Midline Percent
Would sell HH goods or animal before withdrawing daughter from school		
Somaliland	61.7	88.9
Puntland	56.7	67.5
Galmudug	67.7	98
Would support daughter to stay in school rather than accept marriage proposal		

Somaliland	75.2	87.2
Puntland	75.4	83
Galmudug	80.6	95.9

Discussion

The results in this section should be interpreted as cautionary, but not entirely conclusive. In general, the findings are concerning, because according to the quantitative data, awareness-raising efforts have become *less* prominent within EGEP-T communities since baseline. While a fairly substantial share of respondents were familiar with EGEP-T awareness-raising activities at baseline, we would expect a meaningful increase at midline to reflect increased sustainability over time. At the same time, the baseline evaluation is, itself, not a true baseline for this indicator, as project activities of this kind were already underway at the time of the baseline. Further, we believe there are good reasons to suspect biased recall at midline, as the time elapsed from the start of the school year (from August/September 2018) to the period of data collection (April/May 2019) is sufficiently long that respondents may not accurately recall exposure. Nonetheless, we would expect to see, at minimum, consistency in exposure across rounds, a standard which was not met at midline.

Despite the decrease in awareness-raising activities, we find that support girls education among caregivers in the sample is approximately the same when measured with disagreement or agreement to statements about girls' education, and we find that support has actually *increased* when caregivers are provided with hypotheticals when they have to weigh their daughters' education against other exigencies. This finding that support has increased while activities have decreased perhaps suggests that these attitudes about girls' education may persist among community even without the same level of engagement by RI with a community.

Indicator Grade: 2.5 – Emerging

Sustainability Indicator 2 – Male Support for Girls' Education

Percentage of surveyed members of EGEP target communities that report that boys and men are taking action to support girls in attending and completing school

One of the goals of EGEP-T programming is to ensure sustainability of improvements to girls' education made through the first round of EGEP (GEC phase 1) and the transition phase, EGEP-T. A core assumption of EGEP-T programming is that community attitudes and community support for girls' education has an impact on transition rates and learning outcomes for girls. Closely related to overall community attitudes, addressed elsewhere in this report, are tangible acts of support for girls' education by boys and men in the community. To the extent that fathers have influence or control over household finances, their support for girls' education is essential; boys' influence is less direct, but can still promote a positive learning environment for girls, among other benefits. Therefore, in an effort to promote sustainable

change in EGEP-T communities, the project is seeking to promote positive male attitudes toward, and tangible actions in support of, girls' education.

From a measurement perspective, this sustainability indicator is sufficiently broad that it is not possible to utilize a single measure. Attitudinal measures will capture just one aspect of the indicator – attitudinal change from round to round – but will not indicate whether those attitudes produce tangible actions on the part of men and boys in support of girls' education. Data that focuses on tangible actions are of particular interest, but the range of possible actions in support of girls' education is broad, and cannot be captured in a single question. Moreover, the data collection strategy did not include surveys specifically of male family members of cohort girls, though data was collected directly from cohort boys regarding their attitudes and support of girls' education. Owing to these difficulties, we brought together several different measures, with data collected from disparate sources, including caregivers, teachers, head teachers, cohort girls, and cohort boys, supplemented with targeted questioning during qualitative interviews with many respondent groups, the quantitative portions of which are summarized below. The outline of questions below mirrors the structure of this section.

- **Surveys of caregivers**
    - General perception of male support for girls' education in the household and community
        - Agree/Disagree: Men support girls' education
    - Tangible actions taken by men [in the community or household] to support girls' education
        - Agree/Disagree: Men read to their daughters
        - Agree/Disagree: Men help their daughters with their schoolwork
        - Agree/Disagree: Men support girls to complete schooling before marriage
        - Agree/Disagree: Men support girls to continue school after getting married
- **Surveys of teachers**
    - Perception of male support for girls' and boys' education
        - Now I am going to ask you about a number of community members. I'd like you to rate their relative support for [girls'/boy's] education. [Fathers]
    - Tangible actions taken by men to support girls' education
        - If you scheduled a parent/teacher meeting with a girls' parents, which parent or parents do you think would attend the meeting?
- **Surveys of girls**
    - Tangible actions taken by boys in their household to support girls' education
        - How often does your brother help you with your household chores?
        - Now imagine that your mother asked your brother to help you with the chores, so that you could spend more time studying. How likely is it that he would help you?
    - Tangible actions taken by boys in school to support girls' education
        - Boys at school encourage me and other girls to continue our education
        - Boys at school encourage me and other girls to answer questions in class
        -

Perceptions of Caregivers

The first data source we consider is survey reports from caregivers of adolescent girls.[155] Caregivers were asked to assess the extent to which men support girls' education, both within their own households (i.e. their family members or the male family members of their friends) and within their community more broadly. Caregivers – who include a significant minority (20.4 percent) who are men – were asked whether men in their communities tend to read to their daughters when they are young, and help their daughters with schoolwork when they need it. These questions, which focus on tangible actions, were repeated, focusing on male members of the caregivers' households. In addition, caregivers were asked about intangible support, by asking them the extent to which they agreed, for instance, with the statement "men in this community support girls to complete their schooling before getting married."

In each case, caregivers were asked the extent to which they agreed or disagreed with the statement. For the sake of concise presentation, we calculated the share of caregivers who either agreed or agreed strongly with each statement at baseline and midline. The results are presented in the table below, with the first panel reporting overall results across all zones including in the midline evaluation (Somaliland, Puntland, Galmudug, and Hirshabelle). The panels that follow report zone-specific estimates from baseline and midline.

In the aggregate, there is no evidence that male support for girls' education has increased since baseline, at least according to the reports of caregivers. Across the four community-focused indicators, only two increased from baseline to midline, and neither change was statistically significant. At the same time, a more substantial – though still insignificant – decrease was observed in the extent to which men are perceived to support girls to continue schooling after marriage (from 68.9 percent of caregivers agreeing with such a statement, to just 63.7 percent).

Household-focused indicators did not fare any better. We might expect respondents to be more critical of their community at large than the male members of their own household, or perhaps to see the changes made in their households more clearly than the diffuse changes in male attitudes that may have occurred in their community. But this is not the case: across all four indicators, caregivers report a decrease in male support for and tangible actions in support of girls' education.[156]

---

[155] As noted in the methodological discussion in Section 2, the set of respondents asked these questions – which constitute a household survey in project communities – changed from baseline to midline. At baseline, the sample encompassed households in project communities with at least one girl aged 11-18 years, regardless of their enrolment status. At midline, the sample targeted the households of cohort and bursary girls, who were invariably enrolled at baseline. Enrolment status of daughters is naturally correlated with attitudes toward girls' education and support from male household members. In line with our methodological discussion, we limit the baseline household sample utilized in this section to those caregivers of girls who were enrolled in school, to ensure comparability from baseline to midline.

[156] Interestingly, male caregivers are no more likely to report that men in their household and community support girls' education. At midline, male caregivers (who comprise 20.4 percent of the total sample of cohort and bursary girls' caregivers) were actually *less* likely to believe that men in their households and communities act in support of girls' education.

TABLE 63: CAREGIVER PERCEPTIONS OF MALE SUPPORT FOR GIRLS' EDUCATION, BY ROUND

Outcome	Baseline Percent	Midline Percent
Overall		
Agree: Men read to their daughters	77.5	76.3
Agree: Men help their daughters with their schoolwork	74.1	74.9
Agree: Men support girls to complete schooling before marriage	76.9	78
Agree: Men support girls to continue school after getting married	68.9	63.7
Agree: Men in this HH support girls' education	81.5	80.1
Agree: Men in this HH help daughters with their schoolwork	79.2	77.1
Agree: Men in this HH prefer girls to complete schooling before marriage	81.2	79
Agree: Men in this HH prefer girls continue schooling after marriage	74.1	65.4
Somaliland		
Agree: Men read to their daughters	86.6	73.4
Agree: Men help their daughters with their schoolwork	83.9	70.9
Agree: Men support girls to complete schooling before marriage	83.2	77.3
Agree: Men support girls to continue school after getting married	75.2	54
Agree: Men in this HH support girls' education	85.2	84.9
Agree: Men in this HH help daughters with their schoolwork	83.9	83.7
Agree: Men in this HH prefer girls to complete schooling before marriage	84.6	85.5
Agree: Men in this HH prefer girls continue schooling after marriage	79.2	62.1
Puntland		
Agree: Men read to their daughters	69.6	75.3
Agree: Men help their daughters with their schoolwork	67.3	73.9
Agree: Men support girls to complete schooling before marriage	67.8	76.1
Agree: Men support girls to continue school after getting married	66.7	66.2
Agree: Men in this HH support girls' education	80.7	73.7
Agree: Men in this HH help daughters with their schoolwork	77.8	68.4
Agree: Men in this HH prefer girls to complete schooling before marriage	77.2	71.1
Agree: Men in this HH prefer girls continue schooling after marriage	73.1	64.1
Galmudug & Hirshabelle		
Agree: Men read to their daughters	77.4	92.6
Agree: Men help their daughters with their schoolwork	64.5	95.3
Agree: Men support girls to complete schooling before marriage	96.8	88.5
Agree: Men support girls to continue school after getting married	51.6	90.5

Agree: Men in this HH support girls' education	67.7	88.5
Agree: Men in this HH help daughters with their schoolwork	64.5	87.8
Agree: Men in this HH prefer girls to complete schooling before marriage	87.1	87.2
Agree: Men in this HH prefer girls continue schooling after marriage	54.8	83.8

The aggregate results mask trends that diverge dramatically across zones. In Somaliland, caregivers believe male support has declined across the board, in every respect. They believe men, both in their household and their community at large, are less likely to help their daughters with schoolwork, and less likely to support girls to complete schooling before marriage or continue schooling after marriage. In fact, the results in the table actually understate the deterioration in male support in Somaliland: in terms of three household-focused indicators – support for girls' education, helping daughters with schoolwork, and preferring that girls complete schooling before marriage – the results in the table suggest that there was no meaningful change from baseline to midline. However, by aggregating the "agree" and "strongly agree" categories in the table, we have obscured substantial declines in these three indicators as well, as caregivers tended to move from strong agreement at baseline to milder agreement at midline. Overall, there were statistically significant perceived declines in male support across all eight indicators in Somaliland.[157] Of course, as noted earlier in this section, it is possible shifts in caregiver perceptions of male support could be explained by changes in the sample who responded to the question.[158]

The results from Puntland contrast, generally, with those from Somaliland. In Puntland, caregivers report significant improvements in male community members' support for girls' education on three of four metrics measured. The fourth metric, support for girls to continue schooling after marriage, also demonstrates a small net improvement when we consider the full range of possible responses (including strong and mild disagreement), but the change from baseline to midline is not statistically significant. While respondents in Puntland report improvements at the community level, they are less optimistic about male support within their own households, reporting decreased support across all four indicators focused on these men.

The qualitative interviewees held mixed opinions on the topic of male support for girls' education. In the mothers' focus group discussions, mothers almost unanimously agreed that fathers and boys are supportive of girls' education. There were a few cases where mothers explained that males believe girls and women should stay at home, but these beliefs appear to be changing: "…there is a saying which goes,

[157] In these cases, where response options followed an ordinal structure, we employed a chi-square test to test for the differences in the distributions between rounds. In every case in Somaliland, the decline was significant with $p < .05$.

[158] While we do not have a strong theoretical reason to expect a random sample of caregivers to have different perceptions than a sample of caregivers whose girls are enrolled in EGEP-T schools specifically, it is important to emphasize the fact that the specific respondents have changed since baseline (in other words, this analysis is not based on a panel of respondents who appeared in both rounds), and this variation could explain shifts in perceptions, even in the absence of a tangible change in male support for girls' education in Somaliland.

'Although a girl is educated, she will end up in the kitchen,' although now civilization has changed."[159] Interestingly, it appears that the value of girls' education is still framed in reference to the advantages it offers for marriage and childrearing, with one mother explaining that "we believe an educated girl is easy to marry,"[160] and another explaining that "when the girl is educated, she can support her kid in school."[161] It is not entirely surprising that during this period of transition, some community members are framing the values of girls' education within the broader social norms around the role of women.

In the fathers' focus group discussions, there were very active debates regarding the merits of girls' education, the role of girls in the home, and expectations towards girls as they relate to education. In particular, fathers were asked to share their opinions on the household chore burden girls face, as well as societal pressures to drop out of school and marry. In general, the results again suggest that, for some, there is a tendency to frame the value of girls' education within broader social norms around marriage and the role of women in society. The evidence from the fathers' qualitative interviews, in particular the presence of active debate within the focus group discussions, suggests that communities are going through a period of change wherein the traditional role of women in the home and in society is being challenged but that the change is not yet wholly accepted.

For example, one father expressed that he doesn't believe girls should stay in school up to the university level because "if they are mothers, they have responsibilities which don't allow them to continue the education," and another father agreed and recommended that girls "leave the education after secondary school because 30 year old ladies don't have opportunities to deliver more children compared to the lady who gets married at the age of 20."[162] Another father disagreed, explaining that, "girls can study in every level," and "they can attend the school or universities and they can care for their children as well." Another father echoed this sentiment, explaining that he knows "a daughter and mother who learn together" and that "girls have the right to continue their education."[163]

Aside from these more general conversations around theoretical support for girls' education, there were also more specific discussions around actions boys and men are taking to support girls' education in the home and at school. Here, the results differed by group, with fathers offering far more examples of ways in which they and their sons are encouraging girls' education. The only concrete example mothers offered of fathers supporting their daughters' education was that they, as the providers, pay for school fees and materials: "Fathers pay money for girls to study, so men take part in girls' education."[164] Mothers reported that boys protect their sisters from harassment from other boys in school, "learn together [with girls] or

[159] FGD, Mothers, Somaliland
[160] FGD, Mothers, Puntland
[161] FGD, Mothers, Somaliland
[162] FGD, Fathers, Puntland
[163] FGD, Fathers, Puntland
[164] FGD, Mothers, Galmudug

read to them"[165] at home, and are generally "very happy when they see that their girls are learning in school and sitting the same classes."[166]

On the other hand, fathers – particularly those from Somaliland (in contrast to what was observed in the quantitative data results) – offered a variety of examples of ways in which they and their sons support girls' education. The attitudes father hold appear to be predictive of the level of support they offer their daughters, although this may not always hold true in the face of social norms around marriage, as noted above. As one father explains, "In the ignorance period, people used to educate only boys and assign the girls for domestic work. Later, people realized the importance of girls' education…the people are now busy to educate the girls. No one deserves to be an ignorant person."[167] Although fathers are self-admittedly less involved in following their children's education than mothers, there is evidence that some fathers are participating in school committees to motivate girls to stay in school, ensuring their daughters have time to study at home by providing them with "a chance to read and prepare early to attend school,"[168] assigning their sons to help their daughters with lessons and advising their wives to "distribute household tasks equally in order for the daughter to get a chance to read her lessons,"[169] providing emotional support to their daughters by listening to their complaints, and helping them "revise their daily lessons in order to promote their understanding."[170] However, again, the biggest way in which fathers feel they can show their support for girls' education is through paying for their daughters to go to school. One father explained how he paid to transfer his daughter to a university in Hargeisa so that she could continue her learning: "Yes, we support their goals. We give transfer to near Hargeisa. Everybody has a relative in Hargeisa, so we pay for them here to learn. For example, my daughter has finished five years of university in Hargeisa."[171]

It is unclear exactly why the quantitative and qualitative results for Somaliland are contradictory for this indicator on male support for girls' education. The incongruence may be due to generally higher expectations of male support in Somaliland vis-à-vis Puntland or Galmudug.

Teacher Perceptions of Male Support

Expanding on our discussion of male support for girls' education, we also surveyed teachers and head teachers regarding this subject. These individuals have a unique perspective on the topic for two reasons: first, they are also members of the community and observe what occurs in their communities, especially when it impacts their education, their profession, and their students. Second, teachers and head teachers

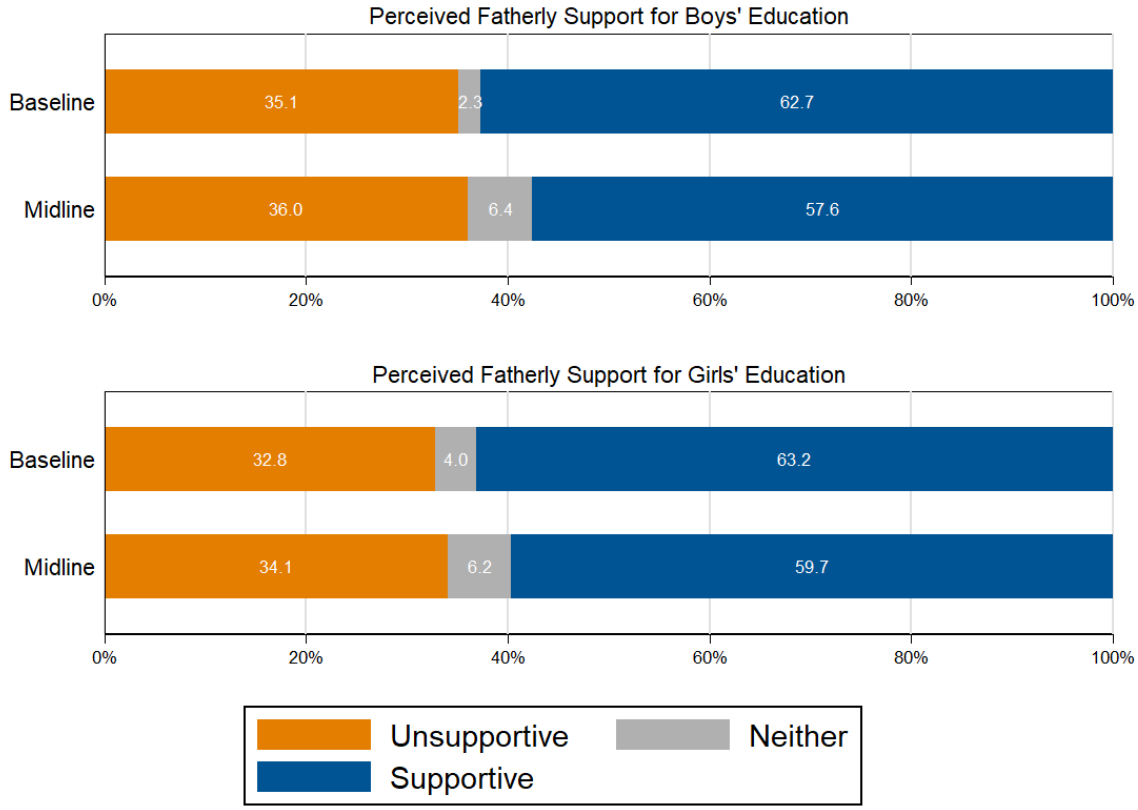
[165] FGD, Mothers, Puntland
[166] FGD, Mothers, Galmudug
[167] FGD, Fathers, Galmudug
[168] FGD, Fathers, Somaliland
[169] FGD, Fathers, Somaliland
[170] FGD, Fathers, Somaliland
[171] FGD, Fathers, Somaliland

interact with male community members who are fathers, uncles, brothers, and husbands of their female students, and often see when these individuals choose to either support or fail to support girls' education.

We asked teachers to rate the support fathers show toward boys' and girls' education, respectively, on a 5-point Likert scale. The results of this exercise, disaggregated by gender (top and bottom panels) and evaluation (round) within each panel, are reported in the figure below. As the figure shows, reported support for boys' and girls' education at baseline was not dramatically different – 62.7 percent of teachers reported that fathers were either somewhat or very supportive of boys' education, while 63.2 percent of teachers reported the same for girls' education. By the time of the midline, teachers were reporting a general decline in fathers' support for education, regardless of the gender of the student being discussed. In the case of boys, 57.6 percent of teachers at midline reported fathers are supportive of boys' education, a 5.1 point decline; in the case of girls, the decline was less dramatic, 3.5 points, but fit the same broad pattern.[172]

[172] We collapsed the 5-point likert scale to three categories to enhance the presentation of results, but this collapsing does not obscure any meaningful differences that are present in the full 5-point scale. In either case, teachers view male support for boys' and girls' education alike to be declining.

FIGURE 19: TEACHER PERCEPTIONS OF FATHERLY SUPPORT FOR GIRLS' AND BOYS' EDUCATION, BY ROUND



There are several reasons why teachers may report a decline in fatherly support for education. The fact that teachers report a decline in support for both boys' and girls' education suggests that gender preference among fathers is not driving this change. It is possible teachers have become more critical of community attitudes or that their expectations for fatherly support have increased over time. It is also possible that a broad shift in fatherly attitudes actually has occurred and that fathers appear less supportive of education in general. It is also possible that sampling variation between baseline and midline is responsible for the shift, since different teachers were surveyed in the two rounds.[173]

Thanks to a separate sample of head teachers, we are able to test the possibility that sampling variation explains the change from baseline to midline, albeit not directly. On average, head teachers reported an *improvement* in fatherly support for both boys' and girls' education from baseline to midline, with improvements of very similar magnitude for both boys and girls. This contrasts with the reports of teachers, and suggests that the trend observed in that sample may not be a true representation of reality.

[173] However, the sample employed here utilizes only schools sampled in both rounds, so any argument based on sampling variation would have to be based on sampling variation between teachers within schools, rather than differences between the set of schools sampled in the two rounds.

More importantly, neither teachers nor head teachers report a *differential* decline in fatherly support for girls' education vis-à-vis boys' education. If fatherly support for education has changed over time, it has not disproportionately benefited or harmed girls.

How has male support changed over time in terms of tangible actions fathers can take to support girls' education? Teachers were asked to imagine a scenario in which they arranged a meeting with the parents of a girl who was not performing well in school. Based on that scenario, they were asked which parent(s), if any, they would expect to attend the meeting. As one would expect, mothers were most commonly expected to attend the meeting by themselves, followed by a joint meeting with both parents. For the purposes of this indicator, we are interested in the share of teachers who believe that a father would attend the meeting, either with or without the girls' mother also in attendance.

As the table below shows, 32.8 percent of teachers at baseline believed that a father would attend such a meeting with their daughter's teacher. At midline, this share of teachers had declined to 29.2 percent, though this decline was not statistically significant. When head teachers were asked to imagine the same scenario, their views were consistent with that of teachers – at baseline, 25.0 percent said a father would attend the meeting, declining to 22.9 percent at midline.[174]

TABLE 64: SHARE OF TEACHERS WHO BELIEVE FATHER WOULD ATTEND A PARENT-TEACHER MEETING

Zone	Baseline Percent	Midline Percent
Somaliland	34.4	24.2
Puntland	30.2	32
Galmudug & Hirshabelle	50	40.9
Overall	32.8	29.2

Perceptions of Girls

The results to this point have not painted a particularly positive picture of changes in male support for girls' education in project locations. In addition to the support of fathers and adult males, we also consider the tangible actions boys can take in supporting girls' education at school and in their households.

We first considered the role of boys in their households. A prevailing pattern in Somalia is for girls to be assigned chores at home, while boys are given more time after school to study and play. This finding arose repeatedly in both the baseline and midline qualitative interviews, where teachers, parents, girls and boys alike cited the fact that boys have fewer household responsibilities and that this can impinge on girls' performance in school.

We were interested in whether boys might be willing to help their sisters with housework if it facilitated their sister's success in school. We asked girls to imagine that they had a pending examination that they

[174] These results contrast slightly with those reported by girls themselves. When asked at midline, more girls were confident reporting that their father would attend a meeting with their teacher, though the change from baseline to midline was substantively small.

needed to prepare for, and that they had asked their brother to assist them with their chores to allow them the time to study. We asked girls the likelihood that their brother would help them with their chores during such a critical time.[175] At baseline, 12.4 percent of girls said their brother was very likely to help, and another 60.5 percent said their brother was somewhat likely to help. These percentages had shifted at midline, with just 7.9 percent of girls saying their brother was very likely to help, and 62.1 percent answering that their brother's help was "somewhat likely." The decline from baseline to midline was statistically significant at the 5 percent level.

In contrast, boys themselves – when presented with the same scenario from their own perspective – reported a slight increase in their willingness to help. Notably, their self-reports of willingness to help their sisters do not differ dramatically from reports by their sisters – at midline, 16.0 percent of boys reported that they were "very likely" to help their sister with her housework if she needed it during an examination period. While boys report a small increase in their willingness to help, girls show a slight decrease in expectations of support, suggesting that the general trend may be effectively flat from baseline to midline, or at least represents too small of a change to be meaningful.

We also asked girls to assess support outside of the home. Specifically, we asked girls about the support they enjoy from their male classmates. We asked girls whether they agreed or disagreed (on a 4-point Likert scale) with the statements:

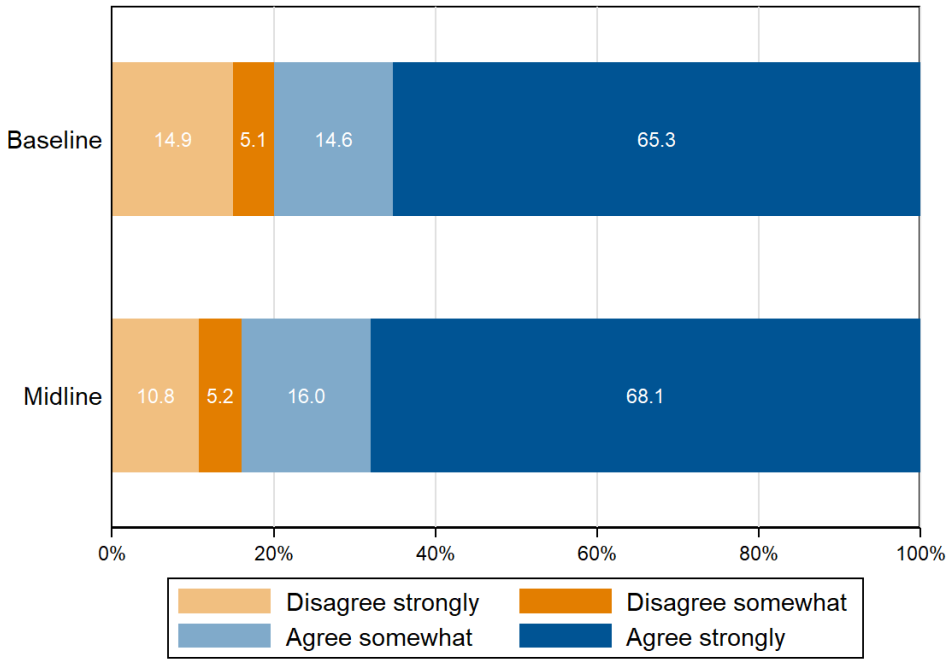
- Boys at school encourage me and other girls to continue our education
- Boys at school encourage me and other girls to answer questions in class

The figure below reports results from the latter question, though responses to both showed a nearly identical trend. At baseline, 79.9 percent of girls agreed or strongly agreed that boys in their school encourage girls to answer questions in class; at midline, this share had increased to 84.0 percent.[176]

[175] We recognize that the responses to this question are unlikely to be accurate in the sense of predicting how many boys would actually help their sisters. We fully expect girls to over-report the willingness of their brothers to help them. However, their responses indicate something meaningful about the extent of support they believe they receive from their brothers, which is nonetheless a useful measure of male support for girls' education.

[176] While this may seem a minor way to support girls' education, it addresses a serious problem for many Somali girls. As girls' enrolment and attendance rates improve, often their participation in class lags behind their male counterparts. Girls are often cited as being too shy to participate in class, especially when taught by male teachers. Therefore, encouragement from male students is a meaningful action that can improve girls' educational outcomes.

FIGURE 20: BOYS ENCOURAGE ME AND OTHER GIRLS TO ANSWER QUESTIONS IN CLASS, BY ROUND



This change represents a qualitative shift – as shown in the figure, the share of girls who strongly disagreed with the statement also declined, and there was a general shift toward greater agreement among girls that their male peers encourage girls' education generally and participation in particular.

As in the findings from the fathers' focus group discussions, there was active debate within boys' focus group discussions regarding the merits of girls' education, the role of girls in the home, and expectations towards girls as they relate to education. Many of the boys who participated in the research appear to largely hold progressive views on the roles of women and on girls' education, but there was evidence that some hold discriminatory views toward girls and women.

For example, in one focus group discussion, one boy expressed the belief that "girls must do household chores alone" and that "women are better in the household than men," but other boys in the same focus group discussion disagreed and argued that boys and girls "must work together as a team" and "work at home equally."[177] In another discussion on whether boys and girls are treated equally in the home, one boy argued that "boys and girls are different, although in this time they are treated equally,"[178] and another boy disagreed, explaining, "I don't think boys and girls are treated equally since every parent loves their girls to stay at home, then after that to get married. The boy will be most of the time outside and then comes back late at night and he expects the dinner to be ready for him. Also when he wakes up in the morning, the same thing - he expects his sister to make the breakfast ready for him. The question

[177] FGD, Boys, Puntland
[178] FGD, Boys, Puntland

is if they are treated the same way, then the boy should wake up early and prepare breakfast for his sister while she gets ready for the school."[179] A heated debated followed, in which the one boy argued that there is nothing wrong with girls doing more work in the household, particularly when boys have other jobs, and others questioned why girls are expected to cook and clean and are discouraged from leaving the house.

Those who believe there should be equitable distribution of chores were unsurprisingly more likely to cite examples of how they have helped their sisters with household chores to lighten the burden. Girls also cited examples of boys assisting them in their learning, both at home and at school: "Yes, my brother helps me at home with the subjects that he is better at than I am, for example at the time of the graduation exams. At school, there is a boys' club that encourages [raises awareness for] girls' education."[180] Fathers also described how boys make sacrifices to help their sisters continue their schooling: "Boys do support girls to stay in school and sometimes when it comes to Eid celebrations, the money that boys should pay and celebrate their Eid, they give to their sister so they can buy school dress or learning equipment."[181]

Girls did not specifically mention their fathers when asked about support for girls' education in the home (when they did specifically mention one parent, it was always the mother), but they did highlight positive changes that have occurred in support for girls' education from parents in general. For example, parents are following their girls' education progress more closely, going to their children's schools "to know the exam results of their children and ask teachers about their educational level."[182] According to some girls, parents are also making efforts to lift household pressures and encourage learning by hiring home teachers, "reducing household tasks and hiring a house maid to do chores,"[183] "sending them to school at an early time,"[184] and allowing them to "attend afternoon classes without saying, 'You attended morning classes, so do not return back.'"[185] In contrast, in the past, "there were no girls returning to take remedial classes if she attended morning class, because she was not allowed to return back to do afternoon revision classes."[186]

The girls explicitly link these changes to improvements in girls' attendance and even self-confidence. As one girl explains, "Girls' attendance is higher because their parents encourage them and they have understood the value of education."[187] When asked whether there have been changes to girls' attendance in the past year, a girl explains, "Yes, changes happened to girls' school attendance. You can see that girls

[179] FGD, Boys, Puntland
[180] FGD, Girls, Somaliland
[181] FGD, Fathers, Puntland
[182] FGD, Girls, Puntland
[183] FGD, Girls, Puntland
[184] FGD, Girls, Puntland
[185] FGD, Girls, Puntland
[186] FGD, Girls, Puntland
[187] FGD, Girls, Somaliland

reach a better level in education and come first in school compared to boys. And this is the result made by donors, school administration, and teachers. Similarly, they got motivation from parents."[188]

Discussion

Our analysis has encompassed five different types of respondents, and a number of different measures of male support for girls' education collected from each. In general, we did not find a substantive increase in male support for girls' education, in terms of either perceptions of general support or expectations of specific, tangible action on the part of boys or men. From the perspective of caregivers, there seems to be a small decline in the support men in their households and in their communities show toward girls' education. Results from teachers and head teachers did not contradict this finding, reporting a decline in general support for both boys' and girls' education, and a decline in the likelihood that a hypothetical father would attend a parent-teacher meeting about their daughter's performance in school. Finally, girls themselves report a small improvement in support from their classmates, but a small reduction in support from their brothers, in terms of concrete actions either group could take.

The qualitative data provides more nuance to these findings, suggesting that there have been some improvements made in male support for girls' education, but that the value of girls' education is still in many cases framed within the broader context of social norms around the role of women – for example, girls' education is often seen as valuable because women are the ones who are in close contact with children and therefore are seen as having a greater impact on the education of their children. Additionally, educated girls are perceived as easier to marry off. It is clear from the qualitative results that there is an active debate in Somali society around the value of girls' education, with many males espousing support for girls' education generally, but that males in particular struggle with the fact that longstanding social norms which affect men are being challenged. This is in some cases reflected by a lack of translation between male perceptions around the value of girls' education into concrete actions to actually support girls' education or remove barriers to girls' education.

Overall, our conclusion is that male support for girls' education has shown some improvement according to the qualitative data, but is still in a transitionary period, which makes it appear stagnant in the quantitative data. Despite some of the findings cited above, we do not believe male support has actually declined since the baseline evaluation; rather, we expect some of the negative results are driven by increased expectations of male support among teachers and caregivers.

Indicator Grade: 1 – Latent

[188] FGD, Girls, Galmudug

Sustainability Indicator 3 – Community Leaders' Advocacy

Percentage of EGEP target communities where community leaders are leading campaigns and advocacy events

In addition to broad-based male support for girls' education, discussed in the previous section, another aspect of sustainability targeted by EGEP-T programming is support for girl's education by community leaders. Community leaders, which we interpret to include religious leaders, clan elders, elected officials, and other prominent community members, have the ability to shape public opinion and, because they may be sought for advice in specific cases, can affect individual decisions regarding schooling, marriage, and other topics of relevance to female educational attainment. EGEP-T programming has included efforts aimed at community leaders, with the goal of creating a sustainable environment for gains in girls' education.

The primary quantitative data for this indicator is drawn from interviews with caregivers and teachers, respectively. Both types of respondents were asked about community events in which local leaders expressed support for girls' education.[189] These events are explicit aspects of EGEP-T programming -- outreach to community leaders was supposed to prompt public appearances and efforts to promote girls' education in this way. We asked respondents whether they recalled any such event occurring in their area over the last 12 months. This metric captures both the occurrence of the event (if no such events occurred, no respondents will recall them) and the broadness of the event's reach (if only a few community members recall the event, it suggests it occurred but included only a small portion of the community).[190]

The table below reports the share of respondents who recall an event in which community leaders advocated for girls' education publicly, broken down by respondent type (top and bottom panel) and zone. Among caregivers, the share exposed to leaders' advocacy fell marginally from baseline to midline, 51.4 to 48.4 percent, though the difference was not statistically significant. But the geographic disaggregation reported for this same group of respondents presents a sharp dichotomy: respondents in Somaliland were much less likely to be exposed at midline, when compared to baseline, while exposure in Galmudug rose dramatically over the same period.

TABLE 65: EXPOSURE TO EVENTS WHERE COMMUNITY LEADERS ADVOCATED FOR GIRLS' EDUCATION

Zone	Baseline	Midline
Caregivers Sample		
Somaliland	46.0%	34.1%

[189] Both caregivers and teachers were presented with the same question: "In the last year, have you heard about any events or efforts to encourage girls education in this community? Please indicate which of the following you have heard happened in the last year: Door-to-door visits about the importance of girls education; back-to-school campaigns encouraging the enrolment of girls in school; e vents where community leaders gave speeches encouraging girls education."

[190] Note that there is significant overlap between our analysis in this section and that of the first sustainability indicator, focused on awareness-raising activities.

Puntland	61.3%	55.6%
Galmudug	22.9%	73.7%
Overall	51.4%	48.4%
Teachers Sample		
Somaliland	69.6%	63.5%
Puntland	82.8%	71.7%
Galmudug	50.0%	62.2%
Overall	75.9%	67.6%

Interestingly, the results from caregivers are broadly, though less starkly, confirmed using the data collected from teachers. Overall exposure declined from baseline to midline, with the 8.3 point decline significant at all conventional thresholds. The zone-by-zone pattern is also confirmed: in Somaliland and Puntland, teachers became less likely to be exposed to or recall events where community leaders advocated for girls' education, while teachers in Galmudug became more likely to have been exposed, over the same time period.

While the focus of this indicator is on public advocacy by community leaders, we find it useful to expand on this discussion by analysing the attitudes of these same leaders. Unfortunately, no systematic sample of community leaders was selected for interviewing. Despite this, both baseline and midline include data *perceived* attitudes of community leaders, as reported by caregivers and teachers. Caregivers were asked the extent of support among community leaders for girls' education generally, for completing schooling before marriage, and for completing schooling even after getting married. Teachers, on the other hand, were asked to rate the strength of support for girls' and boys' education (separately) among mothers, fathers, religious leaders, and clan leaders. While caregiver and teacher perceptions must be taken with a grain of salt, because they are influenced by the individual viewpoints of, for instance, the teachers and their interactions with and perceptions of community leaders, the comparison from baseline to midline gives us the opportunity to see how support among community leaders may have changed. To the extent that caregivers or teachers report increased support for girls' education among community leaders over this time period, it likely reflects those leaders' efforts to encourage education -- precisely the goal of this indicator.

The figure below reports results for questions directed at caregivers (left panel) and teachers (right panel). Note that the questions asked of each group were different. Caregivers were asked the extent to which community leaders:

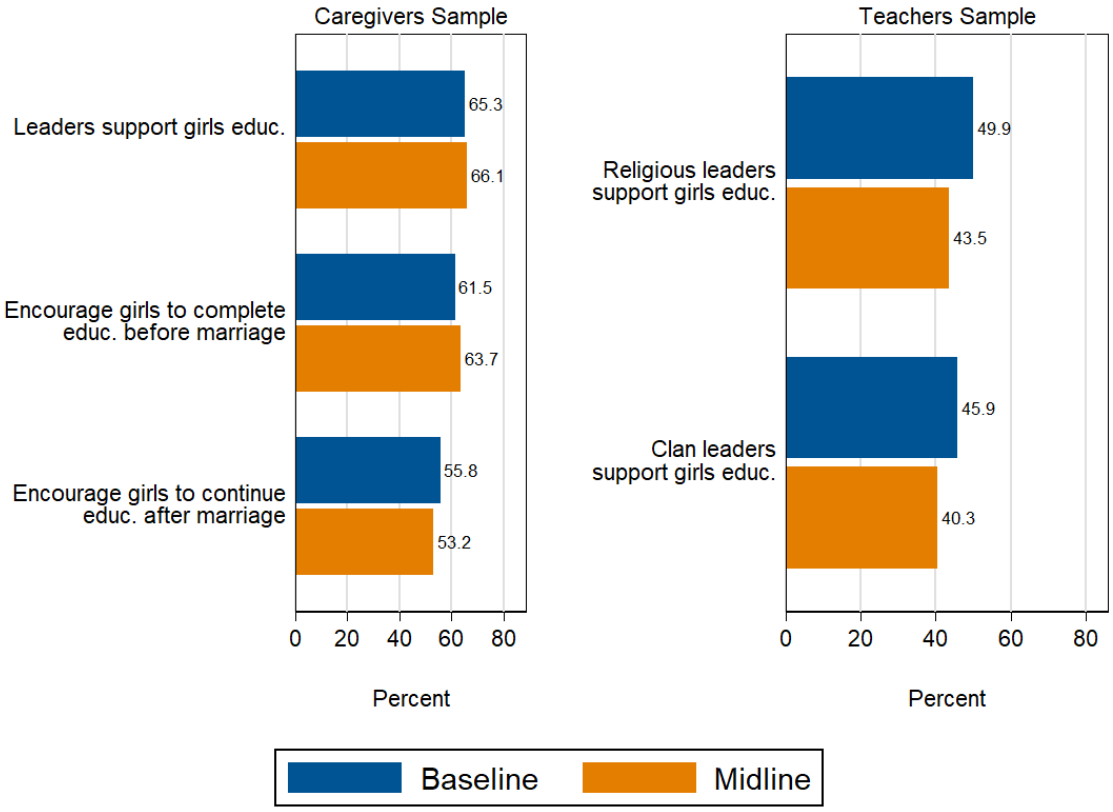
- Support girls' education
- Encourage girls to complete schooling before getting married
- Encourage girls to continue their schooling after getting married

All three are taken as pro-education for girls. Meanwhile, teachers were asked the strength of the following leaders' support for girls' education on a 5-point Likert scale:

- Religious leaders
- Clan leaders

Raw caregiver responses were binary (yes/no) answers; the left panel reports the share of caregivers who agree with each statement. Teacher responses were ordinal, which we convert to a binary measure capturing the share of teachers who believe religious or clan leaders are "very supportive" of girls' education. The right panel of the figure reports this percentage for both clan and religious leaders.

FIGURE 21: PERCEIVED COMMUNITY LEADER SUPPORT FOR GIRLS' EDUCATION, AMONG CAREGIVERS AND TEACHERS



Although our earlier results were focused on actions by community leaders, and this figure is based on *perceived* attitudes of those same leaders, the findings nonetheless confirm and provide additional support for the idea that community leaders have become less supportive of or less engaged with regard to girls' education since baseline. Among surveyed caregivers, there has been effectively no change: caregivers surveyed at midline have very similar views regarding community leaders and their support for girls' education as did caregivers surveyed at baseline. In comparison, teachers' opinions of community leaders have declined over the same period -- fewer teachers are willing to call religious leaders "very supportive" of girls' education, and the same finding applies to clan leaders. Both results are marginally significant ($p < 0.10$). Further, both findings reflect shifts across the distribution of possible responses; in

other words, we observed shifts from "somewhat unsupportive" to "very unsupportive", and not only changes in the share responding "very supportive." This indicates that our finding is not conditional on the coding approach employed, and reflects an actual change from baseline to midline.[191]

There is evidence in the qualitative interviews to suggest that community events have been held to promote the importance of girls' education in most communities, but the involvement of community leaders in particular was seldom mentioned. These events appear to be focused primarily on generally promoting the value of girls' education, encouraging girls' enrolment and attendance, discouraging drop-outs, and promoting EGEP-T project activities.

For example, in one community, an event was organized when the school opened "to increase girls' enrolment instead of staying in their houses,"[192] and in another community, "teachers, students, and some different parts of the community like the Ministry of Education" organized an event "about how to participate in this EGEP-T program – to send to school those who didn't study and bring back to the school those who have dropped out."[193] In another community, where meetings are arranged between school administration, religious leaders, and CEC members to discuss the meaning of education, "awareness has been done to the community by ADRA and local community leaders, so the students must come back for afternoon classes with separate classes for the boys and girls as well, and that makes strong competition among students which can cause education enhancement."[194] Several qualitative interviewees mentioned events that had taken place with the sponsorship of EGEP-T, but little information was offered regarding the extent of support for girls' education from traditional leaders or those outside of education-oriented government ministries.[195]

Discussion

The results in this section suggest a small negative trend in the context of community leaders. Either fewer community leaders are advocating publicly in favour of girls' education, or fewer community members are aware of their advocacy and support. While there are a variety of possible explanations for this decline, it is hard to reconcile both a decline in the leadership of events *and* a shift in the attitudes perceived by others. Since baseline, community members have not sensed an improvement in the support for girls' education provided by community leaders, and this perception has actually regressed among teachers. These findings indicate that additional work is needed to promote girls' education among community leaders, ensure their buy-in to the project, and encourage them to take an active role in efforts to increase girls' educational attainment.

---

[191] Though, we should note that sampling variation in the selection of teachers from baseline to midline within the same set of schools could also drive changes over time.

[192] FGD, CEC, Puntland

[193] FGD, CEC, Galmudug

[194] FGD, CEC, Somaliland

[195] This shortcoming stems largely from the fact that no community leaders were targeted for qualitative interviews. At the endline, additional qualitative interviews could be conducted with traditional and religious leaders, to ascertain their attitudes and avenues of support for girls' education directly.

Indicator Grade: 2 – Emerging

Sustainability Indicator 4 – Bursary Support Provided by CECs

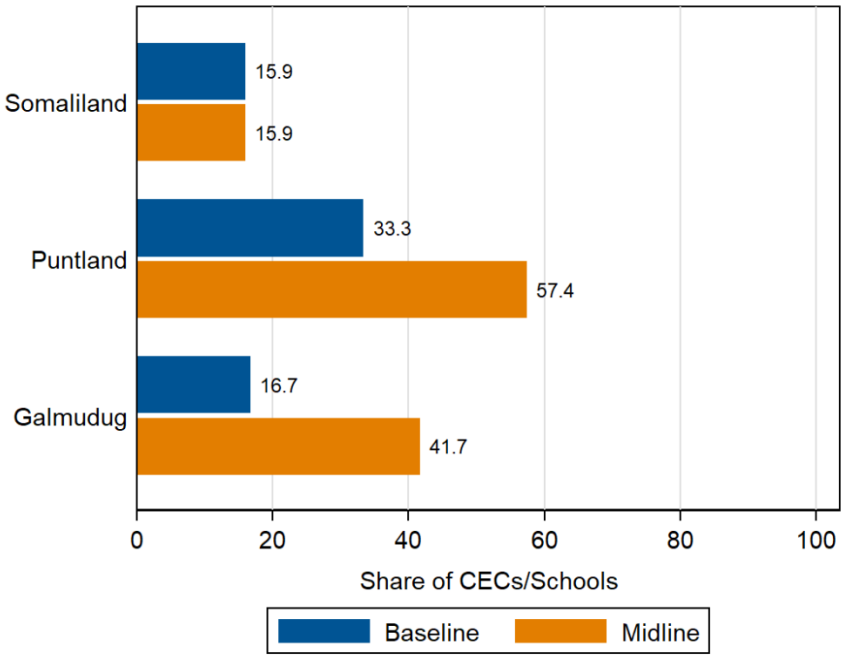
Percentage of CECs providing bursary support to marginalized girls in EGEP target schools

As we have described repeatedly in this report, one of the main interventions employed by EGEP-T is bursary support of marginalized girls. The set of girls tracked from one evaluation to another includes a group of such girls who receive bursary support through EGEP-T. Bursaries are intended to promote enrolment and transition among marginalized girls. As part of the project's efforts to ensure sustainability beyond its own life, CECs have been encouraged to begin providing bursary support to girls in their own communities, to supplement those provided directly through EGEP-T actions. Financial support provided by CECs is particularly valuable because it incorporates deep local knowledge and can be accurately targeted – CEC members are part of the community, and know best where the greatest need lies. CEC financial support to families may often be on a smaller scale than EGEP-T bursaries, but they can be adapted, increasing when a family faces severe need, and being reduced when a crisis has passed, for instance.

At baseline, we found that just under one-third of CECs (30.7 percent) provided bursary support to at least one girl. This finding was influenced heavily by schools in Banadir, where bursary support was more common; in the reduced sample employed at midline – which excludes Banadir and includes only schools that appeared in both rounds – just 24.6 percent of head teachers report that their CEC provides bursary support to at least one marginalized girl.[196] The share of CECs providing bursary support has clearly increased from baseline among sample schools: overall, 39.1 percent of CECs now provide support of this kind, up 14.5 points from baseline. As shown in the figure below, these gains are driven entirely by Puntland and – to a lesser extent, owing to the smaller number of schools surveyed there – Galmudug and Hirshabelle. No discernible change was observed in Somaliland from baseline to midline.

[196] Data was collected from head teachers because they are the individual most broadly familiar with the operation of the school and the CEC. While surveying all CEC members may be slightly more accurate, head teachers are traditionally members of the CEC, and have the added insight provided by their status as head administrator of the school. Note that the survey did not ask explicitly about "bursary support", because respondents may view small-scale or partial financial support as distinct from a bursary – bursaries may be considered more formal and competitive than the assistance provided by CECs in practice. The precise survey question asked head teachers whether the CEC supports any girls with their school fees.

FIGURE 22: SHARE OF CECS THAT PROVIDE BURSARY SUPPORT TO AT LEAST ONE FEMALE STUDENT, BY ROUND



While the number of CECs providing some form of financial support has increased, this does not necessarily mean that the number of girls supported has risen appreciably. A number of schools that provide financial support do so for just 1 to 5 girls in their entire school. Hinting at the possibility that financial support is limited to a few girls is the fact that relatively few caregivers reported that the CEC at their girl's school supports students financially. When asked, just 4.3 percent of caregivers said their CEC supports students financially, meaning that financial support may benefit a very narrow set of students, rendering most caregivers unaware of their CEC's actions.

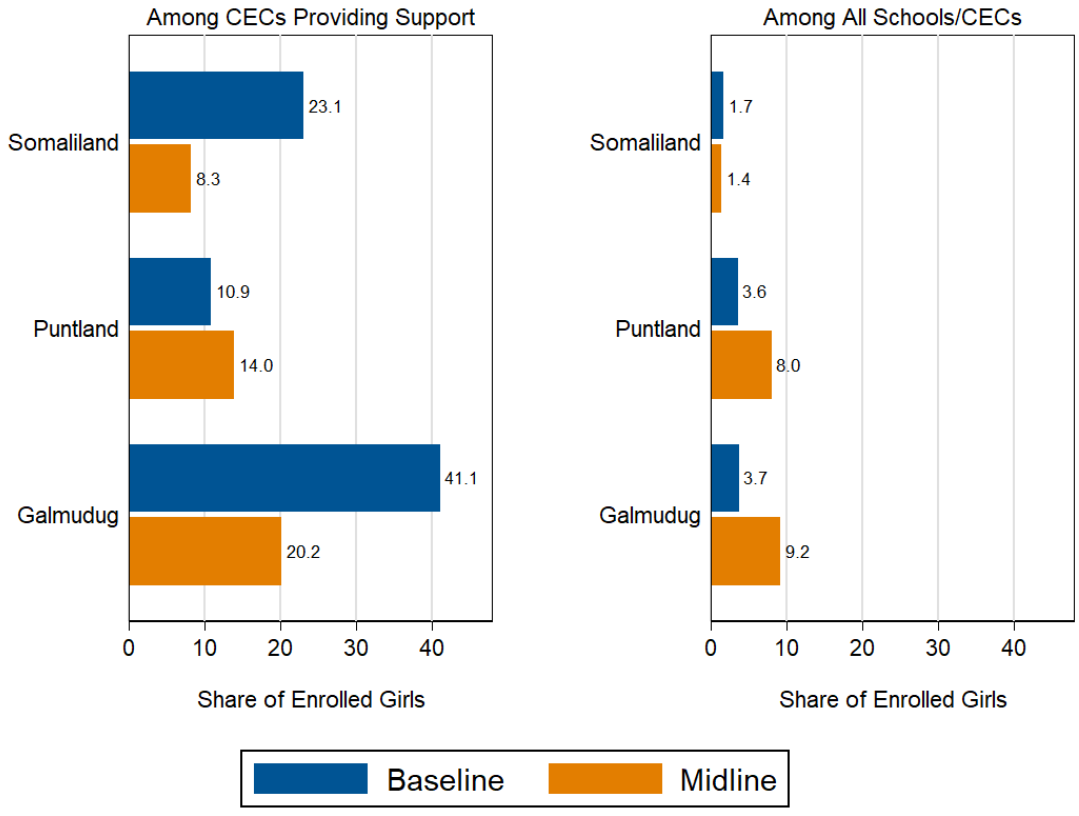
To further assess the extent of support, we asked head teachers how many girls in their school receive assistance for either part or all of their school fees from the CEC. As noted, a number of head teachers (10.9 percent of those citing any support) indicated support for just 1-5 girls. On the other extreme, three CECs provide support to 100 or more girls. On average, the number of girls supported increased from 6.7 to 9.3 from baseline to midline; relative to overall girls' enrolment, this corresponds to an increase from 2.9 to 5.6 percent of all enrolled girls.[197]

As the figure below shows, however, the increase in the share of girls supported is a function of more CECs providing support, rather than an increase in the breadth of support provided by the typical CEC. In the left panel, we report the share of girls supported, by zone, among that subset of CECs that provide support to at least one girl. Overall, this share was steady (13.9 percent at baseline, 13.8 percent at

[197] In other words, at baseline, 2.9 percent of girls in the typical school were being supported by their CEC; at midline, this share had increased to 5.6 percent.

midline), with large declines observed in Somaliland and Galmudug. In contrast, the right panel reports the share of enrolled girls supported by their CEC, among *all CECs*, rather than just those who provide support to at least one girl. In this sample of schools, the share of girls supported rose from 2.9 to 5.6 percent, with meaningful increases in Puntland and Galmudug. This finding should not be viewed as a criticism – more girls are receiving bursary support from their CECs, and that is the most important outcome. However, it is important to understand the source of those gains, so that programming can be adjusted accordingly.[198]

FIGURE 23: SHARE OF ENROLLED GIRLS SUPPORTED BY CEC-PROVIDED BURSARIES



For the sake of contextualizing these results, it is important to note two broad findings regarding project schools. First, for the sake of clarity, these results are focused on bursaries provided specifically by CECs.

[198] At this juncture, there is still room for improvement from promoting bursary support among CECs that do not currently provide such support *and* from encouraging increased activity from CECs already providing some bursary support. A particularly fruitful approach may be to target CECs that have *decreased* their bursary support from baseline, to assess why they were unable to maintain or increase their level of activity.

It does *not* include bursaries provided by EGEP-T or other programs; in practice, every EGEP-T school in our evaluation includes at least one girl receiving a bursary through the project. The broader availability of bursaries is also confirmed by head teachers, the vast majority of whom report that at least one girl in their school receives bursary support from some source. Our focus for this indicator is exclusively on bursaries provided by CECs.

Second, the financial burden of education – on both families and CECs that support them – varies by location and by school level, and has changed since baseline. School fees in Somalia are typically charged annually or monthly; to ease interpretation, we have converted all fees reported by head teachers into a monthly fee, quoted in US dollars. Overall, the share of schools that charge school fees has been stable from baseline to midline, rising slightly from 85.5 percent at baseline to 87.3 percent at midline. At the same time, the value of fees has decreased, with significant variation from zone to zone. Overall, the average monthly school fee has declined from $11.23 to $9.00 per month. But this mean decrease was observed only in Somaliland and Galmudug, where average fees declined from $8.67 to $2.60 and from $12.83 to $9.92, respectively. In Puntland, average fees increased by a little under two US dollars from baseline to midline.

The table below describes the situation at midline. Across all areas, school fees are higher for secondary than primary school, with the highest fees in Puntland, followed by Galmudug. In the baseline evaluation, we noted that most Somaliland primary schools charge fees – according to head teachers – despite the fact that school fees at primary level had been formally abolished by the Somaliland Ministry of Education. This situation remains unchanged at baseline, though the average fee in Somaliland primary schools is extremely low. The fact that, overall, school fees have declined since baseline is a positive outcome; combined with an increase in the share of girls receiving financial support from their school's CEC, the financial burden of education on families should be decreasing significantly.[199]

TABLE 66: SCHOOL FEE FREQUENCY AND AMOUNTS, BY SCHOOL LEVEL AND ZONE

Zone	Share of Schools Charging School Fees	Avg. Monthly Fee, Primary Level	Avg. Monthly Fee, Secondary Level
Somaliland	77.8%	$2.04	$8.57
Puntland	96.3%	$12.82	$19.73
Galmudug	83.3%	$10	$14.75
Overall	87.4%	$8.45	$15.65

In general, the factors that we would expect to predict changes in CEC financial support generated counterintuitive results. For instance, our comparable sample includes 22 schools deemed to have been drought-affected at the time of midline data collection. However, contrary to expectations, drought-affected schools exhibited a greater over-time increase in the share of CECs that provide financial support

[199] Of course, other educational costs still exist and present a financial strain on families. Many students pay for textbooks and other learning materials, uniform, examination fees, and other costs.

than those schools that were not affected by drought. Even trends in the share of girls supported were similar for drought-affected and unaffected schools. Findings were very similar for conflict-affected schools, though the sample of conflict-affected schools at midline was very small (n = 7 in this subsample).

With regard to school level, we found that CECs affiliated with secondary schools were the most likely to begin providing financial support at midline, vis-à-vis the baseline. At baseline, 22.8 percent of primary schools and 22.2 percent of secondary schools had CECs providing financial support; at midline, these shares had increased to 34.5 and 52.0 percent, respectively. The outsized positive trend in secondary schools is also reflected in the share of girls supported, increasing from just 2.8 percent of all enrolled girls at baseline to 8.0 percent at midline, and far outstripping the gains observed in primary schools.

In the qualitative data, there was little mention of CECs providing support for girls via bursaries – CECs most often mentioned using the money they raise to fund school rehabilitation projects. In most cases, CEC members cited impoverishment in their communities as the reason they are often unsuccessful in raising money. As one CEC member explains, "As CEC members we are struggling to pay our school fees. The country is financially poor, and this restricts our ability to make plans or to contribute money to the school, except motivating and raising awareness in the community."[200]

There were some few cases where CECs appear to have been successful in either raising funds amongst community members or creating "rainy day funds" to cover school fees. For example, in one community, CEC members explained that "advocacy has been done in the community to pay fees for girls, and 10 people who will pay the fees of 10 girls have been found."[201] In another community, a teacher explains that the CEC goes to business people in the community when there are issues at the school: "If there are huge problems, we share them with business people. We have developed a financial account. We intend this account for support of poor students or needy teachers who lost their salary."[202] However, it should be noted that this is not the norm, and that the qualitative data suggests that CECs are not currently focusing their efforts on bursary support.

Discussion

Overall, we find consistent evidence that CECs are providing additional support to girls in the form of bursaries compared to baseline. While we do not have the ability to triangulate these results extensively, the data available are clear: more CECs are engaged in providing financial support, and a significant greater share – nearly double – of enrolled girls are receiving such support now, compared to baseline. These is certainly room for additional improvement, as the vast majority of girls receive no support from their CECs, and many of those girls receiving support likely receive partial, rather than full, support. Further improvement at endline might reflect an increased share of girls receiving full support from the CEC, partial support from the CEC coupled with supplemental assistance from alternative sources, or a wider swath of girls supported in the form of bursaries. Nonetheless, it is clear that this indicator is moving in

[200] FGD, CEC, Puntland
[201] FGD, CEC, Puntland
[202] KII, Male Teacher, Somaliland

the right direction over time, with substantial gains in the share of girls supported overall and the share of CECs providing support to at least one girl.

Indicator Grade: 3 – Becoming established

Sustainability Indicator 5 – CEC Financial Support for Schools

Percentage of CECs in EGEP target schools providing match funding (financial or in-kind) for school improvement initiatives

In the previous section, we assessed the extent of support CECs provide to female students in terms of their school fees. In this section, we consider other forms of CEC material support for the schools they oversee and govern. We divide this support into two types: straightforward financial support from CEC members to school needs, and in-kind contributions of their time/labour, materials, etc., for school improvement initiatives, school monitoring, and so forth.

To assess the support provided by CECs, we asked head teachers whether their CECs provided either financial or in-kind support to the school. In the case of in-kind contributions, we also asked them to briefly describe the types of contributions they receive. The main results, disaggregated by zone, are reported in the table below. Overall, the share of CECs providing financial support has declined from 26.9 to 22.5 percent, with the largest declines observed in Puntland and Galmudug. In-kind support, on the other hand, has increased: 38.7 percent of CECs now provide in-kind support, compared to 33.6 percent at baseline. Again, there are meaningful differences across geographic space: head teachers in Puntland reported a fairly significant increase in in-kind support, and in-kind support increased from two-thirds of schools in Galmudug to *all* schools in the area. Meanwhile, fewer head teachers in Somaliland reported in-kind support at midline than at baseline.

TABLE 67: SHARE OF CECS PROVIDING FINANCIAL AND IN-KIND SUPPORT TO SCHOOLS, BY ROUND AND ZONE

Zone	Financial Support Baseline	Financial Support Midline	In-kind Support Baseline	In-kind Support Midline
Somaliland	27.7	26.7	51.1	44.4
Puntland	25	18.5	13.3	20.4
Galmudug	33.3	25	66.7	100
Overall	26.9	22.5	33.6	38.7

It appears that financial and in-kind contributions act as substitutes for one another, in a general sense. At midline, 27 CECs provided only in-kind contributions, and a further 9 provided only financial contributions; just 16 CECs (14.4 percent) provided both types of contributions, according to head teachers.

Whereas CECs operating in drought-affected areas were *more* likely to provide bursary support to girls, as noted in the previous section, we find a clear reversal of this trend when considering financial and in-

kind support. In schools affected by drought at midline, the share of CECs providing financial support to their schools declined from 40.9 percent at baseline to 18.2 percent at midline.[203] This is an interesting pattern, because it suggests – and the available data bear this out – that CECs in drought-affected areas may be opting to allocate their limited financial resources to support girls, and halting, likely temporarily, their financial support for infrastructural improvements, purchases of new equipment, and renovations. Our initial expectation was that drought and other community-level economic shocks would reduce CEC financial support across the board. This may be true, but it also appears that CECs are reacting to greater financial need within their communities in the face of shocks, shifting financial support toward households in need, and away from more general investments in their schools. While the precise value of this approach could be debated, it is a rational reaction in the face of financial limitations.

Beyond trends in contributions, we also asked head teachers to describe the types of in-kind support they received from CECs. Their responses revealed a number of common types of contributions: arguably the most commonly-cited material in-kind contribution (i.e. a physical resource, rather than time/labour) was in support of repairing or replacing school equipment, especially furniture such as desks and chairs. School improvements such as painting and improvements to the school's water or sanitation facilities were also common, with the purchase of learning materials representing the third most-common tangible in-kind contribution. Head teachers less often noted that CECs supported school feeding programs, raised awareness or engaged in pro-education advocacy; even less commonly, some CECs advised students or sponsored school events.

These results are broadly consistent with the reports of caregivers, who were asked to describe the role the CEC plays in their schools. A minority of caregivers indicated that their school had an active CEC, and the share of caregivers who reported different types of actions by their CEC was uniformly low – the most common activity cited was monitoring of student attendance, which 19.5 percent of caregivers asserted their CEC does. However, even caregivers underestimate the relative activity levels of their CEC, the nature of the activities of which they are aware is instructive: according to caregivers, CECs are most likely to monitor student and teacher attendance; their next most common tasks are improving school infrastructure and engaging in advocacy and awareness-raising, such as promoting enrolment. All of the tasks to this point can be supported at least partially through in-kind contributions of time/labour or materials.[204] Raising funds, supporting students financially, and purchasing learning materials – all activities that require cash contributions that cannot easily be substituted – were invariably less commonly cited by caregivers. Of course, it is natural for caregivers to be most aware of activities that are prominent visually (infrastructure improvements) or which may target them or their neighbours (advocacy and monitoring of student attendance), as opposed to activities which may affect just a few

[203] Note that the correlation between schools affected by drought at baseline and midline is low – of 36 schools affected by drought at baseline, just 11 were still affected at midline, according to internal observations provided by Relief International. Of 20 schools affected by drought at midline, just 11 were also affected at baseline. This is important to bear in mind, because it means that the schools affected by drought at midline are often *newly* affected, which would explain why financial support from CECs may have declined relative to the baseline.

[204] Infrastructural improvements are the most cash-intensive of those cited most commonly by caregivers but, even here, CEC members can contribute their time and labor to projects in lieu of cash.

girls (financial support) or occur behind the scenes (raising funds or purchasing learning materials). Nonetheless, the nature of CEC activities suggests that their efforts are often shaped by financial limitations, and that in-kind contributions are almost always easier to make, even if they may or may not target the most important needs of the school.

In the qualitative interviews, CECs were asked to describe their level of involvement with activities like fundraising and developing school plans. The results suggest that CECs are at varying stages in their development in that they are involved in these activities, and in school management, to varying degrees. The most commonly mentioned barrier to CEC involvement was lack of resources, although community attitudes sometimes also serve as a significant barrier for CEC members. CEC members seem to recognize the value of infrastructural improvements, as many report wanting to pursue projects that would make water more available at their schools.[205] At the same time, some of the same CECs cited resource constraints that prevented them from implementing their plans or halted projects midway to completion – one CEC member reported having extensive internal discussions about constructing a tank, but a lack of resources seems to have discouraged any action beyond planning.[206]

CECs most often mentioned using the funds they raise to pay for teachers' salaries or fund school rehabilitation projects, such as building new classrooms, buying new chairs or desks, and building school gates. However, some CECs mentioned that they are not currently fundraising or have not fundraised recently because community members are struggling financially after having lost their livestock to the drought and/or do not want to contribute because they do not understand the value of education. It is unclear exactly what determines whether a CEC is successful in fundraising, as some do seem able to successfully raise funds from fellow community members, but it does appear as though some benefit more than others from support from government, development organizations, diaspora, and small business owners in their areas.

In one community, the CEC members explained that although "people are vulnerable and have no livestock to sell and no money," "the number of people who want to take part in fundraising is more than the number who do not want to take part in fundraising."[207] However, this same CEC also described receiving funds from UNICEF to make repairs to the school and install a water system. In another community, the CEC members describe how most people are too poor to contribute as they are suffering from the drought, but they also describe having a close relationship with the government, which has been beneficial for their school in the past: "We went to the Ministry of Education once and we told them that the school needs chairs, and 30 chairs were given to us. And that was because of the relationship we have with the government and community."[208] In another community, when asked who has contributed money through fundraising efforts, a CEC member responded that the "funds are from business people and

[205] FGD, CEC, Somaliland; FGD, CEC, Puntland; FGD, Boys, Galmudug
[206] FGD, CEC, Puntland.
[207] FGD, CEC, Somaliland
[208] FGD, CEC, Puntland

diaspora in the district," and that they were in the amount of $1,000+, which was used for "toilets for girls, school library repairs, which are still ongoing, and subsidies for the teachers' salaries."[209]

However, in most communities, CECs members explained that the financial situation of their fellow community members makes it difficult to successfully fundraise: "The majority of people are unemployed and just trying to find their daily life…so do they pay for their school fee, water bill, or food?"[210] Even in communities that are not directly experiencing the effects of the drought, "everyone in the community feels pressure from rural relatives who the droughts affected seriously."[211] As a result, CECs tend not to fundraise on a continuous basis, and instead fundraise only when "there is need,"[212] such as when a "teacher gets sick or other challenges which need funds"[213] arise. Lack of success in fundraising appears to be a major demotivating factor in the continuous development of school plans – as one CEC member describes, "We make a lot of plans but few of them are implemented." This same community member described how plans were made with Relief International that did not materialize: "There was a time we did a lot of planning and then we discussed it with the Relief International and they said we can help you with the water tank. After that they said we will change with school repairs. Until now we've done none of it, so there was more involvement but less impact and it will be a good impact on both boys and girls if they implement what they promised."[214] It appears from this quote as though the project is not always being implemented as planned, or at least this is the perception among some CECs – again, this appears to be a demotivating factor, though this also highlights that CECs are in some cases using their expectations of donor action as a substitute for community-driven action. This is understandable in the context described above, wherein CECs face considerable challenges fundraising from poor community members.

In other cases, CEC members face challenging attitudes from community members. There are times when CEC members talk to community members, and they respond in an "aggressive way"[215] or "don't understand the importance of education, so when [the CEC] goes to them, they say [to the CEC] they won't pay."[216] CECs are also at times accused of practicing favouritism in their activities and interactions with students. For example, one CEC member describes the reaction of parents when they were involved in the distribution of sanitary kits: "Parents say to us, you are making favouritism because you gave some girls sanitary kits and you didn't give them to others."[217] Even in their monitoring activities, "some of the parents think that their children are being targeted or pressured in the school."[218]

[209] FGD, CEC, Puntland
[210] FGD, CEC, Puntland
[211] FGD, CEC, Puntland
[212] FGD, CEC, Puntland
[213] FGD, CEC, Galmudug
[214] FGD, CEC, Puntland
[215] FGD, CEC, Somaliland
[216] FGD, CEC, Galmudug
[217] FGD, CEC, Somaliland
[218] FGD, CEC, Puntland

As the qualitative interviews quoted above make clear, the context in which CECs operate is difficult and volatile. Even the best-laid plans of these committees can be disrupted by outbreaks of conflict or drought, which reduce their ability to fundraise among the community in terms of financial or even in-kind support. In some ways, CEC efforts that focus on infrastructural improvements to school – as opposed to bursary provision – may be a partial reaction to this volatility: infrastructure efforts occur at a discrete point in time and then pay dividends down the line, while other forms of support, including the provision of bursaries, require continual effort. In a volatile context, girls and their families may migrate, "wasting" resources that were invested in a bursary for the girl in previous years; it may also be impossible to maintain the level of previous years' fundraising, leading to girls who received bursary support in the past being abandoned by the CEC in the present. These are unlikely to be the only ways in which CECs adjust their efforts to their operating environment, and understanding the environment and their reactions to it more fully would help make sense of the barriers CECs face. We also note that a more holistic approach to measuring CEC financial engagement with schools would take account of the *intentions* and *efforts* of CECs, in addition to their results or successes. In this way, CEC effort could be partially disentangled from the context in which they work, though we would certainly still encourage direct measurement of the primary outcome – financial support for schools – as well.[219]

Discussion

Overall, there is little evidence of increased sustainability, in terms of CEC support for their schools, from baseline to midline. The share of CECs making in-kind contributions has increased marginally, but the share making financial contributions has declined. Even if this is offset by greater support for individuals girls and their families, it does not change the fact that CECs have not broadened the role they play in supporting their schools. It is an open question whether the *extent* of support provided by individual CECs has deepened – our data only allow us to judge how many CECs provide support, but it is possible that such support may have increased in magnitude among those CECs providing it. But, based on the available evidence, the progress on this indicator has been extremely limited.

Indicator Grade: 2 – Emerging

Sustainability Indicator 6 – Continuous Professional Development

Percentage of schools actively participating in the peer-mentoring programme

Teacher training is a central focus of EGEP-T programming. While at baseline this component of programming had not begun implementation, the project is now actively training teachers at midline. The structure of the continuous professional development (CPD) program is multi-layered: teaching coaches are recruited through Teacher Training Institutes (TTIs), organizations which include teacher training colleges in disparate regions; coaches perform face-to-face training in large group settings, visit schools to observe classrooms and provide feedback, and develop training materials, among other tasks. The CPD curriculum is wide-ranging, and includes core competencies around numeracy, gender-sensitive and child-

[219] For instance, head teachers could be asked to describe the efforts undertaken by the CEC to raise funds, and teachers could rate those efforts on a likert scale.

centred approaches, and remedial education; it can also be adapted by coaches to meet specific needs they identify in their teachers.

The midline evaluation incorporated a series of questions regarding CPD into surveys conducted with a sample of teachers at project schools. The indicator selected to measure sustainability surrounding CPD programming asks how many schools have at least one teacher actively participating in the program. This approach is preferable to simply analysing a list of CPD teachers provided by Relief International and its implementing partners, because it differentiates between the *selection* of CPD teachers for inclusion in the program and their *actual* inclusion and participation in the program. That is, we received a list of teachers who had been recruited into the CPD program by Relief International and its partners, but this does not mean that they are actively participating in the program. Therefore, we surveyed teachers – as part of a broader questionnaire directed at a stratified random sample of teachers – regarding their participation in CPD activities in the last year, and whether they had used the skills they learned in the classroom.[220] The sample of teachers was stratified in the sense that field teams selected teachers randomly from a list of both CPD and non-CPD teachers, as identified by the head teacher. CPD status is, therefore, the stratifying variable. Within strata, selection was random among teachers meeting our eligibility criteria, which was based on teaching English, mathematics, or Somali.

As noted above, CPD activities were not fully designed or being implemented at the time of the baseline; as a result, the baseline evaluation did not target teachers or head teachers with CPD-specific questions, because the program design was still in flux. Therefore, we do not make comparisons from baseline to midline for this indicator, as the baseline level of CPD participation at project schools was, by definition, zero. A starting level of zero activity is our baseline for measurement of CPD participation at midline.[221]

Again, our primary measure is whether a school has at least one teacher actively participating in CPD. In the overall sample of teachers, 50.5 percent of teachers reported active participation in CPD in the previous year. Of the 97 schools for which we have data, 82.5 percent (n = 80) had at least one teacher participating. The median school had 3 out of 5 surveyed teachers participating actively. In the figure below, we report the share of schools satisfying the activity indicator by zone. As shown, Puntland lags behind other zones in this respect, with just 77.4 percent of schools including at least one active teacher, though this gap between zones is not statistically significant.[222]

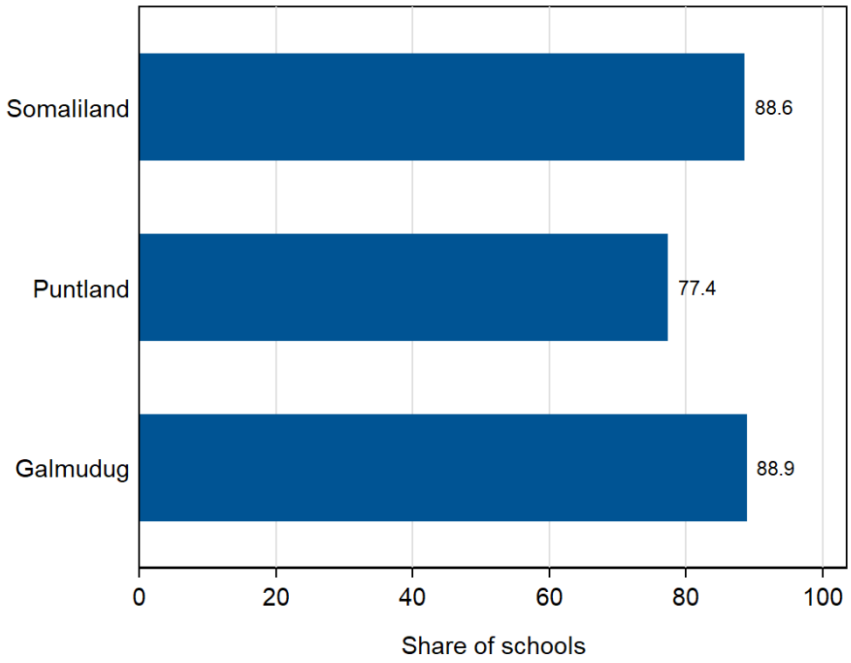
---

[220] Given the stratified nature of the sample, we cannot draw conclusions about the share of teachers who participate in CPD across all schools, because the stratified sample intentionally aims for approximately equal numbers of CPD and non-CPD teachers in the sample, even if their share in the population of all teachers is not split evenly. Moreover, our eligibility criteria, based on subject taught, may produce bias in terms of the share of CPD teachers as, for instance, more mathematics teachers may participate in CPD than other teachers. For this reason, we focus on whether schools have CPD teachers, and the characteristics of those teachers, rather than assessing the proportion of CPD teachers in the population.

[221] Finally, note that questions regarding CPD were added to the tools after fieldwork began, reducing the sample from 516 to 422 teachers, with no data from 16 schools.

[222] It is possible that differences in the number of teachers surveyed across different schools could influence the likelihood of finding at least one CPD-active teacher. However, this does not appear to explain the result for

FIGURE 24: SHARE OF SCHOOLS WITH AT LEAST ONE TEACHER ACTIVELY PARTICIPATING IN CPD PROGRAM



What is the typical CPD teacher like? For the purposes of highlighting aspects of the program, we analysed the characteristics of the CPD teachers surveyed. In total, the sample includes 213 self-declared CPD-active teachers. Consistent with the fact that teachers are overwhelmingly male, CPD teachers are also overwhelmingly male (92.5 percent of CPD-active teachers were men). While this is not surprising, we did find that female teachers were slightly less likely to be either selected for or active in CPD training. As shown in the top-left panel of the figure below, 51.7 percent of male teachers were CPD-active, compared to 39.5 percent of female teachers.

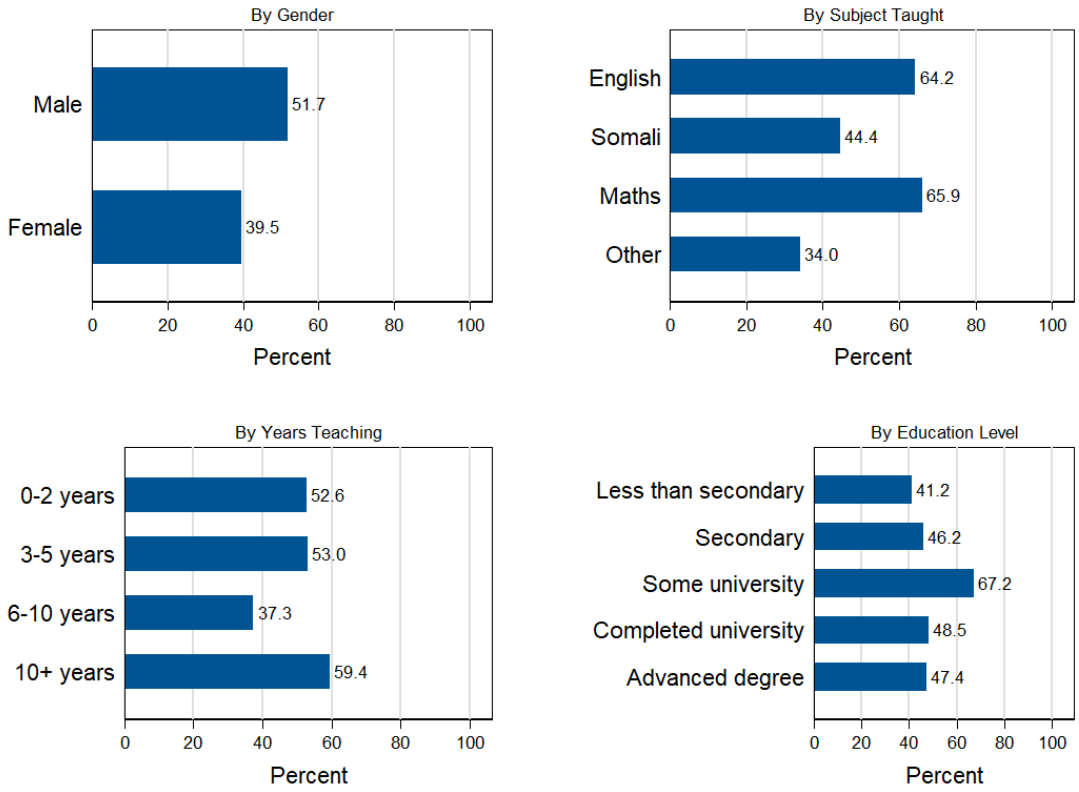
Relatedly, the data show that teachers of Somali are underrepresented vis-à-vis English and mathematics teachers. These two issues are clearly related: the project opted to prioritize the recruitment of English and mathematics teachers into the CPD program, and male teachers are much more likely to teach English or mathematics than their female counterparts. When we consider the likelihood of a teacher being recruited into the CPD program among *only* teachers of English and mathematics, female teachers are about equally as likely to be selected as their male counterparts. Tellingly, there are just six female teachers in English or mathematics in the entire sample, so it is difficult to draw conclusions regarding the

Puntland: schools in Puntland included a higher number of surveyed teachers, on average, than schools in either Somaliland or Galmudug. Therefore, there should be a *higher* probability of finding a CPD-active teacher in Puntland than the other zones, all else equal.

gendered nature of selection CPD opportunities.[223] To test for the relative importance of gender and subject taught, we estimated a linear regression predicting whether a teacher would self-report as active in the CPD program. We included a dichotomous ("dummy") variable for gender, and one for each possible teaching subject, as well as dummy variables for zone to control for differences across zones. The evidence clearly shows that subject taught, not gender, is the main predictor of whether a teacher is CPD-active, after controlling for other factors. The general results regarding subject taught should be interpreted somewhat cautiously, as many teachers cover multiple subjects. Unfortunately, the survey design did not account for this fact, and teachers were only given the option to specify multiple subjects if they responded "other" to the initial question. In many cases, the Somali, English, or mathematics teachers documented in the top-right panel of the graph may also have taught other subjects, but chose their primary subject as their response.

[223] Perhaps more insightful is the fact that female teachers are much more likely to teach Somali, teachers of which were not targeted by the project for recruitment into CPD. In total, 40.0 percent of female teachers teach Somali courses, compared to just 18.6 percent of men.

FIGURE 25: CHARACTERISTICS AS PREDICTORS OF A TEACHER'S PARTICIPATION IN CPD PROGRAM



By comparison, the lower two panels of the figure do not reveal any truly systematic patterns. The most experienced (10 or more years of experience) teachers are most likely to self-report as CPD-active, but the difference between these teachers and their less-experienced counterparts is not statistically significant, nor is the relationship between experience level and likelihood of being CPD-active monotonic.[224] Likewise, teachers who have attended but not completed university, or a technical skills course, are most likely to self-report as CPD-active, but both highly-educated and less-educated teachers are often included in the program.

According to the teachers surveyed, skills gained through the CPD program are extremely likely to be used. We asked CPD-active teachers whether they had used the skills they gained, and 93.4 percent indicated that they had, with some inconsequential variation across zones. We also asked CPD-active teachers

[224] If the relationship were monotonic, i.e. the likelihood of being CPD-active increased in step with each additional year of experience, we would put more stock in a possible relationship, even if no individual category of experience level was more likely than others to be CPD-active. However, the fact that the most experienced teachers have the highest probability and the second-most experienced have the *lowest* probability suggests the result stems from random noise instead.

whether they were satisfied with the training and coaching support available to them over the last year. It is important to be aware that this question did not explicitly reference the CPD program; therefore, respondents may be judging the quality and availability of *all* training sources, not CPD through EGEP-T exclusively. Nonetheless, because the question was asked within a module focused on CPD activities, we expect most teachers to have CPD on their mind when answering the question.

While the vast majority of CPD-active teachers indicated they used the skills gained, satisfaction with the availability of training was more varied. Overall, 72.5 percent of CPD-active teachers were very satisfied with the training available to them – again, noting that this includes *all* training opportunities – and a further 14.2 percent were somewhat satisfied. Satisfaction was highest, by far, among CPD-active teachers in Puntland, where 79.7 percent said they were very satisfied with the opportunities available to them.

Discussion

The available data shows that the CPD program has become fairly active since it began following the baseline, teachers find the skills useful in their day-to-day classroom practices, and teachers are generally satisfied with the training opportunities available to them, which include the CPD program. Thus far, 82.5 percent of schools had at least one teacher who self-reported as an active CPD participant within the last year. Given that the intervention began following the baseline evaluation, the ramping up of activities seems to have been rapid and consistent across most schools.

By any standard, the findings in this section are encouraging. At the same time, we would recommend a more detailed measurement strategy at endline, to ensure we are able to measure *relative* participation levels across schools and distinguish one-off activities from consistent training programs. We would also recommend an approach that surveys a broader set of teachers – or relies on records and input from the head teacher if this is not possible – to capture the total number of CPD-active teachers in each school, as a share of all teachers. This will ensure that large schools, with many teachers, are not blindly counted as being active in the program if just one of their teachers has occasionally received coaching. More nuanced measures are likely to reveal significant variation in the extent and quality of coaching, but our findings at the time of this midline are encouraging. At endline, we would expect the remaining schools to commence documented CPD activities and for the extent of these activities to increase in terms of their breadth and the depth of engagement among CPD teachers. Finally, the measurement strategy going forward could be deepened to more actively investigate satisfaction with CPD activities, their utility to teachers, and how they have specifically changed teaching practices in the classroom.

Indicator Grade: 2 – Emerging

Sustainability Indicator 7 – Child Protection Mechanisms

Percentage of schools with functional child protection mechanism

The next system-level indicator aims to assess child protection mechanisms being applied within schools and connecting schools to appropriate government officials. Child protection, in this context, is focused on adequate and safe reporting processes and the formation of a formal and reliable procedure for

keeping track of cases and referring them to the relevant authorities as needed. Part of the project's child protection approach includes the recruitment and training of a focal point for child protection issues. This individual should be trained to log cases brought to them, should serve as the primary point of contact for students reporting abuse of any kind, and should be trained to refer cases upward as necessary, or otherwise handle cases appropriately. Focal points and school staff should also raise awareness regarding the issue of abuse and how to report it.

We collected rich data concerning child protection at midline, including a number of indicators of how actively the school is handling child protection issues. From among this varied data, we selected two measures which we feel indicate that a school has established at least a rudimentary child protection system:

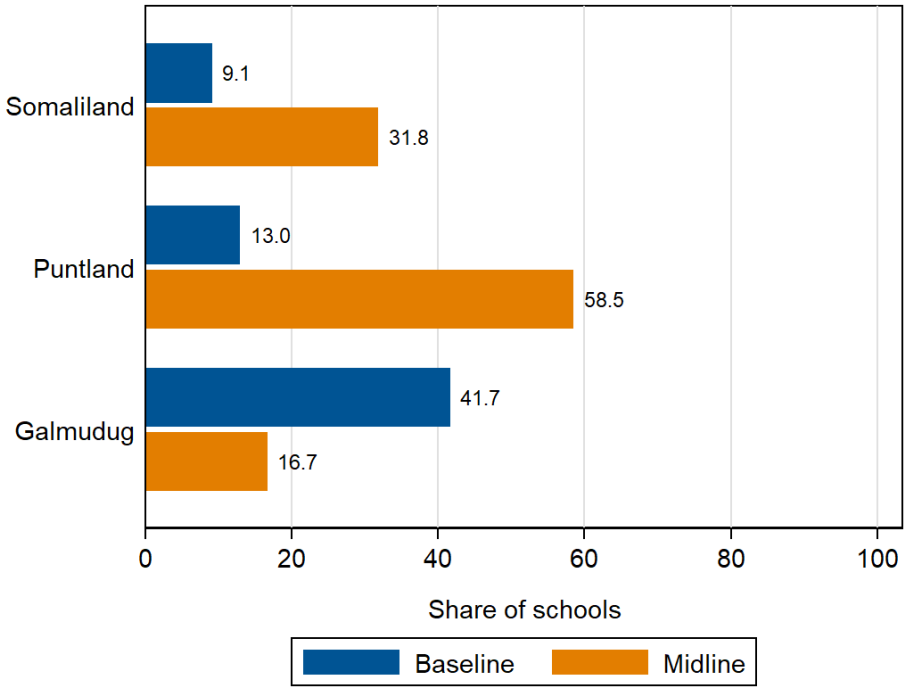
- Does the school have a focal point for child protection cases?
- Does the school have an established procedure for following up on child protection cases?

This is a fairly minimalist approach to defining an established mechanism. Indeed, we could also ask whether the focal point has received training on logging cases, whether teachers have all signed a Code of Conduct that includes child protection provisions, and so forth. However, these questions are less focused on the establishment and use of a formal mechanism, and centre instead on ancillary efforts – important issues nonetheless, but not specific to a mechanism for reporting cases. We utilize both measures listed above because we feel a single measure looking exclusively at the establishment of a mechanism neglects the portion of the indicator specifying the mechanism "is being actively used." By designating a child protection focal point, we feel it is much more likely any established mechanism will be used in practice.

Our data is drawn from head teachers who were surveyed at each school. For both survey questions, we collected the same data at baseline and midline, allowing over-time comparisons. We define a school as having an established child protection mechanism if the head teacher affirms they have a focal point *and* they have an established mechanism.

The results, disaggregated by zone and round are reported in the figure below, which shows the share of schools in each zone that simultaneously meet both requirements at baseline (blue) and midline (orange). Overall, the share of schools meeting this standard rose from 14.6 percent at baseline to 43.1 percent at midline, a difference that is statistically significant at any conventional level. As the disaggregated results show, schools in both Somaliland and, especially, Puntland were much more likely to have an established mechanism at midline than in the previous round.

FIGURE 26: SHARE OF SCHOOLS WITH ESTABLISHED CHILD PROTECTION MECHANISM, BY ZONE AND ROUND



Because the outcome measured above is a composite indicator, it is not immediately obvious from where the major gains in Somaliland and Puntland came from, relative to the baseline. In other words, did schools in these two zones tend to establish child protection mechanisms between baseline and midline, select a focal point, or both? When we split the indicator into its two constituent parts, it becomes clear that schools simultaneously improved on both aspects – in Somaliland and Puntland, the share of schools with an established child protection mechanism rose from 24.5 to 65.0 percent, and the share of schools with a child protection focal point rose from 21.4 to 61.9 percent. Conversely, schools in Galmudug experienced a decline in *both* constituent indicators, with fewer head teachers reporting the presence of a focal point and many fewer reporting the establishment of a formal mechanism. In short, focal points and established mechanisms are strongly correlated, and the trends illustrated in the figure above are not driven exclusively by one or the other.

As noted, the data allow us to investigate many other aspects of child protection at project schools. For instance, we asked head teachers whether the school maintains a log of child protection cases, whether the school has an official child protection policy, whether the head teacher could show our data collection teams a copy of the child protection policy, and whether posters or other information about reporting abuse are publicly posted where students can see them, among other questions. Baseline-to-midline trends for each of these indicators, all of which were measured through questionnaires directed at head teachers, are reported in the table below.

As the results show, the typical school improved in every aspect of child protection measured at both baseline and midline. In some cases, these improvements were small – and not statistically significant – as in the case of the head teacher being able to show our team leader a copy of an official child protection policy. In other cases, the improvements are dramatic: for instance, the share of schools maintaining a log of child protection cases rose from 14.7 percent at baseline to 43.1 percent at midline. In general, the results suggest that schools lag most with respect to writing down and disseminating information: while many head teachers claim to have an official child protection policy, many fewer teachers could show our team leaders a copy of such a policy; somewhat similarly, only around one-quarter of schools have information posted with instructions on how to report abuse, which may reduce students' willingness to report issues.

TABLE 68: CHILD PROTECTION INDICATORS, BY ROUND

Indicator	Baseline	Midline
School maintains record of CP cases	14.7	43.1
Has established procedure for following up cases	31.5	59.6
Has focal point for CP cases	23.9	56.9
Focal point trained on receiving and logging CP cases	N/A	48.6
School has official CP policy	36.1	57.8
Head teacher can show a copy of official CP policy	27.1	31.0
School has posters with instructions for reporting abuse	10.1	26.6
Head teacher can show posters with instructions for reporting abuse	10.0	20.4
Posters for reporting abuse are posted in public place	10.9	24.3
School has official code of conduct	61.5	85.3
Code of Conduct has specific section dealing with CP issues	38.6	61.5
In last 12 months, has school taken any action to improve CP?	N/A	67.9

Our analysis also revealed improvements in the share of teachers required to sign a Code of Conduct when being hired, though there are still substantial gaps in this policy. At baseline, 53.2 percent of head teachers reported that all teachers were required to sign a Code of Conduct upon being hired, a share which rose to 69.7 percent at midline.[225] A similar share (64.2 percent) of head teachers report that new staff – such as watchmen, administrators, or finance staff – are required to sign a Code of Conduct. However, given that only 61.5 percent of schools at midline have a Code of Conduct that includes specific provisions regarding child protection, this means that just 26.8 percent of schools require teachers to sign a Code of Conduct that addresses child protection, and 25.0 percent of schools require the same for new non-teaching staff hires.[226] We found a similar trend in terms of induction training specific to child protection:

[225] These findings include all schools; those that do not have a Code of Conduct are coded as not requiring teachers to sign.

[226] To calculate the share for teachers, we determined the share of midline schools that have a Code of Conduct, whose Code of Conduct includes provisions specific to child protection (verified by a review from our team leaders), and who require all teachers or staff to sign the Code of Conduct upon being hired.

at baseline, 34.6 percent of schools provided such training to new teachers, while 42.2 percent of schools did so at midline. However, this still leaves a majority of schools that do not provide induction training on child protection to new teachers.

To triangulate findings across different respondent types, the midline evaluation also asked regular teachers, male and female mentors, students, and their caregivers about child protection at their schools. In general, our findings from these additional surveys confirmed what we learned from head teachers, but these results also provided additional insight regarding the current status of and gaps in existing child protection procedures.

With regard to the presence of a focal point, findings from teachers confirmed those from head teachers. A similar share of teachers (57.6 percent) reported that their school has a child protection focal point compared to the 56.9 percent of head teachers who reported the same. Importantly, the vast majority of teachers whose schools have a focal point believe that students know who the focal point and that they know how to report abuse. Of teachers reporting their school has a focal point, 87.9 percent believe students are familiar with the focal point and their role. Likewise, 67.2 percent of students surveyed indicated there was an adult at their school to whom they could report abuse at school or in their community, if they were presented with the need. Again, this broadly confirms the findings drawn from interviews with head teachers and teachers, but it simultaneously suggests a large minority of students who may be silent in the face of knowledge about abuse.[227] Adult caregivers of cohort girls were roughly as likely to know a school staff member to whom they could report or address abuse allegations, and were also most likely to report such issues to the head teacher.[228]

In the broader case of codes of conduct, head teachers and teachers broadly agree, but there is still room for improvement in terms of both the content of these documents and the extent to which teachers receive training on them and are required to sign them. As noted above, 69.7 percent of teachers claimed that new teachers were required to sign a Code of Conduct when hired, though this Code of Conduct did not always include child protection clauses. When we surveyed teachers regarding their own experiences at the time of hiring or in the time since, we found that 71.1 percent of teachers had signed a Code of Conduct at some time, with a slight majority having signed the document at the time of hiring.[229]

---

[227] It is also worth noting that the modal adult chosen by students to report abuse in the hypothetical scenario we presented to respondents was the head teacher – 55.6 percent of students would report abuse to the head teacher, while most others would report to a regular teacher. It is important that students have an avenue, any avenue, to report abuse, but it may be preferable for students to report abuse to the designated focal point.

[228] Among caregivers, 52.4 percent indicated they would file a report if they were aware of abuse occurring at the school *and* knew the person to whom they would report it. Approximately two-thirds (66.9 percent) of those caregivers would report the abuse to the head teacher.

[229] We focus on the share of teachers who have signed a Code of Conduct *either* at the time of hiring or since then, because the experiences of teachers hired many years prior are not indicative of current hiring practices. We do not wish to penalize schools for having experienced teachers, but hope that experienced teachers have signed a Code of Conduct since one was established, if they did not sign at the time of their hiring.

The midline evaluation also expanded on the data collected at baseline by including a brief survey of male and female mentors/focal points that have been selected and trained at project schools. While the goal of the mentor program is broader than child protection – mentors lead girls and boys clubs, teach life skills courses, and engage with students on a wide range of gender-specific issues – child protection is also one of their core responsibilities. At midline, 171 mentors were surveyed, with 101 of 119 sampled schools including an interview with at least one mentor (in most schools, one male and one female mentor were surveyed).

Mentors were asked about their role in child protection and their preparation for that role. To the extent that these individuals are a critical component in the implementation of the child protection mechanisms sought by the project, their training and preparation is crucial. As discussed at the outset of this section, if the school lacks a child protection focal point, we do not think the child protection mechanism can meet the standard of being "actively used." Similarly, if the focal point has not received adequate training, it *suggests* the established mechanisms may be used less actively than we would like.

The table below reports results drawn from mentors, disaggregated by zone. Overall, 74.3 percent of mentors surveyed have received training specifically on child protection and child rights, while fewer – 60.2 percent and 56.1 percent, respectively – have been trained on receiving and handling child protection cases brought by children and escalating those cases upward, where appropriate. In general, mentors surveyed in Somaliland tended to lag those in Puntland and Galmudug in terms of their training: they are less likely to report having received any training regarding their role, and less likely to have received specific training on any of the three child protection topics surveyed.

TABLE 69: CHILD PROTECTION TRAINING AND OUTCOMES, ACCORDING TO MENTORS/FOCAL POINTS

Outcome	Somaliland	Puntland	Galmudug	Total
Trained on role as focal point	62.5%	81.2%	92.9%	74.3%
Trained on child protection & rights	68.1%	77.7%	85.7%	74.3%
Trained on handling CP cases	47.2%	70.6%	64.3%	60.2%
Trained on escalating CP cases to appropriate authorities	43.1%	67.1%	57.1%	56.1%
Number of CP cases heard this year	1.9	2.1	2.1	2.0
Keep a log of CP cases	29.2%	61.2%	21.4%	44.4%

The findings from mentors are actually somewhat more optimistic than the responses provided by head teachers from the same schools. For instance, 48.6 percent of head teachers reported that their child protection focal points had been trained on how to receive and log child protection cases, while 60.2 percent of mentors report the same training.

Unfortunately, many of the measures we employ to assess child protection policies and the presence of child protection focal points in schools were not included in the baseline evaluation. Among those that were, we find significant improvement from round to round, especially with regard to the promulgation of child protection policies, the recruitment and training of child protection focal points, and the

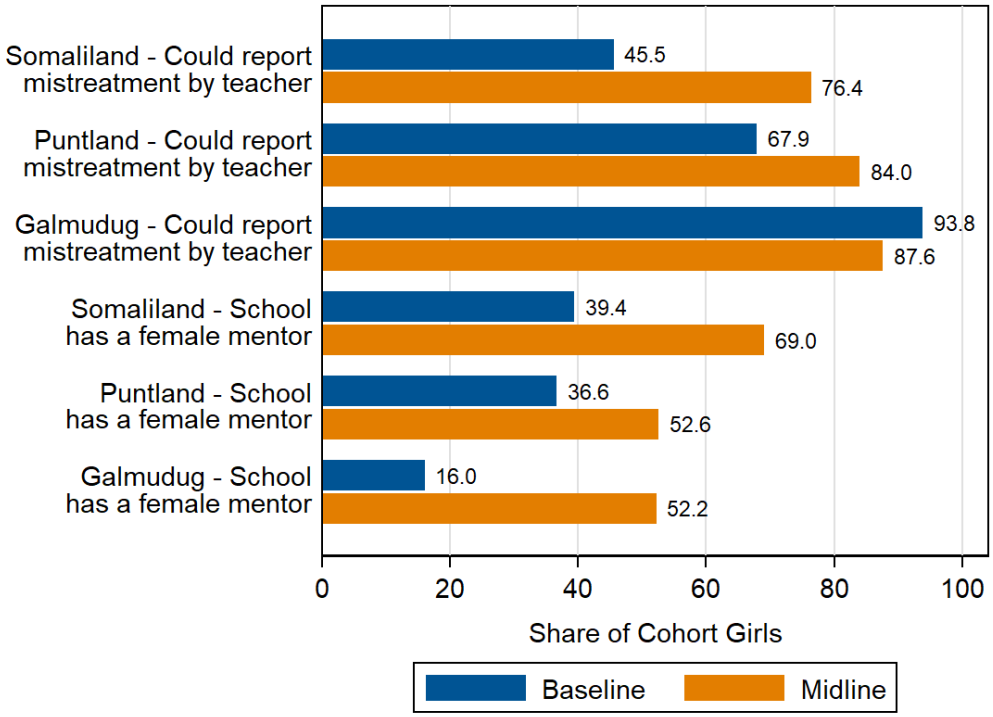
incorporation of child protection issues into schools' codes of conduct. However, in the case of cohort girls, questions explicitly focused on child protection were primarily asked only at midline. While we found that most cohort girls (67.2 percent) had an adult to whom they could report abuse, the lack of a baseline prevents us from tracking progress over time.

Usefully, the baseline evaluation did include questions whose content is adjacent to child protection. For instance, cohort girls were asked to imagine two scenarios in which a friend was being mistreated by a teacher and in which a friend was being bullied by a boy in their class. We asked them whether there was an adult at school to whom they could report such issues. We also asked cohort girls whether there school had a female mentor.

The results are reported in the figure below, disaggregated by zone and round. Overall, we find a substantial increase in the share of girls who feel they could report mistreatment by a teacher at midline, from 61.6 to 81.2 percent. A similarly large increase – not reported in the figure below – was observed in the case of students whose friend was being bullied by a boy in their class. We also observe very large increases in the number of girls who report their school has a female mentor, though a significant minority of students (40.5 percent of girls surveyed) still do not have access to, or are unfamiliar with, a female mentor at their school.[230]

[230] For the purposes of this analysis, we limited the sample of cohort girls to those who constitute a "true panel", i.e. those who were surveyed at both baseline and midline. While this is a more restrictive sample, because it does not include replacements, it provides the most rigorous possible comparison from baseline to midline, because it entirely eliminates sampling variation between baseline and midline.

FIGURE 27: SHARE OF COHORT GIRLS WHO FEEL THEY COULD REPORT MISTREATMENT BY A TEACHER AND ARE AWARE OF THE FEMALE MENTOR AT THEIR SCHOOLS, BY ZONE AND ROUND



Our approach to measuring child protection has been wide-ranging, and it requires some care to aggregate these disparate results. In terms of the primary indicator regarding child protection mechanisms, the evidence is clear: more schools have actively established and are using a formal child protection mechanism now than they were at baseline. Other indicators have also shown meaningful improvements: the share of schools where teachers must sign a Code of Conduct, and the share of schools in which the Code of Conduct explicitly references child protection issues, to take just two examples. Perhaps most importantly, girls increasingly feel willing and able to report issues they encounter at school to adults within the school, suggesting a greater level of comfort with teachers and greater access to sympathetic teachers (especially female mentors) than existed at baseline. This finding, combined with the other improvements outlined, suggest child protection has become significant better at project schools in the last 1-2 years.

Child protection strategies differ on a school-by-school basis. There was evidence in the qualitative interviews that schools have developed codes of conduct, and in some cases that teachers and students have received training on child protection policies. For example, one CEC member reported that "the teachers received training about child protection like using bad words.[231] Also, there is awareness at the

[231] Results from the qualitative data suggest that teachers have received training that encourages them to foster positive relationships with their students by using positive, motivating language that challenges the traditional, more

student level – they have form masters who raise awareness in the classes."[232] A different CEC member in the same region reports that "parents were called in a meeting about how to participate in child's security, and also children were taught if conflict happens between them who they can go to."[233]

However, the level of awareness of the proper reporting mechanisms among different groups differs from school to school, as do the exact reporting mechanisms and the level of success these child protection policies have met. Some respondents explained that teachers report issues to parents, and if the parents cannot solve the issue, it is then reported to the education committee or a school administration committee. In other areas, "there are clan leaders, and they solve the issues."[234] In one school, the head teacher described a recent incident in which two girls who were fighting were taken directly to the police station so that news of their fighting would not reach their families: "For example, if two students are fighting and cannot be reconciled, we take them to the police station, which is close. They stay there for a day or a half day because they are students and can't be there for too long. Recently, two girls from Form 4 in secondary school got into a fight and we sent them to the police station because we were a bit scared that the issue would reach their families. They stayed there for eight hours. We released them afterwards and they resolved their controversy."[235] This quote is by no means indicative of the child protection strategy taken in other schools, and it is unclear whether this extreme approach was used in just this particular instance because the case itself was extreme in nature or whether this type of approach is commonly used in this one school. However, it was a surprising admission from the head teacher that is worth highlighting to showcase the potential need for follow-up training with head teachers in particular around child protection escalation.

Notably, there were also a few instances in which respondents explained that there is no proper enforcement of laws or punishment for perpetrators. One father explains, "There is no government authority or Islamic law working here, so it will shows you how human right is very low in this area. So there is no place you can report any case. Only the community resolves if there's a case."[236] There appears to generally be low trust for government institutions, even when they are available: "First you should go to the police station and then they give you an evidence letter which will prove how much damage he/she has. After that, we go to the elders and they are the ones who solve these problems. We don't usually go to the court."[237]

There were also cases of rape mentioned by respondents in Galmudug. One female teacher reported that there has been a recent problem in girls' attendance because they are "in Galkacayo and a lot of rape cases have happened." When asked whether anything has been done to ensure girls' security, she

strict/adversarial teacher-student relationships of the past. Teachers generally describe this phenomenon by explaining they have 'made friends' with the students.

[232] FGD, CEC, Puntland

[233] FGD, CEC, Puntland

[234] FGD, CEC, Somaliland

[235] KII, Head Teacher, Somaliland

[236] FGD, Fathers, Puntland

[237] FGD, Fathers, Puntland

responded that she does not think that exists and that there has been "no reaction since that case occurred – even about previous cases," but that if a similar case were to occur it "should be submitted to the Ministry of Education through the police."[238] Fathers in Galmudug confirmed this finding, reporting that, "This year in the school, girls were feeling scared from the boys to rape them or kidnap them. But after community co-operation this has been solved." However, they later explain that the man who was responsible for the rape was only fined, which they did not find to be a satisfactory punishment.

Discussion

Overall, progress with respect to child protection mechanisms has been significant. When we define the indicator as the share of schools with both a focal point and an established procedure for child protection cases, the share of schools meeting these criteria has increased from 14.6 percent to 43.1 percent over the last 18 months. But gains are not limited to this fairly narrow definition: the set of schools with an official child protection policy, posters providing instructions for reporting abuse, and an official code of conduct – among other outcomes – has also improved markedly over time. Going forward, these gains need to be consolidated and expanded to the remaining schools, though further linear gains, matching those observed from baseline to midline, would be an excellent result.

Indicator Score: 2 - Emerging

Sustainability Indicator 8 – Gender Development Strategy

Federal level and Galmudug Gender Units develop gender strategy and Federal level strategy is being implemented within the life of the project.

Insufficient data was collected on this indicator at midline to permit assessment. Specifically, no qualitative interviews were conducted with federal-level gender specialists in the Ministry of Education, due to the cancellation of fieldwork in Banadir. In addition, no interview was conducted with gender specialists in Galmudug because the qualitative sample was adjusted to reflect the relatively small proportion of schools in Galmudug relative to Somaliland and Puntland. Ultimately, no interviews with ministry officials took place in Galmudug, while one gender specialist was interviewed in each of Somaliland and Puntland. **Indicator Grade: N/A – not measured**

Sustainability Indicator 9 – Monitoring Visits by Ministry of Education

Percentage of EGEP-T schools receiving follow-up monitoring visits from MoE officials, including Gender Focal Points, District/Regional Education Officers/RES

While most EGEP-T interventions, and the bulk of this evaluation, focus on girls, their families, and the schools in which they learn, EGEP-T also has significant goals with regard to the educational system more broadly. EGEP-T seeks to increase the capacity of the respective ministries of education with which it works, to improve their ability to monitor schools under their remit. While CEC members provide local

[238] KII, Female Teacher, Galmudug

governance of schools and are the frontline actors in terms of monitoring schools for quality, MOE officials can and should provide monitoring as well, especially because they have pedagogical and administrative expertise that is typically lacking on any given CEC.

To promote sustainability, EGEP-T hopes to increase the frequency and quality of MOE monitoring visits to project schools. If successful, this will promote continued accountability of schools, head teachers, and teachers, and will encourage ownership of the schools and their performance by the government.

To evaluate progress on this goal, the evaluation surveyed head teachers regarding their experiences with recent MOE monitoring. We asked head teachers whether their school had been visited by an MOE official for the purposes of monitoring the school within the past year, and how many such visits had occurred. In a few cases, head teachers reported a very high number of visits over the past year (upward of 50, as high as 200 visits in a single year). These cases are likely those in which the school is located very near an MOE office or, in the extreme case of 200 visits, it appears that the MOE official works out of or is affiliated directly with the school in some way. To ensure that these outliers do not influence our findings, we censored their values to 24, which represents two visits per month on average throughout the last year, a high rate of monitoring visits.

In all, the share of schools reporting a monitoring visit in the last year increased by 3.6 percentage points, as shown in the first two columns of the table below. At baseline, 83.6 percent of schools reported such a visit in the year prior, compared to 87.2 percent at midline.[239] Outsized improvements were observed in Galmudug, where the share of schools reporting a visit increased by half.[240]

TABLE 70: MOE VISIT OCCURRENCE AND FREQUENCY OVER PREVIOUS YEAR, ACCORDING TO HEAD TEACHERS

Zone	Received visit Baseline	Received visit Midline	Number of visits Baseline	Number of visits Midline
Somaliland	86.4%	88.6%	7.3	7.6
Puntland	85.2%	83.0%	3.3	2.5
Galmudug	66.7%	100.0%	2.1	2.7
Overall	83.6%	87.2%	4.7	4.5

While a somewhat higher share of schools received visits in the year preceding the midline evaluation, the mean number of visits received has not increased at all since the baseline, as shown in the right-most two columns of the table above. Importantly, the number of visits includes those schools that received *no* visits – meaning that the aggregate number of visits conducted by MOE personnel declined slightly over

[239] The difference from baseline to midline, though substantively meaningful, is not statistically distinguishable from no change, likely owing to the small sample size in the head teacher survey (n = 115 head teachers at midline).

[240] Note that the sample size available for analysis in Galmudug and Hirshabelle (which are combined in all geographically-disaggregated results in this report) is small, consisting of just 12 schools at most. Therefore, we should not draw too strong of conclusions from these results, as they may be the result of random year-on-year fluctuations or statistical noise.

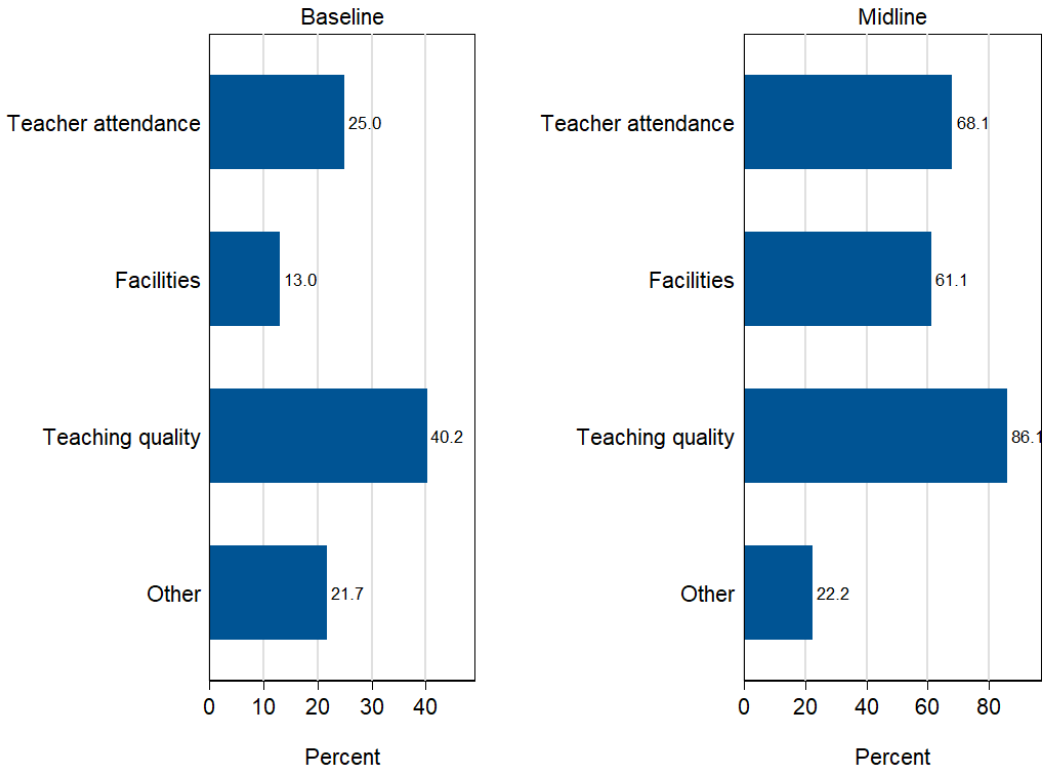
time, even while increasing coverage (i.e. the number of schools visited at least once). Head teachers in Somaliland report, by far, the highest number of visits on average, but neither Somaliland nor any other zone experienced a statistically significant increase in MOE monitoring visits over the baseline. In general, the number of visits is quite small for an entire school year: the median school in Somaliland received four visits in then last year (one per quarter); the median schools in Puntland and Galmudug received just two and three visits in the last year, respectively, working out to less than one visit per quarter.

Interestingly, when we asked head teachers directly whether they felt their school received more monitoring visits this year than last year, the majority agreed that they had. Overall, 66.7 percent claimed more visits this year than last, with the highest rate of agreement in Somaliland. However, when asked how many visits they had received, a comparison to the baseline did not support the idea that more visits occurred this year than last. It is worth noting that many of the head teachers interviewed had been reassigned and were new to their schools this year. Therefore, their perception of visits relative to last year may not reflect the reality of changes from baseline to midline for these specific schools. On the other hand, head teachers may also be unaware of some visits that occur, especially if they are not the target of the visit. MOE officials occasionally visit schools to support life skills training or other efforts, and these visits may escape the attention of head teachers, particularly in larger schools.

School monitoring can be a wide-ranging affair, and could include a brief stop to check that the school is open and functioning, or a much more detailed visit to check school records, observe classrooms, and ask students for their feedback, among other possibilities. Focusing on a fairly narrow set of possible topics out of necessity, we asked head teachers the purpose of the last visit they received. The results are reported in the figure below. Baseline results are reported in the left panel, where respondents were asked to select one option from a list of response options; midline results are provided in the right panel, in which respondents were able to select multiple responses.[241]

[241] It is for this reason that the total percentages sum to 100 in the left panel but sum to over 100 in the right panel -- because multiple response questions allow cumulative percentages greater than 100.

FIGURE 28: SUBJECT OF MOE MONITORING VISITS, BY ROUND



The pattern of monitoring topics has not changed dramatically over time. In both evaluation rounds, teaching quality was the most common topic of monitoring, though the extent of its dominance over other topics changed from baseline to midline, possibly as a result of the change in the structure of response options noted above. Many more respondents indicated that officials were interested in the quality or availability of facilities at midline; it is likely this is an ancillary goal of most visits, and it was de-emphasized at baseline when respondents were forced to select a single response option. Other head teachers reported visits centred on checking student attendance, the backgrounds of newly-transferred students, enrolment figures, and checking the curriculum being used.

While there is little evidence for increased visits by MOE officials, there is suggestive evidence that the quality of MOE visits may be improving, at least along one dimension. We asked head teachers whether they receive feedback or any form of communication *after* a visit by MOE officials. The idea of this metric is that interaction between schools and the government officials that oversee them should be two-way. Schools and head teachers provide the officials with information regarding their school, issues they face, and data on students and their grades, among other topics. But officials should also communicate feedback in a timely and digestible manner, to give schools the opportunity to improve.

At baseline, we found that relatively few (39.4 percent) of head teachers who reported an MOE visit in the last year also reported receiving feedback from MOE officials.[242] By the time of the midline, 57.3 percent of head teachers reporting a visit also indicated that they received feedback following visits, an improvement that is statistically significant.

The qualitative data is mixed on the topic of MOE involvement in schools, and it appears as though ministry involvement differs heavily on a school-by-school basis. Some head teachers report explicitly that the government provides no support to their school. Head teachers suggest that lack of government involvement in their schools is the result of either corruption, challenges the ministry faces in reaching their schools due to lack of transportation, or the ministry employing unqualified people to carry out monitoring.

For example, one head teacher who appears to have been particularly bothered by the lack of ministry involvement explains, "It is very important for the ministry of education to make monitoring. They have monitors but they are not functioning, they are sitting in the offices. I researched and I realized that monitors are former retired teachers or lazy teachers which could not able to teach their schools – these people are not qualified to do monitoring, so I suggest to function the people who have the capacity to monitor. Those who are not active you can replace with very active and honest citizens who serve their country."[243] Another explains, "We are far from the government centres and that is why there is no more monitoring on us. There is a lack of communication between us and the government, so they should come to us frequently to get our information."[244] In another school, the head teacher explains that they did receive some monitoring, but that the feedback was minimal: "Yes, we did not receive any support from him – he was just busy to continue his duties. And he just said 'You are good; you are a good school.'"[245]

However, in other schools, the ministry appears to be heavily involved in project activities and general monitoring, which has in some cases improved teaching quality according to the head teachers. One head teacher reports, "We have seen better quality than in the past year, and the reason could be that most of them are trained teachers. Also, another reason could be that there was more monitoring from the ministry of education and local government in the last year. Every 10 days, the ministry comes to us for a monitoring. They monitor many things, such as the attendance of the teachers and the way they are teaching."[246] In this same school, the head teacher reports, "Always there is supervision from the ministry

[242] This question was worded somewhat ambiguously, but should capture head teachers' perceptions of the frequency of feedback following visits. We asked "after the MoE officials leave, do you receive any feedback or reports about these monitoring visits?" The question did not reference the last visit, or the "typical" visit. Our expectation is that most head teachers would consider the totality of their experience over the last year (i.e. thinking over the set of visits they had received, if they had received multiple monitoring visits), perhaps affected by some degree of recency bias that is common in public opinion surveys. Given that our interest is in the general frequency of feedback, rather than feedback received exclusively after the last visit, this is in line with our goals, though less precisely worded than we would prefer.

[243] KII, Head Teacher, Somaliland

[244] KII, Head Teacher, Puntland

[245] KII, Head Teacher, Puntland

[246] KII, Head Teacher, Somaliland

of education and supervision from the local government project that is implemented by the organizations, which is called GPLG under which the local government made a specific office for education. Under that supervision, they supervise many things such as sanitation, attendance for teachers and the ways of teaching. So in all these things, there is supervision which was not existent in the previous years." In another school in Somaliland, the head teacher also reports heavy involvement from the ministry: "The Ministry of Education works closely with our school because our school is a public school. If we have any issues related to teachers, school management, or any other needs, we report the matter directly to them and they support us. This school has 33 teachers and 19 of them receive their salaries from the government. In addition, both the government and ADRA work with us on a project called GPL. As the head teacher, I presented my development plans to them and am awaiting implementation."[247]

Discussion

The core indicator of sustainability surrounding MOE visits and engagement with project schools is the frequency of visits and the number of head teachers who report at least one MOE monitoring visit. By these metrics, there has not been any improvement over time, when we compare a sample of head teachers and schools that are comparable from baseline to midline. Positive findings regarding the provision of feedback following a visit are encouraging, but cannot and do not replace more frequent visitation by MOE officials.[248] Additionally, it is clear from the qualitative data that improvements are not uniform across regions or schools, and that rather, ministry involvement differs heavily on a school-by-school basis. As a result, the score for this system-level indicator remains latent, as at baseline.

Indicator Grade: 1 – Latent

Sustainability Indicator 10 – TTIs incorporate CPD curriculum

TTI (Teacher Training Institutes) integrate components of the CPD approach into their curriculum for teacher training

Teachers, coaches, and TTI staff all confirm that aspects of the CPD curriculum have been incorporated into teachers' lessons, and students' descriptions of the teaching methods teachers use also confirm that this activity has been largely successful. In particular, teachers are more consistently formulating lesson plans and preparing for their lessons ahead of time, incorporating learner-centred activities into their lessons, engaging in participatory teaching techniques, taking special care to engage both genders equally, and building trusting relationships with students using positive reinforcement techniques.

Through implementing the CPD approach, teachers are experiencing positive results and feel better equipped to fulfil their roles. One teacher who received training under the program explains, "When you

[247] KII, Head Teacher, Somaliland

[248] As noted above, the total number of visits reported by head teachers declined very slightly over time, though the share of schools receiving at least one visit increased from baseline to midline. We consider improved coverage of schools with at least one visit a positive indicator, as the marginal benefit of the first visit by an MOE official to a school is likely higher than the marginal benefit of subsequent visits. Nonetheless, improved coverage at the expense of fewer visits per school is not sufficient to warrant an improved score for this indicator.

plan something and when you do not plan anything, it is not the same - the person who is well planned will implement well. Otherwise it will be disorganized. Before I use to prepare the lessons in the morning but now I prepare all the school year lessons."[249] Another teacher discussed how her training has helped her build a better relationship with students: "Over past years, I did not communicate with students in a friendly way, but now I have changed that and I have a good relationship with students. So I can know that has changed."[250] Others have gained knowledge that guides them in each lesson: "…every teacher must answer three questions during class: 1. How do you teach? 2. Who do you teach to? 3. How much do you want to teach? In addition to this, he/she must provide suitable lessons to each level and prepare supportive tools. Similarly, there must be proper class management with good ethics during class."[251] Specific teaching techniques allow teachers who have been trained under the program to better engage with their students: "The behaviours I have changed include: when I am presenting the lesson, I make connections between previous lessons and current ones and ask questions related to previous lessons to the student…in addition I give homework assignment to present in front of the student by providing them awards to be motivated."[252]

Although these activities have been largely successful, qualitative interview respondents did have a number of suggestions for improving them. First, one of the TTI staff explained that the training is only successful for teachers who have some basic academic knowledge/background: "This program helps more teachers who are academically strong. We know if the teachers do not know basic knowledge about the subject, this cannot help anymore. In the rural teachers, this program was a waste of time and resources."[253] This suggests that some teachers in rural areas do not have a foundation of education upon which the project can build through its trainings. Additionally, one teacher who received training under the program explained that it would be more useful if principals also received training: "It can make it more helpful just to include the school principal in the training to overcome problems. Because once the teacher is being trained, then he needs the administration to have enough information about what the teacher has taken. There may be challenges, or the training can be unsatisfactory, so for the principals to make it easy and solve problems, it is important to train them."[254]

Coaches also had a number of insights into the sustainability of the program and CPD approach. Overall, coaches have an exceedingly positive view of the program, and report positive improvements in teachers' teaching methods in the classroom and levels of organization/preparation for lessons. To improve the sustainability of the program, many suggest that the program simply needs to be continued and expanded – the more teachers who receive training through the program, the more institutionalized the CPD approach will become. There were a number of notable, more concrete recommendations coaches provided related to the sustainability of the program. For example, one coach notes the current lack of institutionalized knowledge and encourages the program to find a way to make these activities

[249] KII, Female Teacher, Somaliland
[250] KII, Female Teacher, Galmudug
[251] KII, Male Teacher, Somaliland
[252] KII, Male Teacher, Galmudug
[253] KII, TTI Official, Somaliland
[254] KII, Male Teacher, Galmudug

sustainable without financial support from the project: “To keep its own financial, so as to keep away breaking off the program. It was something missing from the school and it needs to be increased in the program, because it is very important in the school. The teacher can die or he can go to another school - therefore the new teacher that comes to the school needs this program and training.” A training of trainers approach wherein trained teachers are encouraged to provide additional training to other teachers in their school or even to arrange conferences with teachers from other schools to share learnings might be one way in which this type of institutionalization of knowledge could be encouraged. Another coach recommends that the program provide teaching instruments that can be used even after the project has concluded: “I am just recommending two things: trainings and teaching instruments. We already told ADRA. If these two things are provided to teachers, we will see changes for both students and teachers. For example, if the teacher is given educational tools like audio-visual tools, it will help the students to simply pronounce words and write.”

Discussion

The assessment of this indicator is complicated by the fact that it was not measured at baseline, and our initial measure is taken from a period following the implementation of CPD activities. This is particularly problematic because our data do not reveal the extent to which the curricula of the Teacher Training Institutes *already* incorporated aspects of the CPD approach. In short, although we can assess the extent of incorporation in the present, it is nearly impossible to judge this outcome against an objective baseline. While the resulting score is subjective, it reflects a general view that TTIs are actively engaged with the CPD program, though further refinement at endline may be necessary.

Indicator Grade: 2 – Emerging

Aggregate Sustainability Findings

EGEP-T interventions have been underway in project communities since mid-summer 2017. In addition, the majority of EGEP-T schools were previously included in the first phase of GEC programming, as part of EGEP, which ran for approximately three years and concluded in late 2016. In contrast to many development projects, which consider sustainability only at the conclusion of their activities, seeking to sustain whatever impact they achieved, GEC-T projects incorporate efforts to encourage sustainability from their outset and throughout the life of the project. To the extent that EGEP-T is in a favourable position, with regard to sustainability, it is likely due to this long-run focus on sustainability.

GEC-T employs a scorecard approach to measuring sustainability, applying a uniform scoring scale (0-4) to each indicator and subsequently averaging these scores to generate a composite score on the same 5-point scale. As noted above, possible scores are as follow:

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

It is important to remember a key caveat that applies to the sustainability analysis conducted above and summarized in this section: the scores assigned for each indicator are naturally subjective. While many of the indicators include precise quantitative measurement, analysts must still make judgments when assigning a score on a 0-4 scale; further judgments are made when analysts assess change from baseline to midline. What values constitute an "emerging" indicator? How much improvement is necessary to move from "emerging" to "becoming established"? While we have attempted to interpret the data fairly and triangulate results across multiple sources, subjectivity is inherent in this process.

Throughout the previous sections, we reported scores for individual indicators at the conclusions of their sections. Those scores are reproduced in the table below, with distinct categories of indicators listed in separate columns. 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. Midline and, for reference, baseline scores are reported in the table.

TABLE 71: SUSTAINABILITY SCORECARD RESULTS

	Community	School	System
Indicator 1:	Awareness-raising and attitudinal change **Baseline Score: 2** **Midline Score: 2.5**	CEC bursary support for girls **Baseline Score: 2** **Midline Score: 3**	Gender Development Strategy **Baseline Score: 1** **Midline Score: Not Measured**
Indicator 2:	Male support for girls' education **Baseline Score: 1** **Midline Score: 1**	CEC financial support for schools **Baseline Score: 2** **Midline Score: 2**	Ministry of Education Monitoring Visits **Baseline Score: 1** **Midline Score: 1**
Indicator 3:	Community leaders' advocacy for education **Baseline Score: 2** **Midline Score: 2**	Teacher mentoring programs **Baseline Score: 1** **Midline Score: 2**	TTIs use CPD curriculum **Baseline Score: Not Measured** **Midline Score: 2**
Indicator 4:		Child protection mechanisms **Baseline Score: 1** **Midline Score: 2**	

Baseline Sustainability Score (0-4)	**Mean Baseline Score: 1.67** **Mean Midline Score: 1.83**	**Mean Baseline Score: 1.5** **Mean Midline Score: 2.25**	**Baseline Mean Score: 1.0** **Midline Mean Score: 1.5**
Overall Sustainability Score (0-4, average of the three level scores)		**Overall Baseline Score: 1.39** **Overall Midline Score: 1.86**	

The following sub-section and table below should be completed by the project.

Changes needed for sustainability [to be completed by project]

TABLE 72: PROJECT RESPONSE TO SUSTAINABILITY FINDINGS

School	
Change: What change should happen by the end of the implementation period?	Effective community level school management
Activities: What activities are aimed at this change?	CEC capacity strengthening plays a significant part in the support to schools in general and marginalized girls in particular. Midline findings suggest 39.1% CECs provide bursary support to marginalized children, which is up 14.5% points from baseline[255]. To further enhance the gains observed following the midline and further supported from EGEPT monitoring data, the project will conduct enhanced mentoring support to CECs. This effort will build on the training provided to CECs that include longer term school development planning; mobilization of funds and in-kind contributions to support implementation of the SDPs, community-led initiatives to support the most severely marginalized girls as a community safety-net approach; awareness raising for increase for girls education, follow- up with any girls potentially dropping out of school, tracking teacher attendance, school based Child Protection, and strengthened administration systems. Mentoring support will focus on assessing progress on

[255] EGEPT Midline Preliminary Findings – June 2019.

	the CECs specific identified capacity gaps and actions on their school development plans, which will include resources mobilization as part of the SDP. Further, they will be supported to explore opportunities in their catchment area that would support the education initiatives in their school plans. These are aimed at enabling the CECs to take increased responsibility over the course of the project and to continue functioning effectively in critical school management capacity on exit of the project. A broad-based strategy including other actors in the discussion and planning will increase wider community engagement and localized solutions.
Stakeholders: Who are the relevant stakeholders?	EGEPT aims for strengthened linkages between the CECs and the respective ministries regarding education and child protection to advocate for school needs and to develop systems for holding the respective ministries to account. In addition, stakeholders include like-minded education and CP actors, local businessmen, Somali diaspora.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	CEC training targets a few individual community members in the hope that their commitment lasts and that they cascade the knowledge and skills gained with the rest of the CEC members effectively. Low education levels have been observed to affect CEC capacity to meaningfully engage teachers on the quality of teaching they provide. Head teachers play a significant role in decision making even as they serve as the secretary of the CEC. Lack of distinct institution relationship with MOE & 'umbrellas bodies' providing education.
System	
Change: What change should happen by the end of the implementation period?	Improved teaching quality
Activities: What activities are aimed at this change?	Low teaching quality has been noted from the midline to have statistically significant relationship with learning outcomes. Therefore, enhancing teaching quality and facilitating sustainability of both efforts and gains is important is shifting learning for children. EGEPT envisaged sustainability of teaching quality to be ensured through a two-pronged approach; First, training approach that is built to enable retention of knowledge and skills gained by primary and secondary teachers offering mathematics. Teachers enrolled in the continuous development programme have been trained on subject content, gender sensitive and inclusive pedagogy, child-centred approaches and child protection. Second, MoE and the teacher training institutes with support of the project, develop teacher training content and reviews the approach every 6 months. In addition, trained teachers are mentored through on-job-training by trained coaches employed by the

	mainstream national level teacher training institutes in each zone. The later initiative is aimed at enhancing greater institutionalization of the EGEPT professional development approach for teachers at Ministry level. In due course, head teachers who receive feedback from the coach's support to teachers are expected to learn from them and borrow best practices to integrate with other teachers in their schools. As noted in the midline report, implementation of CPD is in its infancy stage and therefore measure towards sustainability is not currently possible. The project will continue with the approaches mentioned above coupled with the proposals mentioned from the qualitative interviews that include engaging of head teachers, coaches and target teacher sharing of lessons learnt with teachers not enrolled in the professional development program. Additionally, review of the data collection tools and MEL framework is planned to enhance the quality of coaching support as well as generate evidence for effect of EGEPT teacher professional development approach for greater buy in from MoE and other like-minded agencies.
Stakeholders: Who are the relevant stakeholders?	Engagement between MoE, Teacher training institutes and like- minded agencies to share lessons and build upon EGEPT teacher continuous professional development. GPE and Grant Agents e.g. Save the Children International are important stakeholders leveraging the fact that GPE seeks to support establishment of Public TTIs across Somalia/Somaliland.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Lack of a holistic Teachers Development Plan and Strategy across locations led to projectization and fragmentation of teachers training interventions, which is one of the key contributors affecting teaching quality. Lack of clear understanding of the magnitude and relevance of the problem and resultant lack of strategic focus on teachers' development within ESSPs to provide guidance on the pathway to learning and resourcing. Inadequate public teacher training institutes managed by the Ministries of Education.
System	
Change: What change should happen by the end of the implementation period?	Functional School Based Child protection mechanism
Activities: What activities are aimed at this change?	Building on the gains observed from the midline on child protection, EGEP-T project will continue to strengthen the school based child protection mechanism. The project will engage more closely with girls' and boys' leadership networks to ensure they are engaging with their peers as well as with the decision making organs in the school. Leadership networks members supported by the teacher

	mentors will receive refresher training on their roles and responsibilities and life skills. Leaders from the networks will be encouraged to use platforms to enhance sharing with their peers, for example during school assemblies and structured sports events etc. To enhance the voice of children in decision making within the school management, CEC including head teachers who have received training on their roles with regard to child protection will be mentored on receiving and engaging with children in decision making. Girls will be encouraged to raise general child protection concerns (as opposed to specific cases) in meetings with the CECs. This component will close the gap in the feedback loop and institutionalize child protection within the school management system. Additionally, at Ministry level the project will advocate for inclusion of child protection clauses in the teacher code of conduct where possible and its implementation across schools. Efforts will continue to build capacity of respective ministries responsible for child protection on tools to allow them provide supportive supervision on the implementation of child protection action plans. Advocacy on alternative/positive discipline approaches coupled with enhanced engagement of head teachers on approaches to be used in school will be conducted. Moreover, EGEP-T will link the school protection focal points to GBV and CP service providers with support of the MoE for referral of cases reported. Where possible, focal points will be linked to existing Child Protection Focal Points in the community (trained previously through other projects) and will advocate for respective ministries to budget for Child Protection Focal Point person salaries. Efforts to link schools with the community child protection mechanism will be enhanced. EGEPT partners will continue to engage with the Somalia and Somaliland Child Protection Working Group and the Protection Cluster, and will exchange learning with group members.
Stakeholders: Who are the relevant stakeholders?	By the end of the project, a range of stakeholders connected to the school will be pushing for change to be sustained – teachers, head teachers, ministry officials, community members, boys and girls themselves.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Child Protection still remains a low priority objective for the state[256] and communities. Child protection services are limited in Somalia. Referral mechanisms largely rely solely on services provided by other NGOs working in the area through projects. Confusion in terms of roles and remit and lack of leadership among the mandate holding ministry i.e. ministry responsible for implementation of CRC commitments and preparation of state of children in Somalia report, and other relevant ministries is one of the largest systems gap. Ministry of Education has taken lead in integrating child protection into education through the

[256] Somalia ratified CRC in late 2015.

	integrated framework. There exist cultural practices that undermine child rights principles however individuals willing to challenge these social norms across the locations are present. Ministry of Education is gradually take lead on child protection in schools as demonstrated by MOE Puntland state which has established a Child Protection unit to guide and oversee protection in schools. In addition Somalia Education cluster developed education-child protection response framework which will guide implementation by all implementers in Somalia. EGEPT project influenced the development of the integrated framework based on learning garner from implementation.
System	
Change: What change should happen by the end of the implementation period?	Systematic and effective monitoring by respective MoE staff
Activities: What activities are aimed at this change?	EGEP-T will provide capacity building opportunities to MoE in all the project zones to aid in ministries taking increased ownership and management of the target schools and wider education systems. Strategic to this effort is the continued engagement of the Regional Education Officers, Gender focal points from the Gender department and quality assurance officers in the Ministry of Education. To promote greater sustainability, EGEP-T will conduct refresher training to increase the frequency and quality of MOE monitoring visits to project schools. It was noted that while there was a slightly higher share of schools who received visits in the year preceding the midline evaluation, the mean number of visits received had not increased. During the training realistic targets will be set to ensure that the spread of support to all schools is shared irrespective of the distance. Specific training content will be developed to review the current monitoring tools and supportive supervision and providing feedback. Key to the review of the monitoring tools will be inclusion of child protection and building on to existing Government tools for alignment including contribution to the Ministry of Education strategic plan indicators. Gender focal points have received training on child protection, psychosocial support and psychological first aid and are lead trainers and mentors for the teacher mentors. Further, all MoE officials interacting with schools including gender focal points who engage closely with teacher mentors will be encouraged to provide constructive feedback on all their visits to the head teacher.

Stakeholders: Who are the relevant stakeholders?	Training and support to MoE staff will include encouragement to share relevant information with schools and communities, to engage in dialogue, listen to their concerns and act on them where possible. Other education actors and donors e.g. EU, GPE, NORAD, ECW, and the World Bank, are important stakeholders for strategic and harmonized efforts.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	High turnover of MoE staff is one of the factors influencing sustainability. EGEP will therefore concentrate on establishing tools and processes at the institutional level, to ensure that knowledge and practice is retained. Facilitation of MoE to visit far flung schools remains an impediment as well.
Community	
Change: What change should happen by the end of the implementation period?	Increased support for girls education
Activities: What activities are aimed at this change?	Support for education especially by male community members and community leaders are critical in rallying support for education. Community leaders in particular, interpreted to include religious leaders, clan elders, elected officials, and other prominent community members, have the ability to shape public opinion and action. At the household level, there is some evidence from the barriers analysis in the midline that caretakers' beliefs about the relative priority of boys' versus girls' education may have a strong influence on girls' learning outcomes. EGEP-T project seeks to employ a myriad of focused intervention to advocate for support to education for girls both at household, community and national levels. As defined in the projects theory of change, radio advocacy coupled with community dialogues will be carried out in communities. In a bid to enhance saturation of messages, frequency of airing radio shows, variation of show employed e.g. radio drama, live talk show with community leaders and outreach activities by girls and boys and leadership networks and CECs will be employed. EGEP-T will support MoE to work with radio stations to develop scripts to air and work with CEC to progressively lobby for free airing as part of PPP for education. Influential Community leaders will also play a huge part in radio advocacy among other awareness and behaviour change interventions. Moreover, boys and men's' workshops will be held in different zones where they will be trained and encouraged to commit to transformation of male support for girls education. Focused messages identified from the midline findings that prompt support will be highlighted in these avenues.

Stakeholders: Who are the relevant stakeholders?	By the end of the project, a range of stakeholders led by community leaders, community members, ministry officials and girls themselves will be pushing for change to be sustained.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Shift in community members attitudes aspirations towards girls' education including completion at higher levels. Cost of airing radio shows by radio stations.

7. Intermediate Outcomes

7.1 Attendance

Indicator: Improvement in marginalized girls' attendance in schools throughout project's life

TABLE 73: INTERMEDIATE ATTENDANCE OUTCOME INDICATORS AS PER THE LOGFRAME

IO	IO indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Attendance	Improvement in marginalised girls' attendance in schools throughout project's life (weighted average from attendance records)	83.2	85.2	85.3	Y	87.3	Y
Main qualitative findings							
• Girls' most common reasons for missing school are related to household obligations, especially increasing chore burden for older girls.							

TABLE 74: GIRLS' ATTENDANCE RESULTS BY ZONE (FROM ATTENDANCE REGISTER)

Round of Data Collection	Somaliland	Puntland	Galmudug

Baseline	80.1	85.7	79.2
Midline	82.5	88.4	84.3

TABLE 75: BOYS' ATTENDANCE RESULTS BY ZONE (FROM ATTENDANCE REGISTER)

Round of Data Collection	Somaliland	Puntland	Galmudug
Baseline	79.6	86.9	80.3
Midline	79.1	89.5	84.4

Attendance Records

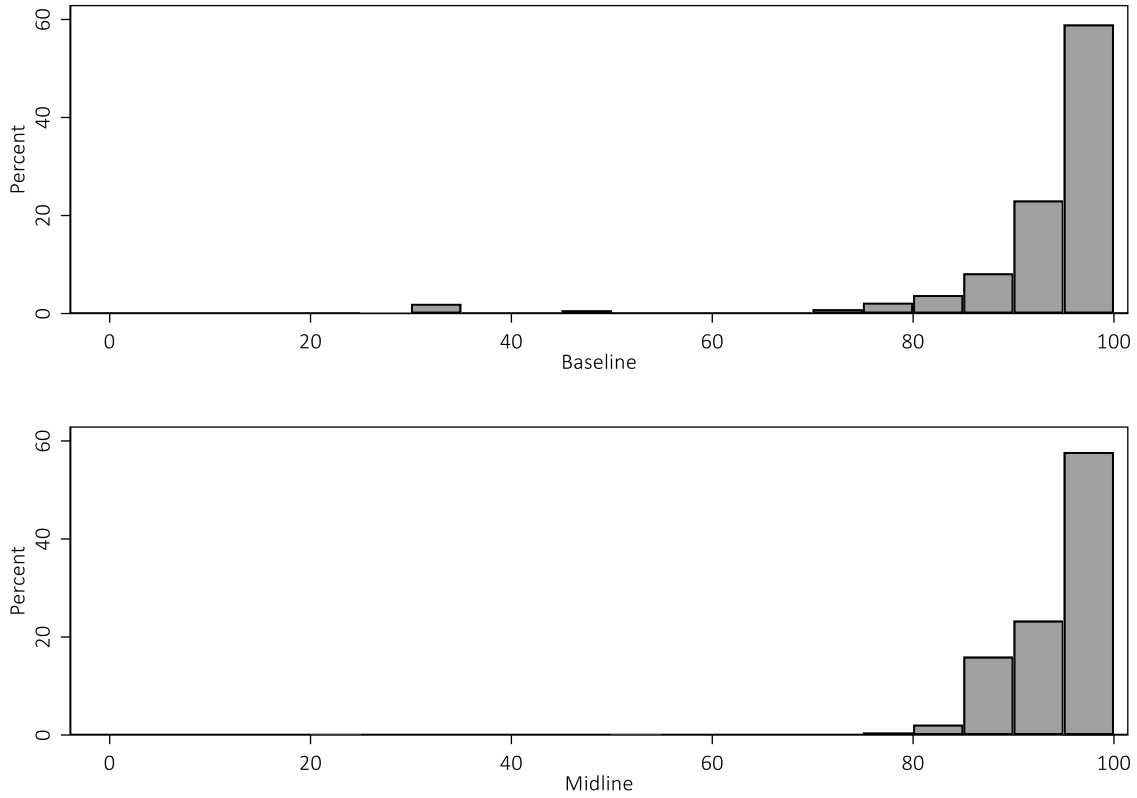
While the baseline and midline evaluations collected data on attendance from additional sources, the primary attendance indicator selected for EGEP-T are the attendance records of cohort girls collected from school records (i.e. classroom registers). These records are aggregated over the academic year to date. That is, the data indicate how many days of school a given girl has attended since the school-year began, and how many possible days of attendance there have been thus far in the school-year.

It is important to note that the sample of girls for whom reliable attendance data for both baseline and midline data is somewhat smaller than the overall sample of cohort girls. There are a total of 613 cohort girls for whom we have both baseline and attendance data and who form the panel for this analysis. This individual-level sample allows us to analyse the relationship between attendance and attributes of individual girls and their families, as well as their individual perceptions of their teachers and school.

The attendance records of cohort girls in the panel collected at baseline and the midline reveal a modest improvement in attendance from, in the baseline, an average rate of 93.2 percent to an average of 94.6 percent in the midline evaluation. This observed positive shift in attendance rates corresponds to the improvement found in the headcount data collected by our town team and which will be discussed in its own section below. Still, this modest gain in attendance following the first year of intervention does not constitute a statistically significant gain.

The improvement observed in the midline seems to be largely driven by fewer girls with attendance rates lower than 85 percent. The figure below displays the distribution of attendance rates in the baseline and in the midline. Both histograms show that attendance rates are left-tailed, and the vast majority of girls have attendance rates above 90 percent. However, in the midline, fewer girls in the midline have attendance rates lower than 85 percent than did so in the baseline.

FIGURE 29: DISTRIBUTION OF ATTENDANCE RATES, BY ROUND



Indeed, as shown in the table below, 18.9 percent of panel girls in the midline have less than a 90 percent attendance rate, while a mere 2.9 percent of girls have less than an 85 percent attendance rate. A similar decline is observed in the baseline among the panel girls, but that decline is less pronounced than that of the midline. Panel girls in the midline are significantly less likely than panel girls in the baseline to attend school less than 85 percent,[257] 80 percent,[258] and 75 percent[259] of the days in the school year. These findings suggest that the EGEP-T intervention may have helped to curb the influence of some of the factors contributing to low attendance rates.

TABLE 76: LIMITED ATTENDANCE RATES, BY ROUND

	Round	
Attendance level	Baseline	Midline

[257] P-value = 0.004 in a cluster robust logistic regression.
[258] P-value = 0.002 in a cluster robust logistic regression.
[259] P-value = 0.013 in a cluster robust logistic regression.

Less than 90% attendance	18.0%	18.9%
Less than 85% attendance	9.8%	2.9%
Less than 80% attendance	6.0%	0.8%
Less than 75% attendance	3.8%	0.3%

There were few subgroups or characteristics of panel girls that had significantly lower attendance rates in the midline. The only subgroups in the midline that were significant predictors of attendance were geographic. Girls in Puntland attended school at significantly lower rates[260] than average, while girls in Galmudug have significantly higher attendance rates[261] in their school records, as shown in the table below. Neither drought nor urbanicity were predictors of attendance rates, which rule out their effects on zone-level differences. Girls from rural areas actually had significantly higher rates of attendance in the baseline. However, conflict *is* a significant predictor of attendance rates, and, though girls in both Galmudug and Puntland live in areas with conflict, a higher proportion of girls in Puntland live in conflict-affected areas across both baseline (12.3 percent) and midline evaluations (7.1 percent).

TABLE 77: ATTENDANCE RATES AMONG DISADVANTAGED SUBGROUPS, BY ROUND

	Baseline Attendance	Baseline Sample Size	Midline Attendance	Midline Sample Size
Characteristics:				
All bursary and cohort girls	93.2%	451	94.6%	613
Somaliland	86.2%	66	95.6%	162
Puntland	95.3%	259	93.4%*	323
Galmudug	92.4%	126	96.4%*	128
Rural	97.3%*	84	94.4%	145
Disability				
Vision impairment	93.5%	3	94.7%	4
Hearing impairment	.	0	.	0
Mobility impairment	.	0	.	0
Cognitive impairment	.	0	100.0%	1
Self-care impairment	.	0	.	0
Communication impairment	85.7%	1	.	0
Any disability	91.6%	4	95.8%	5

[260] P- value = 0.006 in a cluster robust linear regression.
[261] P-value = 0.029 in a cluster robust linear regression.

Serious illness in last yr.	93.0%	37	94.6%	93
Other				
Old for grade	94.2%	46	95.6%	86

Note: Statistically significant differences at the 95% confidence level or higher are marked with an asterisk.

With regard to disabilities and other subgroups, there were none that were significant predictors of attendance. The attendance records of only a few girls with disabilities were collected in both the baseline and the midline, presenting challenges for analysing the effects of disabilities on girls' attendance. In addition, being relatively older for a given grade was not a significant predictor of attendance.

The table below presents the attendance rates of cohort and bursary girls facing potential barriers to attendance. Many of the potential barriers are not significant predictors of attendance. A number of the significant relationships are counterintuitive and are likely the result of a spurious relationship or reverse causality. Variables grouped under the same sub-heading in large part do not have significant relationships that indicate an overall relationship of that sub-heading with girls' attendance. There are however a pair of related variables that point in the same direction: among panel girls, having over 30 minutes of travel time to get to school[262] and feeling unsafe on the way to school are significant predictors of lower attendance rate.[263] While the 92.3 percent attendance rates of girls who have an over 30 minute path to school in the midline is a substantial improvement over the 79.5 percent attendance rate of girls who also had a long travel time in the baseline, these findings suggest that there are yet barriers to be addressed on girls' journeys to school.

The qualitative data offered corroborating evidence that security on the journey to school presented difficulties for girls attending school. One father in Puntland explained, "First, boys and girls are not the same. When girls are going to school, they have drop them off at school because security has recently not been good in the area. Therefore, parents who can afford to rent a taxi for their daughters do so."[264] A female teacher in Galmudug noted that recent rape cases in the area were affecting girls' attendance: "Recently we have a problem with girls' school attendance, because…a lot of rape cases have happened, and sometimes parents are taking the girls to school."[265]

TABLE 78: BARRIERS TO ATTENDANCE, BY ROUND

	Baseline Attendance (%)	Baseline Sample Size	Midline Attendance (%)	Midline Sample Size
Barriers:				
All girls (cohort and bursary)	93.2%	451	94.6%	613

[262] P-value = 0.037 in a cluster robust linear regression.
[263] Feeling unsafe on the way to school is a significant predictor at the 90% confidence level with a P-value = 0.051 in a cluster robust linear regression.
[264] Fathers' FGD, Puntland
[265] Teacher KII, Galmudug

School Infrastructure				
Difficult to move around school	92.5%	95	96.0%	93
Doesn't use drinking water facilities	92.8%	91	92.5%	62
Doesn't use toilet at school	88.1%	62	94.3%	57
Doesn't use areas where children play/socialise	91.5%	171	95.2%	134
School Resources				
No computers at school	93.7%	324	94.7%	423
School does not have learning materials	92.4%	95	95.7%	87
Not enough seats for children at school	93.7%	85	97.5%*	77
No electricity	96.0%*	59	96.2%	85
Only dirt floors	92.6%	48	97.6%*	49
Offers remedial classes	94.6%	301	93.9%	395
Offers life skills classes	95.9%	106	94.0%	148
Separate toilets for boys and girls	93.2%	431	94.5%	550
School feeding program	94.7%	173	94.5%	245
Offers sanitary kits	91.0%	158	95.1%	216
Offers solar lamps	93.2%	156	96.2%*	210
Has girls club	93.2%	415	94.5%	577
Girls club meets at least once per week	91.3%	26	94.0%	62
School has CEC	93.2%	451	94.5%	597
CEC meets frequently	93.3%	102	93.5%	160
Teaching Quality				
Disagrees teachers make them feel welcome	86.1%	9	96.1%*	43
Agrees that they are afraid of teacher	92.1%	180	94.6%	265
Teacher punishes students who get things wrong	92.0%	211	94.6%	271
Classes held fewer than 5 hours per day	94.1%	58	95.1%	86
Agrees teacher is often absent from class	94.4%	76	95.3%	110
Teacher uses corporal punishment	93.3%	107	94.2%	74
Gender Equity				
At least one female teacher	93.1%	291	94.0%	414
Female mentor	93.5%	309	93.9%*	463

Male mentor	95.8%	276	94.6%	392
No mentor	90.2%	71	97.3%*	72
Teachers treat boys differently from girls	92.8%	112	94.9%	184
Family prioritizes boys' education	92.2%	64	94.9%	54
Other Barriers				
Agrees she has no choice in schooling decisions	93.0%	230	94.4%	400
Over 30-minute travel time to school	79.5%	11	92.3%*	10
Feels unsafe on way to school	95.1%	12	92.2%	13
Feels unsafe at school	97.5%*	2	99.1%*	2
Parents are disengaged	93.7%	27	93.3%	13

Note: Statistically significant differences at the 95% confidence level or higher are marked with an asterisk.

Looking specifically at bursary girls, because the sample of bursary girls is substantially smaller, significant predictors of attendance in both baseline and midline samples are few. No significant predictors of attendance emerged for panel girls in the baseline. In the midline, a bursary girl who attended a school which offers solar lamps was significantly more likely to have a higher attendance rate, and a bursary girl who attended a school which had female mentors was significantly less likely to have a lower attendance rate. Both predictors of attendance among midline bursary girls were also predictors of attendance among the sample of cohort and bursary girls. The positive relationship between solar lamps and attendance is along program design as girls who may have inconsistent access to electricity are able to complete their schoolwork and be more engaged in the classroom. However, the negative relationship between attendance and female mentors may be the result of reverse causation in which female mentors are in schools in which girls are more likely to be absent.

TABLE 79: BARRIERS TO ATTENDANCE OF BURSARY GIRLS, BY ROUND

	Baseline Attendance (%)	Baseline Sample Size	Midline Attendance (%)	Midline Sample Size
Barriers:				
All bursary girls	92.4%	91	94.3%	126
School Resources				
No electricity	96.5%	13	95.5%	15
Only dirt floors	91.8%	7	97.6%	8
Offers remedial classes	94.8%	63	93.7%	84
Offers life skills classes	95.1%	23	94.4%	32
Separate toilets for boys and girls	92.5%	87	94.0%	112
School feeding program	94.3%	32	94.4%	45
Offers sanitary kits	88.4%	26	95.4%	38

Offers solar lamps	94.2%	24	96.2%*	32
Has girls club	92.4%	83	94.2%	118
Girls club meets at least once per week	93.8%	7	94.0%	13
School has CEC	92.4%	91	94.2%	125
CEC meets frequently	94.0%	25	92.8%	35
Teaching Quality				
Disagrees teachers make them feel welcome	--	0	96.0%	7
Agrees that they are afraid of teacher	--	0	93.2%	53
Classes held fewer than 5 hours per day	95.4%	12	93.5%	15
Gender Equity				
At least one female teacher	91.4%	62	94.0%	87
Female mentor	92.2%	67	93.7%*	99
Male mentor	95.4%	60	94.3%	88
No mentor	92.5%	11	96.6%	11
Teachers treat boys differently from girls	--	0	94.8%	40
Other Barriers				
Agrees she has no choice in schooling decisions	--	0	94.3%	89
Parents are disengaged	--	0	91.5%	3

Note: Statistically significant differences at the 95% confidence level or higher are marked with an asterisk.

The effect of the project's outputs such as sanitary kits, life skills classes, and girls' clubs on attendance is described above for the panel girls, and below, we also analyse it for all midline girls. The outputs for which we have data primarily address the barriers of household-level economics and lack of confidence, life skills, and psychosocial support. As shown in Table 80 below, there are few statistically significant differences between girls who have not received a project output and those who received project output. This suggests that on the whole, girls who are in schools receiving targeted project interventions and who may have otherwise been more marginalised due to the financial burden of enrolment and lack of life skills are at least keeping pace with girls who did not receive them. Only girls who were in schools that had a female mentor had significantly lower attendance rates than girls whose schools did not receive that particular project output. This does not necessarily suggest that the presence of female mentors in the school has been ineffective in promoting attendance however, and as mentioned above may to some extent be attributed to reverse causation in which schools who have lower levels of attendance receive female mentors.

TABLE 80: ATTENDANCE OF ALL MIDLINE GIRLS RECEIVING PROJECT OUTPUTS

	Midline Attendance (%)	Midline Sample Size
Total		
All midline girls	94.9%	894
Outputs		

Bursary girl	94.4%	159
Sanitary kits distributed at school	95.0%	356
Life skills classes offered for girls	94.6%	221
Girls' club at school	94.8%	833
School has a female mentor	94.3%*	691
School offers remedial class	94.3%	580

Note: Statistically significant differences at the 95% confidence level or higher are marked with an asterisk.

Headcounts

Like the baseline evaluation, we use physical classroom headcounts collected by Forcier Researchers to report the midline attendance levels. In the headcount survey, enumerators arrived at schools to collect data one hour after the beginning of the classes until one hour before the lunch break. This provided the teachers enough time to record attendance and collect data on students who may only attend school for half of the day. By randomly selecting classes from among mathematics, English and Somali classes at a given school, enumerators collected data through two main methods: 1) used the school's attendance register to record students' attendance the day before the visit and the day of the visit and 2) conducted a direct headcount of students in the class on the day of the visit. The list of the headcount survey questions is noted below for reference:

Headcount Survey Questions

B7. Enter the number of GIRLS enrolled in this class

B9. Teacher count on record: Number of girls marked in class YESTERDAY

B10. Teacher count on record: Number of girls marked in class TODAY.

B11. Girls HEAD COUNT in class (done by Enumerator): Enter the total number of GIRLS present in the class by counting

B12. Enter the number of BOYS enrolled in this class

B14. Teacher count on record: Number of boys marked in class YESTERDAY

B15. Teacher count on record: Number of boys marked in class TODAY

B16. Boys HEAD COUNT in class (done by Enumerator): Enter the total number of BOYS present in the class by counting

In the baseline report, the evaluation team discussed the measurement error in measuring attendance in great detail. Concerns were particularly pointed out regarding data quality in school attendance records (i.e. poor quality of school attendance), and therefore, recommended not to use the school attendance (today and yesterday's attendance) as a reliable metric of progress over time. Instead, it was suggested to employ the classroom headcounts as the primary outcome of interest. Following these

recommendations, we will only briefly analyse the school attendance records while we'll put more emphasis on the headcount attendance.

Given the number of students registered differed from class to class, the school attendance rates as well as the headcount rates are divided by the total number of the students registered for that class to make the attendance rates comparable across these measures. In a considerable number of the cases, attendance rates were missing from the today's and yesterday's attendance records, dropping off the total sample size to 927 and 896 for the today's attendance and yesterday's attendance, respectively. Yesterday's attendance rates in the midline are moderate in the sampled schools with an average of 88.5 percent and did not significantly vary compared to the rates in the baseline. Today's attendance rates improved slightly from the baseline (84.8 percent) to midline (86.1 percent), but this change was statistically insignificant as well.

Turning to the question of how headcount attendance has changed since the baseline, the table below summarizes changes over time in boys' attendance rates and in girls' attendance rates across baseline and midline by a number of geographic, demographic and other school-level correlates of attendance.

TABLE 81: CLASSROOM ATTENDANCE RATES, BY GEOGRAPHIC AND SCHOOL-LEVEL FACTORS

Subgroup	Round	Girls' Attendance	Boys' Attendance	Total Attendance
Overall	Baseline	83.2	83.3	83.1
	N=	611	610	614
	Midline	85.3	84.1	84.5
	N=	581	580	583
Somaliland	Baseline	80.4	80.2	80.1
	Midline	82.5	79.1	80.4
Puntland	Baseline	86.4	86.2	86.1
	Midline	88.4	89.5	88.9
Galmudug	Baseline	82.9	86.5	84.7
	Midline	84.3	84.4	84.5
Rural schools	Baseline	78.4	80.4	79.1
	Midline	79.0	75.4	76.7
Urban schools	Baseline	85.7	84.8	85.1
	Midline	88.9	89.2	89.0
Primary schools	Baseline	83.3	83.8	83.4
	Midline	85.0	83.4	84.0
Secondary schools	Baseline	82.6	79.6	80.8
	Midline	87.2	89.6	88.7
Grade 1	Baseline	83.8	85.0	84.1

	Midline	86.3	82.1	82.9
Grade 2	Baseline	81.5	83.9	82.4
	Midline	84.3	83.0	83.8
Grade 3	Baseline	87.5	82.2	85.0
	Midline	83.7	83.2	82.9
Grade 4	Baseline	83.8	84.5	83.8
	Midline	84.9	85.2	85.1
Grade 5	Baseline	81.7	83.0	82.7
	Midline	86.9	84.4	85.6
Grade 6	Baseline	82.4	82.8	82.0
	Midline	85.2	84.4	84.5
Grade 7	Baseline	82.7	83.0	82.4
	Midline	83.0	82.1	82.4
Grade 8	Baseline	82.9	85.6	84.4
	Midline	86.1	82.9	84.4
Form 2	Baseline	82.2	73.3	76.0
	Midline	87.0	91.9	90.3
Form 3	Baseline	85.8	80.0	82.3
	Midline	84.9	89.7	87.4
Form 4	Baseline	78.8	82.5	81.7
	Midline	86.3	87.0	87.1
Teachers absentee	Baseline	82.5	83.6	82.9
	Midline	84.7	82.1	83.2
IDP schools	Baseline	82.9	81.2	82.2
	Midline	92.1	90.7	91.6
Non-IDP schools	Baseline	83.2	83.4	83.3
	Midline	84.8	83.7	84.2
Conflict-affected schools	Baseline	86.5	85.8	86.1
	Midline	93.4	92.2	92.3
Non-conflict affected schools	Baseline	82.4	82.7	82.5
	Midline	84.7	83.6	84.1
Drought-affected schools	Baseline	77.2	81.5	79.5
	Midline	84.2	84.6	84.2
Non-drought affected schools	Baseline	85.7	84.1	84.9
	Midline	85.5	84.0	84.7
Has school feeding program	Baseline	80.7	82.0	81.1
	Midline	80.7	77.5	78.8
Has girls' toilets	Baseline	84.5	84.2	84.1
	Midline	85.6	84.5	84.9

No girls' toilets	Baseline	76.7	78.9	77.9
	Midline	83.5	81.8	82.2

As shown in the table above, the total attendance as well as the attendance rates of both boys and girls gathered by enumerators changed only marginally and the changes were positive from the baseline to midline. More specifically, girls have had 2.5 percentage increase since the start of the project while boys' percentage increase in attendance were less than 1. Similarly, girls who reside in Somaliland, living in the urban areas, and in secondary school have higher attendance rates from the baseline to midline compared to the girls in Puntland and Galmudug, live in the rural areas, and are studying in elementary schools, respectively. While counter-intuitive, the increase in students' attendance rates across the "IDP schools" and those affected by conflict and drought is slightly higher than the girls' attendance in the non-IDP areas and schools whose areas were not affected by conflict or drought. When boys' headcount attendance rates are disaggregated by the geographic, demographic and other school-level characteristics, the only difference that is distinguishable from the girls' rates is that since baseline, boys living in Puntland have attended schools more regularly compared to the boys in other zones.

However, none of these improvements was statistically significant. In order to remove the effect of geographic, demographic and other school-level factors on students' attendance rates from the baseline to midline, we have gradually added them to our regression model to control for those differences among the students. Overall, neither we found any significant changes in the attendance rates of both girls and boys after controlling for different factors individually, nor when we controlled for those differences altogether.

In summary, schools' records as well as enumerators' headcount data revealed that both boys' and girls' attendance rates have insignificantly improved from baseline to midline. As the findings of qualitative data that are discussed in the following section suggest that this improvement might be a result of the EGEP-T activities that have tried to address major causes of absence among students. The percentage increase in girls' attendance is slightly more than that of the boys. Also, girls in Somaliland have attended schools more regularly from baseline to midline compared to the girls in Puntland and Galmudug while boys' attendance rate in Puntland seems to be slightly higher than the two other zones. Some insignificant improvements in students' attendance rates are observed across urban and secondary schools as well as those in IDP areas and became affected by conflict and drought.

Reasons Girls and Boys Miss School

In the qualitative data, there are stark differences between the reasons girls miss school and the reasons boys miss school. Boys largely appear to miss school due to distractions outside of school, such as staying up late to play or watch football, watching TV all night, smoking with their friends, or chewing khat. Employment opportunities also cause boys to miss school, as some work during their free time or school breaks (e.g. driving motor bikes). In some cases, this was blamed on the lack of attention boys are getting from organizations. When asked why boys are attending school less, a CEC member answered, "Because boys don't get awareness like girls. Also, boys are busy of watching films and playing football."

Respondents most commonly mentioned that girls miss school because of their household obligations. Girls are responsible for helping their mothers with household chores, and when the parents are not available to stay at the house because they are working or otherwise occupied, girls are expected to stay at home to handle chores, care for young ones, and cook. One girl explains that "every time mothers go somewhere else, girls stay home and take the responsibility of the house." Another explains that "there are many reasons that force girls to miss school, including household chore load…similarly when moms give birth, girls must stay at home around 40 days." In the qualitative interviews, both boys and girls highlighted that the expectation differs for boys: "Of course boys are different according to Somali culture. As we know, Somali boys don't do house tasks, except small ones who carry small tasks in the house. But there is not any boy who will make breakfast or lunch or clean the house." Respondents also commonly mentioned that girls miss school during menstruation. As one boy explains, "Boys and girls miss school, but mostly girls miss school due to health issues which will make girls not come to school for more than one day."

As a result of these absences, "after consecutive weeks of being absent, she may disappoint the whole learning," and girls "do not know they can refresh their education and gain success after failing." Additionally, although most students expressed that teachers are supportive and understanding of girls' special situations, there appear to still be cases in which girls' tardiness is not tolerated and results in further absence from school: "I hate when the school doors are closed when I'm in late because I'm female and supporting my family in the house so sometimes you can be late because of housework. The boy can be late even though you have already prepared him to school. So the teacher always closes the doors without consideration of the girls' circumstance." This appears to be an issue in multiple locations, as a mother from a different zone explained, "If a student is late at school because of household chores or something like that and she/he comes at 1:30 PM, the school gate will close and she/he will lose two lessons and later fail at examinations."

However, the results of the qualitative interviews also suggest that these major causes of absence are being successfully addressed through EGEP-T activities. First, girls are receiving more motivation from their parents at home, and are also receiving "huge motivation from social organizations visiting their schools," as well as from "donors, school administration, and teachers." In some cases, mothers are taking on some of the household chore burden and in others, girls are speaking up about their educational needs and pushing back on their household expectations—one mother explained that, "Sometimes girls refuse household chores so that they are not late to school." As a result, numerous qualitative respondents from different locations mentioned that girls are "even better in attendance compared to boys" and "are attending school more [regularly] because they have understood the benefit of education."

Girls are also receiving material support through the project—in particular, teachers have noted a decrease in girls' absence due to their receiving sanitary kits: "Before and now students are not same. Before I didn't receive the trainings when girl is in her menstruation period. The girl will feel tired and she might not be able to come to the school. Now we received sanitary kits and it's in the school. We told

committee[266] students females that this thing is normal and natural for all females and it's not something that you have to miss your school for. Most of the girls take that awareness. Even they come with us two by two and ask for these kits when they need."

Progress to Targets

This section summarizes the progress that has been made to the midline targets set out in the baseline evaluation. The targets were set by defining a percentage point increase in attendance that girls and boys should reach by the time of a subsequent evaluation point. These percentage point increases per evaluation were different by gender. All girl headcount targets set the goal of a 2 percentage point increase in attendance over the baseline value by the time of the midline evaluation. The headcount target for boys set the midline goal of a 1.6 percentage point increase in attendance over the baseline value. The more modest goal for boys reflected the project's focus on improving girls' attendance and outcomes and the projects' consideration of boys' learning outcomes as secondary indicators of project efficacy. Lastly, it should be noted that the goal for improvement in girls' attendance records is to improve attendance rates as recorded in attendance records by only one percentage point since the average attendance rate of girls is already rather high, and there is less room for improvement in attendance. The table below presents achieved baseline and midline levels of attendance along with targets and the progress to those targets.

Beginning with the overall targets, both cohort and bursary girls' overall attendance rates targets in the in the attendance records were met. Girls' attendance rates in the headcounts also exceeded midline targets. The boys' headcount target was not met, missing the target by 0.8 percentage points. Among location-specific targets for attendance, girls' headcount attendance rates met their targets in Somaliland and Puntland, but not Galmudug. Boys' headcount targets were met only in Puntland, and in Somaliland and Galmudug, boys were found to attend school at even lower rates than in the baseline evaluation.

Turning to targets for girls in subgroups, we find that girls in urban, IDP, and conflict-affected areas achieved headcount attendance targets. The attendance targets were not met in the contrasting subgroups. That is, girls from rural, non-IDP, and non-conflict areas tended to have lower attendance rates and did not meet the targeted attendance target. The reason for the divergent results in achieving targets may reflect RI's relative strength and or focus in working with urban populations, IDPs, and in conflict-affected areas.

The most substantial increases in attendance rates occurred among girls in IDP and conflict-affected areas. The largest increase in percentage points over baseline values were observed in children in IDP areas in which girls in the midline had a 9.2 percentage point increase in attendance over baseline girls—a 7.2 percentage point improvement over the midline target for girls—and in which boys had a 9.5 percentage point increase in attendance over boys in the baseline—a 7.9 percentage point improvement over the midline target for boys. The next largest increase was for girls in conflict-affected areas in which we

[266] It is unclear what committee the respondent is referring to, but this is likely just the word the respondent was using to refer to a girls' group at the school – perhaps a leadership network or girls' club.

observed a 6.9 percentage point improvement over the baseline level for girls and a 6.4 percentage point increase over the baseline level for boys. Should these attendance rates be maintained to the endline evaluation point, the endline attendance targets for girls and boys in IDP and conflict-affected areas will have been achieved.

TABLE 82: SUMMARY OF ATTENDANCE AND TARGETS

Indicator	Baseline Level	Midline Level	Midline Target	Diff b/w ML Lvl. and Target	Endline Target
Overall Targets					
Girls' attendance rates, headcounts	83.2%	85.3%	85.2%	0.1%	87.2%
Boys' attendance rates, headcounts	83.3%	84.1%	84.9%	-0.8%	86.5%
Cohort girls' attendance rates, school records	93.3%	94.7%	94.3%	0.4%	95.3%
Bursary girls' attendance rates, school records	92.4%	94.3%	93.4%	0.9%	94.4%
Location-Specific Targets					
Somaliland					
Girls' attendance rates, headcounts	80.4%	82.5%	82.4%	0.1%	84.4%
Boys' attendance rates, headcounts	80.2%	79.1%	81.8%	-2.7%	83.4%
Puntland					
Girls' attendance rates, headcounts	86.4%	88.4%	88.4%	0.0%	90.4%
Boys' attendance rates, headcounts	86.2%	89.5%	87.8%	1.7%	89.4%
Galmudug					
Girls' attendance rates, headcounts	82.9%	84.3%	84.9%	-0.6%	86.9%

Boys' attendance rates, headcounts	86.5%	84.4%	88.1%	-3.7%	89.7%
Subgroup Targets					
Urban					
Girls' attendance rates, headcounts	85.7%	88.9%	87.7%	1.2%	89.7%
Boys' attendance rates, headcounts	84.8%	89.2%	86.4%	2.8%	88.0%
Rural					
Girls' attendance rates, headcounts	78.4%	79.0%	80.4%	-1.4%	82.4%
Boys' attendance rates, headcounts	80.4%	75.4%	82.0%	-6.6%	83.6%
IDP					
Girls' attendance rates, headcounts	82.9%	92.1%	84.9%	7.2%	86.9%
Boys' attendance rates, headcounts	81.2%	90.7%	82.8%	7.9%	84.4%
Non-IDP					
Girls' attendance rates, headcounts	83.4%	84.8%	85.4%	-0.6%	87.4%
Boys' attendance rates, headcounts	84.0%	83.7%	85.6%	-1.9%	87.2%
Conflict-affected					
Girls' attendance rates, headcounts	86.5%	93.4%	88.5%	4.9%	90.5%
Boys' attendance rates, headcounts	85.8%	92.2%	87.4%	4.8%	89.0%
Non-Conflict					
Girls' attendance rates, headcounts	83.1%	84.7%	85.1%	-0.4%	87.1%
Boys' attendance rates, headcounts	83.6%	83.6%	85.2%	-1.6%	86.8%

Indicator	Baseline Level	Midline Level	Midline Target	Diff b/w ML Lvl. and Target	Endline Target
Overall Targets					
Girls' attendance rates, headcounts	83.2%	85.3%	85.2%	0.1%	87.2%
Boys' attendance rates, headcounts	83.3%	84.1%	84.9%	-0.8%	86.5%
Cohort girls' attendance rates, school records	93.3%	94.7%	94.3%	0.4%	95.3%
Bursary girls' attendance rates, school records	92.4%	94.3%	93.4%	0.9%	94.4%
Location-Specific Targets					
Somaliland					
Girls' attendance rates, headcounts	80.4%	82.5%	82.4%	0.1%	84.4%
Boys' attendance rates, headcounts	80.2%	79.1%	81.8%	-2.7%	83.4%
Puntland					
Girls' attendance rates, headcounts	86.4%	88.4%	88.4%	0.0%	90.4%
Boys' attendance rates, headcounts	86.2%	89.5%	87.8%	1.7%	89.4%
Galmudug					
Girls' attendance rates, headcounts	82.9%	84.3%	84.9%	-0.6%	86.9%
Boys' attendance rates, headcounts	86.5%	84.4%	88.1%	-3.7%	89.7%
Subgroup Targets					
Urban					
Girls' attendance rates, headcounts	85.7%	88.9%	87.7%	1.2%	89.7%
Boys' attendance rates, headcounts	84.8%	89.2%	86.4%	2.8%	88.0%

Rural					
Girls' attendance rates, headcounts	78.4%	79.0%	80.4%	-1.4%	82.4%
Boys' attendance rates, headcounts	80.4%	75.4%	82.0%	-6.6%	83.6%
IDP					
Girls' attendance rates, headcounts	82.9%	92.1%	84.9%	7.2%	86.9%
Boys' attendance rates, headcounts	81.2%	90.7%	82.8%	7.9%	84.4%
Non-IDP					
Girls' attendance rates, headcounts	83.4%	84.8%	85.4%	-0.6%	87.4%
Boys' attendance rates, headcounts	84.0%	83.7%	85.6%	-1.9%	87.2%
Conflict-affected					
Girls' attendance rates, headcounts	86.5%	93.4%	88.5%	4.9%	90.5%
Boys' attendance rates, headcounts	85.8%	92.2%	87.4%	4.8%	89.0%
Non-Conflict					
Girls' attendance rates, headcounts	83.1%	84.7%	85.1%	-0.4%	87.1%
Boys' attendance rates, headcounts	83.6%	83.6%	85.2%	-1.6%	86.8%

In general, targets for girls' attendance rates were met, while the targets for boys' attendance rates were not. Of the 11 attendance targets set for girls in the baseline, 7 of them were met in the midline evaluation. In contrast, 4 of the 10 attendance targets set for boys were met by the midline evaluation point. Further, when targets were not met, the *severity* of that failure was greater among boys than girls. Among the 6 targets in which boys did not meet their attendance targets, the attendance level of boys fell short of the goal by an average of 2.9 percentage points. In contrast, in the 4 attendance targets in which girls did not meet attendance targets, the attendance level of girls fell short of the target by an average of 0.8 percentage points. The disparity in target achievement between boys and girls is found despite the less ambitious goal of 1.6 percentage point increases between evaluation points for boys. The finding suggests

that the intervention is succeeding in helping girls to not only achieve parity, but in some cases exceed, the school attendance of boys.

Summary Discussion

By the midline evaluation point, the average rate of attendance among panel cohort girls had risen modestly to 94.6 percent from 93.2 percent in the baseline evaluation. Our analysis found that girls with limited exposure to education had significantly decreased. Panel girls in the midline were significantly less likely than panel girls in the baseline to have attendance rates lower than 85 percent, 80 percent, and 75 percent. Girls in Puntland attended school at significantly lower rates than girls from other zones, while girls in Galmudug had significantly higher attendance rates. Few barriers were found to predict attendance, however we did find that among panel girls in the midline, having over 30 minutes of travel time to get to school and feeling unsafe on the way to school were predictors of lower attendance rates. In the qualitative data, we found that girls were said to miss school commonly because of household obligations while boys largely missed school due to distractions outside of school such as staying up late to play or watch football, smoking with friends, or chewing khat. Findings from the qualitative data suggested that the causes of girls' absences from school were being addressed through EGEP-T's intervention. Targets for girls' overall attendance in the attendance record data of cohort girls and in the headcount data were met, but the boys' headcount attendance target was not achieved. Lastly, we find that girls in urban, IDP, and conflict-affected areas achieved headcount attendance targets, however targets were not met in rural, non-IDP, and non-conflict areas.

7.2 Self-esteem and Empowerment

The primary self-esteem and empowerment indicator is the self-esteem index, the construction of which is detailed below, and is composed of the same questions asked in the baseline and midline and will be asked in the endline.

Indicator: Increase in self-esteem and self-confidence of marginalized girl[267]

TABLE 83: INTERMEDIATE SELF-ESTEEM OUTCOME INDICATORS AS PER THE LOGFRAME

IO	IO indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)

[267] The self-empowerment indicator in the baseline was the percent of girls who agree with the statement, "When I have an opportunity, I can organize my peers or friends to do an activity." In the midline, the self-empowerment indicator was changed to be the self-esteem index, which will also be the self-empowerment indicator in the endline.

Self-esteem and empowerment	Marginalised girls' average score on the self-esteem index	0.51	N/A	0.72	N/A	0.76[268]	Y

Main qualitative findings

- Awareness-raising around the importance of girls' education may be a source of positive inspiration for girls.
- Girls also noted that exposure to education and life experiences is an importance source of confidence for them.
- When asked what contributes to girls' confidence, many girls mentioned the importance of interactive class activities, such as presenting in front of the class.

TABLE 84: GIRLS' SELF-ESTEEM RESULTS BY ZONE (INDEX SCORES)

Round of Data Collection	Somaliland	Puntland	Galmudug
Baseline	0.55	0.52	0.54
Midline	0.75	0.71	0.74

Empowerment of girls and greater self-esteem are among EGEP-T's intermediate outcomes. Through the EGEP-T project, EGEP-T has designed various activities such as girls' clubs, female teacher mentors, improvement of teacher quality, and bursary support to vulnerable girls in order to improve their self-esteem. According to the project's Theory of Change, girls with better self-esteem are more likely to ask questions in class when they do not understand the content of lessons and have more confidence to pursue their education and dreams. As a result, girls' improved self-esteem is more likely to help them learn better and have a successful transition.

Due to measurement difficulties of this multidimensional concept, the evaluation team utilized a set of indicators related to self-esteem and empowerment. The list of survey questions related to self-esteem and agency is shown in the table below:

TABLE 85: QUESTIONS USED TO MEASURE SELF-ESTEEM AND AGENCY INDICATORS

	Survey Questions	Respondents
Se lf	I can read as well as my friends.	Everyone

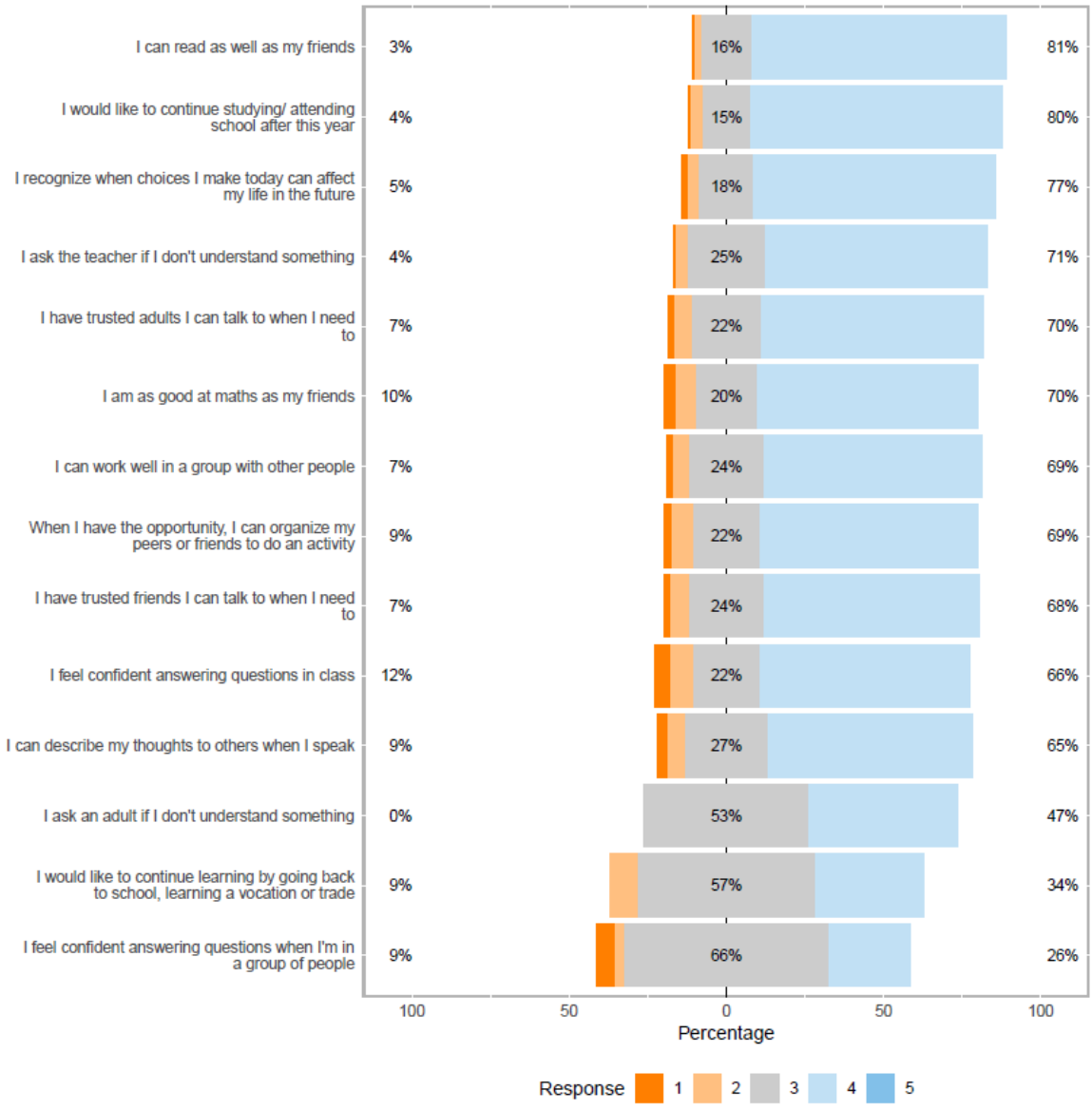
[268] The target has been set by adding 0.25 of the standard deviation of girls' self-esteem index scores, which was 0.16, in both the baseline and midline data to the achieved midline average score of 0.72.

	I am as good at maths as my friends.	Everyone
	I get nervous when I have to read in front of others.	In-school girls
	I get nervous when I have to speak in front of an adult.	Out-of-school girls
	I get nervous when I have to do maths in front of others.	In-school girls
	I get nervous when I have to speak in front of a group of people my age.	Out-of-school girls
	I feel confident answering questions in class.	In-school girls
	I feel confident answering questions when I'm in a group of people.	Out-of-school girls
	I would like to continue studying/ attending school after this year.	In-school girls
	I would like to continue learning by going back to school, learning a vocation or trade.	Out-of-school girls
	I recognize when choices I make today can affect my life in the future.	Everyone
	I can describe my thoughts to others when I speak.	Everyone
	I can work well in a group with other people.	Everyone
	When I have the opportunity, I can organize my peers or friends to do an activity.	Everyone
	I have trusted friends I can talk to when I need to.	Everyone
	I have trusted adults I can talk to when I need to.	Everyone
	I ask the teacher if I don't understand something.	In-school girls
	I ask an adult if I don't understand something (PROMPT, e.g., a community leader, parents).	Out-of-school girls
	Who mostly makes decisions about the following, or if this is in the future for you, who do you expect will make this decision?	
	Whether or not you will go to school.	In-school girls
Agency	Whether or not you can go back to school or vocational training.	Out-of-school girls
	Whether or not you will continue in school past this year.	In-school girls
	When/ at what age you will get married.	Everyone
	If you will work after you finish your studies.	In-school girls
	What type of work you will do.	Everyone

	How you spend your free time.	Everyone
	How often you spend time with your friends.	Everyone

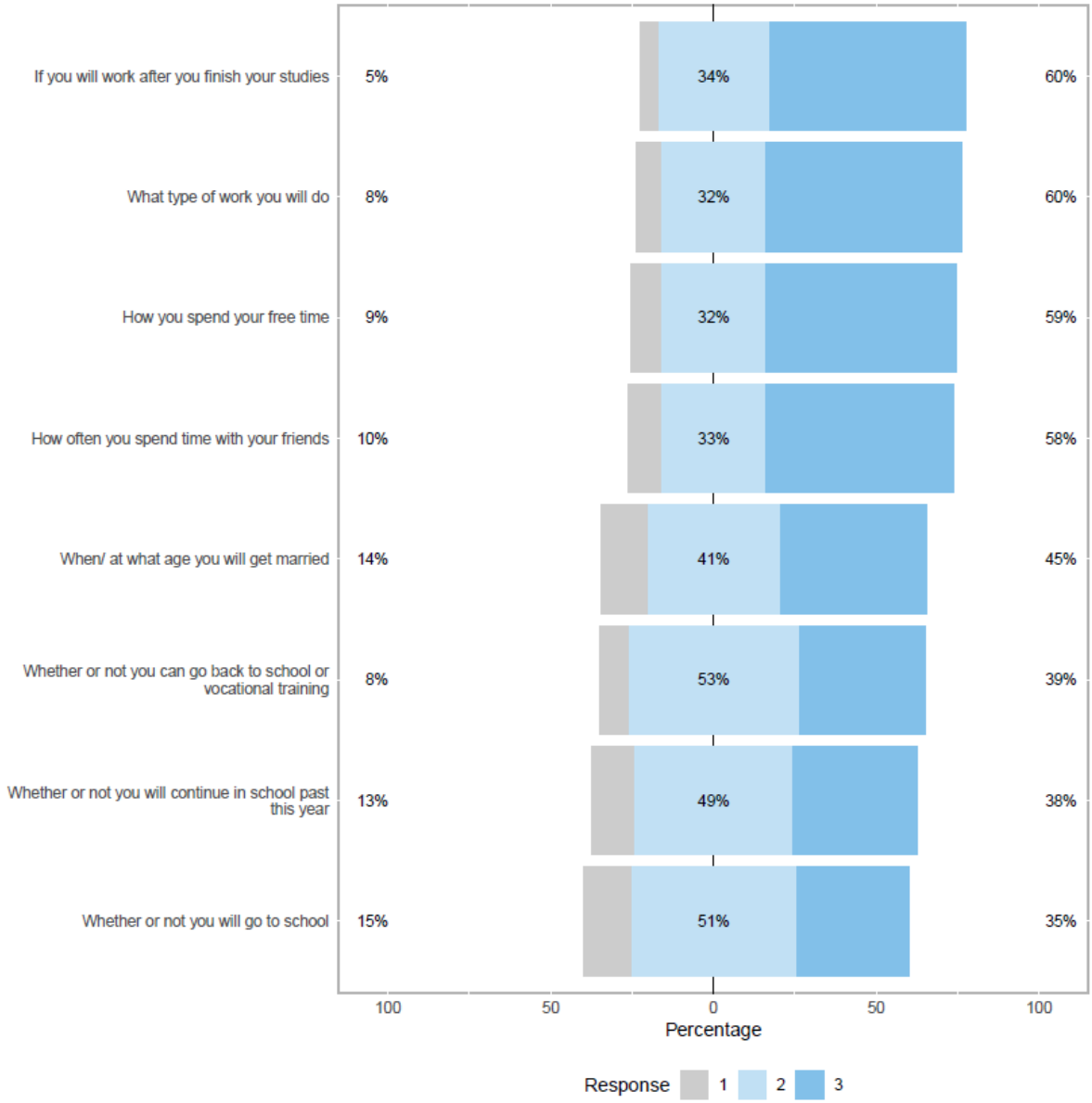
As shown in the table, some of these questions are common across the girls who attend school and those who do not while others were designed exclusively to the school status of each group. However, it is important to note that due to a lack of baseline data for the out-of-school girls and small sample size in the midline (with only 64 OOS girls re-contacted successfully from the baseline), the survey responses related to out-of-school girls' self-esteem are not discussed in the following analysis. The questions related to self-esteem have Likert-type responses ranking between "strongly agree" and "strongly disagree" with a score of 0 to 4, respectively, while questions on agency provide "I decide," "I decide jointly with my family," or "my family decides" as response options. However, in order to make the interpretation easier and more intuitive, the scores have been recoded so that an answer of 0 is the lowest self-esteem answer possible and 4 is the highest self-esteem answer possible. The following graphs summarize all Likert items that comprise the self-esteem index. For ease of analysis, these items have been sorted by their median. In the graph below, dark orange corresponds to strong disagreement, whereas dark-blue corresponds to strong agreement. No respondents strongly agreed with these items so the only colour of agreement shown is light-blue.

FIGURE 30: SUMMARY OF 5-POINT LIKERT ITEMS COMPRISING SELF-ESTEEM INDEX



The graph below presents all 3-point Likert items that also contribute to the self-esteem index. The colour scheme below ramps from the lowest level of autonomy and empowerment for girls to the highest level of autonomy, ranging from grey ("my family decides") to light blue ("I decide jointly with my family") to dark blue ("I decide").

FIGURE 31: SUMMARY OF 3-POINT ITEMS COMPRISING SELF-ESTEEM INDEX



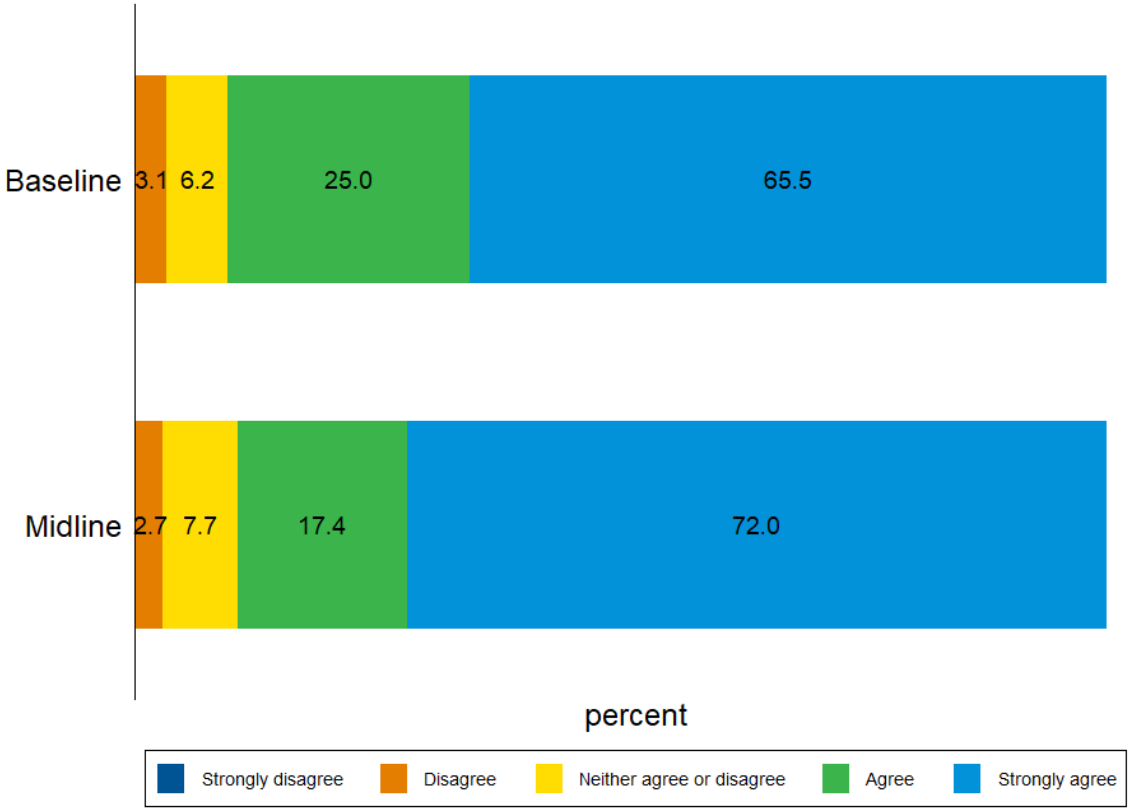
In this section, first, we analyse a primary indicator, selected for targeting by RI, that seeks to evaluate girls' and boys' self-confidence in leadership positions by asking them to rate the level of their agreement on the following statement: "When I have the opportunity, I can organize my peers or friends to do an activity." Next, a standardization method was adopted to create two index scores, "agency index" and "self-esteem index", where each girls received a score between 0 and 1. The first index only includes the agency related indicators while the second index encompasses agency as well as self-esteem indicators. The closer a girl's score is to 1, the higher self-esteem and agency she has, and vice versa. The results are disaggregated by a number of geographic, demographic and school-level correlates of attendance.

Primary Indicator – Girls' Leadership

Girls' willingness to take on leadership positions is the main indicator for evaluating girls' self-esteem and empowerment. The evaluation particularly focused on the peer organization aspect of girls' and boys' leadership skills, assessing whether girls had any willingness in organizing their peer group to achieve a goal or accomplish a task or an activity. More specifically, they were asked to express the level of their agreement or disagreement with the following statement: "When I have the opportunity, I can organize my peers or friends to do an activity."

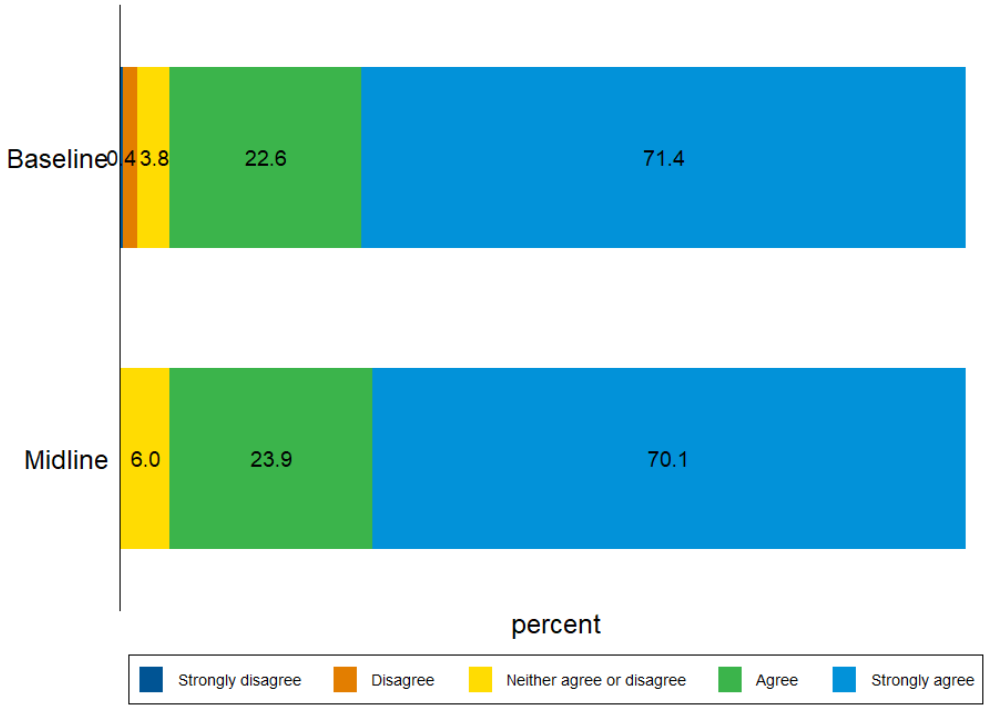
In general, the majority of girls both at baseline (65.5 percent) and midline (72.0 percent) have indicated high levels of confidence in their ability to organize their peer group. Data reveals a statistically insignificant increase from the baseline to midline in the share of girls who agreed with the statement, as shown in the figure below.

FIGURE 32: SHARE OF GIRLS EXPRESSING CONFIDENCE IN THEIR ABILITY TO ORGANIZE THEIR PEERS



Like girls, the majority of boys at baseline (71.4 percent) and midline (70.1 percent) strongly agreed that they can organize their peers to accomplish a work, but the result of regression shows a statistically insignificant decrease in their levels of confidence.

FIGURE 33: SHARE OF BOYS EXPRESSING CONFIDENCE IN THEIR ABILITY TO ORGANIZE THEIR PEERS



The table below lists the percentage of girls' responses to the leadership indicator disaggregated by subgroup. While there is no midline data for girls equal to or younger than 12 years of age, the girls older than 12 seems to have rated their confidence in leading activities less in the midline than baseline. Among zones, Galmudug appears to have the least number of girls with a high level of confidence (85.2 percent) when compared to Somaliland (94.1 percent) and Puntland (86.0 percent) in the midline. Similarly, secondary schools and schools in urban areas are identified as having a bigger drop in the number of girls who believe they have strong leadership skills from baseline to midline, with a percentage difference of 1.8 and 4.9, respectively. As shown in the table below, when examining girls' responses by IDP, conflict, and drought factors, it is observed that the number of girls who agree and strongly agree that they can organize their friends to do an activity has decreased across schools in the IDP areas and those that were affected by conflict and drought, whereas girls' confidence in the non-IDP schools and those that were not affected by conflict and drought either almost stayed the same or slightly increased.

TABLE 86: SHARE OF GIRLS EXPRESSING CONFIDENCE IN THEIR ABILITY TO ORGANIZE THEIR PEERS BY SUBGROUPS

Subgroup	Round	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Overall	Baseline	0.2	3.1	6.2	25.0	65.5

	Midline	0.2	2.7	7.7	17.4	72,0
Age <=12	Baseline	-	6.8	6.1	25	62.1
	Midline	-	-	-	-	-
Age >12	Baseline	0.2	2.6	6.3	25.0	66.0
	Midline	0.2	2.7	7.7	17.4	72.0
Somaliland	Baseline	0.5	2.4	8.7	17.6	70.8
	Midline	0.2	2.4	3.1	14.5	79.8
Puntland	Baseline	-	4.3	5.3	26.1	64.3
	Midline	0.2	3.2	10.5	21.8	64.3
Galmudug	Baseline	-	1.0	1.0	52.0	46.1
	Midline	-	2.0	12.8	8.8	76.5
Rural schools	Baseline	0.3	3.1	7.3	24. 8	64.5
	Midline	.	0.9	4.3	18.4	76.5
Urban schools	Baseline	0.2	3.1	5.7	25.1	65.9
	Midline	0.3	3.6	9.4	16.9	69.8
Primary schools	Baseline	0.1	3.1	6.5	25.0	65.3
	Midline	0.3	3.3	7.0	16.8	72.6
Secondary schools	Baseline	0.4	3.3	5.4	24.9	66.0
	Midline	-	0.8	10.0	19.1	70.1
IDP schools	Baseline	-	2.0	8.2	20.4	69.4
	Midline	2.0	8.2	16.3	8.2	65.3
Non-IDP schools	Baseline	0.2	3.2	6.1	25.2	65.3
	Midline	0.1	2.4	7.3	17.9	72.3
Conflict-affected schools	Baseline	-	3.8	3.1	33.1	59.8
	Midline	-	8.2	6.5	22.9	62.3
Non-conflict affected schools	Baseline	0.2	2.9	6.7	23.4	66.5
	Midline	0.2	2.3	7.8	17.0	72.6
Drought-affected schools	Baseline	0.3	2.1	2.1	30.4	65.0
	Midline	-	1.0	8.6	27.5	62.7

Non-drought affected schools	Baseline	0.1	3.0	8.0	21.8	66.8
	Midline	0.2	3.0	7.5	15.0	74.11

Furthermore, when the effects of school, geographic and demographic factors are controlled in the logistic regression model, the percentage of girls who scored their leadership skills high in the midline is insignificantly higher compared to those in the baseline.

On the other hand, boys' overall confidence in their leadership ability doesn't seem to have changed considerably overtime when the "strongly agree" and "agree" responses are examined together. As shown in the table below, boys residing in Puntland have scored their leadership skills slightly higher than the boys in Somaliland and Galmudug. More specifically, boys in Puntland have had a 3.1 percent increase in the number of people who agreed to the statement, while a 10 percent decrease was observed among boys in Galmudug. When data is further disaggregated by school level, IDP status, and conflict and drought affected areas, the only increase in levels of leadership skills were detected among boys in the secondary schools as well as those located in non-drought affected areas while boys' confidence have decreased or did not change significantly across other schools.

TABLE 87: SHARE OF BOYS EXPRESSING CONFIDENCE IN THEIR ABILITY TO ORGANIZE THEIR PEERS BY SUBGROUPS

Subgroup	Round	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Overall	Baseline	0.4	1.7	3.8	22.6	71.4
	Midline			6.0	23.9	70.1
Age <=12	Baseline			8.3	20.8	70.8
	Midline					
Age >12	Baseline	0.5	1.9	3.3	22.9	71.4
	Midline			6.0	23.0	70.1
Somaliland	Baseline	1.0	1.0	4.0	12.0	82.0
	Midline			6.0	16.0	78.0
Puntland	Baseline		2.8	4.7	29.2	63.2
	Midline			4.7	34.9	60.4
Galmudug	Baseline				37.0	63.0
	Midline			11.1	7.4	81.5
Rural schools	Baseline	1.3	1.3	3.8	11.4	82.3
	Midline			6.3	31.6	62.0
Urban schools	Baseline		1.9	3.9	28.4	65.8

	Midline			5.8	20.0	74.2
Primary schools	Baseline	0.6	1.2	2.3	26.6	69.4
	Midline			6.4	24.9	68.8
Secondary schools	Baseline		3.3	8.2	11.5	77.0
	Midline			4.9	21.3	78.8
IDP schools	Baseline		1.7	5.0	31.7	61.7
	Midline			15.4		84.6
Non-IDP schools	Baseline				38.5	61.5
	Midline			5.4	25.3	69.2
Conflict-affected schools	Baseline		3.2	6.5	32.3	58.1
	Midline				45.5	54.5
Non-conflict affected schools	Baseline	0.5	1.5	3.4	21.2	73.4
	Midline			6.3	22.9	70.9
Drought-affected schools	Baseline	1.3	1.3	1.3	20.3	75.9
	Midline			8.7	45.7	45.7
Non-drought affected schools	Baseline		2.2	5.2	23.7	68.9
	Midline			5.3	18.6	76.1

Agency Index

Before turning into the discussion of indexes, it is important to mention how these indexes were constructed in the first place. We employed the standardization method for creating the indexes because the self-esteem and agency indicators have different scales and some of them were measured in different directions. In order to standardize them, the indicators were, first, re-coded to become consistent in terms of direction. Next, the responses were divided by the standard deviation of the variable where the result is a variable with a mean of 0 and a standard deviation of 1:

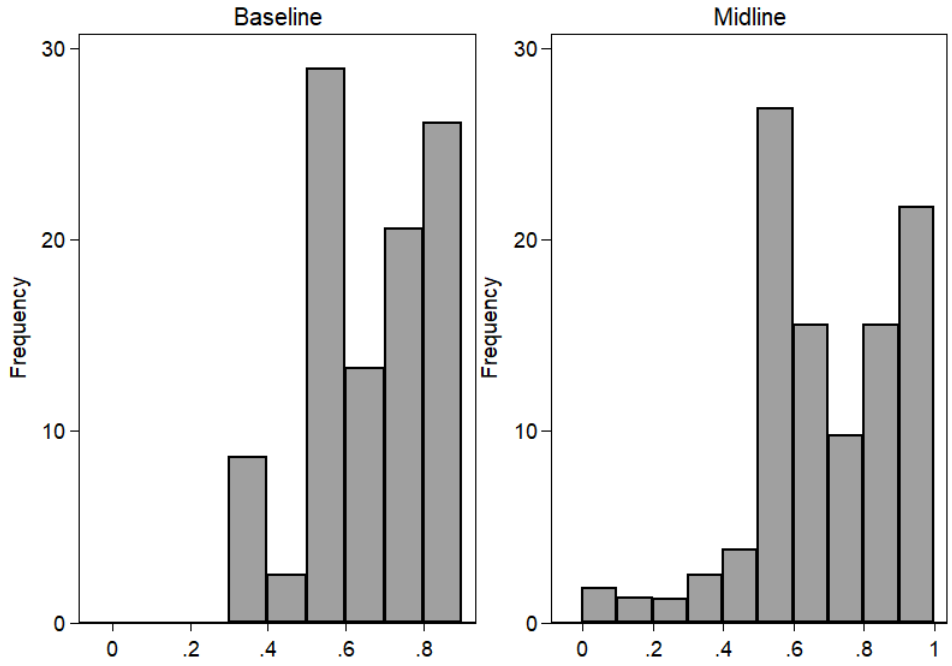
$$ystd_i = (y_i - \mu_y)/(\sigma_y)$$

Then these variables were added up to create a single index, using the following formula.

$$Rescaled\ X_i = \frac{X_i - Min_x}{Max_x\ -\ Min_x}$$

The agency index is developed by combining all the agency related indicators in order to understand how girls' decision-making power have improved as the result of taking part in the girls' clubs.[269] In particular, we examine to see if there has been any positive changes from baseline to midline in girls' ability in making decisions on different aspects of their lives, such as decisions on whether or not to go to school, when to get married, and how to spend their free time and time with friends, etc.

FIGURE 34: HISTOGRAM OF IN-SCHOOL GIRLS' AGENCY INDEX SCORES



Data indicates an overall significant improvement overtime in girls' agency with an index mean of 0.66 at baseline[270] to 0.69 at midline[271]. Yet, when the effects of demographic and zonal characteristics are controlled, the extent of improvements becomes insignificant. In addition, comparing the self-agency index scores of girls who participated in girls' club activities with girls who did not participate in girls' club activities did not yield significant differences.

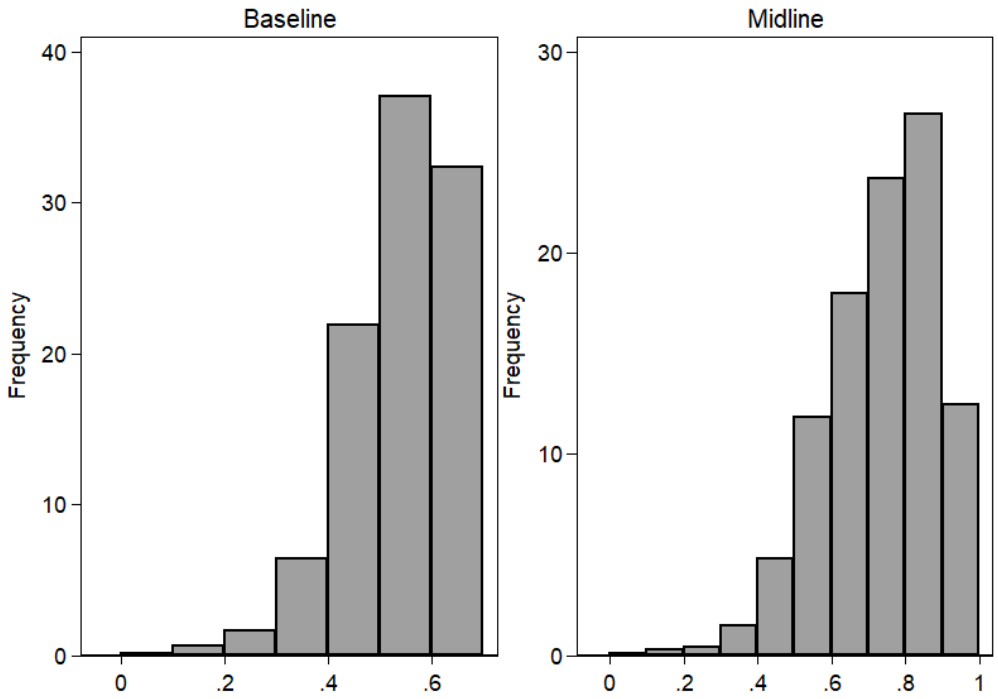
[269] The agency indicators did not have to be standardized (since they are on the same scale), but they are standardized in this case to make the agency index comparable with the self-esteem and agency index.
[270] Girls' baseline index scores have a skewness and kurtosis of -0.7 and 2.65, respectively.
[271] Girls' midline index scores have a skewness and kurtosis of -0.5 and 3.34, respectively.

Next, changes in girls' self-confidence is even further analysed by combining the self-esteem and agency indicators into one index.

Self-Esteem Index

As stated above, girls were asked various types of questions to capture the multidimensional nature of self-esteem. While in-school girls' confidence in organizing their peers seems to have insignificantly improved, girls' index scores show a significant improvement in their self-esteem skills. The histogram below compares girls' score index across the baseline and midline. The girls in the baseline had an average index score of 0.51 with 50 percent of them having a score range between 0.45 and 0.58.[272] The girls in the midline scored their self-esteem significantly higher with an average score of 0.72, and with half of the girls' scores falling between 0.62 and 0.84.[273] When we controlled for girls' demographic and zonal characteristics such as age, grade, and zone, girls' scores still show a significant improvement from the baseline to midline.

FIGURE 35: HISTOGRAM OF IN-SCHOOL GIRLS' INDEX SCORES



[272] Girls' baseline index scores have a skewness and kurtosis of -1.09 and 4.77, respectively.
[273] Girls' midline index scores have a skewness and kurtosis of -0.72 and 3.58, respectively.

In order to better understand why girls' self-esteem index scores have improved significantly, we asked girls through focus groups discussions to share their understanding on what makes a girl have confidence or high self-esteem and what factors they think contribute to girls' self-esteem. Qualitative interview respondents described confident girls as not being afraid to ask questions or present their ideas, whether they are considered right or wrong. Confident girls "always lead,"[274] "help other students,"[275] and are resilient in that they "avoid accepting negative thoughts"[276] from others. Respondents alluded to both internal and external sources of confidence and self-esteem in girls. Some believe that confidence is innate, that "confidence is born to some people while in others it is not,"[277] and that it is evidenced through "personal efforts, self-assuredness, and [her] willingness to study."[278] As one boy described, "Girls must use self-motivation too. She will be fine in education if she tells herself that she can do anything."[279]

However, discussions of internal sources of motivation were always coupled with discussions on the importance of external sources of motivation. Externally, girls draw confidence through the encouragement of their parents/families, their teachers, and their community, as well as the influence of their friends. As one girl describes, "Regardless of how confident she is, if her parents say, 'You are a bad [student] and can't reach your goals,' she will become disappointed and start losing faith in her educational goals. But if her parents motivate her, she will continue learning."[280]

In the quantitative data, there is a significant correlation between having participated in activities with the girls club and greater levels of self-esteem. Indeed, girls who participated in activities with girls club are predicted to have a 0.04 higher self-esteem index score than girls who did not.[281] However, since participation in girls club activities was not randomly assigned, we are not able to say with certainty whether the girls who participate in girls clubs activities have a higher level of self-esteem as a result of their participation or whether the girls who participate in girls clubs activities had a higher level of esteem prior to participation in girls clubs.

Improvements in teaching quality through CPD is another mechanism that may have contributed to increased self-esteem from baseline to midline. If teachers are using more student-centred teaching techniques that reinforce girls' senses of self-efficacy and their confidence in the classroom, this may also be reflected in the positive self-esteem findings here. In order to test this hypothesis, we have merged teacher-level data (on which teachers received CPD training) with learner data on self-esteem and have used difference-in-differences estimation to determine whether learners with CPD-trained teachers were more likely to have higher self-esteem. Learners with CPD-trained teachers are significantly more likely to have higher self-esteem, suggesting that while teacher training has not yet translated into detectable

[274] FGD, Girls, Puntland
[275] FGD, Girls, Puntland
[276] FGD, Girls, Puntland
[277] FGD, Girls, Puntland
[278] FGD, Girls, Puntland
[279] FGD, Boys, Puntland
[280] FGD, Girls, Puntland
[281] The P-value is 0.001 in a cluster robust linear regression.

impacts on learning, there has been a positive impact of teacher training in terms of girls' social and personal development, and these increases in self-esteem may translate into better learning and transition outcomes in the future.[282]

Nevertheless, there was ample evidence in the qualitative interviews that there is a causal link between EGEP-T project activities and increased confidence in girls. First, exposure to awareness-raising around the importance of girls' education appears to be an important source of positive inspiration for girls: "When the girls hear there are NGOs supporting them, they feel confident and will prefer to continue learning because they feel excited to continue learning since they received support from ADRA. I believe if the girls get more support, they will excite learning more than now."[283]

Second, girls appear to be internalizing this encouragement, as many mentioned that exposure to education and life experiences is an importance source of confidence for them. When asked what contributes to girls' confidence, one girl explained, "Girls have received more confidence since they understood the value of learning. You can see girls continuing their education and completing it – girls joining universities. You can also see girls getting skills and leading classes."[284] Another explained that "you have to search for more knowledge – the more knowledge you get, the more you will be a good person."[285] It appears that this change is fairly new and that progress is being built on each year: "Girls have more confidence than ever because they answer questions and everything is changing every new year."[286]

Third, when asked what contributes to girls' confidence, many girls mentioned the importance of interactive class activities, such as presenting in front of the class, in building their confidence. One girl explains, "I really like the interactive activities of the students. Furthermore, I like to stand and present something in front of the students."[287] Another explains the transformative effect exposure to these types of activities and continued practice have had on her: "For example, I was not such confident person previously. I remember that it took me two days to prepare and present the lesson in front of the class. Unfortunately, I forgot all the preparation when I came in front of the class. I surprised myself with how I failed to present in front of the class. I collected all the lesson information even more than the teacher's explanation and it took me more nights to prepare and it costs me to ask my brothers. Finally, I decided to overcome shyness after that day."[288]

In summary, while boys' levels of confidence in their leadership skills appear to have decreased, it was observed that girls feel more confident both in their ability to lead activities and across other agency and self-esteem indicators (agency and self-esteem index score). The changes in girls' confidence from the baseline to midline were more evident when their responses were examined altogether. In particular,

[282] The P-value is 0.015 in a cluster-robust linear regression.
[283] FGD, CECs, Somaliland
[284] FGD, Girls, Puntland
[285] FGD, Girls, Puntland
[286] FGD, Girls, Somaliland
[287] FGD, Girls, Somaliland
[288] FGD, Girls, Puntland

improvement in girls' self-esteem were statistically significant even after controlling for girls' demographic and zonal characteristics such as age, grade, and zone. In the qualitative interviews and FGDs, girls confirmed that they feel confident in their skills and that their self-esteem has improved, particularly as a result of their participation in the EGEP-T project activities. More specifically, they believed that exposure to awareness-raising around the importance of girls' education, education and life experiences, as well as interactive class activities have greatly enhanced their confidence.

7.3 Teaching Quality

TABLE 88: INTERMEDIATE TEACHING QUALITY OUTCOME INDICATORS, FOR FULL SAMPLE OF TEACHERS

IO	IO indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Teaching Quality	Percentage of teachers demonstrating improved teaching practices (Index)	67.3	70.3	63.9	N	73.3	Y
Teaching Quality	Percentage of CPD-trained teachers demonstrating improved teaching practices (Index)	60.4	63.4	67.2	Y	66.4	Y
Main qualitative findings							
• Both girl and boy students report positive changes in teaching techniques used in their own classrooms.							

TABLE 89: TEACHING QUALITY RESULTS BY ZONE (INDEX SCORES), AMONG COMPARABLE SAMPLE

Round of Data Collection	Somaliland	Puntland	Galmudug
Baseline	60.7	71.6	78.2
Midline	55.0	75.8	32.3

Improvement in teaching quality is an intermediate outcome of central importance given the results of the barriers analysis above. Data from both caregivers as well as children indicate that lower teaching quality and low-quality school administration are likely contributing to lower learning outcomes among cohort girls. Teaching quality ultimately affects not only girls' learning, but also their potential to successfully transition to secondary school and post-secondary school. Girls are likely to learn more when

their teachers have teaching methods and practices that are high quality. High quality teaching may also be an incentive for girls to stay in school as they will tend to feel that they are learning and that school is worth their time and money. However, the opposite is also true as well – girls may be more inclined to drop out of school if their teachers are often absent or not prepared, and girls will not learn as much if their teachers are not engaging both boys and girls or have a poor teaching approach.

Teaching Practices

In keeping with the baseline, at midline the primary indicator of teaching quality is an index of teaching practices observed during classroom observations.[289] Fifteen different behaviours were observed for teaching approaches during the classroom observations. These behaviours included participation, referring to previous lessons, and helping students when they did not understand. As at baseline, an index was created on the basis of these fifteen behaviours by assigning every positive behaviour a 0 for not observed and 1 for observed and every negative behaviour a 0 for observed and 1 for not observed. Thus, the combined index is on an intuitive scale ranging from zero to 100, where a score of zero indicates the lowest level of teaching quality theoretically possible, and a score of 100 indicates the highest level of teaching quality theoretically possible.

The analysis of teaching practices is divided into two sub-sections below, which focus on different samples of teachers and different analytical approaches. In the results that follow immediately below, we report results for the full sample of comparable teachers, i.e. the schools that appear in both BL and ML, without replacement schools or replaced schools. This data cannot be considered a panel (in the sense of learning data) because the teachers observed varied from round to round in many cases.[290] The total sample for analysis is thus 205 school-level observations that are matched across baseline and midline. In a separate section, we report analysis focused on CPD teachers and the impact of the CPD intervention on teaching practices. This latter analysis arguably fits more closely with the project's interventions, to the extent that improvements in teaching practices are expected to be concentrated among CPD teachers. Briefly, the analysis in this second section compares a panel of teachers tracked from baseline to midline, comparing changes in teaching practices between teachers exposed to CPD training and those who were not exposed. The outcome studied in both sections remains the same – an index of teaching practices that measures teaching quality – but the reporting is separated for the sake of clarity.

This was noted at baseline, but it is important to remind readers here that the contents of this index are focused primarily on the approach to pedagogy taken by each teacher. Rather than focus on whether students are subject to corporal punishment, for instance, the index asks whether the teacher reprimands students for incorrect answers, with the understanding that reprimanding students in this way discourages future participation in the classroom. Similarly, the index captures the degree to which

[289] Observations were 3 blocks of approximately 15 minutes.

[290] However, the sample is comparable in a cross-sectional sense, insofar as the same schools are included in the sample in each round.

teachers offer alternative explanations to students when there are concepts that they do not understand, a measure of their willingness to tailor the lesson to their students' needs.

It should also be noted that the sustainability indicator related to teacher training and teaching quality uses predominantly qualitative data to establish the degree to which TTIs have adequately integrated components of the CPD approach into their curriculum. These qualitative data from the sustainability analysis present a comparatively positive finding of improvement over time in the degree of integration achieved. On the other hand, the quantitative analysis of teaching practices comprising the index below suggests that no appreciable progress has been made in terms of improving teaching practices in the classroom – at least in the aggregate. While these two results may appear to be contradictory, it is plausible that progress is being made at the level of TTIs in terms of their integration while this progress has not yet translated into measurable improvements in terms of specific practices that teachers employ in the classroom.

Overall Teaching Practices

In this section, we report the main results for teaching quality, based on an index of teaching quality. As noted above, this first set of results focuses on the full sample of comparable teachers, not a true panel, between baseline and midline. All teachers observed at comparable schools – excluding those schools replaced at midline – are included. The table below summarizes the individual indicators that contributed to the teaching quality index, along with the proportions of cases in which a given behaviour was observed at baseline and at midline.

TABLE 90: INDIVIDUAL TEACHING PRACTICES OBSERVED, FULL TEACHER SAMPLE, BY ROUND

Type of Behaviour	Question	Observed – Baseline	Observed – Midline
Positive	Is the teacher confident in their presentation of the material?	93.2%	90.2%
Negative	Students spent most of the time copying from the board.	31.7%	15.6%
Negative	Students spent most of the time repeating teacher's words aloud.	11.2%	11.2%
Positive	Teacher used student-centred activities or games.	47.8%	38.5%
Positive	Teacher allowed students to instruct each other (e.g. come to board to demonstrate something, or explain to classmate).	65.9%	62.0%
Positive	Teacher asked open-ended question (requires more than simple answer) that encourages thinking.	84.9%	75.6%
Positive	After a student gave an incorrect answer, did the teacher explain the concept in a new way?	85.4%	82.4%

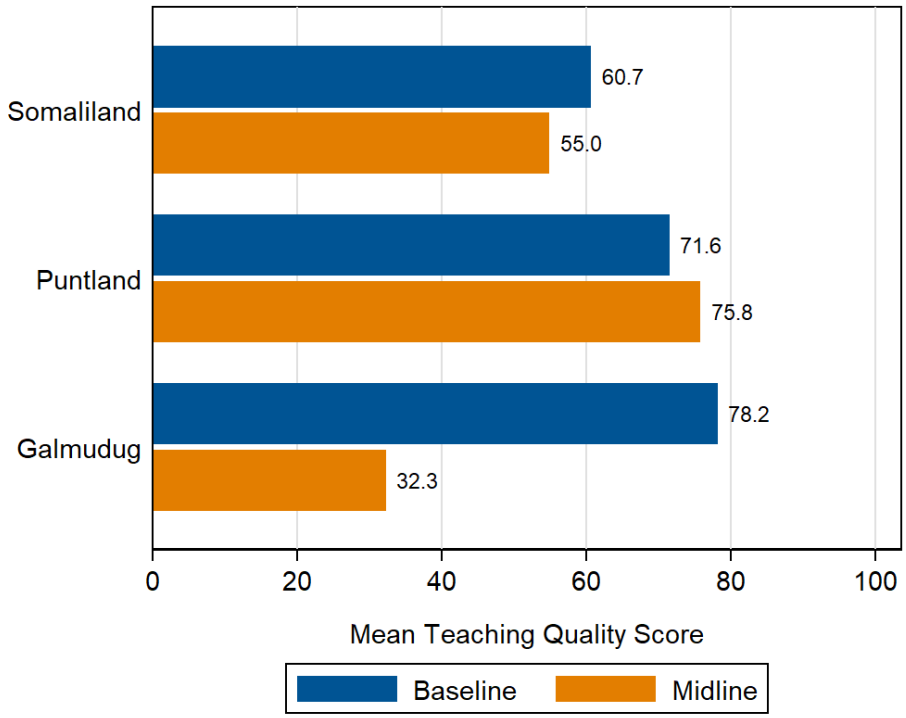
Positive	After correcting a student who gave an incorrect answer, did the teacher verify the student understood the question now?	79.5%	79.5%
Negative	If a student gives an incorrect answer, are they reprimanded (verbally or physically)?	71.2%	88.8%
Positive	Teacher called on or actively tried to involve a student who was not participating.	84.4%	78.5%
Positive	Students worked together in groups.	44.9%	36.6%
Positive	The teacher summarized and clearly stated a key concept or takeaway point from the lesson.	83.9%	81.5%
Positive	The teacher referred back to previous lessons, relating this lesson to previous lessons.	73.2%	74.6%
Positive	The teacher stopped the lesson and invited questions from students.	77.1%	69.3%
Positive	The teacher employs a variety of explanations that differ in difficulty for the diverse learners in the classroom.	75.6%	74.1%

In the aggregate, the average teaching quality score across all midline schools was 63.9, while the baseline average was 67.3. Thus, the aggregate teaching quality score decreased by a total of 3.4 percentage points from baseline to midline, which might suggest a slight decrease in teaching quality across sampled schools over time. However, this difference between baseline and midline is ultimately not statistically significant and thus we cannot infer that teaching quality has actually decreased.

Ultimately, the only conclusion supported by this analysis is that overall teaching quality has not improved over time, and this finding is consonant with our analysis of learning outcomes above, which suggest that there has not been a measurable positive effect of the intervention on learning outcomes, where improvement in teaching quality is one of the most direct mechanisms through which the project might improve learning outcomes. This is an issue which we address further in the next section, which investigates the project's impact on a sample restricted to CPD teachers.

The graph below disaggregates teaching quality scores by zone as well as allowing for comparison by baseline versus midline scores. As expected, the results show minimal differences between baseline and midline for Somaliland and for Puntland. The results for Galmudug are starkly different, but these results are for a comparatively small sample of schools and also are particularly sensitive to issues of inter-rater reliability (from baseline to midline) because one, single researcher carried out all the baseline observations for Galmudug while a different, single researcher carried out all of the midline observations, meaning that (despite intensive training) each researcher might have had different qualitative standards for how they coded a set of indicators, and these differences would have been amplified by the small sample size and by the fact that they were not averaged across multiple researchers or raters.

FIGURE 36: CHANGE IN MEAN TEACHING QUALITY INDEX, BY ZONE

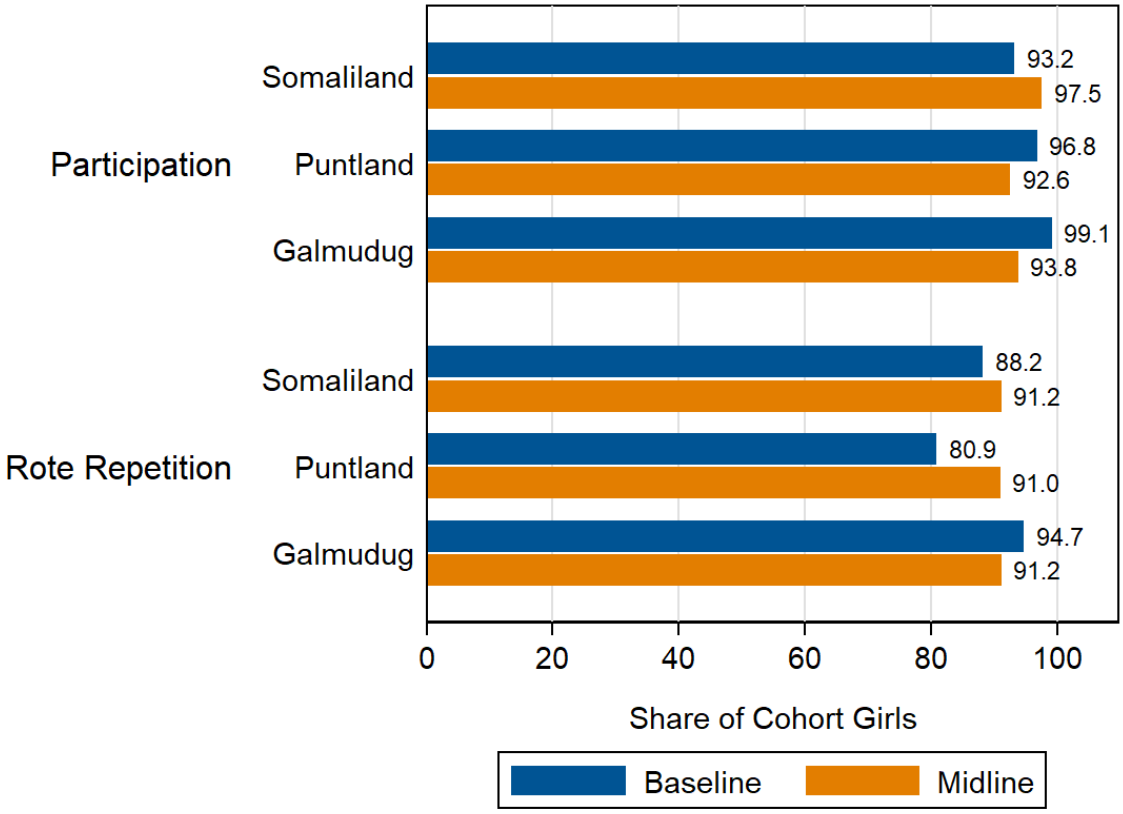


In order to take a second look at zone-level differences, the graph below is presented in order to allow for triangulation of teaching practices on the basis of girls' direct reports of the practices used by their teachers in the classroom. The graph is based on data from the true panel sample of cohort girls only in order to maximize comparability between baseline and midline. The graph presents the share of girls, at a given round of data collection in a given zone, who say their teacher encourages participation often or sometimes (a positive teaching practice) and the share who agree or agree strongly that they spend much of their day repeating what their teacher says (a negative teaching practice). First, it is important to note here that zone-wise differences are exceedingly small and non-systematic, as are differences by round within a given zone. Galmudug does not appear particularly different from the other zones, which suggests that the result above is indeed driven more by differences in raters between rounds than by objective differences in teacher performance between rounds.

It is also worth noting in the graph below that positive, participatory and negative, rote-learning teaching practices both appear to be quite prevalent according to girls' self-reports. Intuition would suggest that these two practices (one typically associated with good teaching and the other typically associated with bad teaching) would be negatively correlated with each other, but this is not the case. A likely explanation for this finding is the fact that much rote learning is done through whole-class, unison recitation of learning points which are written on the blackboard. From a the average learner's perspective this unison recitation of material written on the board may seem participatory (inasmuch as everyone in the classroom is participating in saying things out loud), especially in cases where more truly participatory

methods (such as soliciting individual questions and individualized participation) are so rare that learners have no basis for comparison.

FIGURE 37: STUDENT-REPORTED TEACHING PRACTICES, BY ZONE AND ROUND



However, in the qualitative data, girl and boy students do report positive changes in teaching techniques used. In particular, students report that teachers repeat the lessons, ask questions to gauge student understanding of the content of lessons, arrange teamwork/group discussions, conduct competitions "which makes the lesson less boring,"[291] and asking questions about previous lessons to ensure students have understood them before moving on. However, this really seems to depend on the teacher, as "teachers do not have the same level in education"[292] and "you may not understand some teachers' teaching style at all," whereas "some teachers with small knowledge have great ability to explain and make it simple to understand."[293] One head teacher explains that teachers who have received training are better equipped to apply creative, innovative teaching methods: "As I told you, the teachers are not the same; there are teachers who exceed the expected teacher responsibility quality while some of them fail.

[291] FGD, Girls, Puntland
[292] FGD, Girls, Puntland
[293] FGD, Girls, Puntland

Those who exceed the expected quality are those who received the training to the teachers. There are innovative teachers that create new things and they also motivate the school children."[294]

In the qualitative interviews, teachers outlined a number of challenges they face in the classroom, some of which may be outside the scope of the project to address. One teacher describes how teachers struggle with lack of subject matter knowledge, lack of motivation, and high workloads: "Teachers' challenges include poor knowledge -- the teacher may not have enough knowledge about what he is going to teach). He does not have anywhere to deal this problem, so this is a real challenge in my opinion. Lack of motivation is another factor, since teachers need regular work-related motivations to be active and efficient. In addition, the teacher provides six consecutive periods without resting, so it is heavy. Sometimes, it happens that teachers hurry (do fast teaching of the syllabus) without respecting the standard."[295] Another teacher describes how it is challenging for teachers to manage the high student to teacher ratio: "One of the challenges is the time number of students is more than 70 students and sitting in one class. It's challenging for teachers to manage the students sitting in the class since their number is more than the average number."[296]

Nonetheless, some teachers do describe implementing teaching practices that are consistent with what we would expect to see as a result of the project activities. For example, one teacher describes using a question and answer method to assess knowledge of previous methods and assigning students with presentation of methods to enforce learning of new concepts: "The behaviours I have changed include; when I am presenting the lesson, I make connections between the previous lesson and the current one and ask questions related to previous lessons to the students. I also request them to read previous lessons without books. In addition, I give homework assignments to them to present in front of the students by providing them with awards to be motivated. In other words, I always provide consultation to students."[297] Another explains how she keeps students alert when managing large class sizes: "In a big class, I would say that you prepare the lesson - after that when class starts, the first thing you do is to ask questions for the whole class. Students will be busy on answering these questions. Each student will think that if you talk, then a question is coming right to you. Then in that way we start the lesson."[298] Teachers also describe providing special support to students who are struggling, as well as repeating lessons for their classes when a lesson is not well understood: "If today the lesson is not understood, or one girl is missing today, I try to repeat the lesson."[299]

Taken as a whole, the analysis of teaching practices above suggests that the intervention has not had a measurable positive effect on teaching quality measured in terms of teaching practices in the classroom, at least among the wider sample of all teachers. The qualitative data suggests there have been notable improvements made, but these changes are not uniform and are instead dependent on the teacher in

[294] KII, Head Teacher, Galmudug
[295] KII, Male Teacher, Galmudug
[296] KII, Male Teacher, Puntland
[297] KII, Male Teacher, Galmudug
[298] KII, Female Teacher, Puntland
[299] KII, Female Teacher, Puntland

question. Given that the primary vector by which teaching practices should improve is via CPD training, and not all teachers were targeted for CPD training, it is possible that teaching practices were improved among the CPD-trained subset of teachers. We analyse this possibility further in the subsection below.

Teaching Practices Among CPD Teachers

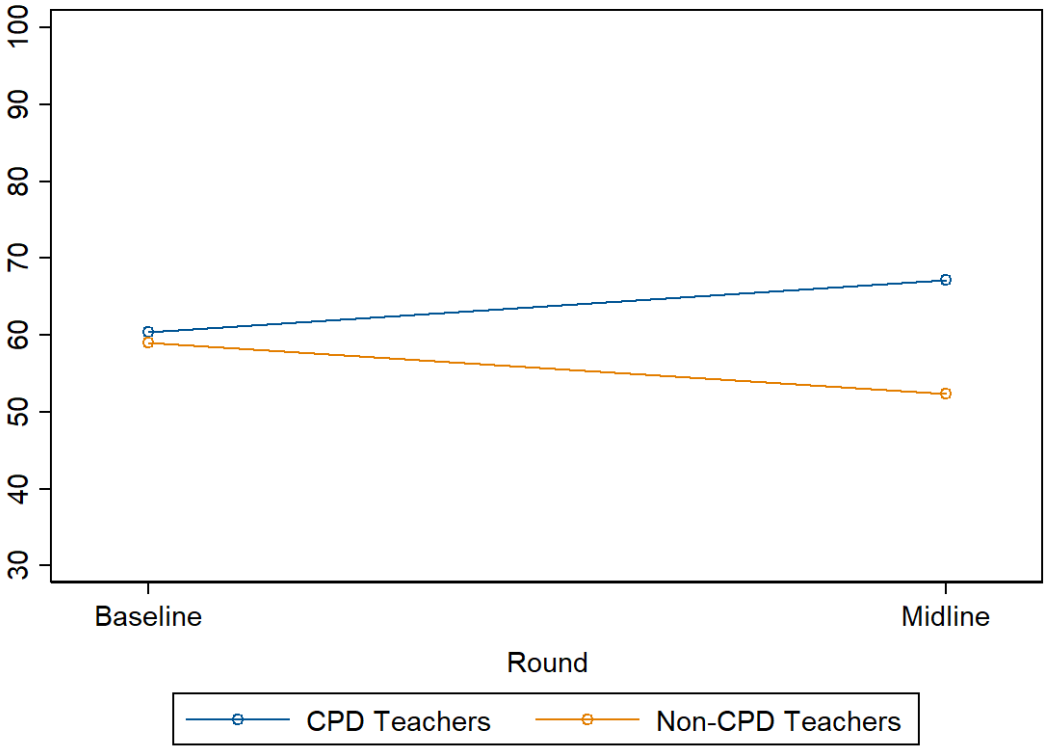
The analysis above suggested that teaching practices have not improved appreciably since the baseline. However, because Continuing Professional Development (CPD) is an important intervention intended to improve teaching quality, this section investigates the impact of CPD training, specifically, on teaching practices. To do this, we analysed the differences in teaching practices between teachers who participated in CPD training and those who did not. We used two complementary approaches to this question. Our first approach relied on the reports of head teachers to identify teachers who had participated or were participating in the project's CPD training. Of the 149 teachers for whom this information was captured at midline, 71.1 percent were identified as "CPD teachers". If CPD training has had a significant and positive effect on the teaching abilities of CPD teachers, one sign of this might be that CPD teachers at midline would have higher teaching quality index scores than non-CPD teachers at midline. A statistical test of this hypothesis reveals that CPD status is not positively or significantly correlated with teaching quality (the coefficient is negative and exceedingly close to zero).[300] Thus, on the basis of this test, there is no measurable positive intervention effect of CPD training on teaching quality.

A shortcoming of this analysis is that it assumes CPD teachers and non-CPD teachers in the midline sample began from identical starting points. In other words, it is possible CPD training improved teaching practices, but that teachers engaged in the CPD programme had *worse* teaching practices prior to training than their comparison group counterparts; in such a case, our analysis would obscure the positive effect of CPD training. In addition, the analysis is heavily dependent on the ability of head teachers to accurately identify CPD teachers in their schools, and does not account for varying levels of exposure to training (some teachers may have been selected for CPD training, but completed only small portions thus far).

To remedy these shortcomings, we performed a second test of the impact of CPD training, by studying a panel of teachers whose classrooms were observed during both the baseline and midline. These teachers allow a more rigorous assessment, because we can analyse impact using a difference-in-differences model. In total, 43 teachers were observed in both rounds; unfortunately, nearly all of these teachers (n = 36, or 83.7 percent) were exposed to CPD training, leaving very few teachers in our "comparison group". The results described below should be interpreted with that shortcoming in mind.

[300] In a regression with cluster-robust standard errors, using CPD status to predict index score, the coefficient is -0.49 with P = 0.913.

FIGURE 38: IMPACT OF CPD TRAINING ON TEACHING QUALITY AND PRACTICES



The results, reported in the figure above, are strongly suggestive, but not conclusive. CPD and non-CPD teachers in the panel scored similarly at baseline, separated by just 1.4 points on the 100-point index. By the time of the midline, scores among non-CPD teachers in the panel sample (n = 7) had declined 6.6 points, while those among CPD teachers had *increased* by 6.8 points. Our best estimate, based on this analysis, is that CPD training results in a 13.5 percentage point increase over and above the untrained group of teachers.[301] It is important to note that this finding is not statistically significant (p = 0.13) at conventional levels; however, given the small number of teachers in the panel, these results should be considered tentative evidence of the project's impact on teaching quality.

As noted in the previous section, overall teaching quality has declined from baseline to midline in the full sample of teachers. This is not true of CPD teachers, as shown in the graph above, who have made small net improvements since baseline. While performance on the index declined from 67.3 to 63.9 over time, CPD teachers' performance improved from an average of 60.4 to 67.2 on this same 100-point scale (n = 36). And, while our earlier analysis showed that small overall gains in Puntland were offset by small overall

[301] These improvements are driven most by changes in four areas: teachers actively seeking participation from non-engaged students; teachers reducing the time students spend copying from the board or repeating the teacher's words; teachers prompting students to teach each other; and teachers making reference to previous lessons, connecting ideas across sessions.

declines in Somaliland, this is not the case for CPD teachers, who show improvement in both regions.[302] Results, restricted to CPD teachers who were tracked over time from baseline to midline, disaggregated by zone, are reported in the table below.

TABLE 91: TEACHING QUALITY INDEX SCORES, BY ROUND AND ZONE, AMONG PANEL OF CPD TEACHERS

Round of Data Collection	Somaliland (n = 18)	Puntland (n = 18)	Galmudug (n = 0)	Overall
Baseline	47.0	73.7	N/A	60.4
Midline	51.1	83.3	N/A	67.2

The observed gains in teaching quality, on this measure, from baseline to midline cannot be explained by changes in the sample, because the results in the table above rely on a panel of teachers who were tracked over time (n = 36). The results are consistent with our expectations regarding the project's impact: improvements in teaching quality are found among those teachers who were exposed to CPD training, but less pronounced or non-existent among the broader population of teachers analysed in the previous section.

Pedagogical improvements are not uniform across all indicators, however. As with the full sample of teachers, changes were observed in some areas of teaching practice but not others. The table below summarizes the baseline and midline frequency for each of the constituent sub-indicators of the index, e.g., the share of classrooms in which students spent most of their time copying from the board, the share of classrooms where the teacher appeared confident in their presentation of the material, and so forth.

The starkest changes from baseline to midline for this specific group of teachers came in their increased use of participatory or interactive teaching methodologies. For instance, teachers had increased their use of student-centred activities or games, made more frequent use of group work, allowed students to instruct one another more often, and made greater efforts to encourage participation among non-participating students at midline than at baseline. While occasional similar improvements were observed in the full sample of teachers, they were not as pronounced or as consistent across multiple indicators. The curriculum and stated goals of the CPD programme itself emphasise participatory teaching approaches, so these improvements can logically be attributed to the training intervention, especially since they are observed specifically in teachers who were trained but not among teachers more generally.

[302] We do not draw inferences regarding the decline in scores in Galmudug in the overall sample, though they were dramatic, because we believe they are driven by the subjectivity of the tool and the fact that a different team leader conducted classroom observations in Galmudug at baseline and midline.

TABLE 92: INDIVIDUAL TEACHING PRACTICES OBSERVED AMONG PANEL OF CPD TEACHERS, BY ROUND

Type of Behaviour	Question	Observed – Baseline	Observed – Midline
Positive	Is the teacher confident in their presentation of the material?	94.4%	94.4%
Negative	Students spent most of the time copying from the board.	25.0%	22.2%
Negative	Students spent most of the time repeating teacher's words aloud.	5.5%	22.2%
Positive	Teacher used student-centred activities or games.	33.3%	38.9%
Positive	Teacher allowed students to instruct each other (e.g. come to board to demonstrate something, or explain to classmate).	63.9%	77.8%
Positive	Teacher asked open-ended question (requires more than simple answer) that encourages thinking.	80.6%	77.8%
Positive	After a student gave an incorrect answer, did the teacher explain the concept in a new way?	83.3%	80.6%
Positive	After correcting a student who gave an incorrect answer, did the teacher verify the student understood the question now?	77.8%	75.0%
Negative	If a student gives an incorrect answer, are they reprimanded (verbally or physically)?	72.2%	88.9%
Positive	Teacher called on or actively tried to involve a student who was not participating.	80.6%	91.7%
Positive	Students worked together in groups.	27.8%	44.4%
Positive	The teacher summarized and clearly stated a key concept or takeaway point from the lesson.	75.0%	80.6%
Positive	The teacher referred back to previous lessons, relating this lesson to previous lessons.	58.3%	80.6%
Positive	The teacher stopped the lesson and invited questions from students.	61.1%	66.7%
Positive	The teacher employs a variety of explanations that differ in difficulty for the diverse learners in the classroom.	66.7%	66.7%

In other ways, CPD teachers showed little or no improvement, however, and mirrored their peers assessed in the previous section. At midline, CPD teachers were more likely to reprimand a student who answered a question incorrectly, potentially dampening the desire or willingness of students to participate actively in class. Teachers were also more likely to use rote repetition as a teaching strategy, in which students repeat the teacher's word aloud for a significant portion of the class time.

Teaching Practices Conclusions

The primary indicator of teaching quality for EGEP-T is an index of 15 teaching practices observed during in-person observations conducted by team leaders. The index is designed to capture a wide range of practices, with a focus on use of participatory methods, organization of the material to be covered, and

the use of methods that are tailored to individual classrooms and students. In the aggregate, there is little evidence that teaching practices have improved from baseline to midline, as overall scores on this 100-point index have dropped from 67.3 to 63.9 over time. Even excluding Galmudug from the analysis for reasons outlined above does not alter this conclusion, as scores also dropped systematically in Somaliland schools.

At the same time, EGEP-T's interventions are not necessarily designed to impact teaching practices among all teachers. CPD training is targeted to specific teachers, and this is where the greatest gains in quality should be observed. Consistent with that idea, CPD teachers – those who were successfully tracked from baseline to midline and were also confirmed as training participants from project documents – show evidence of improvement from baseline to midline of 6.8 points. Moreover, when we compare CPD teachers to non-CPD teachers *who were also tracked from baseline to midline*, we observe an even greater relative improvement in teaching practices among CPD teachers. These findings are consistent with the project's Theory of Change, as teachers exposed to CPD training appear to have improved their teaching practices, especially in terms of participatory pedagogy.

Partly because of how difficult it is to operationalize teaching quality, our analysis extends beyond the teaching practices studied thus far. We developed several ancillary indicators at the start of the evaluation to provide a more nuanced picture of teaching quality and allow for necessary triangulation of data. These ancillary indicators are not components of the teaching quality index analysed above, and thus provide a basis for independent triangulation using separate measures. It is important to note that a number of the aspects of teaching quality analysed below are not being targeted by the project; for instance, EGEP-T does not expect to exert significant impact on teacher absenteeism, because it is a function of systemic shortcomings with respect to teacher pay. These ancillary indicators are the subject of the sections below.

Gender-Equitable Teaching

The first ancillary indicator assesses gender-sensitive teaching and gender equity. This indicator was measured in two different surveys and was selected because inequality may lead to girls learning less and dropping out. The girls' survey asked students whether they thought there was gender equity in the classroom, while the teacher survey measured teachers' attitudes towards gender equity by asking them whether they thought boys and girls should be equally prepared to have professional careers (with the assumption that teachers' views could influence the ways they teach). These results are presented in the table below.

TABLE 93: MEASURES OF GENDER EQUITY IN TEACHING

Indicators Classroom Gender Equity	Baseline	Midline
Cohort girls: Agree a lot, girls are treated equally to boys	72.5%	67.5%
Teacher: Girls and boys should be equally prepared for a professional career	62.4%	68.9%

We find that the proportion of girls who strongly agreed with the statement that "girls are treated equally to boys" in the classroom decreased by 5 percentage points from baseline to midline.[303] On the other hand, the proportion of teachers who agreed that boys and girls should be equally prepared for future careers increased by approximately 7 percentage points from baseline to midline. Both of these results are on the verge of statistical significance (at a standard 95 percent confidence level), but the substantive interpretations of the results run in opposing directions. Girls' reports seem to suggest that equity is somewhat lower in the classroom at midline than at baseline, whereas teachers' reports seems to suggest that they have somewhat more equitable attitudes regarding gender at the midline.

Classroom observations provide an even more direct and potentially objective means of assessing gender equity in teaching. During classroom observations, researchers were asked to log instances of a given phenomenon (e.g. providing encouraging feedback) that was targeted toward students of a certain gender. Each of these behaviours was recorded, either as a binary (occurred or did not occur within a given observation period) or as a count of the number of times that a given behaviour occurred within a given observation period (e.g. number of times the teacher called on a boy or number of times the teacher called on a girl). In order to construct indicators of gender equality in the classroom, we have totalled instances of a given behaviour across all observational periods and have then subtracted the instances of boys being subject to a positive or reinforcing behaviour from the instances of girls being subject to the same behaviour. We thus derive an indicator centred at zero, where zero indicates complete equality of treatment of girls and boys, while a positive number indicates more favourable treatment of girls and a negative number indicates more positive treatment of boys. The table below presents the four indicators of gender equality during classroom observation, with baseline and midline values to facilitate longitudinal comparison, and with the final column showing the theoretical maximum and minimum range of the variable. To briefly illustrate using the example of positive, encouraging feedback: researchers were asked to indicate whether or not positive feedback was given to a girl across three periods, and also to indicate whether or not positive feedback was given to a boy during three periods. Thus, there could be up to three 'instances' of positive feedback for boys and up to three for girls. The number of boys instances is then subtracted from the number of girls instances. The baseline mean of 0.01 in the table below is very close to zero, indicating that the number of instances for girls was very close to the number of instances for boys, but with very slightly more instances for girls. This finding is generally indicative of a high level of gender equality in the classroom. Note that the exception to this coding scheme is for the use of angry tones or harsh language, where instances for girls are subtracted from instances for boys, such that a positive value indicates that this negative or discouraging behaviour happens *more* frequently for boys than for girls.

TABLE 94: MEASURES OF GENDER EQUALITY IN CLASSROOM OBSERVATION

	Baseline Mean	Midline Mean	Range
Teacher provided positive, encouraging feedback by gender	0.01	-0.04	-3 to 3

[303] It is important to note that girls' knowledge or awareness of their rights to equal treatment may have increased since baseline, which would then bias midline results downward (thus helping to explain the 5 percentage point decrease observed).

Teacher directed questions by gender	-0.16	-0.07	-3 to 3
Teacher used angry tone or harsh language by gender	0.01	0.04	-3 to 3
Number of times teacher called on boy/girl	0.06	-0.42	-20 to 20

All of the indicators in the table are close to zero, and there has been a moderate increase in equity in terms of teachers directing questions by gender, as the mean for this indicator has moved closer to zero from baseline to midline. On the other hand, the number of times that a teacher called on a boy versus a girl has become more unequal since the baseline (in a negative direction, thus favouring boys). Ultimately none of the indicators above have shifted to a statistically significant degree from baseline to midline. These indicators were already quite close to zero at baseline, which permitted little room for improvement in terms of equality, which helps to explain why there is no measurable intervention effect on these indicators of gender equality in the classroom.

Nonetheless, it is also worth considering the difference between gender equality (measured in a fairly direct manner here) and the goal of gender equity, which might require a degree of inequality that actually favours girls, since they tend to be disadvantaged. Thus, the goal for these indicators may not be reduction to zero, but rather shifting the balance toward positive values over time, which would mean that girls, on average, are receiving somewhat more encouragement (and somewhat less scolding) than boys.

In the qualitative data, girls and boys almost unanimously report that teachers treat boys and girls equally in the classroom, asking the same questions of both boys and girls – for instance, "during lessons, they ask questions to both boys and girls to know whether they understand the lesson or not."[304] The only cases where girls or boys reported unequal treatment were cases where they were explaining that teachers in fact treat girls more kindly than boys (particularly in punishment methods) and provide special motivation for girls: "If the teacher sees that girls are getting shy to participate in class activities, he tells them that they are equal to class learning, so ask me any question you have. He motivates them."[305] This is supported by the findings from the head teacher interviews, in which many head teachers report that girls are given special motivation. For example, "In the classroom, teachers use class management and treat boys and girls equally. Where some students have a hard understanding when the teacher is explaining the lesson, what we did was teach extra classes. Last year, we helped 30 girls for extra classes in the Math and English subjects, and it was in the afternoon."[306]

As with the main teaching quality score above, the quantitative findings related to gender equity are decidedly equivocal on the question of changes in teaching quality over time. On the basis of this evidence, there does not appear to have been a positive intervention effect on teaching quality, when operationalized as gender equity. However, the qualitative data does suggest there have been positive changes. This discrepancy between quantitative and qualitative results is unclear. One potential

[304] FGD, Girls, Puntland
[305] FGD, Girls, Puntland
[306] KII, Head Teacher, Somaliland

explanation is that social desirability bias was at play during the qualitative interviews, in that the respondents were simply reporting what they thought the interviewer would want to hear.

Preparation

A second ancillary indicator, preparation, was selected to understand how much effort the teachers were attempting to put into their lessons and therefore would be able to teach adequately. This was measured primarily in the classroom observation which included questions on how well-prepared the teacher seemed and if they had the necessary supplies to teach. Cohort girls and boys were also asked whether they thought their teacher was adequately prepared for class. The table below summarizes the results from these different data sources, showing that there has been improvement on all three indicators from baseline to midline.

TABLE 95: MEASURES OF TEACHER PREPARATION

Indicators of Teacher Preparation	Baseline	Midline
Outcomes Based on Classroom Observation		
Lesson Started On Time	92.1%	93.6%
Teacher had Lesson Plan	68.0%	89.1%
Outcomes Based on Student Survey Responses		
Agree: My teacher is prepared for class each day	82.1%	86.9%

There is an increase of 21 percentage points from baseline to midline in the proportion of classroom observations when a teacher had a lesson plan. This increase is substantively large and is highly statistically significant. The increase in the proportion of girls who strongly agreed that their teacher was prepared for class is also substantial and is statistically significant.

These results suggest that there have been measurable improvements in teacher preparation from baseline to midline, a finding that is supported in the qualitative data. Teachers describe prioritizing lesson preparation more so than in the past. For example, one teacher explains, "Nothing will proceed forward without a schedule, so every teacher must answer three questions during class learning: 1. How do you teach? 2. Who do you teach to? 3. How much do you want to teach? In addition to this, he/she must provide suitable lessons to each level and prepare supportive tools."[307] Another explains how organization around lessons has improved: "Every teacher has a lesson plan - also there is a schedule in the school which shows the subjects that every teacher has in that day and which classes the teacher has."[308]. The fact that a larger proportion of teachers are now taking the time to prepare lesson plans is a particularly promising finding, and may be directly related to teacher coaching that has occurred through project interventions.

[307] KII, Male Teacher, Somaliland
[308] KII, Female Teacher, Somaliland

The qualitative data suggests that this may indeed be the case, as head teachers frequently attributed changes in teaching quality, including teacher preparation, to trainings received through the project. When asked whether teachers follow lesson plans, one head teacher explains that "according to the schedule, the teacher will spend 45 minutes in the classroom, and the lessons are planned in a way that the curriculum will finish in the intended time." The head teacher attributes positive changes in teaching quality to training: "I have seen good quality teaching and that is because there was more awareness and trainings that teachers received about that, and the curriculum is implemented the way it is intended."[309] Another explains that, "There are some changes. In the past, teachers did not do any preparation for lessons which will cause a lot of mistakes. Now, they come on time with quality teaching and this preparation of lessons improves the quality."[310]

Teacher Absenteeism

The third ancillary indicator, presented below, focuses on teacher absenteeism, measuring how often the teachers were actually in the classroom teaching. This was measured in two different surveys. Cohort girls and boys were asked about how often their teacher was absent for a whole day or part of a lesson, and the school survey included questions for the head teacher about how many days each individual teacher had missed in the past two weeks. The relevant means and proportions observed in these two different data sources are presented in the combined table below.

TABLE 96: MEASURES OF TEACHER ABSENTEEISM

Indicator of Teacher Absenteeism	Baseline	Midline
Reports from Students		
Average number of times teacher missed lesson, last two weeks	0.53	0.33
Average number of times teacher left class for 30+ minutes, last two weeks	0.68	0.40
Agree: My teachers are often absent	24.8%	29.7%
Reports from Head Teachers		
Avg. Full Days Absent in Previous 2 Weeks	1.25	0.32
Avg. Partial Days Absent in Previous 2 Weeks	0.61	0.18

Across nearly all measures, there have been substantial reductions in teacher absenteeism from baseline to midline. The reductions in absenteeism on the basis of reports from head teachers provide the most precise estimate, and the observed decrease from baseline to midline is highly statistically significant in the aggregate, as well as when disaggregated by zone (for Puntland and for Somaliland). Changes appear to be noticeable enough that it was also noted in the qualitative data that there have been recent

[309] KII, Head Teacher, Puntland
[310] KII, Head Teacher, Somaliland

improvements in teacher attendance: “Teachers have stopped being absent and make more effort in classes. Last year, teachers used to be absent or missing the class.”[311]

The findings above suggest that there has been a real decrease in teacher absenteeism from baseline to midline. While this cannot be directly attributed to project interventions, this is a promising development in terms of the quality of education that students are receiving because it suggests that teachers may be more motivated than at baseline, which is congruent with findings above that teachers are somewhat more prepared at midline than at baseline. If true, this higher level of motivation may also translate (over the longer term) into improved teaching practices in the classroom.

In the qualitative data, teachers describe benefitting from strengthened school systems and administrative practices that keep teachers, students, and school administration accountable to one another. One teacher explains that “warnings have been given to teachers if they miss their schedule,” and “every teacher signs an attendance book in the morning.”[312] Another explains the importance of arriving early for lessons so that students have the allotted time on their schedules for learning: “The lessons start at 7:30am, so the teacher should come before 7:30 and start the class as soon as the time starts because the class is only 40 minutes. If the teacher is absent the student will miss 15 or 20 minutes.”[313] Similarly, teachers also appear to be improving their record keeping around students’ attendance as a result of the project. When asked about attendance record keeping, one teacher responded, “It is the responsibility of the school administration -- they bring it in the morning and return it back to office, protecting it properly. ADRA provides us with attendance books every year.”[314]

In general, the experiences teachers describe having with management highlight the importance of teachers receiving proper support from administration and the benefits they receive from strong school management. One teacher describes how his principal, presumably through the support of the project (given the specific activities described), has established school practices and processes that motivate teachers: “Really our school principal/management is very good. He is strong with his duties. He shares problems to us and we have a role to handle it. Teaching quality is being supported by school administration through inspection. He sometimes enters the classes and shares our weaknesses. As teachers, we have a code of conduct – for instance, we sign a time in and time out attendance sheet. Everybody is known when he comes to school and he is being controlled. They [teachers] have good ethics and are good at motivating students.”[315]

Punishment

A final ancillary indicator of teaching quality involves measuring the extent to which teachers punish students, especially through the use of corporal punishment. Our focus in this section is on the responses

[311] FGD, Girls, Somaliland
[312] KII, Male Teacher, Somaliland
[313] KII, Female Teacher, Somaliland
[314] KII, Male Teacher, Somaliland
[315] KII, Male Teacher, Galmudug

provided by cohort girls from the true panel in order to assess the degree to which their reports of punishment have changed from baseline to midline.

At baseline, it was determined that the most sensitive measures related to corporal punishment that would also be directly related to girls' educational outcomes were likely to be girls' reports of whether teachers had used corporal punishment in the past week and whether or not girls' teachers had used corporal punishment *on them* in the past week.

The table below summarizes these two indicators by baseline and midline, suggesting that there has been a reduction in the frequency (as reported by girls) of general corporal punishment in class, as well as a reduction in the frequency with which teachers punished sampled cohort girls in class.

TABLE 97: FREQUENCY OF PUNISHMENT EMPLOYED BY TEACHERS

Frequency	Teacher used corporal punishment in last week		Teacher used corporal punishment on you in last week	
	Baseline	**Midline**	**Baseline**	**Midline**
Almost every day	4.4%	1.7%	1.1%	0.7%
Once or twice	21.1%	14.0%	11.7%	7.2%
Never	74.5%	84.3%	87.2%	92.1%

For both reports (about general corporal punishment and about specific use of punishment on the respondent), there has been a statistically significant increase from baseline to midline in the proportion of girls who reported that corporal punishment was never used within the past week.[316]

These findings suggest that there has been a substantial reduction in corporal punishment, as reported by girls, from baseline to midline. These reductions are apparent and statistically significant despite the fact that the program may have increased girls' awareness of (and sensitivity to) this issue, which might have been expected to lead to increased reports of corporal punishment at the midline.

In the qualitative data, the disciplinary methods teachers use appears to vary by teacher and across schools. There is evidence in the qualitative interviews that there has been positive change made, in that some teachers are no longer relying on corporal punishment. For example, when asked how teachers discipline students, one student explained, "They treat us well and there are rules in the school. These days, teachers do not use beating with the stick but they ban students for one week of school - then students will behave well."[317] Others described how teachers provide encouragement when students make mistakes. Other punishment methods include suspension, dismissing students from the classroom, giving students extra assignments, sending students home, and transferring

[316] In a regression with cluster-robust standard errors, round is a statistically significant predictor of lower frequency of corporal punishment, with corporal punishment being coded as a dummy variable taking 1 if corporal punishment was reported to have occurred in the past week, and 0 if corporal punishment was reported to have never occurred. For the two measures in question, the p-values are, respectively, P = 0.001 and P = 0.017.

[317] FGD, Boys, Puntland

students to other schools. Teachers themselves also describe how there is increased emphasis on positive reinforcement: “During my school days, discipline and corrections were punishments and intimidation, but nowadays it's different by providing students’ rights in the best manner. Today's discipline is based on friendship and welcoming students, then telling them their faults.”[318]

However, there is also evidence that teachers still beat students in some schools or use other physical punishments to discipline students. For example, teachers use what some students call traditional Islamic punishments, including standing on your knees or holding your ears. These punishments appear to be considered inappropriate for girls, so instead they are in some cases asked to stand under the sun as a punishment: “What other men mentioned (standing knee and holding ears) is what Islamic religion protected. Using this kind of punishment for women is forbidden according to Islamic religion. If they are absent they stand under sunshine.”[319] Although students largely consider punishments to be fair and proportional, it appears that some consider good teachers to be the ones who do not employ these types of punishment methods, as they are considered unbearable in some contexts: “The good person teachers treat the students well. Also they are always late, they don’t punish them because here in Garowe it is very hot, and the students cannot tolerate it because some of the students urinate themselves when they are in the holding the ears position/punishment for a prolonged period of time. I have been a witness to this at times.”[320]

Three additional findings from the qualitative data related to teachers’ disciplinary methods should be noted, as they could inform future programming efforts. First, many students mentioned that girls are treated more kindly than boys, and that they are not punished as harshly. This was mentioned by girls and boys alike. For example, one girl explains, “In my opinion, I don’t think that teachers treat boys and girls equally because they are more kind to the girls. For example, if a boy interrupts in class, the teacher may punish him. Inversely, teachers are so patient with the girls and care about them.”[321] Another girl explains that “they always do more punishments for the boys” because “it is not our culture to be disciplined the same as boys.”[322]

Second, students mentioned that corporal punishment is typically used more often on younger students and that students tend not to tolerate physical punishment as they grow older: “Discipline is when student young but once they join secondary school they are being advised through talk, but the student will oppose you if you take a stick to punish.”[323] Another student explains, “Students in intermediate school are beaten by teachers but students in high school are enough old but when they make mistake they warning him. The second time he does mistake they kick him out from the classroom.”[324] This was also supported by evidence from the teachers’ qualitative interviews: “The way we teach adult students and young students are different because we treat adult students as mature and we don't discipline

[318] KII, Male Teacher, Somaliland
[319] FGD, Boys, Puntland
[320] FGD, Boys, Puntland
[321] FGD, Girls, Puntland
[322] FGD, Girls, Puntland
[323] FGD, Girls, Puntland
[324] FGD, Boys, Puntland

them. They may leave school if you scold adult students a lot. We give advice and show respect to the adult students."[325]

Third, a number of non-physical but potentially harmful punishment methods were mentioned by students, including teachers "deducting some points from their exam grades"[326] or forcing them to "provide pens and sanitation tools to the school as punishment,"[327] which sounds as though students are perhaps forced to give back the materials they received as part of EGEP-T if they misbehave.

Summary Discussion

Taken as a whole, the results in this section suggest that there has not been a measurable positive intervention-effect on teaching quality since baseline in the quantitative data. While there have been some verifiable improvements, most notably in teacher absenteeism and teacher preparation as well as a reduction in use of corporal punishment, not all of these improvements can be directly attributed to project interventions. All of the most direct, observational or reported measures of teaching quality suggest that there has been no substantial change in teaching quality from the time of baseline measurement. The finding that teaching quality has not improved substantially is, in fact, consonant with the findings related to learning barriers analysis, which suggest that teaching quality is still one of the main determinants of lower than average learning among cohort girls at midline. It is also worth noting that one of the principle interventions intended to increase teaching quality – namely CPD training – have not had a full year to take effect, having started only a few months before the midline study. Thus, we can conclude that teaching quality has not improve appreciably, but need not conclude that interventions along these lines have been unsuccessful. It is likely that the CPD training is only beginning to have an effect and that these changes are likely to be measurable at the next evaluation point.

It should also be noted that the qualitative data holds evidence that there have indeed been positive changes in teaching methods used in classrooms, in equitable treatment of students, and teacher preparation that can be attributed to project activities. These changes simply appear to not be uniform enough (across teachers) to have had a substantial positive effect on the quantitative indicators of teaching quality.

[325] KII, Male Teacher, Puntland
[326] FGD, Girls, Galmudug
[327] FGD, Boys, Somaliland

7.4 School Management and Institutional Governance

Indicator: Increase in number of Community Education Committees contributing to effective school management

TABLE 98: INTERMEDIATE SCHOOL MANAGEMENT OUTCOME INDICATORS AS PER THE LOGFRAME

IO	IO indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
School Management and Institutional Governance	Percentage of Education Committees contributing to effective school management (measured with scorecard approach)	47.0	52.2	49.8	N	57.2	Y
	Main qualitative findings						
•	• There was very little mention of CECs providing support for girls via bursaries. • Due to the poor conditions for fundraising, CECs mostly focus on monitoring activities, awareness raising, and other non-financial contributions to schools, such as developing school development plans and organizing school repairs.						

TABLE 99: SCHOOL MANAGEMENT RESULTS BY ZONE (SCORECARD)

Round of Data Collection	Somaliland	Puntland	Galmudug	Overall
Baseline	41.9	49.4	61.0	47.0
Midline	38.5	58.2	66.7	49.8

In the previous section, we highlighted EGEP-T efforts to improve teaching quality. The second school-level aspect of EGEP-T's Theory of Change is school management and governance, which is focused primarily on the Community Education Committees (CECs). Each school is overseen locally by a CEC, comprised of community members and – typically – the head teacher of the school. CECs provide management and oversight to the school, monitoring student performance, teacher attendance, school finances, and other aspects of administration, though the precise parameters of their role varies somewhat from school to school. As we discussed in the sustainability section of this report, effective CECs also support the school financially and in other ways, raising funds for school improvements, distributing support for students who cannot afford school fees, and donating their time and labour.

School management is a critical intermediate outcome for EGEP-T, because it impacts each of the primary outcomes the project seeks to change. At the most fundamental level, well-managed schools are themselves more sustainable, and less likely to be beset by financial problems. Improvements made during the two phases of GEC programming – such as improvements in teaching quality, changes in community attitudes, and support for re-enrolment of girls in school – are more likely to be maintained going forward under an effective CEC. Schools with an effective CEC will spend limited funds more efficiently, targeting the school's and community's needs more precisely; they will be better able to raise funds from community members and more likely to leverage outside sources of support, such as other NGOs; they will be better placed to seek help from the government, since such help is typically managed through or in partnership with the CEC; they will be more likely to continue awareness-raising activities that influence community attitudes; and they will be more likely to continue monitoring schools in a proactive way for teaching quality, student enrolment rates, and learning performance. It is through these mechanisms that effective school management can and should improve learning and transition outcomes *during* the project, and help maintain them after the project has wound down.

As with several of the intermediate outcomes studied in this evaluation, school management is a diffuse, latent outcome. School management is not directly observed; rather, it can be inferred from indicators which are observable. Moreover, school management is multi-dimensional – schools could maintain very good records and pay teachers on time, but their CEC might meet only intermittently and they could provide no active support to the school. For this reason, we developed, in consultation with Relief International, an index to capture several key aspects of school management. The goal was to focus measurement on aspects of school management that EGEP-T actually seeks to impact, such as CEC activity levels and the maintenance of critical school records. In contrast, we do not incorporate aspects of school management that fall outside of EGEP-T's purview, such as the timely payment of teacher salaries, because delays are typically driven by factors outside the control of CECs. It is important to note that these measures are *indicators* and not goals in themselves: they were constructed by asking "what would a well-managed school look like?" In this sense, they are representative of the well-known adage, Goodhart's Law, that "when a measure becomes a target, it ceases to be a good measure." To avoid the possibility that schools have focused on improving on these specific indicators, to the neglect of other aspects of school management, we also triangulate across other measures where possible.

The index we use is analogous to a scorecard approach, in which each school is assigned a single composite score, ranging from 0 to 100. The indicators that comprise the composite index are:

QUALITY OF CEC MANAGEMENT

- Share of schools with a functioning CEC that meets at least once monthly
- Share of teachers rating CEC management “very good”*

CEC SUPPORT FOR SCHOOL AND STUDENTS

- CEC provides bursary support to at least one female student
- CEC makes financial contribution to school
- CEC makes non-financial (in-kind, labour, etc.) contribution to school

ENACTMENT OF FORMAL POLICIES AND RECORD-KEEPING

- Rating of schools' record-keeping for students (0-4 scale)
    - Are records of student grades available? (1 point)
    - Are records of student grades either "mostly" or "extremely complete"? (1 point)
    - Are records of student enrolment available? (1 point)
    - Are records of student enrolment either "mostly" or "extremely complete"? (1 point)
- Rating of schools' promulgation of four formal policies (0-4 scale)
    - Does school have a mission statement? (1 point)
    - Does school have a Code of Conduct, and can they show enumerator a copy? (1 point)
    - Does school have a Child Protection Policy, and can they show enumerator a copy? (1 point)
    - Does school have a School Development Plan, and can they show enumerator a copy? (1 point)

All but one of these measures are drawn from interviews with the head teacher or direct observation of schools with the assistance of the head teacher. The sole exception, denoted with an asterisk above, is the share of teachers rating the quality of CEC management "very good", which is calculated from surveys with regular teachers. The indicators capture three themes, as the list above makes clear: general measures of CEC management quality, measures of CEC support for schools, and promulgation of key policies or policy documents and the accurate keeping of mission-critical records, such as student grades and attendance.

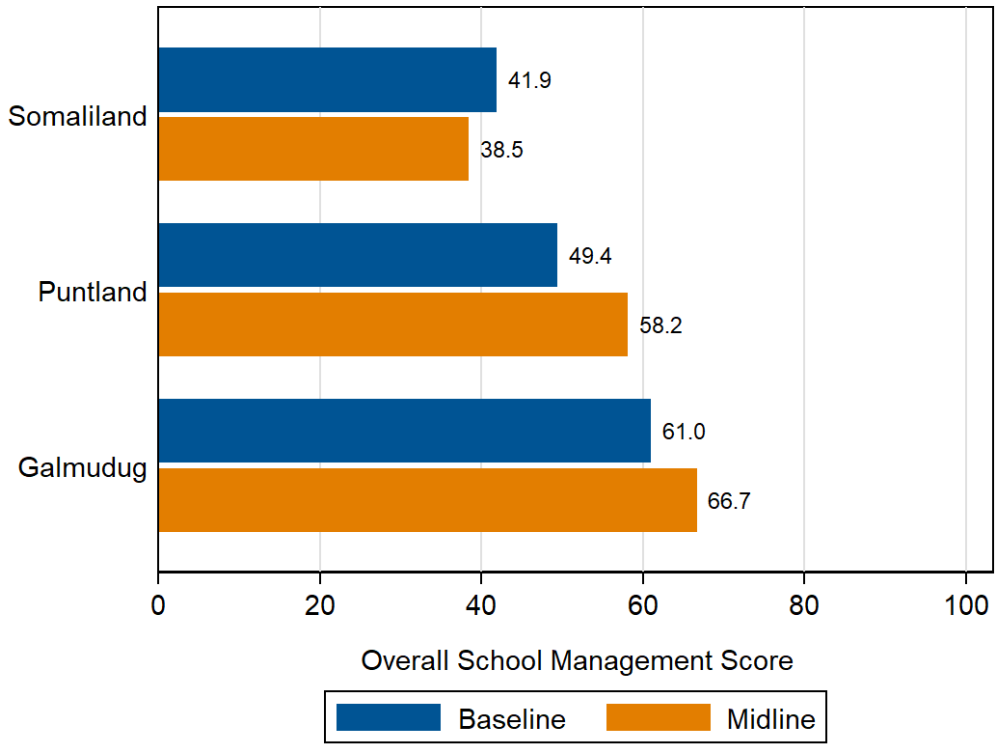
As the description above hints at, the constituent parts of the index are measured on very different scales – some are binary, others are an aggregation of binary outcomes, and others are percentages on a 0 to 100 scale. In every case, higher scores represent more positive outcomes. For each of the 7 top-level indicators (not counting individual sub-indicators regarding record-keeping and policy promulgation), we normalize their scores to a 100-point scale. To illustrate, school record-keeping is scored on a 0 to 4 scale, which we normalize to a 100-point scale; likewise, whether a CEC makes financial contributions to the school is a binary indicator, which we normalize to a 0-100 scale (either 100 if they do make a contribution and 0 otherwise). The normalization process is essential to ensure that each top-level indicator is equally weighted in the composite score. To be clear, the final composite is an unweighted average of the 7 top-level indicators, meaning that it also varies from 0 to 100.[328]

The figure below reports the mean school-level score at baseline and midline, disaggregated by zone. Our analysis employs the set of "comparable schools" – those schools which appear in both baseline and midline, as is typical throughout this report. Aggregating across zones, the net change from baseline to

[328] In the baseline evaluation, the mean school achieved a composite score of 47.2, indicating that there was significant room for improvement, but that the measures employed were not unrealistic. In addition, a number of schools achieved each individual sub-indicator showing, again, that each measure represented an attainable goal.

midline was 2.8 points, from 47.0 at baseline to 49.8 at midline.[329] As the figure shows, Puntland and Galmudug both experienced increases in scores from baseline to midline, while scores declined slightly in Somaliland. Neither the improvement in Galmudug nor the decline in Somaliland is statistically distinguishable from a null effect; the gain documented in Puntland is, however, statistically significant ($p < 0.05$).

FIGURE 39: SCHOOL MANAGEMENT SCORES, BY ZONE AND ROUND



The similar ways in which school management scores evolved over time across zones – even from different starting points – is also reflected in changes observed within key subgroups. The table below reports baseline and midline school management scores for different types of schools, including rural, IDP, drought-affected, and conflict-affected schools. In the case of drought and conflict, we focused on schools that were drought- or conflict-affected at baseline, given that the drought was most widespread and severe during baseline data collection, and had eased by the time of midline data collection, and conflict – while not necessarily easing in severity – was less widespread at midline than baseline. Therefore, the subgroups reported in the table are based on a school's drought or conflict status at baseline; in the case

[329] These values differ from the baseline value discussed above (47.2) because the baseline sample analyzed in this report differs from that of the baseline report itself, due to the exclusion of schools in Banadir, and the exclusion of replacement and replaced schools.

of other characteristics, such as IDP status and urbanity, school characteristics are stable over time. In each subgroup in the table, the changes from baseline to midline are inconsequentially small, typically changing by just 2 or 3 points from baseline to midline. The exception are schools that were affected by conflict at baseline – however, the available sample of such schools is just 6, given that most of the schools afflicted by conflict at baseline were not included in the midline sample.

TABLE 100: SCHOOL MANAGEMENT SCORES AT BASELINE AND MIDLINE, DISAGGREGATED BY SCHOOL CHARACTERISTICS

School Characteristic	Baseline	Midline
Urban	48.4	52.1
Rural	44.9	46.3
Primary Schools	46.7	48.2
Secondary Schools	47.9	55.4
Non-Drought	45.0	48.2
Drought-Affected	50.9	54.0
Non-Conflict	46.7	48.9
Conflict-Affected	51.0	61.8
Non-IDP	46.2	49.0
IDP	60.7	62.1
Total	47.0	49.8

Although composite school management scores have not changed in a meaningful way, overall, from baseline to midline, it is possible that performance on individual indicators may have shifted. To assess this possibility, we calculated baseline and midline values for each of the 7 top-level indicators, which are reported in the table below. These scores are reported in their natural scales – for instance, the share of schools where the CEC provides bursary support to at least one female student, the share of teachers who rate CEC management "very good", and the rating of schools' record-keeping quality on a 0-4 scale.

In terms of stability over time, three indicators stand out: teachers' views of CEC management quality, the quality of school record-keeping, and – to a slightly lesser extent – the promulgation of formal policies. In each case, scores either did not rise at all, or rose very slightly (e.g., 2.66 to 2.74 on a 5-point scale for the quality of school record-keeping) between baseline and midline. Other indicators show more meaningful movement, however. For instance, the share of schools where CECs provide bursaries to at least one female student has risen appreciably, from 21.8 to 40.0 percent, a difference that is statistically significant. The share of CECs that meet consistently (at least once per month) has also risen, from 65.5 to 76.4 percent, a difference that is marginally significant ($p = 0.08$).[330]

[330] Improvements in the activity level of CECs reflect both an increase in the share of schools with an active CEC (from 90.9 to 95.5 percent over time) and an increase in the share of existing CECs that meet consistently. The finding regarding frequency of CEC meetings is supported by other reports from head teachers regarding the timing of the last CEC meeting – the share who reported the last meeting occurred within the previous month rose from 65.5 to 71.6 percent from baseline to midline. Of course, it is important to keep in mind that activity levels in terms of

As we briefly documented in the context of our sustainability findings, CECs have become much more likely to provide bursary support to female students in need of financial support for their school fees, but have simultaneously become somewhat less likely to make financial contributions in support of the school more broadly. These offsetting trends are *part* of the explanation for stagnant school management scores, although they are by no means the only reason school management scores have not improved. Indeed, substantial improvements are not observed in any indicator included in the index, *except* the frequency of bursary support by CECs. No other constitutive indicator changed sufficiently to attain statistical significance.

TABLE 101: BASELINE AND MIDLINE VALUES OF COMPONENTS OF THE SCHOOL MANAGEMENT SCORE

School Management Indicator	Baseline	Midline	Overall
CEC meets at least once per month	65.5%	76.4%	70.9%
Share of teachers rating CEC management "very good"	66.7%	63.4%	63.2%
CEC makes financial contributions to school	27.3%	21.8%	24.5%
CEC makes non-financial **contributions to school**	33.6%	38.2%	35.9%
CEC provides bursary support to at least one female student	21.8%	40.0%	30.3%
Rating of schools' record-keeping for students (0-4 scale)	2.66	2.74	2.70
Rating of schools' promulgation of four formal **policies (0-4 scale)**	1.71	1.89	1.80
Full School Management Index	43.9	44.9	44.4

The findings above are supported by our analysis of additional data collected from head teachers, which was not included in the index and was not referenced explicitly in the baseline evaluation report. These alternative indicators tended to show slight improvements from baseline to midline, similar to those included in the core index. For instance, we asked head teachers to report the number of CEC members who attended the last meeting, and how many CEC members there are in total, as a measure of member engagement. At baseline, average attendance at the last meeting was 85.2 percent, with a small improvement to 86.2 percent at midline.

More positively, the number of schools tracking and documenting aspects of their administration has increased. From baseline to midline, the share of head teachers that report keeping a logbook of teacher attendance has increased (81.8 to 92.7 percent), as have the share of head teachers reporting that they

meetings does not necessarily translate into efficacy in tangible terms, and the analysis in this section suggests that, even though many CECs actively meet, their inherent resource limitations restrict their ability to take on ambitious projects or support students and schools in material terms.

maintain a logbook of expenses throughout the year (61.8 to 69.7 percent) and the share who report maintaining a log of all school fees paid by students (64.6 to 86.1 percent). These changes were not captured by our composite index, which focused on the maintenance of student grade and attendance records. These latter records are mission-critical, but teacher attendance, expenses and school fees should also be recorded faithfully, and the fact that this has improved over time is a positive sign.

Despite this positive finding, however, school management has not changed appreciably since the baseline evaluation, at least among those schools sampled. The share of schools with a functioning CEC has increased slightly, as has the number of CECs that meet at least monthly. However, these gains have not been dramatic, and a functioning CEC is merely a first step in improving school management. In areas of their work that are more difficult than meeting – such as tangible support for their schools and community members – CECs have shown mixed results. And the formal administration of schools has changed very little, with only slightly better overall record-keeping and marginal or negative changes in the extent to which schools have formal policies in place regarding child protection, a code of conduct, or a school development plan.

The qualitative findings largely support the quantitative findings in this section. As was already discussed in the sustainability section, there was very little mention of CECs providing support for girls via bursaries. Rather, due to the poor conditions for fundraising (i.e. community members are impoverished and cannot provide much support), CECs mostly focus on monitoring activities, awareness raising, and other non-financial contributions to schools, such as developing school development plans and organizing school repairs. However, CECs appear to feel their ability to carry out these activities is limited by the circumstances of their communities and, in some cases, the attitudes of parents, who at times accuse them of practicing favouritism or selectively pressuring students in their monitoring activities.

To the extent that school management is an essential ingredient in the sustainability of improvements made through EGEP-T interventions, the results of this section call into question whether CECs are well-placed, at this moment, to ensure sustainability. Additional work will need to be done to reinforce the capacity of, and increase the resources available to, CECs going forward.

7.5 Community Attitudes

Indicator: Increase in caregivers' aspirations for level of schooling they hope their girls will reach

TABLE 102: INTERMEDIATE COMMUNITY ATTITUDES OUTCOME INDICATORS AS PER THE LOGFRAME

IO	IO indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Community Attitudes	Increase in caregivers' aspirations for level of schooling they	89.5	91.5	90.7	N	92.5	Y

	hope their girls will reach						
Community Attitudes	Percentage of girls who disagree strongly with the statement: 'I feel pressure to drop out of school and get married'.	59.6	61.6	62.7	Y	63.6	Y
Main qualitative findings							
• Traditional ideas around marriage and the role of girls in the household are still being actively debated in Somali society.							

TABLE 103: COMMUNITY ATTITUDES RESULTS BY ZONE (CAREGIVER ASPIRATIONS FOR GIRLS' SCHOOLING)

Round of Data Collection	Somaliland	Puntland	Galmudug
Baseline	90.6	88.3	90.3
Midline	88.8	92.3	90.3

TABLE 104: COMMUNITY ATTITUDES RESULTS BY ZONE (GIRLS FEELING PRESSURE TO MARRY EARLY)

Round of Data Collection	Somaliland	Puntland	Galmudug
Baseline	52.4	66.1	57.6
Midline	59.6	62.0	76.4

The intermediate outcomes analysed to this point have tended to focus on outcomes directly related to project schools, and beneficiaries, such as attendance rates and school management or governance. In this section, we shift our focus to a more diffuse set of project outcomes: community attitudes toward girls' education. The goal of shifting community attitudes is based on the idea that individuals' underlying attitudes influence their willingness to invest in girls' education, support it vocally and through tangible actions, and encourage girls' education within their own families and their broader communities. If parents, community leaders, and young men all begin to value girls' education more highly, they will be more likely to pay school fees for their daughters, promote delayed marriage for girls in favour of completing their schooling, and provide incentives for greater educational attainment by girls. Simultaneously, if girls feel that their friends, family, and neighbours support their education, it should

increase their own interest in education and promote better attendance, higher enrolment rates, and – perhaps – better learning outcomes.

In this section, we investigate whether attitudes in EGEP-T communities have become more positive toward girls' education in the 1.5 years between baseline and midline. As we have noted in the context of sustainability – where we analysed support for girls' education among male community members, among other attitudinal outcomes – community attitudes are a latent and multidimensional outcome, which can only be measured indirectly. For instance, while we can ask respondents whether they support girls' education, their responses will typically be anchored to a particular view of what "support" means, which may not incorporate the full spectrum of what we mean. A related issue is that attitudinal measures are heavily influenced by question wording, question content, and the set of response options available to respondents, as well as the context in which the question is asked. Given these well-known issues, we elected to measure community attitudes using a varied battery of questions which seek to measure self-reported general attitudes, more narrowly targeted attitudes (such as a desire for girls to complete a university education), and how respondents would react in practical, but hypothetical, situations that are influenced by their own latent attitudes.

In the discussion that follows, we first report results regarding the EGEP-T logframe indicator for community attitudes, which focuses on the aspirations caregivers hold regarding their girls' education. We then broaden our focus to other aspects of caregiver-specific attitudes. Finally, in an effort to triangulate results from as wide a range of sources as possible, we analyse the perceptions of girls – how much they feel family members and community members writ large support their education – and the views of teachers, as observers of community members with a particular interest in attitudes toward education.

Caregiver aspirations

Our primary measure of community attitudes assesses the aspirations caregivers have regarding their daughters' educational attainment. Caregivers were asked a simple question: what level of education would they like their daughter or girl to achieve? This question reveals differences in the extent to which individuals value higher education, as some respondents indicate that they hope their girl completes primary school or lower secondary, while many hope their girl will complete university.

Despite – or because of – its straightforward nature, this question is also prone to a few methodological problems. The first is that there are potential ceiling effects with regard to responses: at baseline, 89.5 percent of caregivers aspired to a college or university education for their girls. This leaves relatively little room for improvement over time, as no "higher" response options were given, and aspirations beyond a college education are not particularly relevant, or realistic, in the context.

The second problem is one of social desirability bias. In the context of a survey clearly focused on girls' education, respondents may provide misleading answers in an effort to improve their appearance in front

of interviewers who value education, and may also provide inaccurate answers in an effort to improve their self-conception.[331]

The third problem is one occasionally referred to as "cheap talk." When confronted with a costless choice, it is easy for respondents to state that they have high aspirations regarding their girls' education; indeed, most individuals aspire to high achievement, but may not be willing to make the sacrifices or harness the dedication necessary to achieve it. In many ways, our goal is to measure how heavily respondents value educational achievement *relative* to other priorities or needs. By asking about aspirations in isolation, most respondents will likely state high aspirations, even if the weight they place on achieving those aspirations is lower than the weight they place on other life outcomes for themselves and their families.[332]

The table below provides results, disaggregated by zone, regarding caregiver aspirations at baseline and midline. As expected, most respondents stated high aspirations for their daughters – at baseline, 89.5 percent of respondents hoped their daughter would complete university.[333] At the midline, slightly more (1.2 percentage points) caregivers hoped their daughter would complete university, a difference that was not statistically significant at any conventional level.[334]

TABLE 105: SHARE OF CAREGIVERS WHO ASPIRE TO UNIVERSITY EDUCATION FOR THEIR GIRLS, BY ROUND AND ZONE

Zone	Baseline Percent	Midline Percent
Somaliland	90.6	88.8
Puntland	88.3	92.3
Galmudug	90.3	90.3
Overall	89.5	90.7

[331] Social desirability bias is typically associated with respondents who lie to avoid social stigma or to improve their appearance in front of their community or interviews. However, responses can also be shaped by the internal psychology of self-esteem and self-conception. If respondents regard themselves as pro-education or more loosely holding "progressive values," they may exaggerate their stated aspirations to avoid cognitive dissonance (i.e. "how can I view myself as pro-education if I only want my daughter to complete primary school?").

[332] We also reiterate the changes in sampling strategy that took place between baseline and midline with regard to the sample of caregivers. At baseline, caregivers were selected randomly from within EGEP-T project communities if their household included a girl aged 11-18 years, regardless of their enrolment status; for the purposes of this question, only caregivers whose girls were enrolled are included, but the sample includes caregivers of girls enrolled at non-project schools. The corresponding midline sample includes only caregivers of girls from project schools. In Section 2, we investigated the consequences of this shift empirically and found relatively little reason for concern, because girls enrolled at project and non-project schools in the same areas do not differ systematically in terms of their families' attitudes.

[333] While some degree of variation exists in the lower levels of aspirations – for instance, fewer respondents capped their aspirations at primary level in the midline compared to the baseline – these shifts over time were relatively small.

[334] Based on a t-test comparing the share of caregivers who aspired to a university education between rounds, with standard errors clustered at the school level. When we employed the more flexible chi-square test, which takes into account changes in responses at lower aspiration levels, the results were unchanged.

It is worth noting that improvements were observed in Puntland, where the share of caregivers aspiring to a university education increased from 88.3 percent to 92.3 percent, and where improvements were seen across the spectrum of possible responses. In Puntland, the improvements cited were marginally significant, though the precise finding depended, to some degree, on the precise test employed.[335]

In the qualitative data, caregivers, both male and female, almost unanimously report that their daughters wish to continue their studies up to the university level and that they support this goal. The greatest barrier parents appear to face in supporting their girls in continuing their education is providing for it financially. As one mother reports, “There are some students who would like to continue their education but can’t afford that, so they need support to finish their education.”[336] Even boys are said to support higher education for girls – according to CEC members in Puntland, boys and men prefer girls to be educated.[337] These positive results should still, however, be considered within the broader context of the evidence provided in other sections of this report and in the qualitative data. For example, many of the same interviewees who noted community support for girls' education also made a sharp distinction between community members who do and community members who do not provide that support. Their point is clear and can be lost among statements espousing support for girls' education: there are differences in the community, and even within families, over the importance of girls' education, and parents are not uniformly supportive. The take-away is that there is also evidence that traditional attitudes on the roles of girls and women in society still overshadow some of this progress.

Caregiver Attitudes and Actions

Supplementing our data regarding caregiver aspirations, caregivers were also surveyed regarding other attitudes relevant to girls' education. For instance, we presented caregivers with statements regarding the importance of girls' education and asked them to indicate the extent of their agreement or disagreement with each statement. These statements are useful because they tend to move away from the "cheap talk" noted in the previous section; to illustrate, when we ask respondents whether girls' education is a worthwhile investment, even when funds are limited, this frames the question in terms of an implicit trade-off between girls' education and other household priorities.

Taking that logic a step further, we also asked caregivers to consider specific, realistic scenarios that they or people they know might face. For instance, we asked respondents to imagine a situation in which their adult sister is sick and needs money to pay for her hospital bill. Respondents were asked to choose whether they would sell some household goods or livestock to help their sister, or withdraw their

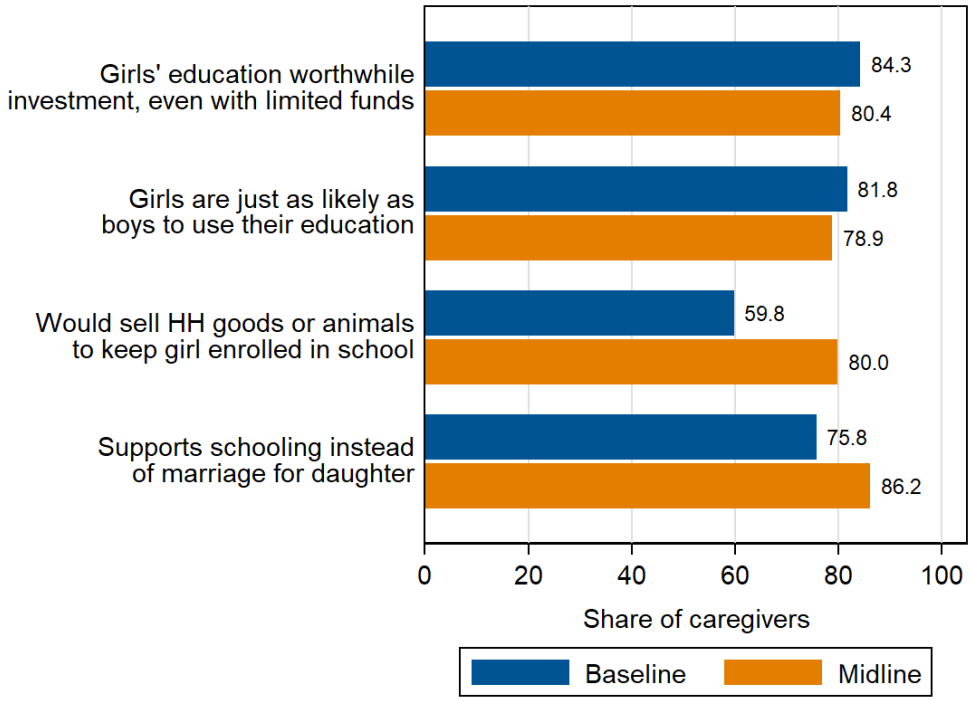
[335] Beyond the differences noted across zones, another geographic factor that appears to shape caregiver aspirations is urbanity. At the midline, urban and rural respondents expressed nearly identical aspirations. However, this represented a substantial shift from the baseline, when rural respondents were much less likely to hope their girls would attend university.
[336] FGD, Mothers, Somaliland
[337] FGD, CEC members, Puntland.

daughter from school and use the savings from school fees to help their sister. This approach makes the trade-off explicit, which we hope will reveal more nuance regarding support for girls' education.

The findings from these supplemental questions are provided in the figure below, disaggregated by the evaluation round. The results from this analysis are somewhat mixed, but generally point to improving attitudes from baseline to midline. On the negative side, we found a 3.9 point decline in the share of caregivers who believe girls' education is worth investing in, even when funds are limited.[338] We also observe a smaller decline in the share of respondents who strongly agree that girls are just as likely to use their education as boys. This is supported by evidence from the qualitative interviews. One boy explained in a focus group discussion that though it is good for girls to pursue their education, for girls who do not excel in school, it is more prudent to get married and save the money she or her family would have spent on her education: “She must not lose the university fee because her husband will take care of her. You can find evidence that 95% of working people are men.”[339] In a mothers’ focus group discussion, one mother explained that “there is saying which goes, ‘Although a girl is educated, she will end up in the kitchen.’"[340]

[338] Specifically, we calculated the share of respondents who strongly agreed with the statement "even with limited funds, it is worthwhile to invest in girls' education." The reported difference between baseline and midline is statistically significant at the 10 percent level only.
[339] FGD, Boys, Somaliland
[340] FGD, Mothers, Puntland

FIGURE 40: SHARE OF CAREGIVERS EXPRESSING SUPPORT FOR GIRLS' EDUCATION



On the other hand, we find significant shifts in the manner in which respondents would handle hypothetical choices regarding girls' education. For instance, we presented respondents with a scenario in which their 15-year old daughter has received a marriage proposal. We asked caregivers whether they would support the marriage, continued schooling, or support a combination of marriage and continued schooling. Relative to the baseline, the share of respondents who chose continued schooling alone – without simultaneous marriage – rose by 10.4 points, and the share of respondents who would support marriage by itself declined from 9.7 to 2.7 percent.[341] Again, this is supported by evidence from the qualitative interviews. In the qualitative data, there were only a few instances where the parents reported that their daughter wishes only to get married. Notably, however, even in cases when girls aspire to marry, it appears to be increasingly accepted for girls to prioritize pursuing an education first before settling down: "Every girl's dreams are to complete university and then get married. I read somewhere that boys can get married late as possible, but girls can't - they have to marry early. But me personally, I believe she can study first then married."[342] There is also evidence that it is becoming more normal for girls to

[341] In the figure, we report the share of respondents who would support continued schooling alone, because it is often unrealistic for a girl to get married and continue schooling, since this would require the support, if not outright approval, of her husband. However, even if we consider individuals who support a combined approach to be supportive of education, the trend from baseline to midline is effectively the same.
[342] FGD, Fathers, Puntland

continue to attend school while starting a family: “She works hard in education to continue her schooling. Girls keep studying now while they have children.”[343]

Importantly, the large improvements from baseline to midline are not restricted to a single zone. With regard to the two hypothetical scenarios, we observe large, and roughly equivalent, improvements in Somaliland, Puntland, and Galmudug alike, as shown in the table below. Interestingly, when it comes to realistic trade-offs, rural respondents are actually more willing to make sacrifices in support of their girls' education, a pattern that holds in both the baseline and midline data.

TABLE 106: CAREGIVER SUPPORT FOR TANGIBLE ACTIONS TO SUPPORT GIRLS' EDUCATION IN HYPOTHETICAL SITUATIONS

Zone	Baseline Percent	Midline Percent
Would sell HH goods or animal before withdrawing daughter from school		
Somaliland	61.7	88.9
Puntland	56.7	67.5
Galmudug	67.7	98
Would support daughter to stay in school rather than accept marriage proposal		
Somaliland	75.2	87.2
Puntland	75.4	83
Galmudug	80.6	95.9

The qualitative interview data also suggests that EGEP-T activities are having a significant positive effect on community attitudes toward girls’ education, but that girls still face societal pressures that are harmful to their success in education. It is clear that communities are living in a time of change, where the merits of girls’ education are actively being debated. As community members adjust to this change, they often resort to drawing sharp comparisons between groups – for example, the merits of girls’ education are very often discussed in direct comparison to the merits of boys’ education, and many described those who hold on to traditional views as “ignorant” and those who understand the value of girls’ education as “educated.”

As one girl explains, “Parents help their children if they are educated people. They help their students with reading and by motivating them to study.”[344] A few girls also cited the fact that their mothers have taken on some of their domestic work, giving them more opportunity to study.[345] Another explains that, “Sometimes when are walking along the streets of the market area where the parents work, they call us to check if the homework of their children was correct. If they are ignorant, they don't do that.”[346] In some instances, it appears that community members are even starting to give girls priority over boys: “Now, people understood it, and the girls are educated well. It has been realized that if the boys is educated, at the end of the day he will start having addiction, but if the girl is educated she will benefit her family. So

[343] FGD, Mothers, Puntland
[344] FGD, Girls, Puntland
[345] FGD, Girls, Puntland
[346] FGD, Girls, Somaliland

now girls are given priority."[347] As a result of these changes, community members note that girls' attendance and enrolment has improved and that girls in many instances surpass boys in terms of school performance. Additionally, girls are increasingly breaking previous cultural barriers to continuing their education: "If I talk about my class, there are married girls who still study and continue their learning, and of course we can say that girls' school enrolment has increased over past years because they understood the value of education for themselves and know an educated girl is an educated community."[348]

However, this change is not uniform, and girls still face significant societal challenges to enrolling in and continuing schooling. Societal expectations are perhaps most significant at the household level, where the attitudes of a girls' parents determine whether or not she is enrolled in school at a young age, whether or not her household responsibilities are considered more important than her school responsibilities, whether or not she continues to be enrolled in school year after year, and in some cases, whether or not she remains in school once she is of an age at which she can get married. When asked whether expectations at home are different for boys than they are for girls, one girls explains, "Of course boys are different according to Somali culture. As we know Somali boys don't do house tasks, except small ones who carry small tasks in the house but there is no any boy who will to make breakfast or lunch or clean the house."[349] Pressure also exists for girls to marry, often at the expense of their education. When asked whether he thinks girls should study up to the university level, one father replied, "If the girls spend their whole life studying the university they miss marriage, and the youth don't like to marry old ladies. So I recommend to leave the education after secondary because 30 year old ladies don't have the opportunity to deliver more children compared to the lady who got married at the age of 20."[350]

With regard to girls with disabilities, the parents of the community were often emphatic in their belief that children of all abilities have the right to go to school. As one respondent said, "The human being needs an education, and they deserve educational opportunities. It is very important to prioritize the disabled students since they may be discriminated."[351] Another respondent added, "If you are sending the children to school, then you have to send the disabled girls to school too. Otherwise, they will think that they are neglected when it comes to education. So they must support them and send them to school and let them take part in their community."[352] Alongside a belief in the rights of girls with disabilities to attend school was a sense that educational opportunities did not meet their needs: "There are no schools for disabled children who are blind or deaf…So we need humanitarian support and community-based projects to help the disabled children….When we talk about disability, the girls are more vulnerable, because they have specific needs and challenges. We need to think about how to educate these disabled children because they have the right to get education."[353]

[347] FGD, CEC, Somaliland
[348] FGD, Girls, Puntland
[349] FGD, Girls, Puntland
[350] FGD, Fathers, Puntland.
[351] FGD, Fathers, Puntland
[352] FGD, Fathers, Puntland
[353] FGD, Fathers, Puntland

Attitudes and Perceptions of Girls

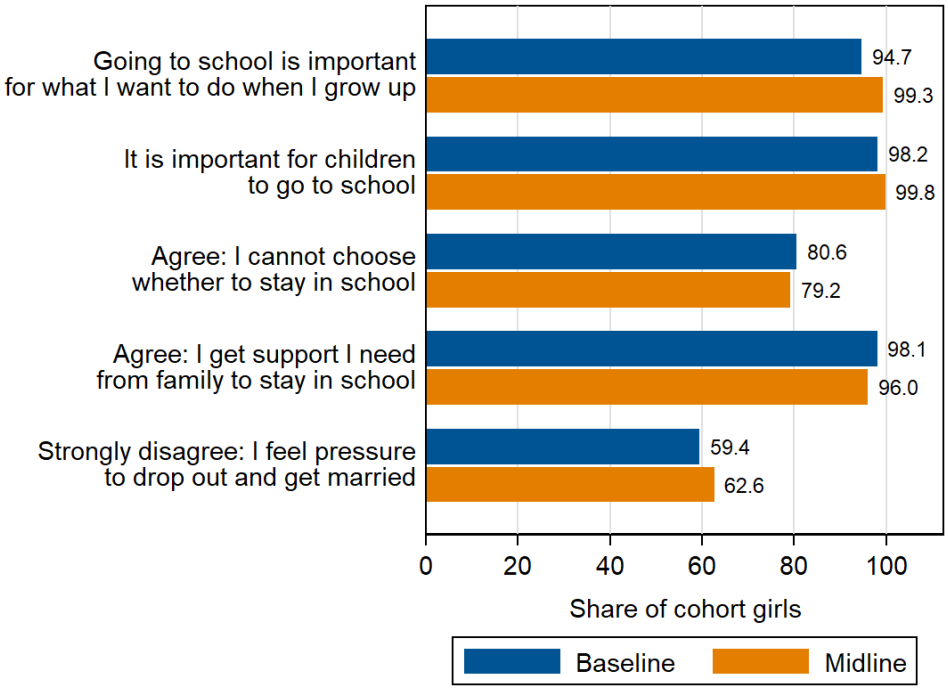
Expanding our interest beyond the narrow group of caregivers analysed thus far, we also surveyed cohort and bursary girls about their attitudes toward education and their perceptions of community and familial support for education. There are subtle differences between these two approaches: we are interested in the attitudes of girls themselves because their continuing and increasing interest in education is critical to encourage enrolment, attendance, and effort on their part. However, we are also interested in girls' perceptions about the attitudes of their families and community members. Girls provide an independent source of information about these attitudes – a source which is generally reliable, because the attitudes in question affect them directly, and which is important regardless of its accuracy, because the extent to which girls' perceive support or opposition to their education is independently important for their educational attainment.

With regard to girls' attitudes, our findings, shown in the figure below, are encouraging. When asked whether going to school is important for their futures (i.e. what they want to do when they grow up), the share of cohort girls who answered yes increased from 94.7 percent at baseline to 99.3 percent at midline.[354] And, while we observe a smaller increase in support for the second statement – which asks whether it is important for children, generically, to go to school – this is not altogether surprising. Baseline support for this statement was very high, and there is a ceiling which the midline results are rapidly approaching, with 99.8 percent of cohort girls surveyed agreeing with the statement.[355]

[354] While it is possible that this finding could stem from sampling variation in the cohort girl sample from baseline to midline (i.e. girls being replaced due to panel attrition), this is emphatically not the case. When we restrict the analysis to cohort girls who appeared in both sample – the "true panel" of cohort girls – our findings are nearly identical, a 4.5 point increase, versus a 4.6 point increase in the full sample.

[355] Despite the relatively small effect size, this finding is a statistically significant improvement from baseline to midline. As with the first statement analyzed, the results are consistent when we limit our attention to the "true panel" of cohort girls.

FIGURE 41: GIRLS' ATTITUDES TOWARD EDUCATION AND PERCEPTIONS OF COMMUNITY SUPPORT



In the case of girls' perceptions, the trend from baseline to midline is less positive. Our assessment of girls' perceptions focus on the extent to which they feel they can control decisions about their enrolment in school, the depth of support they receive from their family to stay in school and perform well, and the extent to which they feel pressure to drop out. The results are mixed – there is no discernible pattern of improvement from baseline to midline. Overall, slightly fewer girls feel powerless regarding schooling decisions, and fewer feel pressure to drop out and get married, but these differences are not statistically significant. In addition, there is a decline in the share of cohort girls who feel they get needed support from their families.[356]

As has been previously noted, there is an abundance of evidence in the qualitative data around changes in attitudes toward girls' education, but traditional ideas around marriage and the role of girls in the household are still being actively debated in Somali society, which may explain these results. Girls still face more pressure than boys to do household chores, help their mothers care for younger siblings, and drop out of school to get married. As one girl explains, "Mostly parents assign us household chores and the boys don't challenge us. On the other hand, they may motivate us to marry and stop education."[357]

[356] Only the finding regarding being pressured to drop out is statistically significant, a baseline-to-midline trend which runs counter to the project's goals and our expectations. Notably, the results are unchanged when we limit our analysis to cohort girls who appeared in both rounds of the evaluation.
[357] FGD, Girls, Galmudug

While we did not find a consistent pattern of changes from baseline to midline, there are interesting patterns in responses across respondent types. When we study differences between cohort girls, bursary girls, and cohort boys, we find that cohort boys report the greatest level of empowerment over educational decisions, which is consistent with cultural explanations in which boys are given greater autonomy and input into decisions that affect them. Cohort boys are also somewhat more likely to believe they receive necessary support from their families, with 86.3 percent of cohort boys and 83.8 percent of cohort girls strongly agreeing with the statement "I get support I need from my family to stay in school and perform well." However, bursary girls outperform both of these groups, reporting the highest rates of familial support, which is perhaps tied to the scholarship support they receive from the project itself.[358]

What do girls think of community attitudes outside their own households? As discussed above, we find this to be an important question, both because girls observe their own communities, but also because their perceptions of community support may influence their own self-esteem and feelings about education. If girls feel they have the support of their communities, they may be more likely to stay enrolled and work to excel in school.

Our approach to this question is somewhat unusual, both because attitudes are latent within the community, and because Somali girls have a tendency to express blanket agreement with positive statements about others (e.g., their teachers, their families).[359] For this reason, we designed a hypothetical scenario in which girls were asked to imagine that a girl and a boy in their community, separately, had both been accepted into university, but lacked the funds necessary to attend. We asked girls to imagine that the two students' families asked for financial support from their community to help them attend university, and asked them to rate the probability that the boy and girl student, respectively, would receive the necessary support from their community. At baseline, this produced a stark finding: 32.6 percent of cohort girls thought it very likely the boy's family would receive the necessary support from the community, while just 12.4 percent thought it very likely the girl's family would receive the same support.

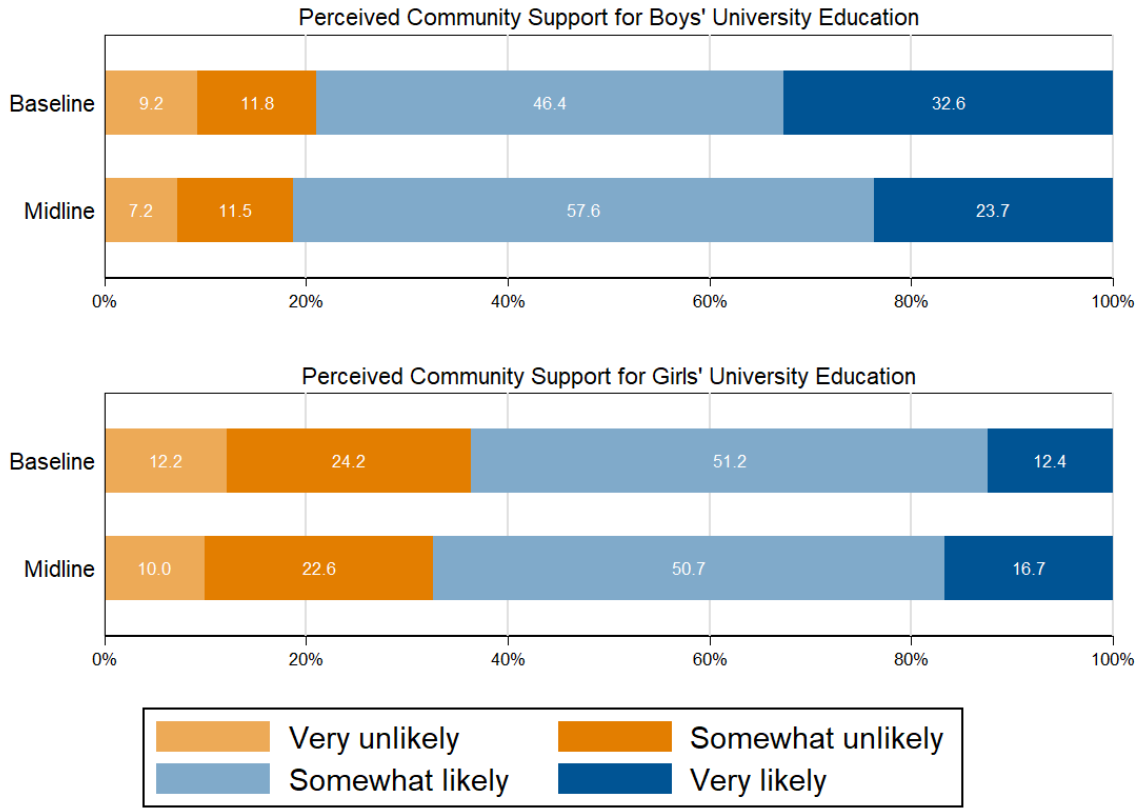
In the figure below, we document shifts in response patterns for this question from baseline to midline. While the results are a bit noisy, the overall trend is that cohort girls perceive greater community support for *both* boys and girls at midline than they did at baseline. From baseline to midline, the share of cohort girls who believe the boy's family is somewhat or very likely to receive support increased from 79.0 to

[358] One possibility is that bursary girls perceive greater familial support because their enrolment does not constitute the same type of financial burden that applies to cohort girls and cohort boys who are not receiving bursary support through EGEP-T. When this financial burden is removed, families may express greater support for the girl's enrolment, and fewer misgivings about their continued education.

[359] A number of counter-intuitive findings in both the baseline and midline EGEP-T evaluation – as well as other educational evaluations we have conducted in Somalia – support this conclusion. Strictly for illustration, we have previously found that girls who report that their teachers are frequently absent or late for class still view their teachers as extremely effective. Similarly, we have found that it is very rare for girls to express negative opinions of their communities, their families, their teachers, and so forth.

81.3 percent. The bottom panel in the figure shows an approximately equal increase in the share of cohort girls who believe the girl's family is somewhat or very likely to receive support.

FIGURE 42: GIRLS' EXPECTATIONS OF COMMUNITY FINANCIAL SUPPORT FOR A BOY OR GIRL TO ATTEND UNIVERSITY



Our takeaway from the figure is that the gap between perceived support for boys and girls has not changed. Girls still believe that the community is much more likely to support a boy's university education than a girl's university education. Again, this may be explained by the qualitative findings outlined above that suggest girls are perceived as less likely to seek employment after marriage, and therefore their education may be prioritized below that of boys. It is also worth noting that the project did not set out to challenge traditional gender roles per se, thus it is not necessarily expected that project activities will result in girls having completely equal life opportunities to boys.

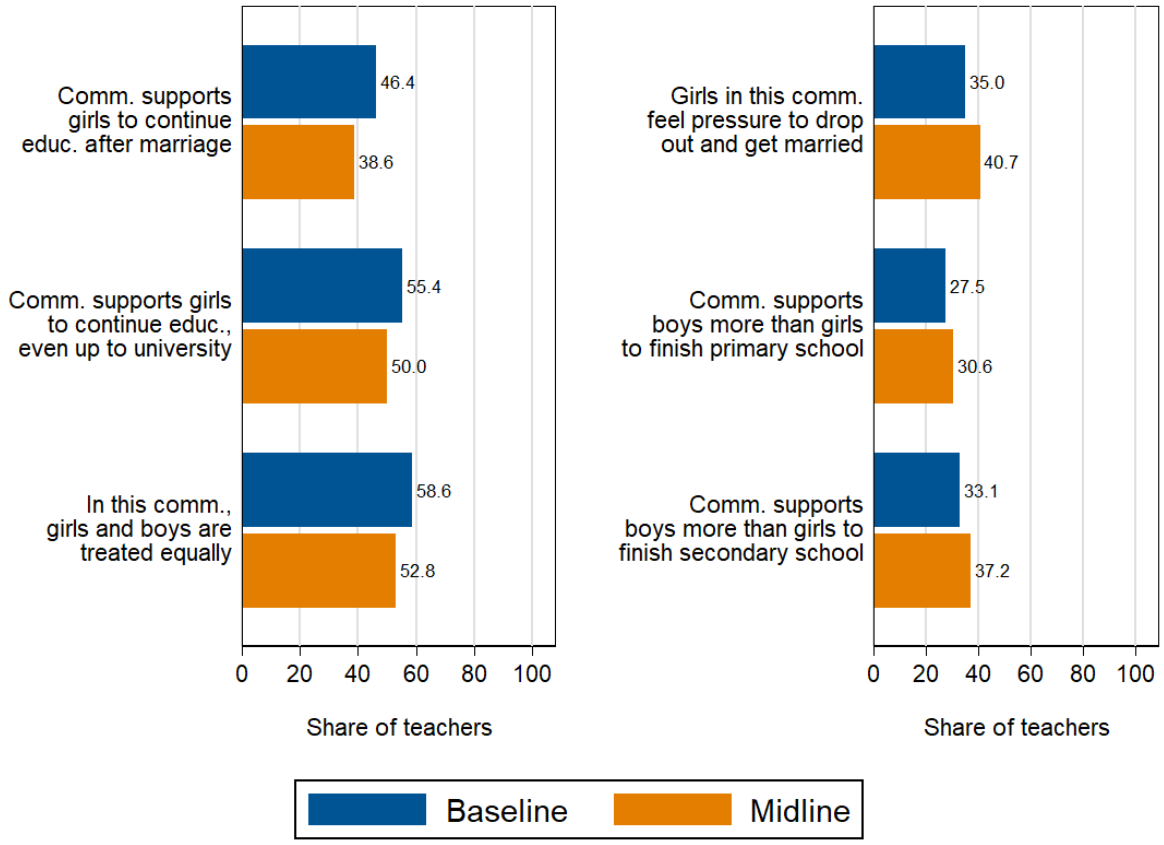
The Views of Teachers

As with cohort girls, we also surveyed teachers regarding the attitudes of the community members among which they live. It should be noted that teachers, being part of the same cultural background as others in the community, have their own sets of biases and hold gender stereotypes which are likely reflected in their responses. Additionally, teachers are trained and may also hold biased expectations in relation to

what they know of the project. These factors serve as important context for the analysis presented in this section.

We asked teachers whether they agreed with a number of statements regarding community attitudes. To illustrate, we asked them whether the community tends to support girls to continue their education after getting married, and gave them four response options (agree a lot, agree a little, disagree a little, disagree a lot). In our analysis, we report the share of teachers who agreed strongly with each statement. The results, in the figure below, are split into two panels: in the left panel, the statements are positive, such that agreement with the statement implies positive community attitudes, and higher values in the graph are a positive outcome; in the right panel, the statements are negative, such that agreement (and higher values in the graph) implies a negative outcome.

FIGURE 43: TEACHER PERCEPTIONS OF COMMUNITY SUPPORT FOR GIRLS' EDUCATION, BY ROUND



The pattern revealed in the figure is consistent: teachers view community attitudes as less positive toward girls' education at midline than they did at baseline. Fewer teachers believe that community members support girls' education through the university level, and fewer agree with the idea that girls and boys are treated equally in their community. In the left panel, all three findings trend in the same direction,

although only the top results – centred on community support for girls to continue their education after marriage – is statistically significant. The lack of statistical significance is not dispositive, however: when multiple, related variables all trend in the same direction, it is strong evidence for a given hypothesis – in this case, that community attitudes, or teachers' perceptions of them, have deteriorated over time – even if individual hypothesis tests are not statistically significant.[360]

As with the results in the left panel, we find a consistent story in the findings reported in the figure's right panel. Teachers are more likely to believe community members favour boys over girls in terms of support for finishing either primary or secondary school. And perceived pressure to drop out and get married has also increased: 40.7 percent of teachers strongly agree that girls in their community face such pressure at midline, compared to 35.0 at baseline.[361]

The geographic distribution of responses reveals that the results reported above are consistent in Somaliland and Puntland, but not in Galmudug. In Galmudug, teachers reported a significant *improvement* in community support for girls' education, while teachers in Somaliland and Puntland reported a decline, in line with the overall results above. The biggest declines were observed in rural areas, but teachers in urban areas also generally reported a decline in community support for girls' education.

Summary Discussion

Drawing conclusions from across these different findings is difficult, due to the disparate data sources used and the gaps in their relative quality. To recap our findings:

- Caregivers expressed slightly higher aspirations for their girls' education, but this finding was not statistically significant and is subject to some concern about sampling variation from baseline to midline
- Caregivers were more likely to take tangible action, in a hypothetical scenario, to support girls' education at midline, compared to baseline

[360] There are a number of methods to test the overall hypothesis of difference from baseline to midline, rather than for differences in individual variables. The most straightforward is Fisher's test of combined probability, but this test assumes the individual tests that go into the test of joint probability are independent, an assumption that is violated in this case, given that the same respondents answered each question. An alternative is to construct an index variable, which aggregates data from each variable into a single, multi-dimensional measure. When we tested this approach using the six variables included in the figure, the finding was stark: an overwhelming difference between baseline and midline, despite the fact that many of our individual findings were not statistically significant in isolated tests. This fact underscores the fact that a consistent trend across multiple individual tests is strong evidence in favor of a hypothesis, even if most of the individual tests are not statistically significant themselves.

[361] In additional analysis, we found that teachers had become more skeptical that their community would support a girl financially to attend university, relative to a boy in the same situation. As with cohort girl respondents, the gap in expected support between a boy and girl facing the same need for financial support remained large between baseline and midline. Likewise, according to teachers, mothers are no more likely to attend a parent-teacher meeting to discuss their daughter's performance at baseline than they were at midline, and the likelihood that a generic, hypothetical father would attend such a meeting had clearly declined from the baseline.

- Girls tend to value education more at midline than at baseline, but do not feel any greater control over schooling decisions than at baseline
- The evidence that girls feel greater community support for their education is mixed and inconclusive, but findings based on girls' perceptions may misrepresent actual community views
- Teachers clearly believe community support for girls' education has declined over time, but this finding could be driven by teachers' own views, such as an increased expectation for community support that they have not seen met in practice

Overall, the evidence for a widespread improvement in community attitudes does not exist. However, small improvements do appear to exist, most meaningfully in the realm of caregiver aspirations, caregiver support for investing in girls' education despite trade-offs, and girls' valuation of education. If these trends continue on the current trend, the endline evaluation will be able to isolate stark differences from baseline to endline, but we also hope to see improvement across a wider range of measures than observed at midline.

7.6 Intermediate Outcomes as Predictors of Learning

This section of the report tests the theory of change by evaluating EGEP-T's five intermediate outcomes – attendance, teacher quality, school management and institutional governance, girls' self-esteem and empowerment, and community attitudes and behaviour – in terms of their relationship to learning outcomes. The project's theory of change implies that there should be positive correlations between improvements in intermediate outcomes and improvements in learning. While there have not been measurable, positive intervention effects on learning in the aggregate, this analysis also provides for a more nuanced examination of why this might be the case.

Before considering intermediate outcomes, it is worth noting that output-level indicators that can be measured quantitatively have not been found to have significant positive effects on learning outcomes. The barriers analysis presented in section 4.5 explores the relationship between a number of output-level indicators and learning outcomes. We ultimately reach the conclusion that none of the school-level interventions provided (e.g. offering of remedial classes, life skills classes, school feeding, or sanitary kits) have had a measurable positive impact on girls' learning. In particular, the analysis above noted that the offering of remedial classes, which is the output-level intervention most directly related to learning, was not a significant predictor of higher learning outcomes. This finding does not necessarily call the project's theory of change into question, but it does suggest that future monitoring of outputs should focus on determining why remedial classes have not had a more direct effect on learning. The other school-level, output-related interventions are more directly tied to the outcome of attendance, and therefore we have weaker expectations that these outputs will be strongly correlated with learning.

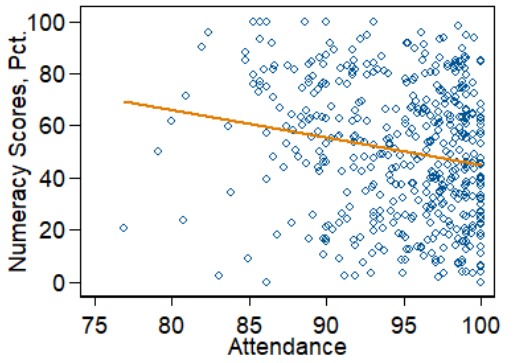
The analysis below also considers participation in Girls' Clubs as a potential intervention that may have affected girls' learning through the mechanism of increasing their self-esteem. While there are some linkages (between self-esteem and learning) to suggest that this hypothesized relationship may exist, ultimately Girls' Club participation is not a strong predictor of learning, and thus this hypothesis remains in question.

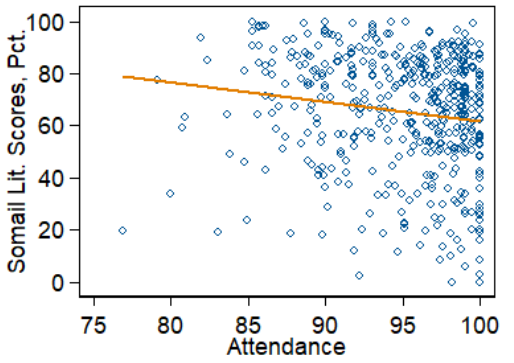
Attendance

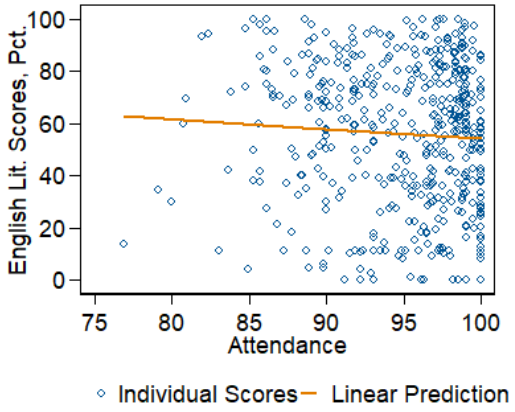
The focus on increasing attendance as an intermediate outcome stems from the simple fact that students need to be in class in order to learn. If teaching in classrooms is effective, higher numbers of days missed should be associated with lower learning scores.

The panel of graphs below shows the relationship between attendance rates (estimated through school attendance data collected as part of the School Survey) and learning outcomes. Readers should note that the graphs below are based on a truncated sample of 475 cohort girls whose data could be matched to the attendance register, and whose attendance rates were above 50 percent. A very small number of girls had attendance rates below 50 percent and these outliers were skewing the attendance distribution as well as making the graphs below more difficult to read. On the graphs, the blue dots represent individual scores, while the orange line indicates the line of best fit in a linear regression.

Contrary to our theoretical expectations, the relationship between attendance rates and learning is negative. On average, higher attendance rates are predictive of lower learning outcomes. This counterintuitive trend may be a result of the fact that older girls (in higher grades) tend to have higher learning scores on average, but also tend to have more household responsibilities that may prevent them from attending school regularly. A parallel explanation for this finding is the fact that teaching quality has not improved substantially since baseline, and thus the potential benefits of higher attendance are not being fully realized because lower quality teaching means that time spent in class is not as productive as it might be with higher quality instruction.





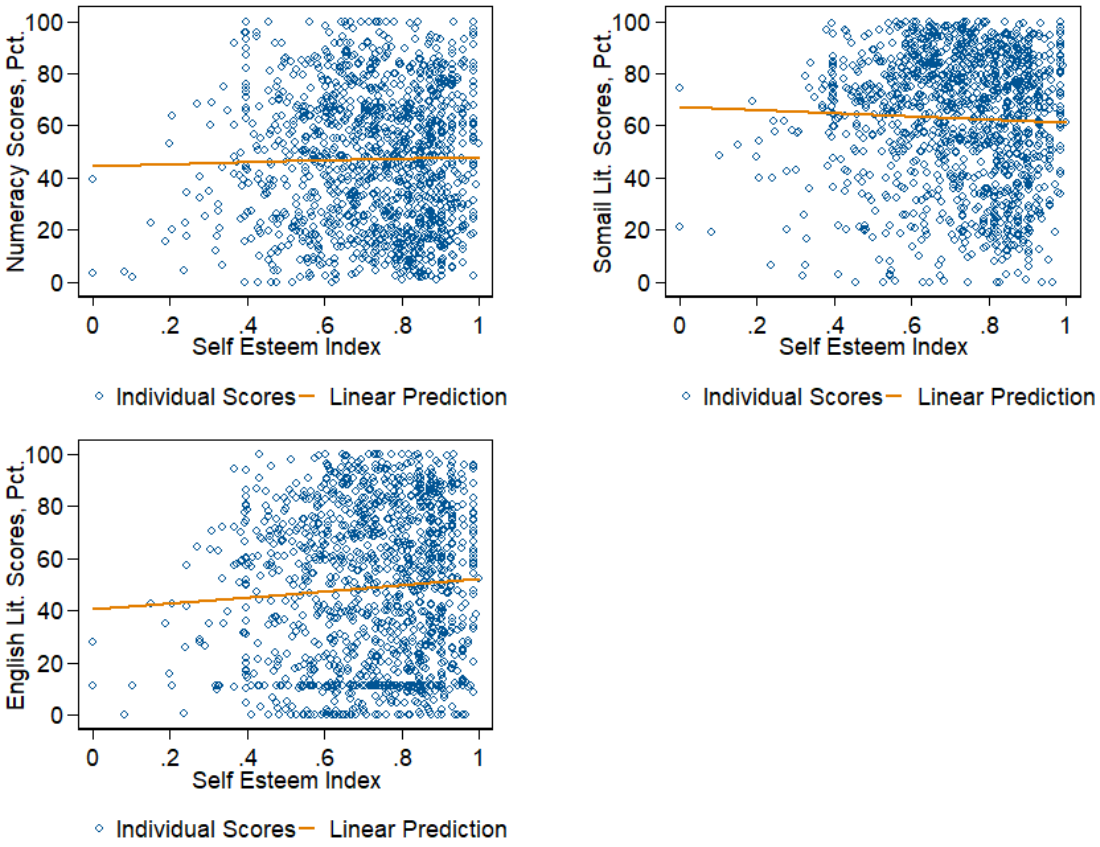


Self-esteem

Self-esteem may affect learning through a number of mechanisms, but the most proximate connection between self-esteem and learning is the fact that girls with higher self-esteem may be more willing to participate in the classroom, thereby learning more through their in-class practice, and thus performing better of learning assessments. More indirectly, girls with higher self-esteem may be able to exercise more control over their home life (including better management of their chore burden, and standing up for their need to attend school regularly), and resultant increases in attendance may thereby lead to increased learning.

The panel of graphs below shows the relationship between the self-esteem index (presented in the section on self-esteem above) and learning assessment scores. As above, the blue dots are individual scores, while the orange line indicates the line of best fit in a linear regression. For numeracy and English literacy, there is a positive relationship between girls' self-esteem as measured through the index, and their assessment scores. For Somali literacy, this relationship is slightly negative. The positive relationship between English

literacy and self-esteem is the strongest correlation observed, and is statistically significant at the 0.05 level.[362]



The analysis of self-esteem in section 7.2 above identified an important correlation between girls participating in Girls' Clubs and their self-esteem, suggesting it might be the case that participation in Girls' Clubs is affecting learning through the mechanism of increased self-esteem. However, participation in Girls' Clubs is not positively or significantly correlated with increases in learning scores. Other output-level indicators related to Girls Clubs (including whether or not a given school has Girls' Clubs and whether or not the club meets at least once a week) are also not predictive of higher than average learning scores. These findings suggest that the linkage between Girls' Club participation and learning is not yet clear. Increasing self-esteem is still a plausible mechanism, but this mechanism is not yet verified in terms of the quantitative evidence at this point.

[362] In a bivariate, cluster-robust regression using self-esteem index to predict English Literacy score, p=0.05.

Teaching Quality

Of all intermediate outcomes considered, teaching quality should have the most direct effect on learning. Analysis of teaching quality above has suggested that teaching quality has not improved measurably since baseline, and this lack of improvement in teaching quality helps to explain the general lack of a measurable intervention effect on learning, as well as the finding above that attendance is not positively correlated with learning (since quality teaching is the mechanism through which higher attendance might translate into higher learning outcomes).

Nonetheless, the findings from the learning barriers analysis above re-emphasizes the importance of teaching quality as a primary mechanism for improving learning. Readers will recall that low teaching quality (measured as girls reporting that their teachers make them feel unwelcome and girls reporting being afraid of their teachers) are both predictors of lower learning outcomes across all three assessments (and to a statistically significant degree across all three assessments in the case of girls reporting that they are afraid of their teachers). Similarly, caregiver reports of low-quality principals and low-quality teachers are also consistently correlated with lower learning outcomes (with the correlation being statistically significant between numeracy and poor teaching quality and between English literacy and poor principal performance).

School Management and Governance

School management bears a far less direct relationship to student learning, but may still affect learning by contributing to teacher motivation and quality, as well as by contributing to key infrastructure and other school resources that are important barriers to learning when absent. The analysis of school management above suggested that school management has not changed appreciably since baseline, and this helps to explain why having an active CEC was not predictive of higher than average learning outcomes in the barriers analysis above.

Nonetheless, one finding from the learning barriers analysis suggests that there are still important connections between school management and learning. Readers will recall that one of the stronger predictors of lower learning outcomes (among the barriers analysed earlier) was schools where classes were held for fewer than 5 hours during the day. Girls in schools with shorter school-days scored substantially lower on all three learning assessments (and to a statistically significant degree across all three assessments). While shorter than average school days is not a direct measure of school governance it is certainly a condition that is produced by school governance, and likely by a lack of adequate funding. This finding provides preliminary evidence that improving school governance may be able to improve learning outcomes for some girls by ensuring that schools do not have to shorten their schooldays.

Community Attitudes

Shifting community attitudes related to girls' education is an important intermediate outcome targeted by the project in order to increase gender equity writ large. However, the potential benefits of shifting

community attitudes are diffuse with respect to learning outcomes. Readers will recall from the discussion above that the goal of shifting community attitudes is based on the idea that individuals' underlying attitudes influence their willingness to invest in girls' education, support it vocally and through tangible actions, and encourage girls' education within their own families and their broader communities. If parents, community leaders, and young men all begin to value girls' education more highly, they will be more likely to pay school fees for their daughters, promote delayed marriage for girls in favour of completing their schooling, and provide incentives for greater educational attainment by girls. Evidence from the analysis of community attitudes above suggested that there have been, at best, modest improvements in some areas, including in terms of caregivers expressing high aspirations for girls in terms of their education.

The learning barriers analysis does provide some evidence of linkage between caretaker attitudes and girls' learning. Readers will recall from above that caretakers' attitudes toward boys' and girls' education were assessed by presenting caretakers with a hypothetical scenario in which they had the choice to choose how to use limited household resources and could choose between prioritizing boys' education over girls', vis-versa, or prioritizing both equally. Girls belonging to households where caregivers suggested that they would prioritize boys' education over girls' (in the hypothetical scenario) scored lower on all three learning assessments (with the correlations for numeracy and Somali literacy being statistically significant).

7.7 Intermediate Outcomes as Predictors of Transition

As with learning, EGEP-T's Theory of Change links intermediate outcomes to the higher-level goal of promoting successful transition. In this section, we consider the evidence in support of this idea, by testing a series of five distinct hypotheses: whether self-esteem, attendance, teaching quality, school management, or community attitudes influence the likelihood of successful transition for individual girls. This analysis builds on and expands similar discussion in the baseline report, where we investigated the evidence for the project's Theory of Change in the context of baseline data. Here, we examine transition outcomes among cohort and bursary girls at the time of the midline.

Attendance

While attendance has a slightly negative impact on learning outcomes – per the discussion in Section 7.6 – we have fewer theoretical expectations regarding the relationship between attendance and transition outcomes. It is natural to assume that poor attendance and unsuccessful transition would be correlated, for at least two reasons: first, we would expect girls with poor attendance to be less likely to advance to the next grade, typically because they have poor grades or do not pass qualifying examinations to enter secondary school. Second, girls with poor attendance may have lower motivation levels, less support at home, or otherwise be predisposed to dropping out – we can consider dropping out entirely as the logical extension of particularly poor attendance. However, these two mechanisms have different interpretations

in terms of the Theory of Change: the former implies a causal relationship between attendance and transition, while the latter simply suggests they are correlated and co-determined. Given the nature of the data available in this evaluation, it is not possible to distinguish between these and other explanations, but we are able to study the extent of correlation between attendance and transition outcomes.[363]

In line with the findings above regarding learning, attendance does not appear to be a predictor of higher transition rates. Among a sample of 595 cohort and bursary girls for whom we have attendance data that can be matched to their transition outcomes at midline, higher attendance rates predict a greater likelihood of being held back in school, in a linear regression model that accounts for clustering at the school level and controls for respondent type (cohort versus bursary girls) and geographic zone. The resulting relationship is quite strong: a 10-point increase in attendance rates is associated with a 7.4 point drop in successful transition. Note that girls who did not successfully transition were held back a grade – no other outcome is possible in this data; thus, our results suggest that girls who were held back tend to have higher attendance rates.[364] To illustrate these findings, consider two broad groups of girls: those who attend school between 80 and 90 percent of days, and those who attend 95 percent of days. Among the former group, transition rates are 95.4 percent; among the latter group, transition declines to 89.2 percent. We interpret this finding to suggest that attendance is not directly related to the likelihood of a girl advancing grade levels, with the important caveat that we are not able to assess the relationship between attendance and dropout rates.

Self-esteem

In our discussion of self-esteem and learning, we noted that higher self-esteem could influence learning outcomes by promoting girls' engagement and participation in the classroom, and their willingness to seek help from their teacher. In the context of transition, we view self-esteem as an even more direct contributing factor. Girls with higher self-esteem are more likely to believe they *should* be in school and perceive themselves as "smart enough" to do well. While a number of factors beyond a girl's self-perception influence enrolment decisions, we expect higher self-esteem to be related to an increased probability of successful transition.

As in our other analysis focused on self-esteem, we employ an index composed of between 17 (for out-of-school girls) and 19 (for in-school girls) individual indicators. These scores capture a broad range of behaviours and opinions, such as a girl's assessment of whether she can read as well as her peers, whether

[363] It is important to note the source of our attendance data, which are drawn from school records assessed at midline. That is, this section examines the relationship between *midline* attendance and transition from baseline to midline. In practice, this means that we are unable to assess the relationship between attendance and dropping out of school entirely, limiting our attention instead to the relationship between attendance and being held back a grade versus successfully transitioning to the next grade. Every girl included in the analysis remains enrolled in school, which is a necessary condition for having attendance data to analyze.

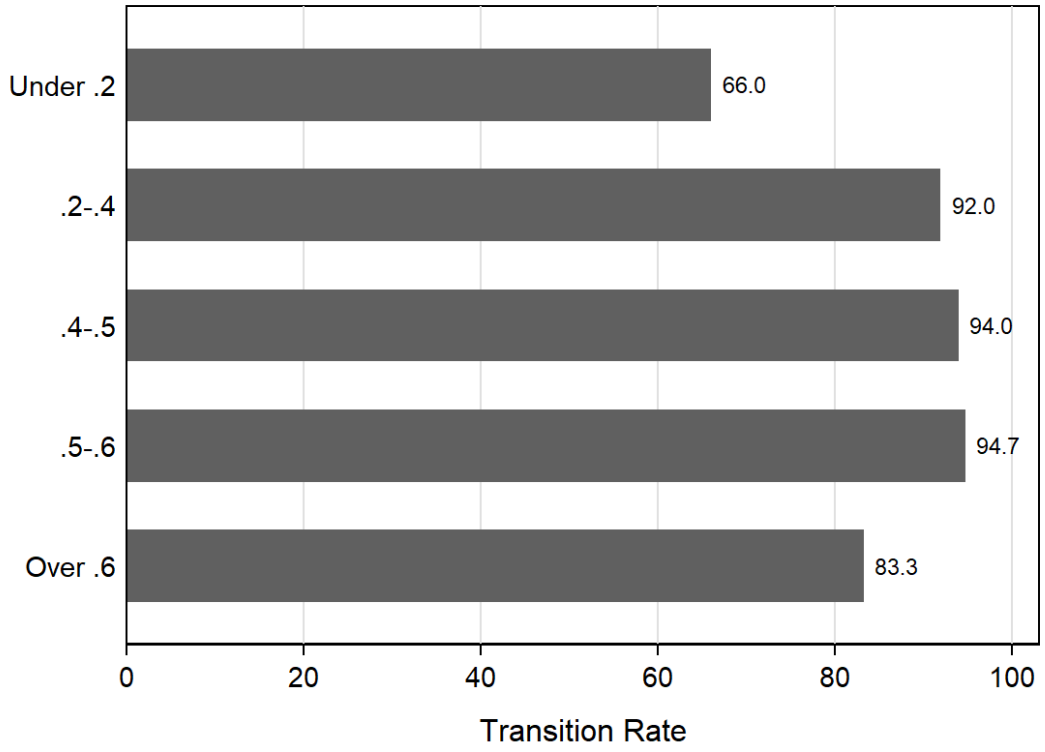
[364] One possible optimistic explanation for this finding is that being held back motivates girls and their families to attend school more consistently in the future, with the goal of advancing grade levels in the next year.

she feels confident answering questions in class, and whether she believes she has control over schooling decisions that affect her. The index ranges from 0 to 1, with higher scores representing greater self-esteem.

Our primary results are based on a linear regression, in which we use the normalized self-esteem index to predict transition rates. The results of this model are striking – a 0.1 point change in self-esteem (on a 0-1 scale) results in a 1.7 point higher likelihood of successful transition ($p = 0.08$). Because an index like the one we use for self-esteem is non-intuitive to interpret, it is easiest to consider what would constitute a substantial but *plausible* change in self-esteem. We define this as a 1-standard deviation change in self-esteem, in line with conventional rules of thumb for interpreting statistical results. Our model predicts a 1-standard deviation change in self-esteem is associated with a 3.2 point increase in transition rates. This finding is robust to clustering and the inclusion of variables controlling for respondent type and zone.

In practice, the relationship between self-esteem and transition is more complicated still. First, our results are driven almost entirely by girls with particularly low self-esteem – those in the bottom quintile. As shown in the figure below, girls scoring below 0.2 on the self-esteem index have transition rates of just 66.0 percent. When we exclude these girls from our analysis, self-esteem is *negatively* related to transition. That is, there is a quadratic relationship between the two variables, as shown in the figure below. In brief, girls with extremely low self-esteem are likely to drop out; among girls above some minimal threshold of self-esteem, it does not seem to factor into transition outcomes in a meaningful way.

FIGURE 44: TRANSITION RATES, AS A FUNCTION OF GIRLS' SELF-ESTEEM SCORE (0-1)



The analysis above is subject to at least one methodological reservation surrounding measurement of self-esteem. In-school and out-of-school girls were asked a slightly different set of questions to gauge their self-esteem, which may yield different values for girls of objectively identical self-esteem levels. For instance, in-school girls were asked their level of agreement with the statement "I feel confident answering questions in class", while out-of-school girls were asked to assess the statement "I feel confident answering questions when I'm in a group of people". In total, six questions were re-worded or adjusted to be context-appropriate for out-of-school girls.

To guard against issues of measurement validity, we constructed an alternative self-esteem index that utilized only the indicators that were asked identically for both types of girls.[365] In total, the measure incorporated 11 indicators, instead of up to 19 used for in-school girls in our main analysis, but is significantly more valid for the purposes of this section. Using this alternative index, we re-estimated our linear model and found results that were substantively unchanged. In the full sample, self-esteem is positively correlated with transition rates – even more so than when using the original measure – but this

[365] Such an approach is not necessary in analyses reported elsewhere, such as the self-esteem section, because those sections do not assess the relationship between transition, or enrolment status, specifically, and self-esteem. The difference in question construction is problematic in the context of studying transition because the divergence in measurement is correlated perfectly with the outcome variable of interest.

effect is caused entirely by very low transition rates among girls with extremely low values of self-esteem. It is also noteworthy that low self-esteem is specifically tied to a tendency to drop out, as opposed to being held back a grade: among girls who dropped out, mean self-esteem was 0.14; among girls held back and girls who transitioned successfully, average self-esteem is nearly indistinguishable at 0.78 and 0.75, respectively.[366] The results suggest that interventions oriented toward improving girls' self-esteem could be most usefully targeted to those exhibiting particularly low self-esteem.

Teaching Quality

In this section, we consider the relationship between teaching quality and transition outcomes. While we would expect teaching quality to have the greatest direct impact on learning outcomes, it could also influence transition rates, either through an indirect effect on transition scores, or because girls and their families may see greater value in education if their teachers are of high quality. To test this possibility, we employed two measures of teaching quality derived from girls' survey responses: the frequency with which teachers encourage participation in class, and how much classroom time they spend on rote repetition. Because girls were only asked these questions if they were *presently* enrolled in school, we aggregate responses to the school level, yielding an indicator that captures, for instance, the average response regarding participation in a given community.[367] These measures vary from 0 to 100 in theory but, in practice, vary from around 50 to 100 percent.

The data do not suggest a relationship between the use of rote repetition and transition rates. While school-level reports of rote repetition are correlated with higher transition rates, this effect is too small and noisy to be differentiated from a null result. However, classroom participation does appear to have a relationship to transition: in a linear model, we find a strong and consistent positive effect of participation on successful transition. For every 10-point increase in the share of girls who report the encouragement of participation by their teachers, transition rates rise 3.6 percentage points, or even more in saturated models that control for additional factors known to predict transition outcomes, such as grade level. This effect is statistically significant at the 5 percent level across multiple models and – to emphasize its substantive size – is approximately as large as the gap in transition rates between the lowest-performing zone (Somaliland) and the highest (Galmudug).

[366] Of course, it is possible that dropping out itself reduces a girl's self-esteem, or that the two are correlated without one causing the other.

[367] Our measure of participation allowed responses of "often", "sometimes", "rarely", and "never." We transform this variable to a binary measure, distinguishing between respondents who responded "often" or "sometimes" and all others. When aggregated to the school level, our measure captures the share of in-school girls in a given community who reported their teacher often or sometimes encourages participation. Similarly, our measure of rote repetition captures the share of in-school girls in a given community who agreed, strongly or somewhat, that they spend much of their time on rote repetition.

School Management and Governance

In many ways, school management is likely to have the most diffuse impact on transition outcomes. While a well-managed school can influence students and their families in many ways, most of these linkages are only related to transition decisions indirectly. For instance, a well-managed school might provide better teaching, better value to parents, or an increased chance of qualifying to attend secondary school or university, and thus shape enrolment decisions. Similarly, a well-managed school might handle its own internal finances more efficiently, charging lower schools, or implement higher-quality pedagogical practices. However, when compared to direct measures of teaching quality or individual-level characteristics, such as self-esteem, school management and the quality of school governance is less likely to influence transition rates.

We assessed three measures of school management and their relationship to transition outcomes:

- The share of caregivers rating school management excellent
- The share of caregivers rating the school's head teacher excellent
- The share of caregivers reporting improved school management over the prior year

In each case, we aggregated the indicator to the school level, because the caregivers of OOS girls were not asked questions about school management. The resulting measures provide an indication of community views of the school, its head teacher, and the changes over the last year.[368]

Based on data for 1,390 students across 115 schools, there is no systematic relationship between the quality of school management and transition outcomes. In general, positive caregiver ratings of school management were associated with slightly higher transition rates at the school level, but these effects were too small to approach statistical significance in any of the regression models we employed. For instance, a 10-point increase in the share of caregivers rating school management as excellent was associated with an increase in transition rates of just 0.7 points ($p = 0.47$). Even weaker results obtained in the case of head teacher quality and improvements in management quality over the past year actually weakly predicted *worse* transition outcomes, possibly owing to poor school management in the recent past.

Community Attitudes

Pro-education community attitudes are the final intermediate outcome, and one that is most closely tied, theoretically, to transition outcomes. Specifically, pro-education attitudes among caregivers and heads of household are expected to promote transition by increasing the willingness of families to invest in girls'

[368] Caregiver assessments of school management and the head teacher are related but not perfectly correlated. Overall, 81.9 percent of caregivers gave the same ratings for school management and the performance of the head teacher, but a substantial minority of respondents distinguished between school management and head teachers, sometimes dramatically so (i.e. rating the school as extremely well-managed but the head teacher's performance as poor).

education, and pro-education attitudes in the broader community – not just within a given household – may have more diffuse, but still positive, effects on enrolment decisions and the desire of girls to remain in school.

As we discussed in Section 5.3, however, where we analysed transition outcomes in relation to key expected barriers, there is little evidence that caregiver attitudes influence transition rates in the midline data. We extended the analysis presented in Section 5.3 by employing linear regression models that controlled for additional factors, such as respondent type and zone, known to influence transition outcomes. Despite this, the results are broadly, though not entirely, consistent with those reported previously.

Caregiver aspirations are correlated with transition outcomes. In cases where caregivers aspire to a college education for their daughter, transition rates are 6.1 points higher than in comparison households, though this finding is only marginally significant, depending on the model. Consistent with the idea that aspirations influence transition, though, girls in households with higher caregiver aspirations are simultaneously less likely to drop out *and* to be held back a grade, which suggests that higher aspirations produce pro-enrolment decisions *and* greater investment in learning during the school year (shown by the relatively smaller number of girls in such households who are held back).

However, as we noted in our earlier discussion, we view caregiver aspirations as potentially problematic in measurement terms, insofar as it is subject to social desirability bias and aspirations are costless to the person reporting them. Our approach was to pose hypothetical scenarios in which caregivers would be forced to choose between schooling for their daughter and other desirable outcomes, such as accepting a marriage proposal. In households where caregivers indicated they would sell household goods or livestock rather than withdraw their daughter from school in the face of a family financial emergency, girls are no more or less likely to transition successfully. In fact, dropout rates among this subset of girls are actually higher than average.

Similarly, support for schooling over marriage does not appear systematically related to transition outcomes. When a caregiver states that they would encourage their daughter to continue their schooling rather than accept a marriage proposal, transition rates are nearly identical to those that obtain when a caregiver supports marriage instead. In a final scenario, we asked caregivers to consider a situation in which their daughter *wanted* to get married, and asked them whether they would encourage their daughter to continue their schooling *despite* being married, or if they would discourage continued schooling in that case. Support for schooling in this scenario is predictive of higher transition rates – about 2.7 points higher among girls whose caregivers would encourage continued schooling – but only marginally significant in statistical terms ($p = 0.10$). The weight of the evidence suggests that caregiver attitudes are only very loosely related to transition outcomes, at least as the former is currently measured.

8. Conclusions & Recommendations

8.1 Conclusions

Learning

The results of arithmetic difference in differences analysis suggest that there have been no measurable intervention effects on cohort learning since baseline. The same findings also hold for bursary girls. When testing for differential intervention effects between cohort girls and bursary girls, we find that bursary girls are keeping pace with cohort girls, and may be closing the gap with cohort girls in numeracy and Somali literacy. When comparing bursary girls and boys, we find that cohort girls are keeping pace with boys, but not closing the substantial gap that has existed since baseline. Finally, we find that girls in baseline grade 8 (midline F1) have markedly attenuated learning vis-à-vis their peers in other grade-levels as well as when compared with benchmark expectations or when compared with boys at the same grade-level. Qualitative evidence suggests that this dip or plateau in learning may be occurring as girls enter secondary school because the transition from primary to secondary school is difficult for all students, but particularly for girls, because this is also at a critical age when girls tend to take on much higher levels of household responsibility and chore burden, all of which detracts from their ability to study at home and attend school regularly.

Among girls, the strongest indicators of marginal status (signalled by lower than average learning outcomes) are membership in a poor household and living in a rural community. In addition, girls' whose caregivers or heads of household have never received any formal education score much lower than average across all learning assessments, as do girls who have a disability, and girls who are old for their grade. The qualitative evidence from caretakers and teachers reinforces the fact that girls with disabilities and girls who are old for their grade are at a significant disadvantage relative to their peers, with qualitative respondents suggesting that having extreme mixes of students (in terms of ages and ability levels) within a single grade is a daunting challenge for teachers, and one for which they do not feel they have been adequately trained. The primary barriers to learning are limited school infrastructure and school resources, as well as low teaching quality.

Transition

Relative to the baseline benchmark sample, transition rates at midline have increased by 0.3 percentage points at midline as compared with the baseline benchmark from 88.7 percent to 89.0 percent. This difference is not statistically significant at any standard level. The observed improvements in transition come from an overall reduction in dropout rates, which had declined from 6.8 percent to 3.7 percent from baseline to midline. This reduction in dropout rates is statistically significant at the 5 percent level. A decline in dropout rates is meaningful, even if much of the shift has been toward girls repeating a grade, because it means girls remain enrolled, continue learning, and may continue to progress in school during the following year.

Subgroup analysis did not reveal strong predictors of successful transition, in general. Substantial individual-level barriers to education, such as motherhood and early marriage, were associated with lower transition rates, as were some forms of physical impairment, though the latter were based on exceedingly small sample sizes. Despite ample qualitative evidence that the financial burden of schooling on families poses a barrier to continued enrolment, girls from economically marginalized households do not have systematically lower transition rates.

Sustainability

Improvements in sustainability have occurred at the level of schools, while community-level indicators of sustainability have stagnated at their baseline values. The project defined sustainability in schools based largely on material support from CECs; CECs are more likely than they were at baseline to provide bursaries to girls in their communities, which should encourage higher enrolment rates and improved transition outcomes even after bursary support from EGEP-T ends. However, increased support in one realm appears to have been offset by decreased financial support from CECs for school improvements -- CECs are less likely to provide this support at midline, possibly because financial resources have been shifted to the provision of bursaries. Perhaps the most significant gains in sustainability have been in the implementation of child mechanism procedures. Schools are much more likely at midline to have appointed a focal point within their school for child protection issues, are more likely to maintain a record of such cases, and are more likely to have an established procedure for handling, escalating, and following up on reports. These improvements are reflected in the comfort level of girls, more of whom feel they could report mistreatment by a teacher or bullying by a male classmate now than at midline.

Where sustainability has not improved over time is in community attitudes toward girls' education. Awareness-raising activities within communities, meant to encourage enrolment and galvanize public support for girls' education, have not increased since the baseline, though the baseline likely captured a concerted increase in awareness-raising at the start of the school year. Unfortunately, male support for girls' education – in the form of tangible actions taken by men and boys in the community to support their sisters and daughters to complete school – has not changed since baseline.

Net improvements in sustainability have been driven primarily by changes in school-level outcomes, as noted above. In general, sustainability indicators based on participation in project interventions – teacher mentoring and the establishment of child protection procedures – have shown the greatest gains, while indicators based on changes in community sentiment or tangible actions by community members have shown no progress.

Attendance

The average rate of attendance among panel cohort girls has increased moderately from baseline to midline. Girls in the midline are significantly less likely than panel girls in the baseline to have attendance rates lower than 85 percent, 80 percent, and 75 percent. Girls in Puntland attended school at significantly lower rates than girls from other zones, while girls in Galmudug have significantly higher attendance rates. This difference in attendance by geographic area appears to be linked to Puntland’s sustained conflict

across baseline and midline evaluation points. While few barriers were found to be strong correlates of attendance, we found that among panel girls in the midline, having over 30 minutes of travel time to get to school and feeling unsafe on the way to school were significant predictors of lower attendance rates.

In the qualitative data, we found that girls were said to miss school commonly because of household obligations while boys largely missed school due to distractions outside of school such as staying up late to play or watch football, smoking with friends, or chewing khat. Findings from the qualitative data suggested that the causes of girls' absences from school were being addressed through EGEP-T's intervention. Targets for girls' overall attendance in the attendance record data of cohort girls and in the headcount data were met, but the boys' headcount attendance target was not achieved. Lastly, we find that girls in urban, IDP, and conflict-affected areas achieved headcount attendance targets, however attendance targets were not met in rural, non-IDP, and non-conflict areas.

Self-esteem

It was found that girls felt more confident in their ability to lead a group of their peers to do an activity when compared to boys. They also indicated significantly higher levels of self-esteem and agency when compared to the baseline data. The changes in girls' levels of self-esteem were still statistically significant even after controlling for girls' school level factors as well as demographic and zonal characteristics including age and grade. The findings from interviews and focus group discussions with girls also confirmed that girls noticed improvement in their confidence and self-esteem as a result of participating in the EGEP-T project activities such as awareness-raising programs around the importance of girls' education as well as interactive class activities.

Teaching Quality

There has not been a measurable positive intervention-effect on teaching quality since baseline. In the aggregate, the average teaching quality score across all midline schools was 63.9, while the baseline average was 67.3. Thus, the teaching quality score decreased by a total of 3.4 percentage points from baseline to midline, but this difference is ultimately not statistically significant. Efforts to train teachers through Continuing Professional Development (CPD) appear to have produced improvements in the teaching practices used by this subset of teachers, but the results are based on a small sample and should be considered with caution. The improvements in CPD-exposed teachers is consistent with the idea that trained teachers use more participatory, interactive teaching styles. Triangulating the teaching quality score with other ancillary indicators of teaching quality suggests that teacher motivation has likely increased since baseline, as reflected in reduced teacher absenteeism and increased teacher preparation, there is broadly no evidence of a measurable intervention effect.

School Management and Institutional Governance

In the aggregate, schools improved only slightly from baseline to midline, improving 2.8 percentage points on a standardized 100-point scale. Improvements were concentrated in Puntland; in fact, Puntland is the only geographic zone that showed improvement in overall school management.

The score for school management captures the quality and activity level of CEC management, the level of tangible CEC support for the schools and girls in them, the enactment of formal school-level policies key to school functioning and improvement, and the quality of school record-keeping. The biggest gains from baseline to midline were seen in the activity levels of CECs. CECs are also more likely at midline to provide bursary support to girls in the schools they manage. At the same time, CECs are somewhat less likely to make financial contributions to manage or improve the school, which may suggest there is a substitution effect, in which CECs can increase their bursary support, but only at the expense of other monetary needs.

Community Attitudes

The share of caregivers who expressed high aspirations for their daughters' education increased modestly from baseline to midline, but this difference was not sufficiently large to be distinguishable from a null effect. Gains on this indicator were concentrated exclusively in Puntland, where the share of caregivers who aspired to send their daughters to college rose from 88.3 percent to 92.3 percent.

Findings related to potential improvements in community attitudes were mixed. Caregivers had only slightly higher aspirations for their girls in terms of educational attainment, but they were more willing to make tangible sacrifices to support their girls' education (when prompted to in a hypothetical scenario) than they were at baseline. Girls themselves appear to be more inclined to stay in school and recognize its importance. However, while girls' views on education seem to have improved, their control over educational decisions that affect them has not increased, at least according to their own self-reports: slightly fewer girls feel they get support from their families to stay in school and do well, and slightly fewer feel they can choose whether to stay in school or not.

Project Approach to Gender Inequality

Gender Analysis

EGEP-T interventions have the potential to be gender-transformative, but this potential has yet to be realized at this evaluation point.

The vast majority of EGEP-T interventions are targeted to specific sources of gender inequality in educational attainment. The following output-level interventions target gender-specific barriers in ways that have the potential to be gender transformative:

- **Provision of bursaries and cash grants –** Limited financial resources at the household level are, arguably, the single largest barrier to education in Somalia. But limited resources disproportionately affect girls – when forced to choose which child they should educate, most families select their sons.

- **Provision of solar lamps –** Girls in Somalia are typically charged with a heavy burden of household chores, in addition to their studies. These responsibilities reduce their ability to study after school, impacting learning outcomes, and often cause tardiness in the mornings.
- **Female teacher-mentors –** Unlike their male counterparts, many girls do not have educated female role models at home or at school. This can contribute to a situation in which girls see less value in education; within schools, it can contribute to discomfort and unwillingness to participate in class. Female teacher-mentors should provide valuable psychosocial support to girls, act as role models, and serve as confidants on topics where girls are more comfortable speaking with women than men.

Even at the higher level of policy development, EGEP-T's efforts are potentially gender-transformative. The project will assist in the drafting and implementation of an explicit gender strategy within, for instance, the federal Ministry of Education. These efforts have, arguably, the greatest potential for influencing gender inequality in education, even beyond project schools.

To briefly rehearse findings that are relevant to measuring gender inequality: In learning, a significant gap between girls and boys was found at baseline and persists at midline. In transition, boys' outcomes are not being tracked, and this is problematic in terms of the potential to assess gender transformative effects on transition within EGEP-T. This is particularly true for the potentially gender-transformative outputs described above, because the principle effects of bursaries and the other interventions listed are likely to be on girls' abilities to remain enrolled. Without being able to draw direct comparisons with boys' outcomes, the effects of these interventions on gender equality will remain unclear.

Analysis of community attitudes suggests that, at best, there have been modest improvements in attitudes toward girls' education, and qualitative evidence suggests that girls still have more challenges than boys with regard to feeling safe on the way to, and at, school. As girls age, they also face more pressure than boys to miss school to perform household duties, and to drop out of school to marry. Qualitative evidence suggests that social barriers, such as early marriage, are being actively contested, and thus more progress may be evident at the next evaluation point.

Continuing Professional Development for teachers is likely to be one of the more impactful interventions in terms of learning outcomes, and this intervention is designed in a way that is gender neutral. CPD emphasizes inclusive and learner-centred pedagogy. While the CPD training curricula is less heavily focused on gender-specific teaching strategies than it was during phase 1 of EGEP, it still includes material on gender-sensitive teaching. Moreover, the curricula has been expanded beyond gender-responsiveness to focus additional attention on catering to many different students' diverse needs. The emphasis on tailoring teaching strategies to students' specific needs should improve the classroom environment for girls, students with disabilities, students with special educational needs, and all other marginalized groups. For instance, teachers are being trained to adjust the level of their lessons to the needs of their students and incorporate activities and explanations of diverse difficulty to help students of all levels remain engaged and able to learn. The use of formative assessments to gauge student progress should also improve inclusivity, as formative assessments naturally take into account students' baseline learning

levels, their particular strengths and weaknesses, their specific needs, and other factors that are particular to each student.

Ultimately, the CPD framework document "EGEP-T CPD Conceptual Framework" states that the project will recruit 1 male and 1 female teacher per school for CPD. This approach is effectively gender neutral. Training more female and fewer male teachers might be more gender-transformative overall, and this may be necessary given the significant pre-existing disparities between girls' and boys' learning.

8.2 Evaluation Recommendations

- In light of the problems encountered with creating learning assessments that were adequately comparable in difficulty between baseline and midline, we would strongly recommend two rounds of piloting of learning assessments at endline. The first round of piloting would lead to adjustments typically made during first-round pilots, while the second round of piloting would specifically be geared toward ensuring that assessments were of comparable difficulty by subtask. Two rounds of piloting and adjustment are likely to lead to far more comparable scores in terms of their difficulty, which will in turn improve our ability to detect intervention-effects using the arithmetic difference in differences method employed in this report.
- It would be ideal to schedule the evaluation at a point in time that avoids potential school closures during the evaluation data-collection period. During this round of data collection, several schools were closed or doing exam preparation by the time we reached them. This meant that the collection of classroom observation data and headcounts was not possible at those schools, and in some cases surveys conducted with teachers were also not possible. In communities where schools were closed, the evaluator still performed child surveys, learning assessments and household surveys, and re-contact rates do not appear to have been impacted negatively. However, in some cases, re-contact rates could suffer if schools are closed when fieldwork takes place.
- Operationalization and measurement of teaching quality is a major methodological challenge. We have addressed this challenge by using multiple measures and triangulation, but there are still limits to our ability to measure teaching quality in ways that are most directly relevant to student learning outcomes. The foundational skill gaps analysis for learning outcomes above highlights key areas where students are falling behind in key skills within a given subject matter. The causes of these skill gaps are seldom obvious from an analysis of the learning outcomes on their own, but it is sometimes the case that students' skill gaps are a reflection of weaknesses in their teachers' own understandings of certain key skills and subjects. Unfortunately, we are not in a position to test this hypothesis because we do not have data on teachers' own grasps of the skills they are teaching. There might be the potential in future evaluations to match data from assessments carried out during teacher training with evaluation data at the school level. Having a more direct measure of teachers' subject-matter knowledge would allow for a better understanding of how teachers' skill gaps may be contributing to skill gaps among their students.
- The analysis of teaching quality has been weakened by the fact that only a small sample of teachers were successfully tracked from baseline to midline and confirmed to appear in both

rounds of data collection. While teaching quality for the full, aggregate sample of teachers was not impacted, a larger sample for the "panel" analysis would improve the quality of inferences that can be drawn regarding project impact. At the endline, selection of teachers for observation and/or teacher surveys should emphasise re-contacting teachers who were surveyed last year (in contrast to random sampling of eligible teachers), and these teachers' information should be scripted directly into the survey tool to facilitate and encourage their recruitment, providing a potentially much larger panel sample for analysis. Also, each teacher's self-reported CPD status should be checked against project documents, a check which can also be scripted into the survey tool if the necessary participation records are available in advance of endline kickoff.

- Panel attrition among grade 8 students is systematic and unlikely to change in future evaluation rounds. Students in this grade are most likely to fall out of the sample because they may relocate to a new community for school, and because their previous teachers may lose contact with them as they move into a new secondary school or non-school alternatives. Ideally, grade 8 students who cannot be located in a follow-up round of data collection should be replaced with girls in a different grade (e.g., Grade 7 or Form 1), to avoid repeated high attrition rates in each subsequent evaluation round.
- A few of the sustainability indicators presented difficulties in terms of data collection and reporting. In some cases, indicators were oriented exclusively toward qualitative data collection; in other cases, indicators could be measured in quantitative terms, but the survey instruments captured only general information about the topic. In other cases, indicators were well-captured but the data provided few opportunities for triangulation. Prior to the endline, the data collection tools should be revised, with time dedicated for consultation between the evaluation team and RI staff. The goal of these revisions should be to more fully capture data on sustainability, making questions more specific to project interventions wherever possible. With more specific information about project interventions available, sustainability can be better captured through revised qualitative tools as well. Specific indicators and areas of improvement are described below, though this list is not exhaustive, and should be considered the starting point for revisions at the endline.
    - Participation in continuing professional development (CPD) training by teachers (sustainability indicator #6) could be improved by expanding beyond a focus on participation, to include both the depth or frequency of participation, as well as the utility of or satisfaction with that participation among teachers. Results could be better triangulated, including by assessing whether head teachers have witnessed improvements in teaching quality as a result of CPD participation; other approaches to measuring tangible impacts of CPD participation should also be explored.
    - At midline, two indicators (sustainability indicators #7 and #10) were exclusively qualitative in nature, but the evaluation plan did not include sufficient qualitative interviews to draw firm, or any, conclusions about their status at midline. In addition to possibly adjusting the measurement approach, the endline should – at minimum – make space for a sufficient number of qualitative interviews to address these indicators in a concrete and convincing way.

- A key methodological limitation of the midline analysis was the shift in sample composition, relative to the baseline, when measuring community attitudes. The baseline survey was conducted among a random sample within communities, while the midline targeted the households of girls who were enrolled in EGEP-T schools in a previous year. Going forward, the sample composition is more likely to stay constant from midline to endline. Nonetheless, effort should be made to ensure comparability between the household samples for the purpose of assessing changes in community attitudes. Prior to the start of endline data collection, an explicit decision should be made regarding which sub-sample of household survey respondents will be comparable from midline to endline, and how comparisons will be made to baseline. The ideal approach would be to conduct a survey with the households of cohort and bursary girls, as at midline, *and* a shorter survey with a random sample of households in EGEP-T communities. The former sample could be used to make midline-to-endline comparisons, in addition to capturing critical data on cohort girls and their households; the latter sample could be used to facilitate baseline-to-endline comparisons with regard to community attitudes alone.
- The qualitative data captured rich insights regarding a number of outcomes, including barriers to transition, and community attitudes. Two areas were neglected in qualitative data collection, which reduced the quality of the sustainability analysis. The first concerns the views of community leaders. Rather than focusing exclusively on reports of their attitudes and/or support for girls' education, qualitative interviews should be completed with traditional and religious leaders in a few communities, to assess how and to what extent they support girls' education in abstract principle and in tangible practice. The second concerns government officials. Where sustainability indicators touched on MOE gender policy and curricula development, the qualitative data was very limited. This issue will only be partially resolved by completing fieldwork in Mogadishu in the future; the allocation of qualitative interviews to government officials should also be increased. Effort should also be made during endline tool development to determine whether some of the sustainability indicators in question could be addressed through targeting questioning of head teachers.

8.3 Programming Recommendations

- In light of the substantial learning gaps that exist for both girls and boys in terms of their reading comprehension in both Somali and English, there is a clear need for additional remedial classes focusing on the skill of reading comprehension. It may also be prudent for the project to consider providing teachers with additional training on techniques for teaching reading comprehension. While we do not have direct measures of teachers' skills in teaching a given subject matter, finding that students have consistent gaps and are performing well below their grade level in skills such as reading comprehension suggests that teachers themselves may be struggling to adequately teach this skill. At minimum, further investigation of this possibility is recommended.
- Girls with long commutes as well as girls who feel unsafe on their way to school have significantly lower learning outcomes and significantly lower attendance rates than their peers, suggesting that these girls are particularly marginal vis-à-vis their peers and are likely to be entangled in a

vicious cycle of dropping attendance and lagging learning. To improve the attendance and learning of these girls, the project should consider creating a transport program that targets girls with long commutes in conflict-affected or unsafe areas. Such a program could involve busing or a number of drivers who have been vetted in order to offer shorter and safer commutes to these girls. An even less resource-intensive option would involve school CECs coordinating such that caretakers could take turns escorting groups of girls to school when they were traveling from the same remote areas.

- Limited school infrastructure is one of the institution-level factors that is the strongest predictors of lower than average learning outcomes. There may be the potential to make significant improvements to school infrastructure by empowering CECs to raise money from communities to invest school improvements that will provide basic health and sanitation facilities where they are missing and will enable girls to feel more comfortable using such infrastructure where it exists. In particular, attention should be focused on mobilizing CECS around these issues in rural and under-resourced schools because these school-level factors are some of the most consistent barriers to learning according to the analysis above.
- While the two recommendations above focus on encouraging CECs to support schools and girls financially, there is an important caveat to this approach. Recognizing that CECs have limited time and resources to invest, it is important to not fragment CECs efforts in a way that lead support in one area to falter while CECs turn their attention to other areas. CECs should be trained to target their financial support to the areas of greatest need in their particular schools and communities, and project activities should focus on identifying alternative revenue sources they can tap for the purpose of expanding the depth and breadth of their financial support. Increased and more consistent training for CECs should also promote greater engagement among the remaining schools with defunct CECs at midline.
- While there have been observable improvements in schools tracking attendance, it seems as though some of the measures taken to track attendance may have unintended negative consequences that impact girls more negatively than boys. For example, some schools close the gate to students who arrive late to school. This policy makes it more difficult for girls (who usually take on household chores such as making breakfast for their family in the morning) to attend if they arrive late, and it potentially discourages them from putting in the effort to come to school even if they know they will be late. As such, we would encourage schools to allow girls who are finishing household work in the morning to still attend even if they are tardy.
- The project's efforts to encourage enrolment prior to the 2017-2018 school year were well-known among community members. But the scope of awareness-raising activities needs to be expanded if community attitudes are to be improved. The focus in these activities should be on recognizing that investments in girls' education require hard choices, but that such investments can pay dividends in terms of community economic empowerment, family financial management, and other outcomes. Community leaders can and should be participate in these efforts, but regular men and women also need to be included, as community members may believe community leaders support girls' education but "regular" people do not.

- Teacher preparation shows some signs of improvement, but teaching quality is still a principle barrier to learning that needs to be addressed and that was also identified at baseline. By most measures, teaching quality has not improved significantly since baseline, and qualitative evidence suggests that part of the problem is that teachers are not thinking explicitly about the methods they are using the classroom and are not using student performance as a means of evaluating the teaching methods being used. Changing the way that teachers think about teaching and intensifying their focus on issues of pedagogy will almost certainly require additional training, but more importantly it will also require the establishment of sustainable mechanisms (including teachers observing one another's classes and in-service days focused on pedagogy) that encourage teachers to evaluate themselves, reflect on their methods, and learn from one another.
- While significant progress has been made in terms of many outputs – CECs are more active, more girls' clubs have been established, and sanitary kits have been distributed widely – the rollout of interventions often appears to be inconsistent. For instance, only a minority of head teachers report that their schools provide life skills courses. Rollout also often appears to be delayed: most schools have an established girls' club, but around one-third of schools either do not have a girl's club or have a club that was formed within the last six months. Given how slow-moving or "sticky" outcomes like transition, community attitudes, and self-esteem are, short-term exposure is insufficient to produce meaningful changes in these outcomes. Additional effort should be dedicated to ensuring widespread uptake of project interventions, in a timely fashion.
- The project has made significant improvements to child protection policies and mechanisms within schools, including the designation of child protection focal points, the development of Codes of Conduct, and training around these topics. However, important gaps in this progress remain. For instance, a minority of new teachers sign a Code of Conduct with explicit reference to child protection issues. Children and parents alike state that they would report a child protection issue to the head teacher, rather than the focal point. These two gaps suggest that child protection training needs to be very broadly available and required at schools – children may prefer to report an issue to a teacher they know and trust, or to the head teacher, and it is essential that those individuals be trained on child protection issues, regardless of whether they are the formal focal point.

Annex 1. Midline evaluation submission process

Please submit all Midline reports and accompanying annexes via Teamspace, an online file-sharing platform. Both the External Evaluator (EE) and Project should have access to their respective Teamspace folders, however please reach out to your EO if you do not.

Please note, Annexes can be uploaded to Teamspace for FM review separately and before the midline report analysis is completed. We advise Projects and EEs to follow the sequence outlined below to speed up the review process and avoid unnecessary back and forth. Where possible, we also advise that projects and EEs do not begin their ML report analysis until Annex 13 is signed off by the FM.

Annexes to submit for FM review any time before the ML report is completed:

- Annex 2: Intervention roll-out dates.
- Annex 3: Evaluation approach and methodology.
- Annex 4: Characteristics and barriers.
- Annex 7: Project design and interventions.
- Annex 9: Beneficiaries tables.
- Annex 10: MEL Framework.
- Annex 11: External Evaluator's Inception Report (where applicable).
- Annex 12: Data collection tools used for midline.
- Annex 13: Datasets, codebooks and programs.
- Annex 14: Learning test pilot and calibration.
- Annex 15: Sampling Framework.
- Annex 16: External Evaluator declaration.
- Annex 17: Project Management Response (this can be revisited following feedback from the FM).

Annexes to finalise after Annex 11 "Datasets, codebooks and programs" is signed off by the FM:

- Annex 5: Logframe.
- Annex 6: Outcomes Spreadsheet.
- Annex 8: Key findings on Output Indicators.

Annex 2. Intervention Rollout Dates

This section has been completed by the project.

Interventions implemented across all the project sites target hard to reach girls in both primary and secondary schools. ADRA, CISP and RI have slightly different dates in start and end dates however most of the interventions run within the same period. Difference is largely for CPD and child protection initiatives.

Intervention	Start	End
Bursary support	August 2017	May 2020
Safety net support (uniforms)	September 2017	October 2019
National examination fee payments	April 2018	May 2020
Menstrual hygiene management support for adolescent girls	September 2017	February 2020
Provision of Solar lamps	October 2017	October 2019
Dry food rations for students and teachers	September 2017	December 2019
Training on Life skills & career guidance for Girls and Boys Leadership Networks	February 2018	December 2019
Girls and Boys leadership networks' sessions	February 2018	Ongoing – end date April 2020
Remedial classes for low performing boys and girls	January 2018	Ongoing – end date April 2020
Teacher trainings (Numeracy, English Literacy, Inclusive education, gender inclusive and child centred pedagogy, child protection)	March 2018	Ongoing – end date October 2019
Provision of school supplies, teaching and learning materials	September 2017	Ongoing – end date Dec 2019
Teacher incentives	October, Aug 2017	Ongoing – end date May 2020

Strengthen technical, management and accountability support to CPD (coaching support to teachers and reflection meetings at various levels)	March 2018	Ongoing – end date April 2020
Engagement with head teachers on continuous professional development for teachers and child protection	September 2018	April 2020
Support MoE on gender based planning and monitoring of education interventions	October 2017	Ongoing – end date April 2020
Support quality supervision by quality assurance department of MoE	April 2018	Ongoing – end date April 2020
Contextualize positive discipline approaches in school	August 2019	March 2020
Strengthen and support to MoE on child protection	September 2019	Ongoing – end date April 2020
Strengthen linkage for child protection support to schools	August 2019	Ongoing – end date March 2020
Establishment and strengthening of community education committees (CEC)	August 2017	Ongoing – end date April 2020
Matching grant support to implementation of School development plans	July 2018	March 2020
Community dialogues to promote behaviour change	October 2017	Ongoing – end date April 2020
Radio advocacy to promote behaviour change	April 2018	Ongoing – end date April 2020
Proactive engagement with Men and Boys to promote girls education & transition	July 2018	Ongoing – end date April 2020
Back to school campaigns	September 2017	September 2017

Annex 3. Evaluation approach and methodology

Outcomes and Intermediate Outcomes

The table below defines the primary outcomes and intermediate outcomes selected by EGEP-T for evaluation. For each outcome and intermediate outcome, the indicator(s) utilized are operationalized in line with the project's logical framework and documentation in this report.

Outcome	Level at which measurement will take place, *e.g. household, school, study club etc.*	Tool and mode of data collection (please specify both the quantitative and qualitative tool used)	Rationale, *i.e. why is this the most appropriate approach for this outcome*	Frequency of data collection, *i.e. per evaluation point, annually, per term*	Who collected the data?	Discuss any changes from BL (including whether this indicator is new)
Outcome 1: learning – Number of marginalised girls supported by GEC with improved learning outcomes						
Somali literacy	Individual	SeGRA, adapted for Somali	SeGRA is an appropriate adaptation for the project's grade range	Per evaluation point	External evaluator	Additional subtasks added to learning assessment to prevent ceiling effects midline and endline; comparable subtasks updated to prevent recollection from baseline, but kept equivalent in difficulty
English literacy	Individual	SeGRA	SeGRA is an appropriate adaptation for the project's grade range	Per evaluation point	External evaluator	Subtasks were updated to prevent recollection from baseline; no other changes to subtasks
Numeracy	Individual	SeGMA	SeGMA is an appropriate adaptation for the project's grade range	Per evaluation point	External evaluator	Additional subtasks added to learning assessment to prevent ceiling effects midline and endline; comparable subtasks updated to prevent recollection from baseline, but kept equivalent in difficulty
Outcome 2: Transition - Number of marginalised girls who have transitioned through key stages of education, training or employment						
Transition indicator	Individual	Household survey, child and caregiver modules	Child responses regarding enrolment and grade prioritized, as	Per evaluation point	External evaluator	No changes to measurement protocol

			reliability is higher than found in school records or from caregivers			
Outcome 3: Sustainability at community level						
% of surveyed members of EGEP target communities, who have been exposed to project awareness-raising activities report having changed their opinion positively in relation to the importance of girls' school completion	Individual	Household survey; teacher survey FGDs with men, women, boys and girls	Triangulation across respondent types	Per evaluation point	External evaluator	
% of surveyed members of EGEP target communities, that report that boys and men taking action to support girls in attending and completing school	Individual	Household survey; teacher survey FGDs with men, women, girls, and boys; KIIs with teachers	Triangulation across respondent types	Per evaluation point	External evaluator	
% of EGEP target communities where community leaders are leading campaigns and advocacy events	Individual	Household survey; teacher survey FGDs with men, women, girls, and boys; KIIs with teachers	Triangulation across respondent types	Per evaluation point	External evaluator	

		and head teachers				
Outcome 3: Sustainability at school level						
% of CECs providing bursary support to marginalised girls in EGEP target schools	School/CEC	Head Teacher Survey FGDs with CEC members; KIIs with head teachers	Head teachers are the most straightforward source of information, in lieu of a survey of CEC members	Per evaluation point	External evaluator	
% of CECs in EGEP target schools providing match funding (financial or in-kind) for school improvement initiatives	School/CEC	Head Teacher Survey FGDs with CEC members; KIIs with head teachers	Head teachers are the most straightforward source of information, in lieu of a survey of CEC members	Per evaluation point	External evaluator	
% of schools actively participating in the peer-mentoring programme	School	Head Teacher Survey FGDs with CEC members; KIIs with head teachers; KIIs with teachers	Head teachers are the most straightforward source of information, in lieu of a survey of CEC members	Per evaluation point	External evaluator	
Outcome 3: Sustainability at system level						
Federal level and Galmudug Gender Units develop gender strategy and Federal level strategy is being	System	KIIs with gender unit and MOE officials; KIIs with head		Per evaluation point	External evaluator	

implemented within the life of the project.		teachers; FGDs with CEC members				
% of Head Teachers that report the established child protection mechanism to be actively used.	School or Head Teacher	Head Teacher Survey and Mentor Survey FGDs with teachers; KIIs with head teachers	Head teachers are aware of policies in their schools; mentors responsible for implementing aspects of CP mechanism	Per evaluation point	External evaluator	Indicator adjusted at midline to focus on use of CP mechanism, rather than its existence
% of EGEP-target schools receiving follow-up monitoring visits from MoE officials, including Gender Focal Points, District/Regional Education Officers/RES	School	Head Teacher Survey KIIs with head teachers	Head teachers primary source, given they receive MOE visits as representative of school	Per evaluation point	External evaluator	
TTI (Teacher Training Institutes) integrate components of the CPD approach into their curriculum for teacher training	System	KIIs with coaches, mentors, remedial teachers, and teachers		Per evaluation point	External evaluator	Outcome and indicator added at midline
Intermediate outcome 1: Improved girls attendance at primary and secondary school						
Attendance rate of marginalised girls' in schools throughout the life of the project	Individual	School records; classroom headcounts; household survey		Per evaluation point	External evaluator	
Intermediate outcome 2: Positive community attitudinal change						

Caregiver aspirations for girl's education – desire college/university	Individual	Household survey, caregiver module FGDs with men, women, girls, and boys		Per evaluation point	External evaluator	
Percentage of girls who feel pressure to drop out of school to get married	Individual	Household survey, child module FGDs with girls and boys		Per evaluation point	External evaluator	
Intermediate outcome 3: Girls feel more empowered with greater self confidence						
Marginalised girls' average score on composite self-esteem index	Individual	Household survey, child module FGDs with girls and boys	Accounts for multidimensional and latent nature of self-esteem to capture many aspects of outcome	Per evaluation point	External evaluator	
Intermediate outcome 4: Improved school management and institutional governance						
CEC's level of performance in effective school management (scorecard approach)	School/CEC	Head teacher survey; teacher survey FGDs with CEC members; KIIs with head	Takes into account multidimensional nature of school management	Per evaluation point	External evaluator	

		teachers and teachers				
Intermediate outcome 5: Improved teaching quality						
Indicator: Percentage of teachers demonstrating improved teaching practices	Teacher or classroom	Classroom observation tool KIIs with teachers	Classroom observations preferred to teacher surveys for documenting actual behaviour in classrooms	Per evaluation point	External evaluator	Adapted from baseline, with an additional indicator included and distinction between the two teaching quality indicators clarified
Indicator: Percentage of teachers using student centred approaches	Teacher or classroom	Classroom observation tool KIIs with teachers	Classroom observations preferred to teacher surveys for documenting actual behaviour in classrooms	Per evaluation point	External evaluator	Adapted from baseline, with an additional indicator included and distinction between the two teaching quality indicators clarified

Evaluation methodology

This section provides an overview of the overall research design employed by the EGEP-T programme. The EGEP-T evaluation design deviates in important ways from the standard GEC-T design established and recommended by the FM. The first major point of difference is the lack of a set of control or comparison schools in the evaluation. Due to security concerns, RI and its partners elected against including a set of comparison schools in the evaluation. The standard GEC-T evaluation design employs comparison schools and a difference-in-differences design; the lack of comparison schools means that the EGEP-T evaluation utilizes a strictly pre-post design – as opposed to difference-in-differences – in which changes are tracked in the treatment or intervention schools exclusively. This design choice has significant implications for the types of analysis that can be expected at the midline and endline evaluations, as well the strength of any conclusions that can be drawn.

The second major point of difference is in the sampling approach employed at baseline. The approach recommended by the FM has been labelled a "joint sampling" approach, in which the same students who complete learning assessments at schools are also included in the household survey sample. That is, after completing learning assessments at a sample school, enumerators follow up at the same children's residences, recruiting their households into the household survey sample, including a survey module that applies specifically to children. In contrast, the EGEP-T evaluation employs a hybridized version of the joint sample approach and what the FM calls a "split-sample approach". In the split sample approach, the "learning sample" and "household sample" are comprised of different respondents. In practice, children are sampled randomly from program schools, while households are sampled using a random walk strategy in communities that include program schools.

The sampling approach employed at baseline has consequences for the design of the midline and its learning and transition samples. Girls from the learning sample drawn via schools at baseline were tracked for learning outcomes, with girls replaced in the learning sample if they were no longer enrolled in an EGEP-T project school. Meanwhile, the same set of girls were tracked for transition outcomes, but girls were re-contacted at midline for this purpose regardless of their enrolment status. That is, the transition and learning samples are de-coupled at midline, with the transition sample continuing to include girls who were replaced in the learning sample.

The pre-post design employs benchmarked comparisons for both learning and transition. A full description of these benchmarks, their composition, and the comparisons being made at midline is provided in Sections 4 (learning) and 5 (transition) of the main report.

Midline data collection process

In this section, outline the process to collect midline data (both quantitative and qualitative). Provide details on the following areas. Highlight changes since baseline and why they occurred.

Training and Tool Development

Prior to the start of data collection, the data collection tools were updated extensively. With respect to quantitative tools, a new survey tool targeted at male and female mentors was developed. Mentors are focal points within EGEP-T schools, who serve as role models for female and male students, among other responsibilities. The tool was developed by the external evaluation team, with input from RI's Monitoring & Evaluation team and the FM. The goal was to capture information about male and female mentors, when they were recruited, their access to and satisfaction with training, the implementation of their school's child protection mechanism, and their contribution to project outputs, such as the teaching of life skills courses.

Beyond the mentor survey, the quantitative tools were revised extensively to reflect changes in the sampling structure, new indicators, revised understanding of indicators, and other changes that occurred at midline. New questions were added to capture aspects of child protection that were not incorporated fully during the baseline. The most significant revisions centre on the household survey and its child, caregiver and head-of-household modules. These modules were combined into a single tool, reflecting the fact that the midline household sample – unlike that of the baseline – is targeted at cohort and bursary girls, rather than a random household sample.

Qualitative tools underwent the most drastic changes at midline. A number of new respondent groups were targeted for qualitative interviews, with new tools designed to accommodate this change. In addition, the qualitative tools were revised to include participatory exercises meant to generate more useful and insightful data for qualitative analysis. Revisions to the qualitative tools were led jointly by RI's Monitoring & Evaluation team and Forcier's qualitative analyst. As noted below, the approach to qualitative training was also updated; in line with the development of new tools, the training time was increased and made more participatory.

An overview of the quantitative tools employed, with notes on sample size and replacement rates, is provided in the table below. For reference, an overview of the sample size and respondent groups targeted for qualitative interviewing is provided in Table 5, in the main report.

TABLE 107: DETAILS OF QUANTITATIVE TOOLS

Tool and Outcome	Beneficiary Group	Sample Size Agreed in MEL Framework	Actual Sample Size	Notes
Learning Tests (SeGRA, SeGMA), used for learning outcome	ISG (cohort girls) age 11-18 at baseline Bursary girls, age 11-18 at baseline In-school (cohort) boys age 11-18 at baseline OOS girls age 11-18 at baseline	1,638 ISG cohort girls 350 bursary girls 384 cohort boys No target sample size for OOS girls	1,609 at BL; 1,449 at ML 428 at BL; 369 at ML 398 at BL; 324 at ML 120 at BL; 64 at ML	Attrition due to loss of Banadir schools and inability to replace some girls (no eligible replacements). Replacement rates: Cohort girls: 24.3% Bursary girls: 19.2% Cohort boys: 27.1% OOS girls: no replacement
Child Survey (used for self-esteem, teaching quality, transition outcomes)	ISG (cohort girls) age 11-18 at baseline Bursary girls, age 11-18 at baseline In-school (cohort) boys	1,638 ISG cohort girls 350 bursary girls 384 cohort boys	1,609 at BL; 1,449 at ML 428 at BL; 369 at ML 398 at BL; 324 at ML	Same as for learning tests above

	age 11-18 at baseline			
Household survey (used for transition, community attitudes)	All cohort girls and bursary girls re-contacted at midline. Replacement girls excluded. All OOSG surveyed at BL	No set sample size; dependent on enrolment rates	1,070 cohort girls 287 bursary girls 64 OOS girls	HH survey only completed with re-contacted cohort, bursary and OOS girls. Replacement girls were not given HH survey.
Classroom headcount (used for attendance)	Classrooms in EGEP-T schools	No sample size specified in MEL Framework	1 classroom per grade level in each school (e.g., 8 in most primary schools) 893 at BL; 755 at ML	Attrition due to loss of Banadir schools.
Classroom observations (used for teaching quality)	Classrooms in EGEP-T schools	No sample size specified in MEL Framework	2 classrooms per school, chosen based on subject and grade taught. 263 at BL; 215 at ML	Attrition due to loss of Banadir schools.
Teacher survey (used for teaching quality and community attitudes)	Teachers in EGEP-T schools	No sample size specified in MEL Framework	5 per school 516 at ML; 516 at BL	Attrition due to loss of Banadir schools; offset by higher completion rates in remaining schools
Head teacher survey (used for teaching quality, community attitudes, and school management)	Head teachers in EGEP-T schools	One per school, expected 140 at BL	140 at BL; 112 at ML	Attrition due to loss of Banadir schools and non-response by 7 head teachers
Mentor Survey (used for tracking of	Female and male mentors	Two per school targeted, contingent on	N/A at BL; 171 at ML	No mentor survey completed at BL; non-response and

mentorship implementation)	availability within each school	lack of mentors in some schools yielded lower-than-expected sample size

Training for fieldwork took place over five days, from March 24 to March 28, at the Ambassador Hotel in Hargeisa. The first four days of training consisted of traditional, classroom-based training with the field teams. In total, 14 team leaders and 42 enumerators attended training, as each team consists of a team leader and three enumerators.

The teams were all highly experienced: only three or four enumerators had never participated in a GEC or GEC-T evaluation prior to this engagement. In contrast to previous years, when EGEP and SOMGEP evaluations have been conducted simultaneously, the staggered timing of the two evaluations allowed us to draw from a wider range of experienced enumerators and team leaders. In addition, 10 of the 14 team leaders from the baseline EGEP-T evaluation returned as team leaders at the midline; an additional team leader from the baseline was promoted to serve as Fieldwork Manager at midline.

In an effort to improve data quality, two main changes were made to the training plan. First, the training sessions were organized to more aggressively target training topics only to those who needed them. For instance, during the baseline, all team members were trained on the classroom observation tool; during the midline, only team leaders and a select group of enumerators were trained on this tool. Even within the group of team leaders, training was targeted: newer team leaders were given dedicated sessions on fieldwork procedures and the headcount tool, while more experienced team leaders spent additional time learning about the qualitative tools and participating in the training of their teams during this same time. The additional time dedicated to the qualitative tools included a general training session on qualitative interviewing, and significantly more time focused on individual tools, which proved essential for interviewers' understanding of the participatory exercises.

Second, the pilot day was expanded. At baseline, the pilot was a half-day. Moreover, only a minimal debriefing was possible after the pilot, because teams were delayed in returning to the training facility due to political rallies, and ensuing traffic snarls, in Hargeisa prior to the Somaliland elections. During the midline training, no similar logistical difficulties were encountered. In addition, the teams were split into two groups, with one group completing a three-quarters day of piloting and participating in a debriefing session that evening. The second group completed a full day of piloting (from 8 am to approximately 5 pm), participating in a debrief session on the following day. The consequence of this decision was that the teams received greater exposure to collecting data "in the field"; in addition, the smaller debriefing sessions – which often involved just 1-2 teams at a time – promoted a franker discussion, encouraging questions from all team members, and a more personalized review of the tools, problems the teams encountered, and fieldwork procedures going forward.

During training with the data collection teams, RI's Monitoring & Evaluation team conducted a detailed training on research ethics and child protection in the context of evaluation exercises. All enumerators and team leaders reviewed and signed child protection and research ethics protocols in line with RI's standard policy; enumerators were also briefed on Forcier's internal standards for the same.

Re-contact and Fieldwork Procedures

Training for the data collection teams concluded on March 30, and data collection began on April 6, following a short delay and time required for teams to travel to their field locations. Fieldwork began first in Somaliland, on April 6; fieldwork was delayed in other locations due to a lack of permissions from relevant government ministries. For instance, fieldwork was delayed several days in Galmudug and Hirshabelle. Primary fieldwork concluded on May 9, though qualitative interviews with select interviewees continued for several weeks afterward, depending on their availability. In general, all tools – qualitative and quantitative – were completed in the same time period (i.e. qualitative interviews were conducted concurrently with quantitative surveys). The few exceptions were those qualitative interviews that required significant advanced scheduling.

The importance of maintaining the panel's size from baseline required a robust re-contact protocol. Moreover, the complexity of the sampling design – especially the distinction drawn between learning and transition samples – required careful specification of respondent types and the data collection tools they should complete. The data collection teams were trained on the following rules for assigning respondent types to different tools and samples.

TOOLS AND TYPES OF CHILDREN

The types of children and the surveys they complete has not changed.

- Cohort girls – complete Child Survey, Learning Assessments, and Household Survey (full survey)
- Bursary girls – complete Child Survey, Learning Assessments, and Household Survey (full survey)
- Cohort boys – complete Child Survey and Learning Assessments
- Out-of-School girls – complete Child Survey and Household survey. Do not complete learning assessments, even if girl is currently enrolled in school

COHORT GIRLS

For every cohort girl in your list, you complete a Child Survey, a Learning Assessment, and a Household Survey.

- If the girl is enrolled at the school, she completes all 3 surveys.
- If the girl is not enrolled at the school, she completes the HH survey. Replace her with a new girl to complete the Child Survey and Learning Assessments.
- If the girl cannot be found, replace her with a new girl to complete the Child Survey and Learning Assessments. Do not complete a HH survey.

BURSARY GIRLS

For every bursary girl in your list, you complete a Child Survey, a Learning Assessment, and a Household Survey.

- If the girl is enrolled at the school, she completes all 3 surveys.
- If the girl is not enrolled at the school, she completes the HH survey. Replace her with a new girl to complete the Child Survey and Learning Assessments.
- If the girl cannot be found, replace her with a new girl to complete the Child Survey and Learning Assessments. Do not complete a HH survey.

COHORT BOYS

For every cohort boy in your list, you complete a Child Survey and a Learning Assessment. Do NOT complete a Household Survey.

- If the boy is enrolled at the school, he completes Child Survey and Learning Assessment.
- If the boy is not enrolled at the school, replace him with a new boy to complete the Child Survey and Learning Assessment.
- If the boy cannot be found, replace him with a new boy to complete the Child Survey and Learning Assessment.

OUT-OF-SCHOOL GIRLS

If the girl can be found, she completes Child Survey and Household Survey. If the girl cannot be found, do NOT replace her.

For the purposes of guiding enumerators in this process, a reference manual was developed, with a flowchart for each type of child targeted (cohort girls, bursary girls, cohort boys, and out-of-school girls), as different procedures applied to different groups.

Beyond clarifying the respondent groups and the tools to be applied to them, the project also developed a rigorous protocol for re-contacting girls, while also accounting for the limited fieldwork time allotted to each school/community. After arriving at the school and meeting with the head teacher, some children may not have been located at the school. In these cases, the following steps were required:

- Call the phone number listed for her household in the tracking sheet, if there is one
- Ask the head teacher and teachers in her grade level for contact information (phone number and location of household) at the girl's household
- Ask other girls in her grade whether she still lives nearby and how she can be reached
- For every phone number available for the girl, you must make two contact attempts, separated by at least 6 hours. This is why it is important to try to contact the cohort girls early on the first day of your visit to the school, or you will not have enough time to do two callbacks before it is time to move to another school.
- If you are able to locate the girl's household, but she is not there, you must make at least one re-visit attempt, 6 hours after the first.

These procedures were tracked explicitly by team leaders prior to making replacement decisions. Team leaders were the only individuals with the authority to replace girls in the learning sample, and only did so after these procedures were followed and documented.

Sampling and Data Quality

The sampling approach for children re-contacted from baseline is outlined above and in Section 2.2 of this report, which describes the technical sampling approach for selecting schools at baseline, replacing schools – as needed – at midline, selecting children at baseline, selecting replacements at midline, and constructing the distinct learning and transition samples.

Beyond respondents to the household/child surveys and learning assessments, the remaining data collection tools employed alternative second-stage sampling approaches, where the first stage consisted of the selection of schools/clusters. These second-stage approaches are described in Section 2.2, with a separate description for the head teacher survey, teacher survey, mentor survey, classroom observation, and classroom headcount.

Given the complexities introduced by the sampling decisions – and the fact that these complexities are often context- or outcome-specific, we review sampling considerations at length throughout the report. For instance, in our analysis of transition outcomes, we discuss how comparable baseline and midline transition subsamples were constructed from the available data, and the manner in which our analysis is influenced by the nature of those samples.

During fieldwork, the data underwent significant quality assurance and quality control. Prior to the start of fieldwork, the evaluation team developed a set of custom quality checks, supplementing the team's standard quality control script. During fieldwork, Forcier's Quality Assurance Officer downloaded the data from Ona's server and ran the quality control script; a representative but not exhaustive list of the daily checks includes:

- Long and short duration surveys
- Assessment of the composition of key demographic variables, to ensure they are in line with expectations (e.g., grade and age distribution)
- Surveys with unlikely GPS locations or GPS locations that change during the interview
- Logical consistency checks among related questions (e.g., word-per-minute scores among children who could not identify letters, etc.)
- Checking response or re-contact rates
- Identifying surveys with non-response or large numbers of "don't know" responses

In each case, the Research Officer identified observations that the script flagged, checked them manually, contacted the field teams to verify the information, and made corrections to the data in a canonical and replicable cleaning script (.do file). On a less frequent basis, Forcier's Global Technical Team undertook additional, more extensive quality control checks, such as checking for "enumerator effects" in learning scores and other key outcomes. Both sets of checks typically required a number of callbacks or follow-up visits to households to verify or correct the data collected.

In the field, team leaders accompanied enumerators on a set number of interviews, to verify the quality of their work. Team leaders also conducted random callbacks for the same purpose. In addition, team leaders verified the tracking sheets completed by team members and added details to those tracking sheets. Finally, team leaders were responsible for completing a report after the completion of fieldwork in each school, documenting the tools completed, the re-contact and replacement of cohort members, and the contact information of cohort members and their households, among other details.

Sample Characteristics

In this section, we describe the characteristics of the sample achieved during midline data collection. The table below reports the attrition rates in four cohorts of children re-contacted at midline. Note that the target sample size referred to in the second column is equal to the number of eligible baseline respondents in that cohort, among the baseline schools visited at midline. The third column, re-contact rates, refers to the share of the target sample that comprises the panel (i.e. children who were successfully re-contacted at midline). School-level attrition refers to children who fell out of the panel because their entire school was replaced; note that every child lost through school-level attrition was replaced in the sample by an equivalent child in the two replacement schools. The final two columns separate out individual-level attrition in which the child was replaced by another child in the same

school and individual-level attrition in which no replacement was selected, typically because no equivalently eligible child was available.[369]

TABLE 108: ATTRITION RATES BY COHORT GROUP

Cohort	Target Sample Size	Re-contacted	School-level Attrition (replaced)	Individual Attrition (replaced)	Individual Attrition (not replaced)
Cohort Girls	1,370	996 (72.7%)	24 (1.8%)	320 (23.4%)	30 (2.2%)
Bursary Girls	365	274 (75.1%)	7 (1.9%)	65 (17.8%)	19 (5.2%)
Cohort Boys	337	234 (69.4%)	6 (1.8%)	84 (24.9%)	13 (3.9%)

The tables below report characteristics of the sample achieved during midline data collection. In the tables focused on children respondents, we first discuss the learning sample, reporting age, grade, and geographic breakdowns of the cross-sectional and panel learning samples, respectively. We then report the same breakdowns for the midline transition sample. The remaining two tables report the geographic breakdown of the samples achieved for other data collection tools, referring to both the full midline sample and the "comparable" midline sample (i.e. the sample that shares overlapping schools with the baseline) and the rate of disability status in the cohort girl and bursary girl samples at midline.

TABLE 109: SAMPLE CHARACTERISTICS, MIDLINE COHORT GIRLS (LEARNING)

Age (Midline)	Cross-Sectional Sample	Panel Sample
11	0.8% (10)	0.1% (1)
12	3.0% (39)	2.8% (28)
13	7.8% (102)	7.0% (70)
14	17.4% (229)	16.5% (164)
15	22.8% (300)	22.3% (222)
16	20.0% (263)	19.9% (198)
17	14.7% (193)	15.3% (152)
18	9.5% (125)	11.1% (111)
19	4.0% (53)	4.8% (48)
20	0.2% (2)	0.2% (2)
Total	100% (1,316)	100% (996)
Cohort Grade (Baseline)	**Cross-Sectional Sample**	**Panel Sample**
Grade 6	27.3% (359)	30.2% (301)
Grade 7	26.1% (344)	29.9% (298)
Grade 8	25.2% (331)	15.6% (155)
Form 1	11.5% (151)	12.2% (122)
Form 2	10% (131)	12% (120)
Total	100% (1,316)	100% (996)
Zone	**Cross-Sectional Sample**	**Panel Sample**
Somaliland	40.9% (538)	41.7% (415)
Puntland	48.2% (634)	47% (468)
Galmudug	10% (132)	10.2% (102)

[369] A lack of eligible children in the same cohort group occurred most frequently in the context of bursary girls, as some schools have a limited number of female students receiving bursaries.

Hirshabelle	0.9% (12)	1.1% (11)
Total	100% (1,316)	100% (996)

TABLE 110: SAMPLE CHARACTERISTICS, MIDLINE BURSARY GIRLS (LEARNING)

Age (Midline)	Cross-Sectional Sample	Panel Sample
11	0.6% (2)	0% (0)
12	1.8% (6)	1.1% (3)
13	7.1% (24)	7.3% (20)
14	15.3% (52)	13.5% (37)
15	22.7% (77)	24.1% (66)
16	21.8% (74)	20.4% (56)
17	15.3% (52)	17.2% (47)
18	10.6% (36)	11.3% (31)
19	4.1% (14)	4.4% (12)
20	0.6% (2)	0.7% (2)
Total	100% (339)	100% (274)
Cohort Grade (Baseline)	**Cross-Sectional Sample**	**Panel Sample**
Grade 6	30.4% (103)	33.9% (93)
Grade 7	26.3% (89)	30.3% (83)
Grade 8	23.3% (79)	15.3% (42)
Form 1	9.7% (33)	10.6% (29)
Form 2	10.3% (35)	9.9% (27)
Total	100% (339)	100% (274)
Zone	**Cross-Sectional Sample**	**Panel Sample**
Somaliland	38.3% (130)	39.4% (108)
Puntland	52.5% (178)	51.8% (142)
Galmudug	8% (27)	7.7% (21)
Hirshabelle	1.2% (4)	1.1% (3)
Total	100% (339)	100% (274)

TABLE 111: SAMPLE CHARACTERISTICS, MIDLINE COHORT BOYS (LEARNING)

Age (Midline)	Cross-Sectional Sample	Panel Sample
11	0% (0)	0% (0)
12	1.6% (5)	1.7% (4)
13	7.5% (24)	7.7% (18)
14	12.3% (39)	8.5% (20)
15	13.8% (44)	14.1% (33)
16	21.1% (67)	21.8% (51)
17	15.4% (49)	14.1% (33)
18	15.4% (49)	16.7% (39)
19	11.9% (38)	14.1% (33)
20	0.9% (3)	1.3% (3)
Total	100% (318)	100% (234)
Cohort Grade (Baseline)	**Cross-Sectional Sample**	**Panel Sample**

Grade 6	26.1% (83)	25.6% (60)
Grade 7	27.4% (87)	32.1% (75)
Grade 8	24.8% (79)	16.2% (38)
Form 1	11.3% (36)	13.7% (32)
Form 2	10.4% (33)	12.4% (29)
Total		
Zone	**Cross-Sectional Sample**	**Panel Sample**
Somaliland	42.1% (134)	42.7% (100)
Puntland	46.5% (148)	45.3% (106)
Galmudug	10.4% (33)	11.5% (27)
Hirshabelle	0.9% (3)	0.4% (1)
Total		

TABLE 112: SAMPLE CHARACTERISTICS, MIDLINE COHORT TRANSITION SAMPLE

Age (at *t-1*, i.e. baseline)	Midline Cohort Transition Sample
11	3.8% (41)
12	8.5% (93)
13	18.5% (202)
14	24% (261)
15	19.5% (212)
16	14.4% (157)
17	11.3% (123)
Total	100% (1,089)
Zone	**Cross-Sectional Sample**
Somaliland	41.4% (451)
Puntland	47.3% (515)
Galmudug	10.3% (112)
Hirshabelle	1% (11)
Total	100% (1,089)

TABLE 113: SAMPLE CHARACTERISTICS, OTHER DATA COLLECTION TOOLS

Zone	Total Midline Sample	Comparable Midline Sample
	Classroom Headcounts	
Somaliland	43.8% (331)	45.2% (323)
Puntland	48.6% (367)	49.2% (351)
Galmudug	6.9% (52)	5.6% (40)
Hirshabelle	0.7% (5)	0.0% (0)
Total	100% (755)	100% (714)
	Classroom Observations	
Somaliland	42.8% (92)	43.9% (90)
Puntland	49.3% (106)	49.8% (102)
Galmudug	7% (15)	6.3% (13)
Hirshabelle	0.9% (2)	0.0% (0)

Total	100% (215)	100% (205)
Teacher Surveys		
Somaliland	41.3% (213)	41.2% (199)
Puntland	48.6% (251)	51.1% (247)
Galmudug	9.1% (47)	7.7% (37)
Hirshabelle	1% (5)	0.0% (0)
Total	100% (516)	100% (483)
Mentor Surveys		
Somaliland	42.1% (72)	N/A
Puntland	49.7% (85)	N/A
Galmudug	7% (12)	N/A
Hirshabelle	1.2% (2)	N/A
Total	100% (171)	N/A
Head Teacher Surveys		
Somaliland	40.2% (45)	40% (44)
Puntland	49.1% (55)	49.1% (54)
Galmudug	9.8% (11)	10% (11)
Hirshabelle	0.9% (1)	0.9% (1)
Total	100% (112)	100% (110)

TABLE 114: DISABILITY STATUS IN THE MIDLINE COHORT SAMPLE

Zone	Cohort Girl Sample	Bursary Girl Sample
Difficulty seeing	0.5% (6)	0.9% (3)
Difficulty hearing	0.1% (1)	0.0% (0)
Difficulty walking or climbing steps	0.0% (0)	0.3% (1)
Difficulty remembering or concentrating	0.1% (1)	0.6% (2)
Difficulty with self-care	0.1% (1)	0.0% (0)
Difficulty communicating	0.1% (1)	0.0% (0)
Any impairment	0.7% (8)	1.5% (5)

Data Cleaning and Analysis

Data Quality Assurance and Cleaning

Following the conclusion of data collection, the data were cleaned and checked for consistency, including a more intensive set of quality checks. This process identified a small number of issues that were not already identified during quality control checks employed during the fieldwork period. In these cases, team leaders performed callbacks to verify information from households and head teachers, as needed.

Throughout data collection, there was constant communication between team leaders and enumerators in the field, and the Research Officer in the national office. This communication allowed the Research Officer to keep track of issues that arose during fieldwork, and record the researchers' insights and experiences. At the conclusion of fieldwork, team leaders were debriefed and their observations recorded.

During fieldwork, the quantitative data was uploaded daily – or else whenever network service would allow – to Ona's server. The data was downloaded, checked, and basic cleaning tasks were performed daily. After fieldwork was over, the evaluation team conducted a more thorough cleaning, including merging the midline data with the baseline data for consistency checks and to prepare for analysis. All cleaning and quality control actions were performed in Stata, using scripts (.do files) to facilitate replication.

All qualitative interviews were audio recorded. Qualitative researchers took notes immediately after each interview, to ensure that their recollections were fresh. Audio files were returned to the national office, where experienced staff transcribed the audio recordings and, separately, translated, verbatim, the transcriptions into English. Each transcription and translation was checked by a second Somali member of Forcier's core national staff, to verify its quality. A final check was performed by one of Forcier's Research Officers or Research Managers, during which they reviewed the English translations for coherence; translations with probable mistakes were returned to the translator and staff member who originally performed its quality check. The English translations were provided to the evaluation team's dedicated qualitative analyst. As discussed in our description of the evaluation design, the qualitative and quantitative analysis were kept strictly separate at the start, to allow themes to emerge naturally from the qualitative data. Only after thoroughly reviewing the interview transcripts and writing a first set of results on that basis did the qualitative analyst begin to review the quantitative findings.

In the next round of data collection, cohort tracking will proceed much as it did at midline. For every child interviewed at midline, enumerators sought consent from the child's caregiver to re-contact the household in the future. Enumerators recorded contact information about the household to facilitate re-contacting them at endline, and took GPS coordinates at the location of the interview, ideally their house, prior to finishing the interview. This information, as well as the physical tracking sheets filled in for each cohort respondent, have been compiled into a midline tracking database that can be used at endline.

Qualitative Analysis

Qualitative data was collected from a wide range of respondent groups. As noted in the discussion of tool development, above, the qualitative tools underwent dramatic changes prior to the start of the midline evaluation, including the incorporation of participatory exercises. Several new tools were developed collaboratively with RI, and additional time was dedicated to training team leaders on the administration of qualitative interviews. As previous mentioned, qualitative interviews were audio-recorded, transcribed verbatim, and translated into English.

The primary approach to utilizing qualitative data involved the exploratory coding of emergent themes, with the goal of generating insights from the bottom-up – through respondents' own emphases – rather than from the top-down – by fitting evidence to the project logframe. Similarly, the approach was completed in isolation from the quantitative data and results; that is, the qualitative data was reviewed without reference to quantitative findings, with the goal of allowing themes to emerge, rather than simply using qualitative data to contextualize quantitative results or fill in gaps in quantitative data.[370] In particular, insights were identified that can speak to future

[370] Of course, the qualitative data also served these purposes later in the process. For instance, during later stages of the analysis, it was necessary to revisit the qualitative data with specific project outcomes in mind, and to check whether the qualitative data could shed light on unexpected findings from the quantitative analysis. But these uses followed after an initial exploration of the data.

program/intervention priorities, potentially unforeseen causes or consequences of interest, as well as previously unidentified gaps in programming.

As a second layer of analysis, qualitative data has also been examined for narrative evidence that can make sense of the historical processes and lived experiences behind quantitative findings, including prevalent social and gender norms (which were not necessarily expressed in quantitative surveys as a result of social desirability bias, but surface in qualitative narratives). Counter-narratives or minority narratives (that potentially contradict or qualify quantitative findings) were also given voice. Finally, in the last stage of analysis and in response to specific feedback from the analysis team, RI and the FM, qualitative data were queried selectively to make sense of quantitative outliers, unusual findings, or gaps in the quantitative data.

Methodological Limitations

As noted previously, a number of methodological issues arose during the midline evaluation. Given the importance of many of these issues and the solutions we employed, we have elected to discuss them at length in the main body of the report, beginning in Section 2.3 and proceeding through Section 2.6. These sections document the target and achieved sample sizes for different tools and cohort groups, analyse replacement rates and the predictors of replacement versus successful re-contact, analyse the baseline and midline household samples for comparability, and generally discuss the methodological limitations of the evaluation design in principle and in practice. In addition, outcome-specific methodological discussion and empirical analysis of methodological decisions are provided in the context of learning (Sections 4.2 and 4.3) and transition (Section 5.1).

Annex 4. Characteristics and barriers

The tables below describe the characteristics of the baseline and midline samples of cohort girls, respectively. Note that, to facilitate comparisons between the baseline and midline sample, we have limited the baseline sample to the portion that is comparable to the midline. For some characteristics and barriers a baseline proportion is not presented, because a household survey was not completed in the baseline survey.

TABLE 115: GIRLS' CHARACTERISTICS

Characteristic	Learning Sample (Baseline)	Learning Sample (Midline)	Transition Sample (Baseline)	Transition Sample (Midline)
Single orphan	---	11.1%	12.1%	10.3%
Double orphan	---	0%	0.4%	0.0%
Living in female headed household	44.4%	65.4%	61.3%	65.2%
Married	---	2.3%	8.2%	3.4%
Household doesn't own land for themselves	---	2.1%	47.2%	38.6%
Home uses poor roofing material*	---	0%	14.9%	14.0%
Gone to sleep hungry for many days in past year	---	15.5%	12.3%	4.3%
LoI different from mother tongue	---	0.8%	56.3%	17.1%
Girl doesn't speak LoI	---	16.1%	26.5%	8.4%
HoH has no education	---	8.4%	53.8%	51.2%
Primary caregiver has no education	26.7%	50.7%	53.7%	52.8%
Primary caregiver has no education	---	52.9%	28.9%	35.3%

TABLE 116: GIRLS' BARRIERS

Barrier	Learning Sample (Baseline)	Learning Sample (Midline)
Fairly or very unsafe travel to schools in the area	---	4%
Doesn't feel safe travelling to/from school	3.2%	2.2%
Girl travels more than 30 minutes to school	3.8%	2.4%
Doesn't get support to stay in school and do well	0.3%	4.8%
Girl has no choice in whether to attend school	67.1%	69%
Attends school half the time	---	10.1%
Attends school less than half time	---	4.3%
Doesn't feel safe at school	2%	0.6%

No seats for all students	20.1%	11.9%
Doesn't use drinking water facilities	36.3%	18.6%
Doesn't use toilet at school	26.6%	17.2%
No computers in class	87.3%	90%
Cannot use books or other learning materials at school	29.1%	16.7%
Disagrees teachers make them feel welcome	3.4%	6.6%
Agrees teachers treat boys and girls differently in the classroom	37.3%	39.8%
Agrees teachers often absent from class	12.8%	21.3%
Afraid of teacher	47.8%	56.5%
Uncomfortable asking teachers question	3.9%	---
Teacher punishes/disciplines when students gets lesson wrong	59.4%	61.6%
Physical punishment witnessed last week	25.9%	16%
Caregiver rates quality of teaching as poor	---	0.5%

Beyond comparisons of baseline to midline sample composition, an important metric for assessing the comparability of the sample over time is the similarity between re-contacted and replaced girls. This issue is also addressed in Section 2.4 of the report. In that section, we analyse the probability of successful re-contact among cohort girls from the baseline, studying whether some characteristics are associated with a higher probability of successful re-contact versus replacement. To the extent that girls who fall out of the sample are systematically different from those who remain in the sample through re-contact, this can produce bias in comparisons of learning scores – and other outcomes – over time. The table below provides a breakdown of sample characteristics for baseline cohort and bursary girls who were successfully re-contacted (middle column) and replaced at midline (right-most column). Note that all girl characteristics were drawn from the baseline survey, as replaced girls were, by definition, not contacted at midline. As a result, the analysis is limited to the data collected directly from girls at baseline, and the availability of household-level information is sparse. The sample analysed includes both bursary and cohort girls, as both were included in re-contact efforts at midline; however, some of the variables in the table were not asked of bursary girls, and those variables have a smaller sample size as a result.

TABLE 117: SAMPLE CHARACTERISTICS OF RE-CONTACTED VERSUS REPLACED COHORT AND BURSARY GIRLS

Characteristic	Re-contacted Girls	Replaced Girls
Living in female headed household	43.8%	50.1%
HoH does not have an occupation	30.7%	28.5%
HoH does not have any education	28.8%	30.7%
HoH has religious education only	18.4%	14.8%
Vision impairment	1.4%	0.3%
Hearing impairment	0.1%	0.3%

Mobility impairment	0.0%	0.3%
Cognitive impairment	0.1%	0.6%
Self-care impairment	0.1%	0.0%
Communication impairment	0.1%	0.0%
Any disability	1.7%	1.0%
Serious illness	10.4%	9.7%
Girl is old for grade	10.0%	16.9%
Difficult to move around school	31.6%	27.0%
Doesn't use drinking water facilities	35.2%	32.8%
Doesn't use toilet at school	29.8%	25.9%
Doesn't use areas where children play/socialise	46.1%	47.3%
School does not have learning materials	26.1%	24.7%
No computers at school	89.7%	87.2%
Not enough seats for children at school	23.2%	20.6%
Disagrees teachers make them feel welcome	4.0%	3.1%
Agrees that they are afraid of teacher	52.6%	52.8%
Teacher punishes students who get things wrong	61.3%	58.8%
Agrees teacher is often absent from class	14.2%	13.4%
Agrees she has no choice in schooling decisions	70.2%	65.9%
Over 30-minute travel time to school	4.5%	3.4%
Feels unsafe on way to school	3.6%	2.2%
Feels unsafe at school	1.9%	1.9%
Parents are disengaged	9.1%	10.9%

Annex 5. Logframe

The project's logical framework is attached as a separate annex.

Annex 6. Outcomes spreadsheet

The Outcomes Spreadsheet for the project at midline is attached as a separate annex.

Annex 7. Project design and intervention

The project's design and intervention is attached as a separate annex.

Annex 8. Key findings on output indicators

The project's findings on output indicators is attached as a separate annex.

Annex 9. Beneficiary Tables

This annex was completed by the project.

Project Beneficiary Counts

EGEPT targets marginalized students especially girls in 227 schools (173 primary schools and 54 secondary schools) in Somalia/Somaliland. The Project categorizes the students we work with into two categories as follows i.e. direct and indirect participants. Below we detail out the two categories

1. **Direct participants:** The direct program participants are divided into two categories

    a. Direct cohort participants: are girls studying in cohort grades i.e. Grade 6, 7 and 8 in target primary schools and girls studying in Form 1 and Form 2 in Secondary. Since EGEPT is a 3-year project, we assume that girls in non-cohort grades in lower primary school e.g. Grade 4 and 5 will progress and be studying by the second and third year in one of the cohort grades of EGEPT. In addition, girls in lower secondary i.e. Form 1 and Form 2 will be progressing to Form 3 and Form 4. Therefore, they are added to the direct program participants.

    b. Additional direct participants: EGEPT provides extra support in the form of bursary (school fee and/or cash to caregivers~~)~~ to some of the most marginalized girls within the target schools selected under some pre-defined criteria. These girls are part of the direct program participants. Some of them are studying in cohort grades and already categorized under point (1a) and few are studying in lower grades such as Grade 2, 3 and 4.

Program participants are selected based on marginalization criteria. Below is a list of criteria used for different program participants:

- Direct beneficiary: students in all target schools are either poor and/or marginalized.
- Girls who receive bursary support were selected because they are most marginalized i.e. they met at least one of the following criteria: can't afford paying school fee; is an IDP, is from an ethnic/clan minority group; is from an impoverished female –headed household; is a disabled girl, and/or is an orphan;

Direct program participants are exposed to different activities including but not limited to: bursary support to marginalized girls in the form of school fee in Puntland, Banadir and Galmudug and pocket money in Somaliland; provision of TLMs; school uniforms to girls; solar lamps for marginalized girls in rural and IDP settings; remedial classes for underperforming students; life skills through Girls' and Boys' Leadership Networks; and school based Child Protection interventions. Other set of activities that are implemented in schools in order to indirectly affect the learning outcomes of the students include but not limited to: Teacher training and coaching through Continuous Professional Development (CPD); provision of English as a Second Language (ESL) training of teachers; monthly incentives to teachers; provision of grants to support implementation of School Development Plans; provision of dry food ration to students in drought-affected schools; provision of water to drought-affected schools.

2. **Indirect participants:** are all boys studying in all grades in target school and girls who are studying in grades other than cohort grades; excluding all girls categorized under point (1a) as some girls in non-cohort grades are direct program participants. This category of participants is exposed to general activities that benefit the

whole school such as CEC training, support to school development plans, teacher training, dry ration distribution and water trucking

Further details on the target program participants and expected exposure is provided in the MEL framework.

Table 30: 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) – girls in the intervention group who are specifically expected to achieve learning outcomes in line with targets. If relevant, please disaggregate girls with disabilities in this overall number.	26,560	[This may equal the total project number in the outcomes spreadsheet and in the column to the left, or may be less if you have a staggered approach]	Please see the ***EGEPT Midline Beneficiary Numbers Calculation Methodology*** and the ***EGEPT Midline Beneficiary Numbers calculations.xlsx*** for the methodology behind the calculations which is based on actual 2018/19 enrolment figures for 227 schools.

Table 31: Other beneficiaries

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.	234[371]	We do not count any boys as part of the learning cohort. We track evaluation cohort boys' learning for comparison purposes
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 change, etc. but not necessarily achieve improvements in learning outcomes.	61,090	
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.	23,704	
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.	Total: 3,078 M=2,603; F= 474	Head teacher are added into the calculations
Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.	Total: 1,534 M=902; F=632;	CECs

[371] Sample size of boys at re-contracted successfully at midline

- Tables 32-35 provide different ways of defining and identifying the project's target groups. They each refer to the same total number of girls, but use different definitions and categories. These are girls who can be counted and have regular involvement with project activities.
- The total number of girls in the last row of Tables 32-35 are the same – these are just different ways of identifying and describing the girls included in the sample.

Table 32: 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	Y	387	
Upper primary	Y	14,656	1,321
Lower secondary	Y	40	316
Upper secondary	Y	11,477	
Total:		26,560	[This number should be the same across Tables 32-35]

Table 33: 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[372]
Aged 6-8 (% aged 6-8)			
Aged 9-11 (% aged 9-11)	Y	3,992	(61) 94
Aged 12-13 (% aged 12-13)	Y	8,895	(424) 254
Aged 14-15 (% aged 14-15)	Y	8,597	(673) 566
Aged 16-17 (%aged 16-17)	Y	3,552	(412) 547
Aged 18-19 (%aged 18-19)	Y	47	(39)226
Aged 20+ (% aged 20 and over)	Y	1,477	(0) 3
Total:		**26,560**[373]	[This number should be the same across Tables 32-35]

[372] The sample provided in bracket is for baseline. Figures outside bracket is for midline.

[373] The age-range numbers are based on estimated percentage driven from the midline cohort sample percentages

Table 11834: 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 domain of difficulty)	Y	156	74
Orphaned girls	Y	2,433	154
Pastoralist girls	Y	3,958	420
Child labourers	Y	0	24
Poor girls	Y	14,263	1,609[374]
Other (please describe): Minority	Y	1,193	Not collected
Other (please describe): IDPs	Y	4,197	173
Total:		**26,560**	[This number should be the same across Tables 32-35]

Table 35: 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	NA		
Out-of-school girls: have attended school, but dropped out	NA		
Girls in-school	Y	26,560	
Total:		**26,560**	[This number should be the same across Tables 32-35]

Table 36: Beneficiaries matrix

[374] This is the total sample size. We assume that all girls in the sample are poor. Please note that the sum of the numbers in the sample size column exceeds the total sample size. This is due that some numbers are overlapping

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/CEC	Local government
Learning	✔			✔				
Transition	✔			✔				
Sustainability	✔				✔		✔	✔
IO 1: Attendance	✔			✔				
IO2: Self-esteem and empowerment	✔			✔	✔		✔	✔
IO3: Quality of teaching	✔			✔	✔			✔
IO4: School management and governance	✔			✔	✔		✔	✔
IO5: Community attitudinal change	✔			✔	✔		✔	

External Evaluator Discussion

Based on the documentation provided by RI and the data available from the baseline and midline evaluations, we briefly analyse the expected accuracy of RI's beneficiary count. As noted by the project's discussion above, the number of expected EGEP-T beneficiaries is 26,560 at the time of the midline evaluation. We separate our comments into two parts: a discussion of the quality of the underlying data on which the beneficiary counts are made, and a discussion of the assumptions that go into those calculations.

The beneficiary numbers reported are based on direct enrolment counts from EGEP-T schools, compiled by RI and its implementing partners, as was the case at baseline. During the baseline, we noted that enrolment numbers had fallen slightly from the year prior, in a sample of schools (n = 176) where enrolment numbers for two consecutive years were available. This fact made us hesitant to assume stable enrolment numbers over time, which were part of RI's original beneficiary calculations – to the extent that enrolment numbers were declining, they may overestimate the direct beneficiary count.

While this is still a concern, the evidence available at midline no longer supports the idea that enrolment numbers are dropping in EGEP-T schools. Our analysis is based on classroom enrolment counts from 1,334 classrooms, divided between baseline and midline (in 105 schools visited in both periods). This sample is the same used for classroom headcounts in our attendance analysis, but here we focus on enrolment numbers based on classroom registers. Drawing from a comparable sample of schools, it appears that both girls' and boys' enrolment have increased slightly from baseline to midline. This is reinforced by increased numbers of students physically in class (i.e. attending on the day of the headcount) and by counts conducted during classroom observations. Therefore, we feel it is safe to assume relatively stable enrolment over time for the purposes of calculating beneficiary counts.

Beyond enrolment stability, the remaining question surrounding data quality concerns the accuracy of enrolment counts. Forcier did not collect school-wide enrolment numbers at midline, but did collect this data at baseline, for 114 schools in total. This data was collected from records provided by head teachers, and provides an opportunity to compare enrolment numbers from third-party data collection to those provided by RI.

In the 114 schools for which we have overlapping data, RI's enrolment numbers captured a total of 28,690 girls and 36,813 boys, across all grade levels (i.e. grade 1-8, forms 1-4).[375] In comparison, Forcier's data includes 25,115 girls and 34,212 girls in the same set of schools. While these discrepancies are fairly large, they may stem from the lag between data collected by RI (prior to the baseline) and data collected by Forcier (during the baseline), especially since documented declining enrolment over that period during the baseline evaluation. In general, our assessment is that the data used for RI's beneficiary counts are likely accurate; at worst, they may represent a very slight overestimation of the count, but it is equally possible that RI's enrolment counts are more accurate than those generated by Forcier's data collection teams.

Beyond the quality of the underlying data, we find that RI's approach to calculating beneficiaries is relatively conservative. As we noted in the baseline evaluation report, "RI includes only in-school children in their beneficiary counts." The count of direct beneficiaries includes only girls in project-targeted grades, and indirect beneficiaries include only girls and boys in the same school, and targeted community members. RI does not count out-of-school girls as potential beneficiaries, despite the fact that back-to-school campaigns and efforts to change community attitudes will also potentially benefit them. Also of note, given the fact that EGEP-T includes a number of urban schools: the project does not count beneficiaries in non-EGEP-T schools. To the extent that the project changes community attitudes, these changes are likely to benefit girls and boys at other schools in the same communities. By this standard, the project's reach in terms of indirect beneficiaries is likely underestimated. At the same time, we would not recommend revising the project's approach to calculating beneficiaries, because the alternative – calculating expected reach within communities beyond the scope of EGEP-T schools – would be technically difficult and fraught with assumptions, and because conservatism is generally warranted with regard to indirect beneficiaries.

[375] It is important to note that the numbers being compared come from RI's enrolment data at baseline, which have been updated for their beneficiary counts reported here. Therefore, any discrepancies discussed are merely suggestive of problems in enrolment counting, not in RI's specific current numbers.

Annex 10. MEL Framework

The project's Monitoring, Evaluation, and Learning (MEL) Framework is attached as a separate annex.

Annex 11. External Evaluator's Inception Report

At midline, the external evaluator produced a revised evaluation/analysis plan, which is attached as a separate annex. Note, however, that the evaluation plan was written prior to the start of fieldwork, and the analysis approach has undergone extensive revision since that time, in response to discussions with Relief International's Monitoring & Evaluation team and guidance from the Fund Manager.

Annex 12. Data collection tools employed at midline

The data collection tools – quantitative and qualitative – are attached as a separate annex.

Annex 13. Datasets, codebooks, and programs

Datasets and analysis programs

Datasets used in this evaluation are attached in a separate annex, in Stata (.dta) and Excel (.xlsx) format. Datasets provided in Stata format are labelled. Replication code is provided in the form of Stata .do files to support the replication of key baseline learning and transition findings, including all outcomes spreadsheet tables.

A number of datasets, including those that combined baseline and midline data, were produced during the analysis stage. Three types of datasets have been provided the Relief International to accompany this report:

- **Midline-only datasets** – these are cleaned versions of the midline data, including only variables specifically included in the data collection tools employed at midline
- **Merged datasets** – these are datasets that combine baseline and midline observations drawn from the same tools. For instance, the merged datasets TeacherSurvey.dta includes teacher surveys conducted at baseline and midline, with each interview occupying a single line, and the round of data collection identified by the variable *round*. Without exception, datasets are in "long" format, i.e. with baseline and midline datasets appended together (rather than merged), such that an individual interviewed at both baseline and midline will occupy two rows of the data.
- **Trimmed datasets** – these are datasets constructed from the merged datasets, but reduced in scale or scope to ease the replication process. For instance, the merged ChildSurvey includes every cohort member – cohort girls, bursary girls, and cohort boys – interviewed at either baseline or midline as a separate row in the data. However, the complex sampling and replacement procedures employed at midline can make replication of our learning or transition findings difficult with these datasets, as not all respondents are included in, e.g., the learning sample. The trimmed datasets are called Learning.dta, Learning_wide.dta, and Transition.dta, and can be used to facilitate replication. Where specified, data are provided in wide format (i.e. with one individual interviewed at both baseline and midline represented on a single row of the data).

The full list of datasets provided includes:

- Midline-only
    - ChildSurvey – the child survey, learning assessments, and household surveys conducted at midline
    - ClassObs – Classroom observations
    - Headcount – Classroom headcounts
    - HeadTeacherSurvey – surveys with head teachers and physical observations conducted at schools (formerly called the "School Survey")
    - MentorSurvey – survey with male and female mentors in schools
    - TeacherSurvey – survey with teachers
- Merged
    - ChildSurvey – child surveys (BL and ML), learning assessments (BL and ML), and household surveys (ML only) conducted with cohort girls, bursary girls, and cohort boys
    - ClassObs
    - Headcount

- HeadTeacherSurvey
- MentorSurvey – includes only midline data, as no mentor surveys were conducted at baseline
- TeacherSurvey

- Trimmed
  - Learning – child surveys, learning assessments, and household surveys conducted at baseline and midline, in long format. Restricts sample to girls included in the learning analysis at either baseline or midline. Girls re-contacted at midline exclusively for tracking transition are dropped.
  - Learning_wide – same as primary learning dataset, but reshaped to a "wide" format, such that one girl interviewed at both baseline and midline is represented on a single row.
  - Transition – child surveys and household surveys conducted with the midline transition cohort and strictly essential data from the baseline household survey Kish grid necessary for calculating transition among the baseline benchmark transition sample.

Codebook

The codebook below provides a summary of key variables for the merged household and learning assessment dataset. Comprehensive codebooks for all questions included in the midline tools are also available and have been provided in a separate attachment, along with Stata scripts for calculating key outcomes, as Annex 13.

Variable	Variable name	Comments
		Typical of All Datasets
Zone	zone	Coded by enumerators
Region	region	Coded by enumerators
District	district	Coded by enumerators
School	school_name	Name of school, coded by enumerators
School	school_code	Numeric code of school, coded by enumerators
Round	round	Round of data collection (baseline/midline)
Comparable Sample	comp_sample	Identifies the comparable sample for baseline-to-midline comparisons. Included in every merged dataset. Excludes data collected from replaced or replacement schools. Some analysis employs more restrictive comparable samples (e.g., the school management index, which employs the comparable sample of schools which have data on a given indicator).
		Child Survey Variables
Respondent type	resp_type	Identifies cohort to which child belongs (cohort girls, bursary girls, cohort boys, etc.)
Replacement status	rep	Identifies midline children as replacements or those successfully re-contacted
Replacement status	rep_full	Full classification of replacement/re-contact status of respondent

Original unique ID	bl_uniqueid	Unique ID assigned at baseline. For baseline respondents, this is their only unique ID. For midline respondents successfully re-contacted from baseline, this is their only unique ID. For midline respondents who are replacements for baseline respondents, this variable identifies *who* they replaced; their midline unique ID is survey_uniqueid.
Midline unique ID	survey_uniqueid	Unique ID assigned at midline. For midline respondents who were successfully re-contacted from baseline, this ID will match bl_uniqueid. For midline respondents who are replacements for a baseline child, this ID will uniquely identify them; it consists of their bl_uniqueid with an "R" concatenated to the end to identify them as a replacement.
Learning Outcomes		
Cross-sectional sample	learning_cross	Indicator variable for inclusion in the cross-sectional learning sample. Includes cohort girls, bursary girls, and cohort boys who appeared in the baseline and who were sought for re-contact at midline. Includes those successfully re-contacted at midline and those replaced. Excludes respondents whose schools were replaced at midline and those replacements.
Panel sample	learning_panel	Indicator variable for inclusion in the "true panel" learning sample. Includes cohort girls, bursary girls, and cohort boys who appear – the same child, without replacement – in both baseline and midline.
Panel weights	panel_tweight	Survey/sampling weights for the learning panel. Corrects for unequal cluster sizes and applies post-stratification weights to produce a sample that mirrors population characteristics at the school level.
Cross-section weights	cross_tweight	Survey/sampling weights for the learning cross-sectional sample. Corrects for unequal cluster sizes and applies post-stratification weights to produce a sample that mirrors population characteristics at the school level.
Grade	cohort_grade	Grades range from grade 6 to Form 2 among baseline respondents and grades 3 to Form 4 in the midline. Grade level is drawn from child responses primarily, with responses given by caregivers as an alternative in some cases.
Somali literacy score, baseline	som_agg_bl	Total baseline Somali literacy score, full assessment
Somali literacy score, midline	som_agg_ml	Total midline Somali literacy score, full assessment
English literacy score	eng_agg	Total English literacy score, full assessment (includes baseline and midline scores, depending on round of respondent observation)
Numeracy score, baseline	num_agg_bl	Total baseline numeracy score, full assessment

Numeracy score, midline	num_agg_ml	Total midline numeracy score, full assessment
Comparable Somali literacy score	som_comp	Total Somali literacy score, baseline/midline, including only the subtasks comparable between the two rounds
Comparable English literacy score	eng_comp	Total English literacy score, baseline/midline, including only the subtasks comparable between the two rounds
Comparable numeracy score	num_comp	Total numeracy score, baseline/midline, including only the subtasks comparable between the two rounds
Numeracy subtask 1	num_sb1	Midline subtask 1 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 2	num_sb2	Midline subtask 2 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 3	num_sb3	Midline subtask 3 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 4	num_sb4	Midline subtask 4 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 5	num_sb5	Midline subtask 5 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 6	num_sb6	Midline subtask 6 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 7	num_sb7	Midline subtask 7 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 8	num_sb8	Midline subtask 8 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 9	num_sb9	Midline subtask 9 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 10	num_sb10	Midline subtask 10 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 11	num_sb11	Midline subtask 11 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Numeracy subtask 12	num_sb12	Midline subtask 12 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 1	som_sb1	Midline subtask 1 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 2	som_sb2	Midline subtask 2 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 3	som_sb3	Midline subtask 3 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 4	som_sb4	Midline subtask 4 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 5	som_sb5	Midline subtask 5 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 6	som_sb6	Midline subtask 6 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Somali literacy subtask 7	som_sb7	Midline subtask 7 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number

Somali literacy subtask 8	som_sb8	Midline subtask 8 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 1	eng_sb1	Midline subtask 1 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 2	eng_sb2	Midline subtask 2 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 3	eng_sb3	Midline subtask 3 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 4	eng_sb4	Midline subtask 4 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 5	eng_sb5	Midline subtask 5 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 6	eng_sb6	Midline subtask 6 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 7	eng_sb7	Midline subtask 7 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 8	eng_sb8	Midline subtask 8 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
English literacy subtask 9	eng_sb9	Midline subtask 9 and comparable baseline subtask scores. Note: may not correspond to baseline subtask number
Transition Outcomes		
Transition pathway	transition	Derived from enrol (enrolment status) and grade variables from baseline, midline, and retrospective at midline, where appropriate. Note: to facilitate transparency and review of its construction, this variable is coded in the replication .do files and is not included in the datasets themselves. This allows replicators to see the exact coding decisions made by reviewing the .do file.
Sample	comp_sample	The comparable sample for baseline-to-midline transition analysis.
Subgroup sample	sub_sample	The sample utilized for subgroup analysis of transition outcome (midline only).
Binary transition indicator	transition_bin	Derived from transition_path, this is a binary variable indicating transition success (1) or failure (0). Note that it is also coded in the replication .do files included in this annex.
Enrolment, current	enrol_thisyr	Enrolment at time of data collection (baseline for benchmark transition girls; midline for midline cohort girls). Assembled from multiple sources.
Enrolment, previous year	enrol_lastyr	Enrolment at time *t-1* (one year prior to baseline for benchmark transition girls; baseline for midline cohort girls). Assembled from multiple sources.
Grade, current	grade_thisyr	Grade at time of data collection (baseline for benchmark transition girls; midline for midline cohort girls). Assembled from multiple sources.

Grade, previous year	grade_lastyr	Grade at time *t-1* (one year prior to baseline for benchmark transition girls; baseline for midline cohort girls). Assembled from multiple sources.
Age	trans_age	Age of respondent at time *t-1*, for the purposes of transition analysis. For baseline benchmark girls, *t-1* refers to the year prior to the baseline data; for midline cohort girls, *t-1* refers to the baseline.
Weight	trans_weight	Weight for overall and comparative (baseline to midline) transition calculations. Adjusts for differences in age distribution between benchmark transition and midline transition samples. Also adjusts sample distribution across strata, via post-stratification, to recover school-level population characteristics in the sample.
Subgroup weight	sub_weight	Weight for subgroup analysis of transition. Respondents are weighted such that the sample matches the school-level characteristics of the EGEP-T population. No age adjustment. Weights are applied to the full midline transition sample, including those falling outside the eligible age range for baseline-to-midline comparisons.

Score and Index Construction

The EGEP-T evaluation employs a number of composite indices to measure diffuse and multidimensional concepts. Where an outcome or indicator of interest is vaguely defined, latent, not directly observed, or consists of multiple dimensions, constructing a composite index allows for a more consistent measure than individual indicators alone. For instance, high-quality school management and governance consists of a wide variety of actions, policies, and procedures that head teachers, teachers, CEC members, and others might take. No single survey question or metric can capture all aspects of such a nuanced concept, which motivates our use of composite indices. In the subsections that follow, we describe the scoring and construction of each index used in the report.

Self-Esteem and Agency

We employed the standardization method for creating the indexes because the self-esteem and agency indicators have different scales and some of them were measured in different directions. In order to standardize them, the indicators were, first, re-coded to become consistent in terms of direction. Next, the responses were divided by the standard deviation of the variable where the result is a variable with a mean of 0 and a standard deviation of 1:

$$ystd_i = (y_i - \mu_y)/(\sigma_y)$$

Then these variables were added up to create a single index, using the following formula.

$$Rescaled\ X_i = \frac{X_i - Min_x}{Max_x\ -\ Min_x}$$

The agency index is developed by combining all the agency related indicators in order to understand how girls' decision-making power have improved as the result of taking part in the girls' clubs.[376]

Teaching Quality

Because the teaching quality index is derived from a specific set of coding rules, we provide full replication code for the generation of the index, including the coding of the underlying items the contribute to the construction of the score.

*Coding score items:

```
replace g8 = g8_bl1 if round == 0

gen z1 = 0

replace z1 = 1 if g8 == 1


gen z2 = 1

replace z2 = 0 if e2_1 == 0 | e2_2 == 0 | e2_3 == 0


gen z3 = 1

replace z3 = 0 if e3_1 == 0 | e3_2 == 0 | e3_3 == 0


gen z4 = 0

replace z4 = 1 if e4_1 == 1 | e4_2 == 1 | e4_3 == 1


gen z5 = 0

replace z5 = 1 if e5_1 == 1 | e5_2 == 1 | e5_3 == 1


gen z6 = 0

replace z6 = 1 if e6_1 == 1 | e6_2 == 1 | e6_3 == 1


gen z7 = 0

replace z7 = 1 if e7_1 == 1 | e7_2 == 1 | e7_3 == 1


gen z8 = 0
```

[376] The agency indicators did not have to be standardized (since they are on the same scale), but they are standardized in this case to make the agency index comparable with the self-esteem and agency index.

```
replace z8 = 1 if e8_1 == 1 | e8_2 == 1 | e8_3 == 1

gen z9 = 1
replace z9 = 0 if e8a_1 == 1 | e8a_2 == 1 | e8a_3 == 1

gen z10 = 0
replace z10 = 1 if e10_1 == 1 | e10_2 == 1 | e10_3 == 1

gen z11 = 0
replace z11 = 1 if e11_1 == 1 | e11_2 == 1 | e11_3 == 1

gen z12 = 0
replace z12 = 1 if e13_1 == 1 | e13_2 == 1 | e13_3 == 1

gen z13 = 0
replace z13 = 1 if e14_1 == 1 | e14_2 == 1 | e14_3 == 1

gen z14 = 0
replace z14 = 1 if e15_1 == 1 | e15_2 == 1 | e15_3 == 1

gen z15 = 0
replace z15 = 1 if e16_1 == 1 | e16_2 == 1 | e16_3 == 1

*Generate raw score and re-scaled score
egen raw_index = rowtotal(z1-z15)
sum raw_index
local min = r(min)
local max = r(max)
gen index = ((raw_index - `min')/(`max' - `min')) * 100
```

School Management and Governance

As noted previously, school management is a latent variable that can be measured in countless ways. In consultation with RI during the baseline evaluation, we selected a set of indicators that met two criteria: they captured what we consider key aspects of school management that are sufficiently widespread that they should be observed at all well-managed schools and they were areas of specific interest to EGEP-T, targeted at least partially by EGEP-T interventions. The latter point was discussed extensively during the baseline, as initial efforts to measure school management focused heavily on aspects of school management – such as the timeliness of teacher pay – that RI felt, and we concurred, fell outside the project's purview or ability to change. The resulting set of indicators was simultaneously meets both of the criteria above.

The school management index is scored on a 0-100 scale. In practice, the index ranges from a low of 7.1 to a high of 85.9 in the data analyzed here, which includes scores for both baseline and midline. The index consists of responses to 13 distinct survey questions; in the report, we group these questions into three overarching categories, but these categories do not imply unequal weighting or weighted aggregated. The three overarching categories, with their seven individual sub-indicators – and, in the case of record-keeping and policy promulgation, their four base values – are listed below.

- Quality of CEC management
    - Share of schools with a functioning CEC that meets at least once monthly
    - Share of teachers rating CEC management “very good”*
- CEC support for schools and students
    - CEC provides bursary support to at least one female student
    - CEC makes financial contribution to school
    - CEC makes non-financial (in-kind, labor, etc.) contribution to school
- Enactment of formal policies and record-keeping
    - Rating of schools’ record-keeping for students (0-4 scale)
        - Are records of student grades available? (1 point)
        - Are records of student grades either “mostly” or “extremely complete”? (1 point)
        - Are records of student enrolment available? (1 point)
        - Are records of student enrolment either “mostly” or “extremely complete”? (1 point)
    - Rating of schools’ promulgation of four formal policies (0-4 scale)
        - Does school have a mission statement? (1 point)
        - Does school have a Code of Conduct, and can they show enumerator a copy? (1 point)
        - Does school have a Child Protection Policy, and can they show enumerator a copy? (1 point)
        - Does school have a School Development Plan, and can they show enumerator a copy? (1 point)

For the purposes of calculating the score, there are three levels of aggregation. The first or lowest level of aggregation involves coding individual survey question(s) to arrive at the base scores. The base scores are all measured on a 0-1 scale at first, and constitute the lowest level within the structure outlined above (i.e. the third level for record-keeping and policy promulgation and the second level in all other cases).

TABLE 119: SCHOOL MANAGEMENT SCORE CONSTRUCTION

Question Text	Response Options	Response Coding (lowest aggregation)	Aggregation (2nd-level aggregation)
Quality of CEC Management			
How often does the CEC typically meet when schools are in session?	• Once every two weeks (or less) • Once per month • Once every two months • Once every three months • Once every 6 months • Less often than every 6 months • Don't know	"Once every two weeks" or "Once per month" → 1 point Otherwise, 0 points	Individual indicator (0/1 binary scale)
How would you rate the overall performance of the CEC in managing this school? Note: asked of teachers, not head teachers	• Very good • Good • Neither good nor bad • Bad • Very bad	Code teacher response as 1 if "Very good", 0 otherwise; calculate mean for each school (ranges from 0 to 1).	Individual indicator (0/1 continuous scale)
CEC Support for Schools and Students			
Does the school management (and/or the CEC) support girls with their school fees? How many girls' school fees are being paid, partially or completely, by the school management and/or CEC?	• Yes • No • Don't know	Yes (supports girls) and pays school fees for 1+ girls → 1 point Otherwise → 0 points	Individual indicator (0/1 binary scale)
Is your school receiving cash contributions from CECs?	• Yes • No	Yes → 1 point No → 0 points	Individual indicator (0/1 binary scale)
Is the local CEC making non-cash contributions (such as commitment of their time and skills,	• Yes • No	Yes → 1 point No → 0 points	Individual indicator

through for example painting the building, building a fence, maintenance of furniture, management of school feeding) to your school?			(0/1 binary scale)
Rating of School's Record-Keeping			
Baseline Are grades for Somali, English and mathematics for girls available? Are grades for Somali, English and mathematics for boys available? **Midline** Enumerator: could you get access to the grades records?	• Yes • No	**Baseline** Yes for both questions → 1 point Otherwise → 0 points **Midline** Yes → 1 point No → 0 points	Composite indicator – sum of four record-keeping sub-indicators (0-4 ordinal scale)
Enumerators: Rate the completeness of the grade records	• Extremely incomplete • Somewhat incomplete • Mostly complete • Extremely complete	Grades are available (see previous indicator) and enumerator rates grade records as "Extremely complete" or "Mostly complete" → 1 point Otherwise → 0 points	
Enumerator: could you get access to enrolment information?	• Yes • No	Yes → 1 point No → 0 points	
Rate the completeness of the enrollment records.	• Extremely incomplete • Somewhat incomplete • Mostly complete • Extremely complete	Enrolment records are available (see previous indicator) and enumerator rates records as "Extremely complete" or "Mostly complete" → 1 point Otherwise → 0 points	

Rating of School's Promulgation of Four Formal Policies			
Does this school have an official mission statement?	• Yes • No	Yes → 1 point Otherwise → 0 points	Composite indicator – sum of four policy promulgation sub-indicators (0-4 ordinal scale)
Does this school have an official code of conduct? Can you show me a copy of the code of conduct?	• Yes • No	Yes to both questions → 1 point Otherwise → 0 points	
Does your school have an official child protection policy? Can you show me a copy of your child protection policy?	• Yes • No	Yes to both questions → 1 point Otherwise → 0 points	
Does this school have a School Development Plan? Can you show me a copy of the current School Development Plan?	• Yes • No	Yes to both questions → 1 point Otherwise → 0 points	

In the table above, this step is represented by the column "response coding", which in every case produces a sub-score on a 0-1 scale. Note that, for some indicators, this is a simple coding of a single yes/no question; in other cases, it involves constructing a new variable based on two or more distinct survey questions. To illustrate, consider cases at the two extremes:

- School receives cash contributions from their CEC: the information necessary to code this indicator is based on a single survey question, which asked head teachers whether their school receives cash contributions from their CEC. Their yes/no answer is coded into a 1/0 binary scale, which comprises the sub-indicator.
- Completeness of school enrolment records: this base score is generated from two distinct questions. The first concerns whether the enumerator was given access to the enrolment records (which is, itself, a base score). If the enumerator was able to access the enrolment records, they were asked to rate the completeness of those records. For a base score of 1, the enumerator needed to be given access to the records *and* rate the records as "extremely complete" or "mostly complete." All other schools were assigned scores of 0.

Base scores, of which there were 13, were then aggregated up to 7 sub-indicator scores. In the outline provided above, these are represented by the second-level bullet points. The process for each indicator is described in the right-most column of the table above, labeled "2nd-level aggregation". Sub-indicator scores in this level were calculated on different scales, as noted above, with some represented on a 0-1 binary scale, others on a 0-4 ordinal scale, and one sub-indicator score measured on a 0-1 continuous scale.[377]

[377] The sub-indicator measured on a continuous scale measures the quality of CEC management as reported by teachers. Rather than ask head teachers to assess CEC management quality, we opted to use the responses of

In the next step, we took measures to account for the different scaling observed at the sub-indicator level by normalizing the variables to a 0-100 scale. Normalization is a linear transformation of a variable from one scale to another, without reference to the variable's standard deviation, as in the case of standardization. We applied the following equation to each sub-indicator *i*:

$$vnorm_i = \frac{v_i - 0}{max_i - min_i} \times 100$$

where *v* represents the input sub-indicator, or its current score, for sub-indicator *i*; *max* represents the maximum value observed for sub-indicator *i*; *min* represents the minimum value observed for sub-indicator *i*; and *vnorm* represents the resulting normalized value for the variable. This approach produced, from 7 sub-indicator scores measured on different scales, 7 sub-indicator scores measured on normalized 0-100 scales. The final step in the construction of the school management index score was to calculate the simple mean of these sub-indicators, itself naturally measured on a 0-100 scale.

regular teachers, because head teachers are typically members of the school's CEC. However, utilizing teacher responses complicated the calculation slightly, because multiple teachers were surveyed in each school. This fact necessitated calculating the mean teacher rating, where each teacher's response was recoded to 1 (rated "very good") or 0 (all other responses) and teacher responses for each school were averaged to a 0-1 continuous scale.

Annex 14. Learning test pilot and calibration

Marking of the learning assessments

The scoring methodology ensured that each subtask was weighted equally in the final aggregate score. Specifically, each subtask was scored as the percentage of items correct out of the total number of items (hence ranging from 0 to 100). In keeping with FM guidance, the reading tasks that involved a word-per-minute (WPM) score were censored at a cap of 100 WPM, with individuals who scored above 100 WPM being assigned a score of 100 WPM. The result is that all subtasks were individually standardized to range from 0 to 100. The total score for the numeracy and literacy assessments was then generated by taking the average of the subtask scores for that assessment (with each subtask being given equal weight), presenting the total percentage score based on the averaged subtasks, ranging between 0 and 100. This procedure ensured that each subtask (and the associated skills) made an equal contribution to the final score for a given assessment, and that the final scores for each assessment have a comparable range from 0 to 100.

Revisions of learning assessments

At the midline, adjustments to the learning assessments have occurred for two primary reasons. The first reason is to address shortcomings in the design of the baseline tools, in response to feedback and suggestions from the external evaluator (Forcier). During the baseline assessment, we found mild ceiling effects in Somali literacy and moderate ceiling effects in numeracy. At the baseline, 3.2 percent of girls achieved a perfect score on Somali literacy and 11.8 percent achieved a perfect score in numeracy. These ceiling effects were of particular concern because this occurred during the baseline – as the cohort girls progress over the life of the project, the ceiling effects are likely to become more pronounced and inhibit our ability to demonstrate improvements in learning outcomes.

In response, RI designed a series of more difficult tasks in numeracy and Somali literacy, which will be introduced at the midline. While these new tasks mean that the midline assessment, as a whole, is not directly comparable to the baseline assessment, the shared components of the baseline and midline will still allow valid comparisons over this period. Moreover, if the endline is designed in the same way as the midline, portions of the baseline and endline assessments will be directly comparable, and the entire midline and endline assessments will be comparable. By proactively addressing the ceiling effects between baseline and midline, the project has set itself up to be able to draw valid conclusions regarding learning between any two rounds of evaluation.[378] At the same time, a number of easier subtasks were removed from the tools, because they did not adequately distinguish between girls of different skill levels (i.e. they were too simple and almost all girls got them correct) and to avoid an assessment that, with their inclusion, would have become too long.

The second reason for adjustments is comparatively straightforward. Girls completing the exact same assessment in consecutive years may recall the questions asked, especially those built around a passage or story they must read. To ensure that girls do not recall the answers to questions, subtasks that were otherwise unchanged – i.e. no changes in difficulty were intended – were, nonetheless, adjusted to include different words, different names, etc.

[378] For a more detailed discussion, see the baseline evaluation report.

Piloting and Calibration of Learning Assessments

Immediately prior to the start of data collection training for the midline evaluation, RI and its implementing partners conducted a pilot test of the midline learning assessments. The pilot was conducted in mid-March, 2019. For each learning assessment, the pilot included a sample of 50 girls, each of whom completed the baseline and midline subtasks for their subject (i.e. Somali literacy, English literacy, or numeracy). All of the analysis that follows relies on paired comparisons, because the same girl completed both baseline and midline subtasks.

The data from the pilots were provided to Forcier in mid-March, at which time we analyzed the overlapping subtasks between baseline and midline for their comparability. We also analyzed the likelihood of ceiling effects in the full, non-comparable midline sample, with the goal of guarding against ceiling effects in the midline-to-endline stage.

Given the importance of maintaining equivalent difficulty of the learning assessments from baseline to midline, we discuss this issue at length in the report itself. We also make adjustments, in at least one case, for subtasks that – despite the piloting and calibration efforts employed prior to fieldwork – still appear to be of unequal difficulty between the two rounds. For complete documentation of assessment equivalence, see Sections 4.2 and 4.3 of this report. Our analysis of the pilot test, including the recommendations made for adjustment of specific subtasks, are provided in a separate attachment, labeled Annex 14.

Preliminary Analysis of Non-Truncated Learning Scores

The following tables present analysis of non-truncated scores for cohort and bursary girls. These initial findings provided the basis for our concern with ceiling effects and serve as a basis for comparison with the truncated scores presented as part of the main findings in this report.

TABLE 120: COHORT GIRL NUMERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Numeracy	Midline Numeracy	Difference Baseline to Midline	Baseline Benchmark Numeracy (Baseline grade)	Midline Benchmark Numeracy (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	50.8	54.6	3.7	50.8	58.0	7.2	-3.5
7 (8)	58.0	61.5	3.5	58.0	62.7	4.6	-1.1
8 (F1)	62.7	62.4	-0.2	62.7	71.2	8.5	-8.8
Form 1 (F2)	71.2	67.5	-3.7	71.2	75.2	4.0	-7.7
Form 2 (F3)	75.2	74.0	-1.1	75.2	80.6	5.4	-6.5

TABLE 121: COHORT GIRL SOMALI LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Somali Literacy	Midline Somali Literacy	Difference Baseline to Midline	Baseline Benchmark Somali Lit (Baseline grade)	Midline Benchmark Somali Lit (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	61.6	62.3	0.7	61.6	69.3	7.7	-7.0
7 (8)	69.3	67.5	-1.8	69.3	73.2	3.9	-5.7

8 (F1)	73.2	69.6	-3.7	73.2	79.0	5.8	-9.4
Form 1 (F2)	79.0	73.2	-5.8	79.0	81.1	2.1	-7.9
Form 2 (F3)	81.1	79.4	-1.7	81.1	86.5	5.4	-7.1

TABLE 122: COHORT GIRL ALTERNATE SOMALI LITERACY SCORES (NO ST3) – PROGRESS AGAINST BENCHMARK

Grade	Baseline Somali Literacy	Midline Somali Literacy	Difference Baseline to Midline	Baseline Benchmark Somali Lit (Baseline grade)	Midline Benchmark Somali Lit (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	20.5	20.8	0.2	20.5	23.1	2.6	-2.3
7 (8)	23.1	22.5	-0.6	23.1	24.4	1.3	-1.9
8 (F1)	24.4	23.2	-1.2	24.4	26.3	1.9	-3.1
Form 1 (F2)	26.3	24.4	-1.9	26.3	27.0	0.7	-2.6
Form 2 (F3)	27.0	26.5	-0.6	27.0	28.8	1.8	-2.4

TABLE 123: COHORT GIRL ENGLISH LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline English Literacy	Midline English Literacy	Difference Baseline to Midline	Baseline Benchmark English Lit (Baseline grade)	Midline Benchmark English Lit (Baseline grade +1)	Benchmark Difference	Arithmetic Difference in Difference
6 (7)	24.4	36.4	12.0	24.4	29.1	4.7	7.3
7 (8)	29.1	44.0	14.9	29.1	40.9	11.9	3.0
8 (F1)	40.9	55.2	14.3	40.9	56.6	15.6	-1.4
Form 1 (F2)	56.6	67.7	11.2	56.6	59.4	2.8	8.3
Form 2 (F3)	59.4	70.1	10.7	59.4	70.5	11.1	-0.4

TABLE 124: BURSARY GIRL NUMERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Numeracy	Midline Numeracy	Difference Baseline to Midline	Baseline Benchmark Numeracy (Baseline grade)	Midline Benchmark Numeracy (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	53.2	58.6	5.4	53.2	52.9	-0.3	5.7
7 (8)	52.9	56.1	3.3	52.9	65.7	12.9	-9.6
8 (F1)	65.7	59.4	-6.3	65.7	79.5	13.8	-20.1
Form 1 (F2)	79.5	81.6	2.1	79.5	65.0	-14.5	16.6

Form 2 (F3)	65.0	74.3	9.2	65.0	80.6	15.5	-6.3

TABLE 125: BURSARY GIRL SOMALI LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline Somali Literacy	Midline Somali Literacy	Difference Baseline to Midline	Baseline Benchmark Somali Lit (Baseline grade)	Midline Benchmark Somali Lit (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	62.5	61.7	-0.8	62.5	61.3	-1.2	0.4
7 (8)	61.3	64.7	3.4	61.3	70.9	9.6	-6.2
8 (F1)	70.9	62.5	-8.4	70.9	82.7	11.8	-20.2
Form 1 (F2)	82.7	83.5	0.9	82.7	75.3	-7.3	8.2
Form 2 (F3)	75.3	84.4	9.1	75.3	86.5	11.2	-2.1

TABLE 126: BURSARY GIRL ALTERNATE SOMALI LITERACY SCORES (NO ST3) – PROGRESS AGAINST BENCHMARK

Grade	Baseline Somali Literacy	Midline Somali Literacy	Difference Baseline to Midline	Baseline Benchmark Somali Lit (Baseline grade)	Midline Benchmark Somali Lit (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	20.8	20.6	-0.3	20.8	20.4	-0.4	0.1
7 (8)	20.4	21.6	1.1	20.4	23.6	3.2	-2.1
8 (F1)	23.6	20.8	-2.8	23.6	27.6	3.9	-6.7
Form 1 (F2)	27.6	27.8	0.3	27.6	25.1	-2.4	2.7
Form 2 (F3)	25.1	28.1	3.0	25.1	28.8	3.7	-0.7

TABLE 127: BURSARY GIRL ENGLISH LITERACY SCORES – PROGRESS AGAINST BENCHMARK

Grade	Baseline English Literacy	Midline English Literacy	Difference Baseline to Midline	Baseline Benchmark English Lit (Baseline grade)	Midline Benchmark English Lit (Baseline grade +1)	Benchmark Difference	Arithmetic DiD
6 (7)	26.5	38.1	11.6	26.5	29.2	2.7	8.9
7 (8)	29.2	41.5	12.3	29.2	35.9	6.8	5.6
8 (F1)	35.9	49.2	13.3	35.9	62.8	26.9	-13.6
Form 1 (F2)	62.8	63.4	0.6	62.8	56.4	-6.4	7.0
Form 2 (F3)	56.4	59.1	2.7	56.4	70.5	14.1	-11.4

Annex 15. Sampling Framework

The project's updated Sampling Framework (revised in March 2019) is attached as a separate annex, as well as the revised sampling plan produced by the external evaluator at the outset of this midline evaluation.

Annex 16. External evaluator declaration

Name of Project: Educate Girls, End Poverty – Transition (EGEP-T)

Name of External Evaluator: Forcier Consulting

Contact Information for External Evaluator: 301 W Platt Street, Suite 388, Tampa, Florida, 33606, USA; +1 239 297 0771

Names of all members of the evaluation team: Jonathan Forney, Samuel Ha, Zakia Nouri, Brenton Peterson, Elena Zambetti

_____ (Name) certify that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received.

Specifically:

- All of the quantitative data was collected independently ((Initials: ____)
- All data analysis was conducted independently and provides a fair and consistent representation of progress (Initials: ____)
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed (Initials: ____)
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by _____(Company) (Initials: ____)
- All child protection protocols and guidance have been followed ((initials: ____)
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols (Initials: ____)

_____

(Name)

_____

(Company)

_____

(Date)

Annex 17. Project management response

Relief International response to EGEP-T Midline Report

The Midline report submitted by Forcier consulting has provided the EGEPT consortium with rich information on the progress on the outcomes and intermediate outcomes since the baseline. ADRA, CISP and Relief International (lead) have read the report and are committed to reflecting upon the findings to inform programing in the remaining life of the project. Although the project already conducted an annual reflection workshop and drew upon some of the preliminary findings of the midline report a considerable amount of findings had been analysed and did not provide the project sufficient amount of information to truly reflect on the findings and make requisite adaptations. The consortium therefore plans to conduct further consultative and participatory process on the implications of the findings for the program. The reflections below however highlight below our some deductions of from the Midline findings and response to the recommendations made by Forcier. Upon the completion of the final report after review by the FM, the project will present more nuanced response to the midline findings and any additional changes in the upcoming RAM meeting.

Comments relating to the Theory of Change

As a whole the findings from the midline are generally consistent with RI's Theory of Change. In most cases, barriers influencing factors to educational achievement were correlated with the outcomes and identified to be consistent with theoretical expectations. Interventions identified including the addition of drought interventions informed by findings from baseline had a positive effect on progress of outcomes. The theory of change was validated and therefore EGEPT is not looking to make any radical changes to it.

Comments relating to Learning

Presentations of the Learning results from the midline were particularly interesting for the project in light of the complexities affected by the level of comparability from baseline to midline as detailed in the report. Measurable changes in aggregate scores for learning for girls' in particular for numeracy and English literacy may constitute positive intervention-effects although the effect is minimal when actual progress is compared against the benchmark. We noted with considerable enthusiasm that bursary girls have reduced significantly the learning gap between them and the cohort girls.

A surprising finding from the analysis was the primary trend that showed the sizable and statistically significant increase, from baseline to midline, in English literacy for all learner types. The project acknowledges that direct exposure on English subject content begun after the midline and hence direct contribution in this regard would not necessarily be from the project.

The continuation of the gendered learning gap observed during baseline although expected to change over time was expected to have shown considerable improvement. Analysis confirmed that "all learning scores increase monotonically by grade for cohort girls, and girls' learning scores lag behind those of boys in the same grade, across all grade levels and across all three assessments".

The analysis of the skills gaps for both girls' and boys' was interesting and will prove useful in enhancing content for teacher training, coaching support and remedial classes. For example, findings suggest that "additional or remedial efforts should be made to target skill-development in the reading of unfamiliar word and in reading comprehension for both Somali and English". Both boys and girls were having the same foundational gap in reading comprehension.

Of note is the consistent lower learning levels in Somaliland compared to the other zones from GEC 1. Different factors explain this finding including the high proportion of schools within the area (60.4%) being in rural areas compared to Puntland with only 14.3% and Galmudug with none in the sample. "Living in a rural (as opposed to comparatively urban) community was a significant geographic predictor of lower learning outcomes at both baseline and midline." Whilst comparison across zones can be sensitive to raise with the respective ministries, it is important that the project makes it known at ministry level that learning in Somaliland is not at the level expected.

In general, the consortium will need to deepen its reflections on the extent of implementation of the interventions for learning.

Comments relating to Self-esteem and Empowerment

The project confirmed as expected from the findings that it had a positive effect on the girls' self-esteem and confidence. Reference from the report that "girls confirmed that they feel confident in their skills and that their self-esteem has improved, particularly as a result of their participation in the EGEP-T project activities. More specifically, they believed that exposure to awareness-raising around the importance of girls' education, education and life experiences, as well as interactive class activities have greatly enhanced their confidence'. Girls' self-esteem significantly higher with an average score of 0.73, and with half of the girls' scores falling between 0.63 and 0.86.[379] However, boys' levels of confidence in their leadership skills appeared to have decreased (although level of statistical significance is quite low). The projects interventions although geared more towards the girls also aims at ensuring boys self- confidence and esteem. EGEPT has adapted the PSI life skills manual for boys and will train both male teacher mentors with support from the MoE Gender focal points. Key to this training will be to ensure that boys do not feel disempowered by supporting a safe and conducive environment for girls. Additionally, efforts to further enhance self- esteem and confidence for girls will be continued through the leadership networks and greater outreach activities with peers and the community at large.

Secondary schools and schools in urban areas are identified as having a bigger drop in the number of girls who believe they have strong leadership skills from baseline to midline, with a percentage difference of 1.8 and 4.9, respectively. Of particular note is the drop for schools in urban areas who having higher exposure would be assumed to have higher leadership skills. Galmudug which is largely urban also trailed on this trend when compared with the other zones. The project will explore more closely on the implementation of intervention targeting self- esteem and empowerment to further increase confidence of girls in the urban areas.

Comments relating to Community awareness and attitudes

The project aims to strengthen and diversify efforts to raise awareness and change attitudes and behaviours of community members. However, findings indicate that the progress in the shift in attitudes is still marginal from baseline. This is not entirely unexpected as behaviour change is a slow process and measurable results in this regard would be seen over longer periods of time. Key adjustments that the project seeks to address especially with regards to awareness, are detailed in the projects comments on sustainability and the comments regarding programmatic recommendations here below.

Comments relating to Child Protection

[379] Girls' midline index scores have a skewness and kurtosis of -0.72 and 3.58, respectively.

Overall, the midline established that there was significant increase from 14.6 percent at baseline to 43.1 percent at midline, a difference that is statistically significant at any conventional level, for schools that had established a child protection mechanism. Improvements were observed in both Somaliland and Puntland whereas the opposite was observed in Galmudug. From the sub group analysis conducted for learning, girls reported not feeling safe at school and provided reasons ranging from discomfort or even fear because of poor and unsafe school infrastructure, as well as threats of violence and bullying at school, and possibly sexual violence while at school. Qualitative information indicate that Galmudug as a project location provided the most direct testimony from a father. This findings especially in Galmudug are disturbing for the project. However, it is worth mentioning that child protection interventions have not begun in this location as the child protection rapid assessment that informs child protection action planning was recently completed. Training on child protection and psychological first aid was conducted for the gender focal points from the MoE in early September before the schools opened and is being cascaded to the male and female mentors in schools.

It was concerning to note from the findings that although corporal punishment had been noted to reduce, one of the qualitative interviews highlighted a case where "two girls from Form 4 got into a fight and were sent to the police station because we were abit scared that the issue would reach their families. They stayed there for eight hours….". EGEPT noted from the findings that in this case, girls were taken out of school without the consent of their parents for such an extended duration and likely suffered both emotional and psychological harm. This potential risk to girls attending school is the first one of its kind that the project has been aware of. The project intends to conduct follow up with the particular school in question to ascertain other emerging protection concerns and discuss with the school the response measures to these concerns and other indiscipline cases within the framework of the principles of Convention on the Rights of the Child. Additionally, in line with the focus on child protection in the remaining duration of the project, strengthening school based and community child protection increases in priority.

Comments regarding evaluation recommendations in the report

Two rounds of piloting learning assessments: The findings on the problems encountered with the learning assessment was quite disappointing considering the depth of effort to develop a comparable assessment. The recommendation is however noteworthy and the project will require support from the fund manager to further improve the assessments.

Evaluation time that does not coincide with exam period: The project plans to begin preparations for the Endline soon after the submission of the midline report. The proposed data collection period is planned for March 2020 which is two months before examination preparation or conduct.

Matching of teacher knowledge and skills with learner assessments to measure teaching quality: This is an interesting and more robust recommendation to assess teaching quality yet rather fairly complex in the design. The project is open to this consideration and would propose for a discussion with the FM on how best to do so. The project currently conducts pre-post-tests from teacher trainings, will automate observation visit tools from coaches and will collect data on use of ESL digital platform.

Replacement of 'lost' grade 8 students with Grade 7 or Form 1: The recommendation is well noted for subsequent evaluation rounds. However, further discussion is required considering the next evaluation round is an endline for EGEPT project.

Comments regarding programmatic recommendations in the report

Creation of transport program: The project noted that there was a small percentage of students who reported feeling unsafe on the way to school. However, the project notes that some of the CECs in some schools have initiated a transport program. The project is unlikely to take up an active role in providing transport to schools but will encourage its consideration during the follow up on child protection initiatives in school especially in conflict affected and remote areas.

CEC investment in school improvements: This recommendation is well aligned in the projects approach. As noted in the sustainability comments from the project, greater emphasis and support will be provided to CEC to implement the school development plans. In consideration with the need to focus on school infrastructure and provision of basic health and sanitation facilities, CECs will be informed of the impact of this on learning to aid in prioritization. However, care will be taken to ensure they do not only focus on this areas at the expense of other barriers to learning. Support to expand their resource mobilization avenues will be enhanced to allow them meet the depth and breadth of the school requirements. For example, the project will consider conducting an assessment of available revenue sources to share with CECs.

Opening gates for late girls: The project has initiated quarterly review meetings with head teachers in a bid to enhance engagement with them for buy in of CPD and child protection issues. This notable finding will be discussed with them drawing attention to its influence to the attendance of girls to school. It is anticipated that the misuse of this leeway for girls coming to school late may be raised by the head teachers. The project will also address the concern using the behaviour change platforms indicating that parents and caregivers should support their girls to go to school on time.

Attitudinal change on investing in girls' education: This recommendation is in line with the projects approach. Emphasis will be laid in all avenues utilized by the project for behaviour change such as radio advocacy, community dialogues, mobile radios etc. to further challenge communities' attitudes making hard choices in investing in educating girls. Borrowing from midline design on the use of different scenarios that families prioritize to fund vis. a vis. investing in girls' education, behaviour change scripts will be developed and used in those avenues. Moreover, the project will be keen to identify and support more influential community leaders to participate in efforts to advocate for girls education. Regular women and women will also be encouraged to further participate in this initiatives. It is worth noting however that the project has consistently involved both community leaders, men and women in advocating for girls education through the avenues mentioned above and hence it is unclear why the effect of awareness raising by community leaders in particular, is not felt by the midline respondents.

Teaching quality with focus on Pedagogy: The recommendation is in line with the projects approach. One of the core elements of the continuous professional development program for the project is gender- sensitive, inclusive and learner- centred pedagogy. Following training, teachers receive ongoing support by coaches using supportive tools during the observation visits to identify areas of strength and improvement in the classroom. Teachers thereafter are supported to improve on the specific area to be addressed. In addition, teachers recently received training on the conduct of formative assessments to shape their teaching approaches of which they are to roll out in the coming school calendar. However, the project will reflect upon this finding during the upcoming reflection meeting in September with MoE, TTI, Coaches and EGEPT staff on how best to improve the efforts to address this recommendation. This may include review of observation tools to include this

components; encouraging more self -reflection of teachers themselves on their teaching approaches and peer learning.

Improved efficiency in roll out of interventions: Measures to further improve the efficiencies in project implementation will be conducted. Informed by the feedback from the midline for project inputs, the project will assess roll out achievement against plans for timeliness using the recently developed project implementation monitoring tool (PIMT). Additionally, EGEPT will enhance reflection on monitoring data by the implementation team to ensure widespread uptake of interventions including full coverage of leadership networks (girls and boys clubs) in schools

Annex 18. Methodological Annex

In order to improve the readability of the report, some of the methodological details have been moved from the main body of the report into this methodological annex.

Data Collection Tools

The primary focus of the methodological discussion above concerned samples of children being tracked over time as part of either the learning or transition cohorts. It is important to note that the EGEP-T evaluation includes a number of data collection tools, only some of which are completed with children specifically. In many cases, these tools require alternative sampling approaches, which are outlined below. Before turning to a discussion of sampling, the data collection tools employed at midline are:

- Learning Assessments – SeGMA and SeGRA assessments designed to assess students' numeracy (mathematics) and literacy in both English and Somali.
- Household Survey – Includes survey modules for heads of household, and the caregiver of eligible girls (aged 11-18 years).
- Child Survey – formerly known as the Girls School Survey, Bursary Girls School, and Boys School Survey, respectively. This survey is given to every child contacted the midline, and captures information on transition, teaching quality, and self-esteem, among other topics.
- Head Teacher Survey – Collects information regarding the administration and governance of schools, teacher performance, child protection procedures, school facilities, and other key school- and community-level outcomes.
- Teacher Survey – Self-administered by teachers; collects information about teacher attitudes, teacher performance, and community attitudes.
- Mentor Survey – new at midline, this brief, self-administered survey collects information about trained mentors and their role in schools.
- Headcount Tool – Provides attendance spot checks for comparison to attendance and enrolment registers.
- Classroom Observations – Observation of classes while in session; provides data on teaching practices, teaching quality, and gender equity within classrooms.

In practice, the first three tools listed – learning assessments, household survey, and child survey, were consolidated into a single integrated tool at midline. This approach improved data quality by ensuring that we were able to match the same child from the learning assessment and child surveys together with their family members that completed the household survey. It is easiest to think of these

three tools as modules within a broader survey, not all of which were completed by each child surveyed. The table below specifies which types of respondents completed which modules in the overall child survey. Note that out-of-school girls referenced in the table are those girls who were out-of-school at the time of the baseline, located via the randomly-drawn household sample discussed in the previous section.

TABLE 128: SURVEY MODULES COMPLETED BY DIFFERENT COHORTS OF CHILDREN AT MIDLINE

Respondent Type	Learning Assessments	Child Survey	Household Survey
Cohort Girl	Yes (if enrolled)	Yes	Yes
Bursary Girl	Yes (if enrolled)	Yes	Yes
Cohort Boy	Yes (if enrolled)	Yes	No
Out-of-School Girl	No	Yes	Yes

Beyond the tools targeted at children and their households, the remaining tools focus on data collected at schools and with teachers or other school staff. Each tool-specific sample included the same first-stage sample draw (i.e. the selection of schools from RI's sample frame) at baseline. Likewise, all tools follow the same first-stage sampling strategy at midline and endline: visit all the same baseline schools, to the extent possible, and replace those that cannot be visited for accessibility or other reasons in line with the replacement procedures outlined above.

Beyond the first stage, samples varied in their second-stage structure. For instance, teachers within sampled schools were selected using stratification, while mentors (male and female mentors) were sampled randomly within each school (though most schools have no more than one male and one female mentor, in which case probabilistic sampling was not necessary or possible). The second stage for individual tools – outside of those tracking cohorts of children over time – as designed and applied at the midline, is described below:

- Head Teacher Survey – conducted at every school, it constitutes a census of head teachers in sampled schools. No specific second-stage sampling approach is required, because every head teacher is included.
- Teacher Survey – Eligible teachers are defined as those teaching English, mathematics, or Somali. Sample is stratified by participation in the EGEP-T Continuing Professional Development (CPD) training program – at least one CPD teacher was selected if a school had such a teacher available. The target sample was five teachers per school.
- Classroom Observations – A target of two classrooms were observed in each school, targeting classrooms where English, mathematics, or Somali were being taught and where a lesson was actively underway. Sample was stratified by CPD status where possible.
- Headcounts – One headcount was conducted per grade level available in the school (typically 8 grades for primary schools and 4 grades for secondary schools, with occasional deviations). Where schools had multiple classes in the same grade, the classroom was selected randomly.
- Mentor Survey – One female and one male mentor was selected in each school, if available. If multiple female or multiple male mentors were available, one was selected randomly.

The quantitative tools outlined above are supplemented by a number of qualitative tools designed for the EGEP-T's midline evaluation. Qualitative tools targeted a wide variety of populations, ranging from girls to CEC members to female remedial teachers to teacher coaches. A full breakdown of the qualitative sample, showing the number of interviews with different respondent types, by zone, is provided in the table below.

TABLE 129: QUALITATIVE SAMPLE, BY ZONE

Type of Interview	Somaliland	Puntland	Galmudug	Total
FGD - Boys (Grades 6-8)	1	1	1	3
FGD - Boys (Forms 1-4)	1	1	0	2
FGD - Girls (Grades 6-8)	2	2	1	5
FGD - Girls (Forms 1-4)	2	2	0	4
FGD - CEC	3	3	1	7
FGD - Fathers	3	3	1	7
FGD - Mothers	3	3	1	7
KII - Female remedial class teacher	0	1	0	1
KII - Female Teacher	1	1	1	3
KII - Head Teacher	3	3	1	7
KII - MOE (CP)		1	0	1
KII - MOE (GE)	1	1	0	2
KII - MOE (QA)		1	0	1
KII - Male Teacher	2	2	1	5
KII - Male Remedial teacher	2	1	0	3
KII - Coaches (trainers of teachers)	2	2	0	4
KII - Teacher Training Institute (TTI) officials	1	1	0	2
KII - Female mentor	1	1	0	2
KII - Male Mentor	1	1	0	2
Total	29	31	8	68

Sampling and Survey Weights

The baseline evaluation of EGEP-T employed survey weights for analysis of learning and transition outcomes. These weights served two different purposes, and the result was a compromise between two ideals. On one hand, a goal of weighting was to make the sample represent the EGEP-T target population more accurately. The baseline, therefore, employed post-stratification weights at the school level, with schools weighted so that the sample of selected schools matched the population of EGEP-T schools in terms of geographic zone, urbanicity, and IDP status. As the baseline report describes, Somaliland comprised 34.6 percent of the raw cohort sample; after post-stratification, Somaliland made up 38.1 percent of the sample, mirroring the 38.4 percent of the school-level population made up of schools in Somaliland.

On the other hand, the baseline also weighted individual cohort girls to ensure that each school was given equal weight in the analysis. In short, we sought weights that allowed a school with 10 cohort girls to count equally in the analysis as a school with 12 cohort girls, so that differences in school-level sample sizes did not affect the results. Importantly, these two goals are slightly contradictory, which resulted in a compromise approach: within each strata (defined by the crossing of zone, urbanicity, and IDP status), schools were weighted equally, regardless of their specific sample size; re-weighting of the strata ensured a sample distribution that matched the population characteristics of schools as closely as possible.

The learning and transition analysis in this report employs the same broad approach to weighting. In the learning analysis, the approach is identical: we define strata based on zone, urbanicity, and IDP status. Within strata, individual respondents are re-weighted so that each school within the strata is

weighted equally. Across strata, weights are applied such that the sample characteristics mirror those of the population at large, in terms of geographic spread, urbanicity, and IDP status. To illustrate the improvement in representativeness from the unweighted to the weighted sample, the table below reports the share of the population in each of nine strata represented in the panel sample of children analysed for the learning outcome. The right two columns report the unweighted and weighted sample distributions. As the table shows, the weighted sample mirrors the characteristics of the population much more closely than does the unweighted sample.

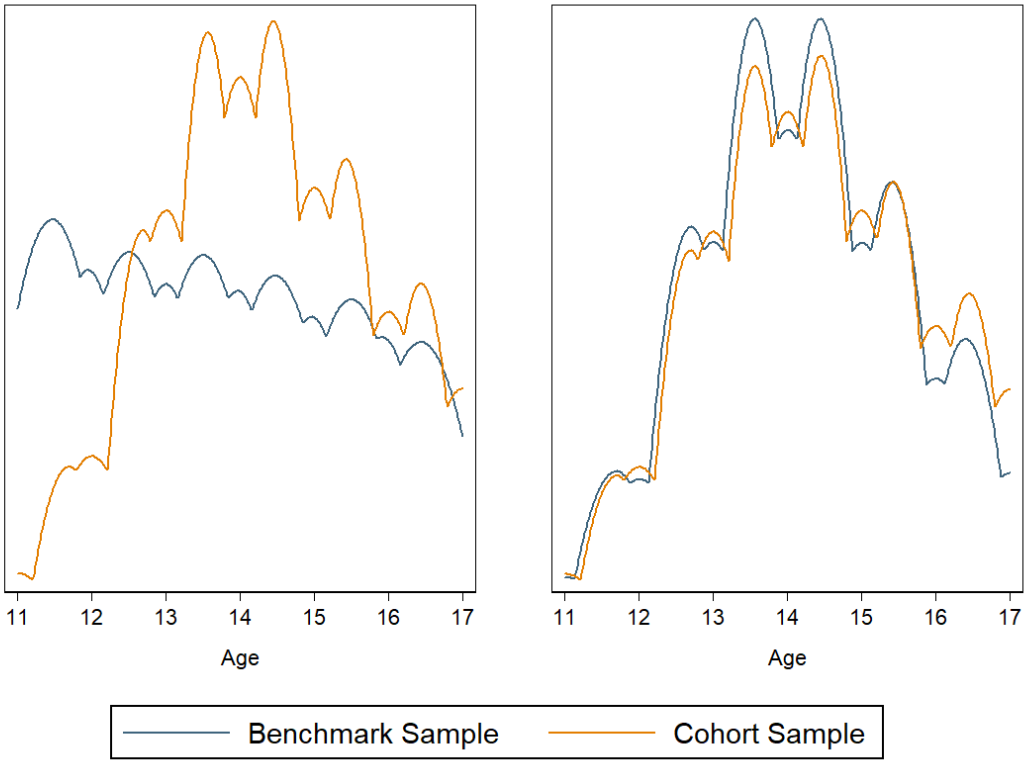
TABLE 130: CHARACTERISTICS OF LEARNING PANEL SAMPLE, WEIGHTED AND UNWEIGHTED

Strata	Population Distribution	Raw Sample Distribution	Weighted Sample Distribution
Somaliland, rural, non-IDP	24.3%	24.9%	25.0%
Somaliland, urban, IDP	2.0%	0.8%	2.0%
Somaliland, urban, non-IDP	16.8%	15.7%	17.4%
Puntland, rural, non-IDP	8.4%	7.3%	8.7%
Puntland, urban, IDP	7.4%	2.8%	7.7%
Puntland, urban, non-IDP	28.2%	37.5%	29.1%
Galmudug, urban, IDP	3.5%	1.6%	3.6%
Galmudug, urban, non-IDP	5.4%	8.4%	5.6%
Hirshabelle, urban, non-IDP	1.0%	1.0%	1.0%

In the context of the transition analysis, a further demand is made of the weights. In line with guidance from the FM regarding the transition outcome, weights must be applied to correct for age imbalances between the cohort and benchmark samples. Because transition outcomes are highly correlated with age, differences in the age distribution between the cohort and benchmark samples can produce significant bias in our analysis. In practice, this clearly applies to the EGEP-T evaluation, because the cohort sample is older, on average, than the benchmark sample. The weights calculated for the transition analysis, therefore, correct for age imbalances between the baseline and midline samples, and attempt to mirror the population characteristics of EGEP-T schools.[380] The figure below plots the age distribution of the benchmark and transition samples before (left-hand panel) and after (right-hand panel) weights were applied. Although the weighting scheme attempts to satisfy multiple goals, it is able to produce a well-balanced sample with respect to age in the context of transition.

[380] Given the complexity of the weighting scheme, the transition weights do not ensure equal weighting of schools with different numbers of girls. The fact that schools had different numbers of girls in the baseline benchmark and midline transition samples, combined with the fact that the weights are already designed to accomplish two other statistical goals, would mean that an approach which also corrected for unequal cluster sizes would produce weights of dramatically different sizes, potentially decreasing precision and statistical power.

FIGURE 45: RAW (LEFT) AND WEIGHTED (RIGHT) AGE DISTRIBUTION, TRANSITION SAMPLE



Fieldwork Challenges

Data collection for the EGEP-T midline evaluation ran from early April to approximately mid-May, 2019. Given the context in which the evaluation was conducted, fieldwork challenges were relatively limited. However, two major challenges were encountered by the field teams:

CLOSED SCHOOLS

A number of schools were closed, typically for exam preparations, the examination period itself, or following examinations. The timing of the evaluation was such that it was being completed just before the end of the school year, after which schools would be closed from May through August or September, depending on the area. Therefore, it was critical that data collection be completed before schools closed for the long break. A number of schools were closed prior to our planning fieldwork dates, with the earliest school closing April 22. In total, nine schools were closed for exams or exam revisions, with the complete list of affected schools provided in the table below.

TABLE 131: SCHOOLS CLOSED OR CONDUCTING EXAM PREPARATION DURING FIELDWORK

School Name	Location	Data Collection Tools Affected by Closure/Exam Revisions
Galkacyo Primary	North Mudug, Puntland	• No classroom observations
Hawlwadaag Primary school (Alanly)	South Mudug, Galmudug	• No classroom observations • No headcounts
Mudug Primary	South Mudug, Galmudug	No tools affected

Godob-Jiraan Primary School	Nugal, Puntland	• No classroom observations • No headcounts
Daawad Primary School	Nugal, Puntland	• No classroom observations • No headcounts
Sunijiif Primary school	Nugal, Puntland	• No classroom observations • No headcounts • No teacher surveys
Kalabayr (Nageeye) primary school	Nugal, Puntland	• No classroom observations • No headcounts
Al-Mamum Secondary School	Galgadud, Galmudug	• No classroom observations
Eyl Secondary School	Nugal, Puntland	• No classroom observations • No headcounts

In schools that were closed, field teams were required to locate children at their homes, rather than through the school itself (though typically with the assistance of the head teacher). This presented an additional challenge, as our field teams had not previously visited students' homes at baseline. An overriding focus during fieldwork was the maintenance of the panel of cohort and bursary girls, and there was concern that girls in schools that were closed would be less likely to be located successfully and more likely to require replacement in the learning sample or fall out of the transition sample altogether. In practice, this concern was unjustified – as part of the evaluation of re-contact rates in the previous section, we studied re-contact rates specifically in the nine schools that were closed or conducting exam preparations. Re-contact rates were actually higher in these schools.

Nonetheless, school closures impacted other aspects of data collection. Often, field teams were able to complete surveys with teachers and mentors, despite the school closure, because teachers were still at school or could be located by the head teacher; however, at times, teachers had travelled away from the school and were not available. The most problematic aspect of visiting schools that had closed was the classroom observation and headcounts, which could not be completed at all. In instances in which key school-level characteristics – such as the activity level of a female mentor – are derived from teacher or mentor surveys, the inability to collect data from teachers can impact our ability to study learning or transition outcomes as a function of those school-level characteristics.[381]

RAMADAN

The midline evaluation also overlapped, in a small way, with the start of the Ramadan holy month. Fieldwork was originally scheduled to be completed for the final team by May 2, with Ramadan beginning on May 5. However, delays following the completion of training led part of the fieldwork take place during the Ramadan holiday, with four teams conducting fieldwork on May 5 or after (as late as May 9 in the case of two teams). Other teams finished fieldwork on May 4 and travelled during the first few days of Ramadan. While Ramadan did not appear to present special challenges in terms of respondent availability – typically, we would expect potential respondents to become more tired and less available later in the month – completing fieldwork while fasting presented an extra challenge for the teams required to do so.

[381] In practice, most school-level characteristics are derived from survey with head teachers – who were generally available despite school closures – or from data provided by Relief International.