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:	Baseline
Report date:	June 2018

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

Baseline Evaluation

JUNE 13, 2018

FORCIER

Abbreviations

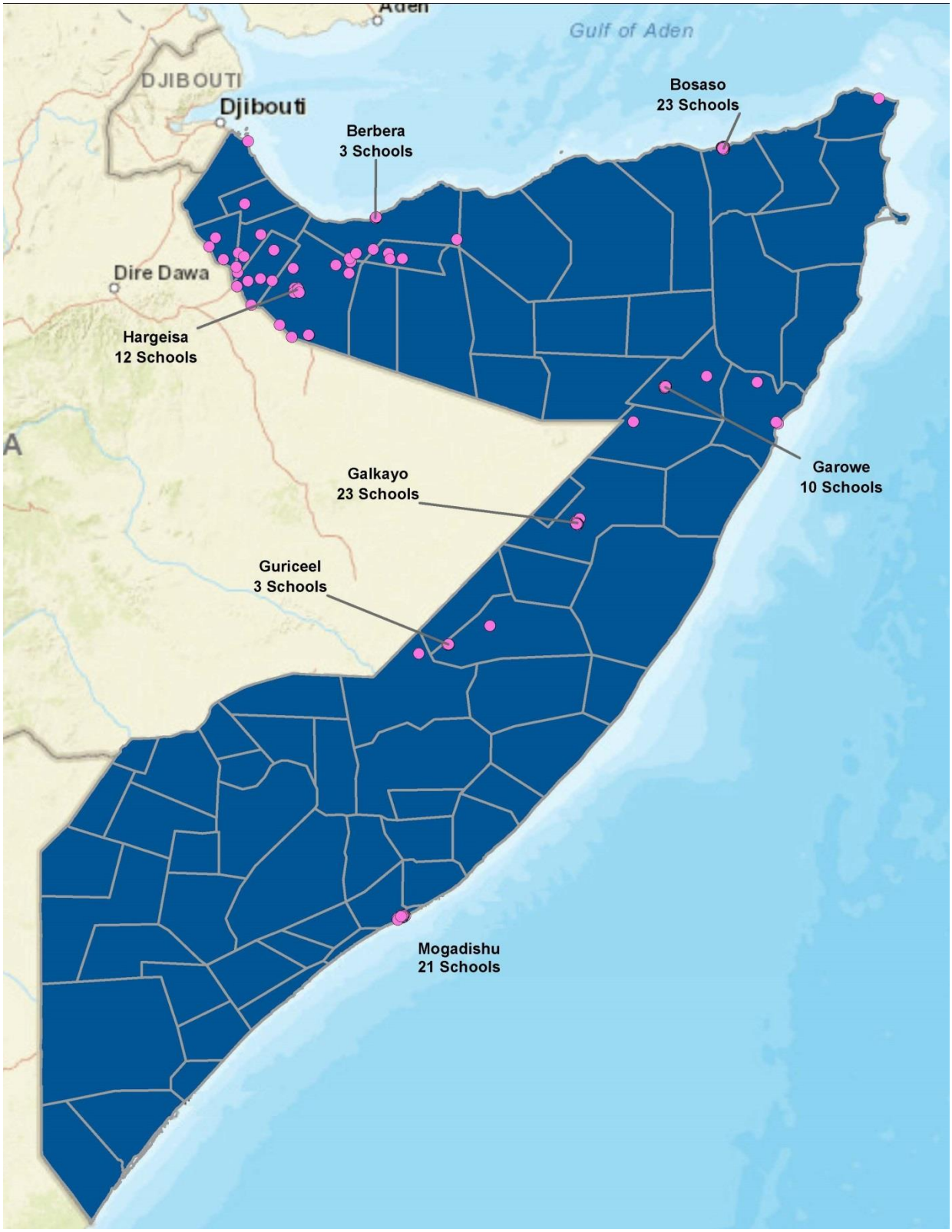
ADRA	Adventist Development and Relief Agency
CEC	Community Education Committee
CISP	Comitato Internazionale per lo Sviluppo dei Popoli
DFID	Department for International Development
EGEP	Educate Girls, End Poverty Project
EGEP-T	Educate Girls, End Poverty – Transition Project
EMIS	Education Management Information System
FGD	Focus Group Discussion
FM	Fund Manager
FRS	Federal Republic of Somalia
GEC	Girls Education Challenge
HH	Household
ISG	In-School Girls
KII	Key Informant Interview
MELF	Monitoring, Evaluation and Learning Framework
MoE	Ministry of Education
OOS	Out-of-School
PSU	Primary Sampling Unit
PWC	PriceWaterhouseCoopers
YLI	Youth Leadership Index

Table of Contents

List of Figures

List of Tables

Evaluation Fieldwork Sites



Executive Summary

This report constitutes the baseline 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, establishes baseline levels of key indicators against which future evaluations will be compared, and makes recommendations for future evaluations and project implementation.

To say that the educational environment in Somalia[1] is challenging would be a significant understatement. Since 1991, Somalia has been embroiled in continuous conflict of various types. Civil war, regional separatist movements, and extremist militias – the most prominent of which is Al-Shabaab – have left Somalia's economy and society in a desperate state. By most metrics, Somalia is among the ten poorest countries in the world on a per capita basis, and it is in a nearly continual state of political crisis.[2] These problems are exacerbated by its susceptibility to drought conditions and its vulnerability to climate change-related shocks to the environment and economy. These systemic problems have impeded the development of an effective Somali educational system. The current system is fragmented across jurisdictions and severely under-funded, with poor schooling infrastructure and a lack of qualified teachers at all levels. Girls face particular challenges in this environment, as tradition and cultural beliefs sometimes dictate that girls do not attend school or drop out of school early in order to get married.

In this context, RI and its partners – Adventist Development and Relief Agency (ADRA) and Comitato Internazionale per lo Sviluppo dei Popoli (CISP) – are implementing the second phase of their earlier EGEP project. EGEP-T builds on the foundation laid by EGEP by taking a multilevel, holistic approach to strengthen girls' educational attainment, with a focus on girls, households, schools, communities, and the educational system overall. EGEP-T utilises the lessons learned from EGEP by expanding and targeting interventions that were shown to work most effectively during the previous project, including, for example, the distribution of solar lamps and sanitary kits, the promotion of girls' clubs, and the provision of bursary support to girls.

EGEP-T's Theory of Change targets key barriers to girls' learning and their willingness and ability to stay enrolled in school. Specifically, it targets: financial obstacles through interventions, including the provision of bursary support to severely marginalised girls; obstacles related to girls' available time, by providing solar lamps to facilitate studying at night; and obstacles to girls' attendance, including the provision of sanitary kits. The project also takes a multifaceted approach to improving girls' self-esteem and confidence by promoting girls' clubs, providing remedial courses, and leading workshops designed to impart valuable life skills.

At the school level, one of the project's primary focuses is on teacher quality. EGEP-T will train teachers in gender-responsive pedagogy, subject content, and remedial teaching practices. Teachers will also be provided incentives and learning resources. Beyond teacher quality, EGEP-T seeks to improve school management by building the capacity of local Community Education Committees

[1] The writers promote no particular views with regards to the issue of the claimed independence of Somaliland. Where 'Somalia' is written throughout the article, this includes Somaliland, but should not be seen as a political statement of support or otherwise on the issue.

[2] See, e.g., per capita GDP, as calculated in the World Bank's World Development Indicators in 2016.

(CECs). Drawing on system-level interventions with relevant ministries, EGEP-T will also promote better monitoring of schools. Finally, EGEP-T includes a number of interventions designed to promote attitudinal and behavioural change at the community level, with the goal of increasing support for girls' education.

In total, the project is expected to reach 31,411 direct beneficiaries. Direct beneficiaries are marginalised girls from Grade 6 to Form 4 and severely marginalised girls from Grade 2 to Form 4. A further 98,232 indirect beneficiaries will be impacted by teacher-, school-, and community-level interventions that influence their schools.

EGEP-T builds on extensive previous learning, making use of lessons from the evaluation of EGEP to focus on gender-transformative interventions. The vast majority of EGEP-T interventions are targeted to specific sources of gender inequality in educational attainment in Somalia. For instance, EGEP-T will provide solar lamps to the most marginalised girls, reducing the damage that an outsized chore burden does to girls' educational achievement. The same set of girls will also receive bursary and other financial support, since girls and their education are disproportionately affected by household financial constraints. Similarly, the project will sponsor school-level interventions, such as the development of female teacher mentors and girls' clubs, which will improve the psychosocial support available to girls at school. Many of the selected interventions go beyond promoting girls' education alone, and seek to challenge existing stereotypes and structural inequities at the level of educational policy and teacher hiring, offering the promise of broad progress.

This baseline 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 will be based on comparisons of student performance over time, vis-à-vis benchmarks established in this baseline where appropriate. Respondents in the primary evaluation sample were in Grade 6 through Form 2 at the time of the baseline.

EGEP-T's core outcomes are learning and transition, where transition is defined as successful promotion from one grade to another without dropping out. At the baseline, literacy and numeracy levels were uniformly low, given the grade level of the respondents. The mean numeracy score among the cohort of girls to be tracked in future evaluations was 68.9 per cent. Meanwhile, the mean aggregate English and Somali literacy scores were 38.6 and 76.4 per cent, respectively. Given the divergence between English and Somali scores, and the context of language-of-learning policy in Somalia, we focus on the two scores separately throughout this report.

Baseline transition rates were not established among the cohort of girls to be tracked over time, because they were selected from within intervention schools (and therefore have a de facto transition rate of 100 per cent). Instead, a benchmark transition rate was established using a random sample of households in the areas around EGEP-T schools. Among these households, girls 11-18 had a baseline transition rate – meaning promotion from their previous grade – of 72.7 per cent.

The third primary outcome of the project concerns sustainability. Using a scorecard which aggregates nine distinct sustainability indicators, EGEP-T achieved a baseline rating of 1.45 on a scale from 0 to 4. With regard to sustainability, the most promising results came from indicators concerning attitudes within communities, and school-level outcomes, such as the extent of CEC financial support for their schools. More pessimistically, sustainability at the level of the system – promoting child protection mechanisms and gender development strategies – and sustainability focused on community actions or behaviours, rather than attitudes, was virtually non-existent at the baseline, and promises challenges for the project over the next three years.

Beyond learning, transition, and sustainability, the evaluation measured the baseline levels of five intermediate outcomes: attendance, girls' self-esteem and empowerment, teaching quality, school management and institutional governance, and community attitudes and behaviours. Due to the multidimensional nature of these outcomes, the evaluation employed a wide range of indicators to capture their varied aspects. Attendance, based on in-person headcounts, showed attendance rates of 84.2 per cent overall, with girls' attendance rates of 83.8 per cent. Girls' self-esteem was captured using an index of approximately two dozen self-reported items, with just 0.4 per cent of girls exhibiting high self-esteem, and another 24.7 per cent exhibiting moderately high self-esteem. Girls' empowerment – their role in decision-making that affects them – was relatively high in most respects; it was lowest in the case of schooling decisions, where just 25.6 per cent of girls are solely responsible for decision-making, and another 51.5 per cent make decisions jointly with their families.

At the school level, teacher quality is uneven. Students broadly perceive gender equity in their classrooms. But teachers are not uniformly prepared for class, and a large number of classes observed during the baseline involved students copying down information from the board for much of the observation period. Absenteeism is the single-biggest shortcoming in teacher quality – on average, teachers miss 1.35 days every two weeks of school. School management is also highly uneven. Many CECs meet only infrequently, and are perceived as poor managers by teachers. Schools generally keep good records of student grades and attendance, but other aspects of management, including the promulgation of school-level policies for child protection, a staff code of conduct, and so forth, is lacking.

Finally, community attitudes are almost exclusively positive regarding girls' education. But this positivity is most common in the abstract. For instance, 89.0 per cent of community members wish for their daughter to stay in school through university. However, 22.2 per cent would be more likely to withdraw their daughter from school than to sell household assets when faced with a pressing financial need. Girls themselves perceive gender bias in the tangible support communities would be willing to provide to students in need, suggesting the need for continued emphasis on attitudinal and, especially, behavioural change.

1. Introduction

1.1 Project Context

Somalia/Somaliland has been in a state of civil war since 1991. In 2012, the first permanent government was formed in Somalia since the start of the civil war. However, the situation is still unstable as militia groups and Al-Shabaab are active in several regions of Somalia. Somaliland is the only region in Somalia to experience peace and stability. However, Somaliland is still developing and has recently experienced an influx of internally displaced people because of the ongoing drought. As of 2017, The Fund for Peace's Fragile States Index places Somaliland, Puntland, Hirshabelle, Banadir, and Galmudug together as the second most fragile state in the world, based on indicators such as economic inequality, the volume and treatment of refugees, human rights violations, and the availability and quality of public services.[3] Corruption is also endemic, further undermining the quality of public services and other outcomes, with Somalia ranked as the most corrupt country in the world last year.[4]

This situation is further exacerbated by security risks and unfavourable climate conditions. Most recently (October 16, 2017), an unknown terrorist group carried out an attack that claimed the lives of over 300 people, including at least one EGEP-T beneficiary, in Mogadishu.[5] Terrorist activities in Banadir, Galmudug, and Puntland make it difficult for parents to ensure their children's safety at home and at school. In addition, schools have been the site of conflict in the past – though not at EGEP-T schools – and schools have been forced to close for long periods while they are either rebuilt or until conflict in the area subsides. Importantly, conflict is not limited solely to terrorist attacks and the activities of militant groups, such as Al-Shabaab. Rather, inter-clan conflict often forces schools to be closed for weeks or months at a time. For instance, during EGEP implementation, several schools in Galkayo were shuttered for one to two months each year due to such fighting and the internal displacement of teachers and students that it caused.

Furthermore, Somalia occasionally experiences prolonged drought leading to school closures and famine that affects millions. The most recent drought started in 2015 and was still in effect at the time of baseline data collection, having severely affected nearly 1.4 million people at the time of writing.[6] Despite better-than-expected rains in late 2017, the Food Security and Nutrition Analysis Unit expects 2.7 million people to require food assistance between now and June 2018.[7] The nation is particularly vulnerable to droughts as approximately 47 per cent of Somalia's general population does not have access to clean water, while weak government regulation has allowed private water suppliers to charge exorbitant prices. Consequently, households that cannot afford to purchase water from private suppliers obtain water from potentially unclean sources[8]. The lack of access to drinking water in homes or close by burdens girls and women in particular. In addition to spending hours traversing often dangerous roads to fetch water, girls are investing less time in the classroom.

[3] Fragile States Index 2017, available at http://fsi.fundforpeace.org/, last checked 5 November 2017

[4] Corruption Perceptions Index 2017, Transparency International.

[5] https://www.theguardian.com/world/2017/oct/15/truck-bomb-mogadishu-kills-people-somalia

[6] https://reliefweb.int/disaster/dr-2015-000134-som

[7] Somalia Food Security Outlook, February to September 2018, Food Security and Nutrition Analysis Unit.

[8] https://www.unicef.org/about/execboard/files/2017-PL13-Somalia_draft_CPD-EN-2017.06.19.pdf

Drought and conflict, among other factors, have contributed to significant internal displacement and the social, educational, and financial dislocation that accompanies migration and displacement. A substantial share, 12.7 per cent, of EGEP-T schools specifically serve Internally Displaced Persons (IDPs).[9] Even among schools not specifically targeting IDPs, many EGEP-T schools serve a student population that consists of a large number of IDPs.[10] There are also orphans and children with disabilities in many of the EGEP-T school populations. According to counts of students conducted by RI and its partners, there are 249 children with disabilities amongst the direct beneficiaries in EGEPT-target schools (representing 0.8% of that group).[11]

For instance, several schools in Mogadishu serve student populations that include over 25 per cent IDPs, and others include over 50 per cent orphans. The effect of serving such a disadvantaged population is that barriers, which would be significant for other schools (i.e. the financial strain drought causes), are overwhelming. This undermines educational quality, prompting large drops in enrolment, and reducing the effectiveness of the educational system and programmatic interventions alike.

The combination of these conditions has also made these regions among the worst locations in the world to be a girl. Save the Children's Girl's Opportunity Index[12] places these four regions at rank 140 out of 144 (the lower the rank, the worse the conditions). Girls have been historically treated unfairly in both the job market and schools in Somalia. Unemployment is far higher for women, and girls often have worse health and economic outcomes than men in Somalia. Historically, a key

---

[9] Relief International defines IDPs as per the UN definition: 'According to the Guiding Principles on Internal Displacement, internally displaced persons (also known as "IDPs") are "persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized border."' See definitions provided by the Office of the High Commissioner on Human Rights at the United Nations, available at http://www.ohchr.org /EN/Issues/IDPersons/Pages/Issues.aspx).

Schools that specifically serve IDP populations are ones that are either in an IDP camp, or are in an urban area but have been designated by the government to specifically serve IDPs.

[10] As we discuss in greater detail elsewhere in the report, the evaluation sample actually underrepresents the number of IDPs impacted by EGEP-T's interventions, for two reasons. First, the sample included a disproportionately low number of IDP schools relative to the number of IDP-specific schools in which EGEP-T works. This is because many IDP schools lack grades 6 through 8 that are targeted for evaluation, meaning that the project's impact among IDPs is likely greater than that described here. Second, the evaluation was unable to reliably identify IDP students enrolled in non-IDP schools, because no household survey was conducted with respondents in the learning cohorts; that is, the evaluation was limited to questions that could be asked of students themselves, and many are unable to reliably identify themselves as IDPs. Throughout this report, we describe results disaggregated by the IDP status of *schools*, rather than the IDP status of *individuals*, which severely underestimates the share of EGEP-T beneficiaries who are actually IDPs.

[11] These figures are derived from a count of children with disabilities in which head teachers or principals were asked to indicate the number of children with disabilities enrolled in their school. As a result, the definition of disability in this exercise is unlikely to conform precisely to the Washington Group standard. The baseline evaluation employed the Washington Group questions and found that 1.9 per cent of the girls learning cohort experienced moderate or severe disability (i.e. experienced "a lot of difficulty" or could not do tasks at all as a result of their disability). This figure may understate the extent of disability among beneficiaries, because girls were asked the questions directly – rather than collecting the information from their caregivers. Questions regarding disability were not posed to the project's "bursary girls", who almost certainly have higher rates of disability, owing to their severely marginalised status (these girls are not all part of the core learning cohort but it is nonetheless interesting to note).

[12] https://assets.savethechildren.ch/downloads/index_only_every_last_girl_print_version_inside_pages_3_10_16 _3_.pdf

obstacle to improving conditions for youth, particularly girls, has been low school completion rates. According to the Federal Republic of Somalia's (FRS) Education Management Information Systems Unit (EMIS), only 26.0 per cent of school-age children were enrolled in school in 2013/2014, with school-age girls ranking even lower, at 24.2 per cent.[13] Overall, enrolment rates were higher in Puntland, but the gap between boys and girls remained, with an enrolment rate of 42.0 per cent among girls and 51.6 per cent among boys.[14] Meanwhile, the World Bank reports that 79 per cent of secondary-age girls drop out of school.[15] Girls are married young in Somalia and often do not continue their education afterwards. Even when they stay in school, they have to deal with teaching quality in Somalia that is well below international standards.

It is also important to note the drastic differences in project context across regions. A unique aspect of project implementation in Somalia is the extremely varied institutional and security context from area to area. EGEP-T is being implemented in five overarching project locations: Somaliland, Puntland, Galmudug, Hirshabelle, and Banadir. Each project location has a myriad of unique factors that affect implementation and eventual project outcomes. Recognizing the importance of regional context, the analysis in this report is frequently disaggregated by location, in addition to external factors – such as drought, and conflict – that vary across space.

Levels and types of conflict differ immensely by project location. Somaliland is peaceful with occasional conflicts on the Somaliland-Ethiopian border. However, other project locations suffer far more extensively. In Puntland, one school in Qandala was closed completely due to violence and fieldwork was momentarily halted due to violence in North Galkayo. Al-Shabaab and the Islamic State of Somalia have a large presence in Puntland and tensions between Puntland and Somaliland seem to be on the rise. The extent of conflict in Puntland is reinforced by EGEP-T's own analysis of the issue: the vast majority, 90.5 per cent, of the 21 schools in the sample deemed to be conflict-affected at present are located in Puntland. Moreover, all of the conflict-affected schools in Puntland are classified as moderately- or highly-affected, per analysis by EGEP-T's Monitoring & Evaluation team.

Other areas bordering Puntland are also affected. Galmudug, which borders Puntland, is affected by violence, particularly in South Galkayo. There is an on-going dispute surrounding North and South Galkayo and the two clans that reside in that area. As mentioned above, this temporarily halted fieldwork and schools were temporarily shut down during the violence. Conflict in Hirshabelle is usually focused on land disputes between clans and sub-clans. While only one school was selected in Hirshabelle for data collection, conflict in the area may prevent girls from attending that or other schools. Lastly, Banadir, composed of the capital of Mogadishu, suffers from an onslaught of terrorist attacks from Al-Shabaab and other militia groups. The security protocols put in place by the Mogadishu government heavily restricted movement to and from schools during fieldwork and at least one bursary girl supported by EGEP-T passed away in the October 14 bombing.

Likewise, although drought has significantly affected all five project locations, some locations have been affected much more severely than others. Large swathes of both Somaliland and Puntland

[13] Note that this data covers Banadir and unspecified portions of South-Central Somalia under FRS jurisdiction and control. The figures do not include all of the Federal Republic of Somalia. See: Education Statistics Year Book, 2013/2014, Federal Government of Somalia, Education Management Information Systems (EMIS) Unit, Mogadishu.

[14] These figures include enrolment in Integrated Quranic Schools and Alternative Basic Education, which may partially account for the higher rates found in Puntland. See: Education Statistics Year Book, 2013/2014, Puntland State of Somalia, Education Management Information Systems (EMIS) Unit, Garowe.

[15] World Bank 2014, http://www.epdc.org/sites/default/files/documents/EPDC%20NEP_Somalia.pdf

have been severely affected by drought and are considered to be food security emergencies for the international community. These are primarily rural areas: Awdal, Sool, and Sanaag in Somaliland and Nugal and North Mudug in Puntland. Likewise, Hirshabelle and Galmudug have suffered extensively from the drought. Hirshabelle specifically is in need of aid as the Shabelle River has recently dried to the point that its remaining water is not considered suitable for use.

Since it is entirely urban, Banadir has been less directly affected. However, nearby regions have lost crops and livestock, which has impacted the Banadir markets and food availability. The next rains, the *Gu* rains, are predicted to be lower than average and would be the fourth consecutive rainy season below average in the northern part of Somalia.

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 per cent, of which more than 90 per cent is paid to support personnel and on-going operations.[16] As a result, teachers are insufficiently trained and unqualified. Many teachers do not have any training in education and teach because they cannot 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, concomitantly, caused teachers to transition to private schools or transition to other occupations because of worsening working conditions. In addition, the federal government in Mogadishu is only now developing a national curriculum to be used in schools. There is little standardization or overview of the education happening 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, Somalia lacks a unified curriculum for either primary or secondary school, owing both to varied jurisdiction over educational policy from area to area and to a more general failure of the federal government in establishing a single curriculum for schools that fall under its jurisdiction. The situation, in which schools often import 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.[17]

The formal institutional structure across jurisdictions is similar, but masks 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 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 an 6-2-4 system, with six years of primary education, followed by two cycles of secondary school of two years (lower secondary) and four years (upper secondary), respectively.

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.

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

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

Puntland, in contrast, does not 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: at the primary schools in Somaliland studied at baseline, a significant share (69.2 per cent) of head teachers indicated that their school charges fees to students.

Similarly, policies regarding language of instruction, though generally prioritising Somali in primary school and English in secondary school, do not completely align across locations. For instance, the Puntland Ministry of Education indicates that Somali serves as the language of instruction for Grades 1-4, while English or Arabic are used in Grades 5-8, and English exclusively in higher, secondary school, grades.[18] In contrast, Somaliland's current policy suggests that Somali act as the language of instruction in primary school and English in secondary school, but does not clearly specify the role that Arabic should play.[19]

These official policies mask even starker divergence within locations, however, where differences in implementation mean that students in the same regions or districts may experience different language of instruction regimes in practice. Language policies are most consistently implemented – at least according to head teachers and in the context of EGEP-T schools – in Somaliland, where *all* primary school head teachers surveyed reported the use of Somali as the language of instruction; even here, however, only 25 per cent of secondary schools reported the use of English, with 75 per cent reporting the use of Somali as their language of instruction. But divergence from official policy is much starker in other locations: in Puntland, 22.2 per cent of primary schools use English, and a further 11.1 per cent use Arabic; in Banadir, 94.1 per cent of primary schools report the use of English; and in Galmudug, 85.7 per cent of primary schools report the use of English and none report the use of Somali – those that do not use English report the use of Arabic as the language of instruction instead.[20] Divergence is less severe at the secondary level, where the majority of schools in all locations report using English as the language of instruction though, even here, there are a few secondary schools the report the use of Somali instead. Differences such as these may have important effects on learning outcomes, especially in terms of Somali and English literacy.

Ownership or control over schools – such as the public or private nature of schools – is another factor driving policy and institutional incoherence across areas. In Somaliland, all sampled schools were public; indeed, this high prevalence of public schools may explain the stronger correlation between official language policies and implemented policies in Somaliland described above. While there are private schools in Somaliland, the Somaliland government has built a relatively robust system of public schools. By comparison, private schools make up a much larger share of the sampled schools in the other locations, especially Puntland, where nearly half of all schools were described by their head teachers as specifically private in nature. In Banadir and Galmudug, head teachers described 33.3 and 27.3 per cent of schools as being specifically private. These figures reflect structural differences in the educational systems across locations. To illustrate, consider

[18] See discussion in the Puntland Education Policy Paper (Puntland Education Policy Paper, 2012. Ministry of Education, Puntland. Available at http://www.moepuntland.com/moe/wp-content/uploads/2011/10/REVISED-FINAL-PEPP.doc). Importantly, these policy statements do not preclude the teaching of Somali, English, and Arabic *as subjects* at all levels. Rather, this discussion focuses on the language in which instruction in other subjects, such as science, mathematics, and social studies takes place.

[19] Somaliland's Education Sector Strategic Plan 2012-2016. 2012. Republic of Somaliland Ministry of Education and Higher Education.

[20] While a subset of EGEP-T schools provide both primary and secondary education, these results are broadly similar among schools that exclusively provide primary education, where the language of instruction should be especially consistent with official policies.

Puntland – although the government has been rebuilding the public school system in recent years, a large share of schools remain privately-controlled. As noted above, ownership of schools has consequences for policy implementation: private schools are much less likely to adhere to the official language-of-instruction policies promulgated by their respective ministries of education, for instance.[21] The extent to which ownership differences impact learning outcomes is not clear from existing research, as it depends on the quality and oversight of public and private schools, respectively, but it is reasonable to expect that a school's ownership structure would influence its management, teacher quality, and the resources available to teachers and students, at the least.

School staffing and teacher qualifications are additional areas of divergence across areas. Somalia is in the throes of a well-known shortage of teachers and, especially, of qualified teachers. A number of reports have documented the high student-teacher ratio of Somali classrooms, and teachers who are teaching subjects in which they have no training.[22] Our own findings from the baseline confirm these large class sizes and poorly-qualified teachers. Out of 494 teachers surveyed at 140 schools, 21.9 had never completed any education beyond secondary school. The shortage, within EGEP-T schools at least, appears most pronounced in Somaliland, where – despite the area's relative stability and other advantages vis-à-vis other project locations – just 58.1 per cent of teachers have completed education beyond secondary school.

Such wide divergence in regional contexts makes one-size-fits-all programming in Somalia nearly impossible – while some schools need assistance for drought response, others are in greater need of teacher training, and still others require formal support from ministry officials and a systematized approach to education. Variation of this kind also complicates evaluation, because barriers to education that may be overwhelming in a few schools are easy to miss in the aggregate, if their impact is concentrated geographically. This context is essential to keep in mind when considering the disaggregated results presented in this report, as well as when establishing targets for indicators in future evaluation waves.

1.2 Project Theory of Change

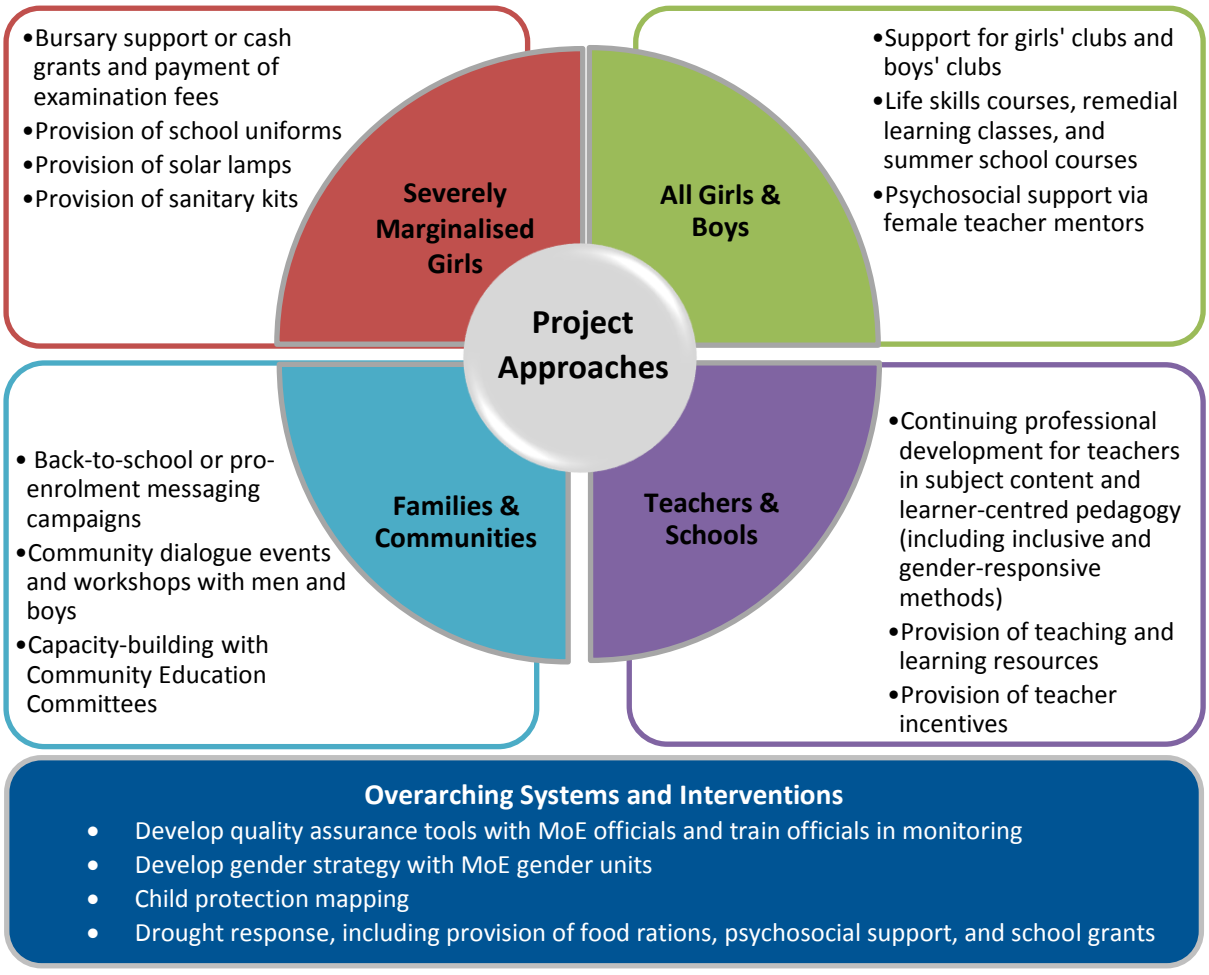
Relief International (RI) recognizes that the challenges associated with girl's education are multi-dimensional with no single "magic bullet" solution. As such, the EGEP-T will take a multi-level and holistic approach similar to EGEP (Phase 1), building on the lessons learned and the work accomplished 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.

[21] Private primary schools are far more likely to use English or Arabic as the main language of instruction, while public primary schools are much more likely to use Somali. This fact drives the differences between Somaliland and Puntland, as examples, in terms of language of instruction: sampled Somaliland primary schools are universally public and universally use English; Puntland's sampled primary schools are mixed public and private and – while Somali is used in all of Puntland's public primary schools, English and Arabic are often used in the area's private primary schools.

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

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.

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



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.

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: schooling imposes a financial burden on households.[23] In Somalia, universally high unemployment rates dampen the incentives to complete secondary-level education. This is especially true for girls, for whom the default expectation is that they will get married and leave the labour force shortly after exiting school. To 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.[24] 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, they might stay home when menstruating, so the distribution of sanitary kits aims 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, showed that girls with higher psychosocial wellbeing had higher attendance rates. As this baseline documents, girls indicate

[23] 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.

[24] 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.

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.[25]

To target this individual-level barrier to girls' achievement, EGEP-T has planned multiple activities to be implemented in schools, including the formation and promotion of girls' and boys' clubs that will engage in leadership and confidence-building activities, and provide children with valuable training on life skills, including 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 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.

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 marginally qualified teachers.[26]

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.[27] 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.[28] 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 to promote a more positive learning environment, EGEP-T plans a number of activities focused on training teachers and providing them with additional resources and incentives to promote high-quality teaching. Somali teachers are often unqualified and untrained, leading to poor teaching quality. Those who have been trained often go through the Strengthening Capacity of Teacher Training (SCOTT) course. By design, this is a two-year course which focuses on all the different subjects in the primary curriculum. It is a teacher training course to fill the gap, rather

[25] 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.

[26] To illustrate, 6.5 per cent of teachers surveyed at the baseline had not completed secondary school, and a further 21.9 per cent 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.

[27] Among teachers surveyed in the baseline evaluation, 32.8 per cent indicate that they do not have the basic supplies necessary to teach. In addition, 75.7 per cent 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.

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

than a continuous professional development approach. Over the past 12 years it has ceased being implemented wholly and consistently, and different providers often deliver different variations of the course, varying in quality, length, and intensity. The course was rolled out in Somaliland and Puntland from 2006 onwards. Unfortunately, the security situation meant that Galmudug and Benadir teachers were largely excluded. EGEP-T will train teachers in subject content in mathematics, subject-specific pedagogical techniques, remedial teaching practices, English proficiency, and gender-responsive techniques. The teacher training programme will take the form of a Continuous Professional Development approach through the use of coaches. The coaches will provide training and also ongoing mentoring support. Additionally, teachers and pupils will have access to a digital learning platform.

The project's efforts regarding teacher training are necessarily focused on intervention schools and improvements that can be made among the existing teacher cohort.[29] Teachers and school administrators routinely request this type of training, as documented by qualitative interviews conducted at the baseline. This suggests 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 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. Head Teachers often play a lead role in the CECs, and in general lead the school management and administration. There is no standardised additional training for Head Teachers aimed at strengthening the associated required skills. Some Head Teachers will have received training in areas such as leadership from NGOs, though this is generally limited to one or

[29] 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.

two days and of no specific standardized content. Capacity of Head Teachers to manage schools effectively varies heavily from school to school. 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. EGEP-T is providing cash grants to schools in communities that are feeling financial strains from droughts. Additionally, 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

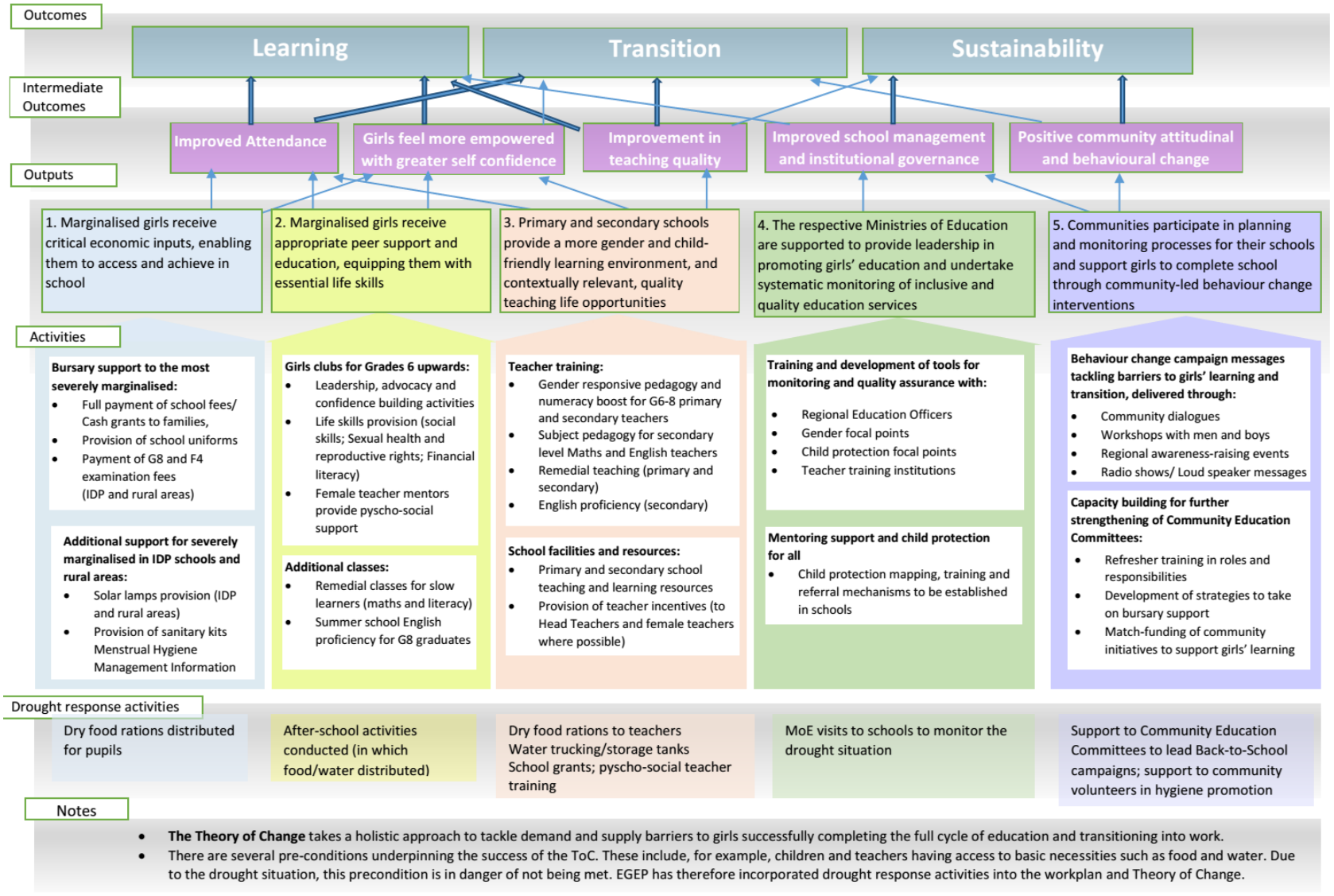
The sixth barrier targeted by EGEP-T activities is a lack of community support for girls' completion of the full education cycle. In the first phase of the EGEP, the project found a high level of community support for girls' education in general. However, this did not always translate into behaviours 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 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 to promote the value of girls' education, 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.

[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



External Factor: Drought

In addition to the barriers to education targeted by EGEP-T programming, the project has also taken stock of the current situation on-the-ground in Somalia, and identified external factors that are likely to influence its ability to improve educational outcomes and affect lasting change. The most prominent external factor considered by the project is drought, with its multifaceted impact on families, schools, and communities. Drought can exacerbate already significant financial obstacles to enrolment for children and their families, but it also has numerous pernicious but less obvious effects. Specifically, drought impacts all community members, reducing the ability of the school to raise funds for improvements, pay teacher salaries, and other critical activities. Teachers' livelihoods can also be affected, causing, for instance, an increase in teacher absenteeism.

Given the potential impact of drought, drought response activities have been integrated into EGEP-T 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.

GESI Standards in EGEP-T

In line with GEC-T's emphasis on alleviating gender and other forms of inequality in the areas where projects are implemented, the design of EGEP-T conforms to the minimum standards of Gender Equality and Social Inclusion (GESI) outlined by the FM. During the design phase of EGEP-T, RI conducted a gender analysis, which informed the development and selection of interventions intended to address the practical needs facing marginalised girls and boys. The project's approach to gender and other sources of inequality were also shaped by the past experiences of RI and its implementing partners in the first phase, EGEP, of this project. Evaluation of EGEP yielded valuable insights in terms of which interventions were effective in reducing specific, tangible barriers to girls' education. For instance, the evaluation highlighted the impact of solar lamps, and the manner in which they allowed girls, who are responsible for significant household chores, to study after dark.

The vast majority of EGEP-T interventions are targeted to specific sources of gender inequality in educational attainment. Below are highlights of a few of these interventions and the gender-specific barriers they seek to address:

- **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.

Overall, the project meets the GESI minimum standards in terms of both gender and broader social inclusivity. The project is – in the views of the evaluators – least transformative in the realm of non-gender gaps in inclusion. For instance, the project does not target interventions at barriers to learning and retention of children with disabilities. Rather, the project is accommodating of such inequality – in the terms employed in the GESI standards, it "acknowledges, but works around" disability to achieve project objectives. Given the difficult context in which EGEP-T is working, the project's accommodation of sources of inequality, like disability, are understandable; nonetheless, they fall in line with the GESI standards.

In other respects, EGEP-T is more clearly transformative. Aspects of EGEP-T are both gender-accommodating and gender-transformative, in line with GESI minimum standards. The project clearly focuses on girls' practical needs, as the set of example interventions above make clear. These efforts alleviate specific, tangible barriers to girls' education, without overturning gender relations in Somalia wholesale. At the same time, there is also a broader focus on girls' strategic needs, particularly in the development of ministry-level gender policies, and the promotion and training of female teachers, which challenges existing inequalities in society and can help reshape – at least in EGEP-T communities – the status and expected roles of women in society.

TABLE 1: PROJECT DESIGN AND INTERVENTION

Output	Intervention Types	Intervention Summary
1: Marginalised girls receive support from their community and peers, equipping them with life skills, for improved life opportunities	Economic support; material support; psychosocial support	EGEP-T will expand direct economic support bursary and safety-net support (school fees, cash grants, and uniforms) for hard to reach girls to cover more secondary schools. EGEP-T will support girls from economically weaker sections with national exam fees. The project will support adolescent girls, especially in remote locations, with quality reusable sanitary kits and mentoring support to enable them to attend school during their menstrual cycle. The project will provide solar lamps to promote home-based group learning for the severely marginalised groups in IDP camps and rural areas.
2. Marginalised girls receive support from their teachers and peers equipping them with education for improved life opportunities	Life skills; female voice; learning support	EGEP-T will proactively engage youth using Girls' and Boys' Clubs to advocate and campaign for inclusive and quality education, and act as platforms for peer exchange and knowledge sharing. The project will support girls with Life Skills (self-confidence, safe motherhood, finance, career guidance, etc.) for effective transition to life beyond school. EGEP-T will also offer a range of remedial classes to both primary and secondary girls and boys.
3. Primary and secondary schools across Somalia provide a more gender and child-friendly and inclusive environment for learning, and contextually relevant teaching	Teachers' professional development; EdTech; school facilities; learning materials	EGEP-T will operate a teachers' continuous professional development programme. The project will also develop a digital platform for teachers' use. EGEP-T will improve school water supply and sanitation facilities in secondary schools to address gaps in conditions for learning. EGEP-T will supply standard teacher learning materials to primary and secondary schools based on a needs assessment.
4: The respective Ministries of Education are supported to provide leadership in promoting girls' education and undertake systematic monitoring of inclusive and quality education services	Capacity building with MoEs; Child Protection support	Designed to build decentralised MOE capacity to plan and monitor education services and outcomes, the support to MOEs includes gender-based monitoring and planning for REOs, as well as training to Quality Assurance Officials. The project will strengthen child protection mechanisms in coordination with child protection stakeholders, embedding principles in education documents and practices, and sensitizing communities, government, teachers, education officials UN conventions in order that children are protected from violence in school.
5: Communities participate in planning and monitoring processes for their schools, and play a lead role in supporting girls to complete school through community-led behavioural change interventions	Capacity building with communities; awareness raising and behaviour change	EGEP-T will continue to strengthen CECs to improve accountability in school governance; support CECs to develop School Development Plans that articulate community contribution; and facilitate community dialogues. Awareness raising interventions will support positive behaviour change. The project will pursue proactive engagement strategies with men and boys to promote girls' education.

Pathways of Change from Intermediate Outcome to Outcome		
Output	**Intermediate Outcome**	**How will this contribute to achieving the learning, transition, and sustainability outcomes?**
1;2;3 ⇨	Improved attendance ⇨	Increase in **learning** with girls' increased presence in the classroom; Improved **transition** rates, as increased learning will increase motivation and confidence to progress to next level.
1;2;3 ⇨	Greater self-esteem and empowerment of girls ⇨	Increase in **learning**: EGEP midline found that girls' increase in psychosocial wellbeing was directly linked to an increase in learning; improved **transition** rates as girls gain confidence to continue.
3; ⇨	Improved teaching quality ⇨	Increase in **learning**: EGEP midline found a 5.4 fold increase in girl-centred teaching, associated with learning gains; improved transition as quality teaching leads to high motivation and learning. Girls are also expected to be more likely to transition if they and their families recognise that they will be receiving high quality **transition**.
4;5 ⇨	School management and institutional governance ⇨	Increased **sustainability** of project interventions as CECs and MoEs take increased responsibility for ensuring girls learn effectively and transition through key learning points.
5 ⇨	Community attitudinal and behaviour change ⇨	Increased **sustainability** as support of caregivers and others develops and continues in the long term; improved **transition** rates as communities increasingly work to remove barriers to girls' education

1.3 Beneficiaries

The EGEP learning cohort and transition cohort are the same set of girls. In phase II, EGEP is targeting severely marginalised girls in Grade 2 to 5, and all marginalised girls in G6 to G8 and Forms 1 to 4 at secondary school. Additionally, EGEP-T will target out-of-school (OOS) girls – both those who haven't been to school, but also those that have dropped out, particularly due to drought – in an aim to get them into school.

All girls in Somalia face intersecting challenges. Save the Children's Girls' Opportunity Index rates Somalia as the fifth-worst place in the world in which to be a girl and one of the key contributing factors is the low school completion rates;[32] 79% of female youth of secondary school age are out of school.[33] The barriers to girls' effective learning and transition are significant, and range across the cultural, psychosocial, political and economic spheres.

Somalia/Somaliland has been experiencing a protracted drought. Due to the inter-reliance of families and communities in rural and urban areas, most EGEP-T beneficiaries are reported to be affected indirectly if not directly. The drought has claimed many livestock and forced families to migrate.

The hardest-to-reach groups face even more severe challenges. Girls who are especially hard to read include mothers and those who have married early; girls who are out of school; girls in highly conflict-affected or remote areas; IDPs; refugees; returnees; ethnic and clan minorities; disabled girls; and girls from impoverished female-headed households. These girls face complex combinations of context, social and economic factors and EGEP-T will provide interventions tailored to these specific groups.

EGEP will also benefit boys at primary and secondary level through various activities including: establishment of boys' clubs in schools; implementation of remedial classes; distribution of teaching and learning materials to schools; training of their teachers.

EGEP-T will be implemented in 172 primary and 56 secondary schools over the next three years.[34] RI has estimated the number of EGEP-T beneficiaries based on school enrolment data collected from project schools.

Overall, the project expects to reach 31,411 direct beneficiaries and 98,232 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 total number of direct beneficiaries includes girls in grades G4 through F4, who make up the vast majority of direct beneficiaries; a small group of 234 girls outside these grade ranges are included in the beneficiary count, as they are receiving active bursary support through EGEP-T.

[32] http://assets.savethechildren.ch/downloads/index_only_every_last_girl_print_version_inside_pages_3_10_16_3, October 2016

[33] World Bank 2014, http://www.epdc.org/sites/default/files/documents/EPDC%20NEP_Somalia.pdf

[34] 17 targeted schools combine primary and secondary school, and are therefore counted in both categories.

TABLE 2: SEVERELY MARGINALISED GROUPS

Severely marginalised group	Current expectations from community, parents, teachers and girls' themselves	Approach and intended results
Girls with disabilities	Low to no expectation that these girls will enrol in school. For those who are enrolled they are not expected to achieve at the same level as their peers.	The teacher training module will be developed focusing on basic approaches to inclusive education, enabling teachers to better engage with disabled children in the classroom. Additionally, disabled girls will be targeted for bursary support.
IDPs, refugees, returnees	Girls may have fled a conflict situation; violence and upheaval may have a negative impact on psychosocial wellbeing. Due to limited mobility, girls miss school and fall behind. These factors can lead to a general expectation, of the girl and her parents, that girls will not progress and succeed in school.	Girls in these groups who are extremely poor will be identified for bursary support ensuring economic barriers do not prevent attendance. Additionally, psychosocial support from trained teacher mentors will aid an increase in girls' wellbeing.
Orphans, minorities and other girls from poor households	Girls are often expected to leave school early, if they enrol at all, to look after siblings and contribute to family income, with school-going costs understandably a lower priority.	Bursary support will help enable girls from poor households to attend school, despite the economic challenges they face.
Children of lower learning ability/demonstrating low performance in the classroom	The percentage of caregivers stating that a child's ability should be an important factor determining whether a child should attend school reduced from 88% to 43%.	Specifically designed messages promoted through BCC activities will increase community support for children of all abilities to attend school. Remedial classes will help those children keep up with the learning levels of their peers.

TABLE 3: ESTIMATED BENEFICIARIES OF EGEP-T PROJECT

Group	Expected Beneficiaries
Direct Beneficiaries – Girls[35]	
Primary School Girls	20,318
Secondary School Girls	10,859
Indirect Beneficiaries – Girls	
Primary School Girls	25,354
Secondary School Girls	0
Indirect Beneficiaries – Boys	
Primary School Boys	55,558
Secondary School Boys	17,320

Table 4, below, describes the structure of EGEP-T in relation to target groups of beneficiaries. The rightmost column in the table lists the types of support that beneficiaries in each category will

[35] The count of direct beneficiaries listed here excludes 234 direct beneficiaries – being supported through the provision of bursaries – who fall outside the standard grade range of other EGEP-T activities. As a result, the direct beneficiaries listed sum to 31,177, while the total direct beneficiaries number 31,411.

receive. Note that Table 4 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 but live in an EGEP-T community.

As part of its programming, EGEP-T collected information from project schools on the number of students with disabilities and IDPs in their schools. It is important to note that this data was collected from school administrators, while data reported elsewhere in this evaluation was collected by the evaluation team directly from girls or their caregivers using the Washington Group questions. Therefore, the prevalence of disabilities is likely to differ significantly; nonetheless, the number of disabled students, as recognized by school staff, is a useful indicator of the prevalence of impairment in EGEP-T schools. Overall, the schools document 448 girls and 600 boys with disabilities enrolled in project schools. Assuming that the distribution of disabilities is proportional across grade levels, EGEP-T expects to directly benefit 249 girls with disabilities, and indirectly benefit the remaining 799 children with disabilities noted. The project has also identified a total of 5,519 direct female beneficiaries who are IDPs, representing 17.4 per cent of total direct beneficiaries.

Table 4 describes specific beneficiary groups, according to their enrolment status, grade, and the urbanicity of their residence. The rightmost column indicates the types of support each group will receive.

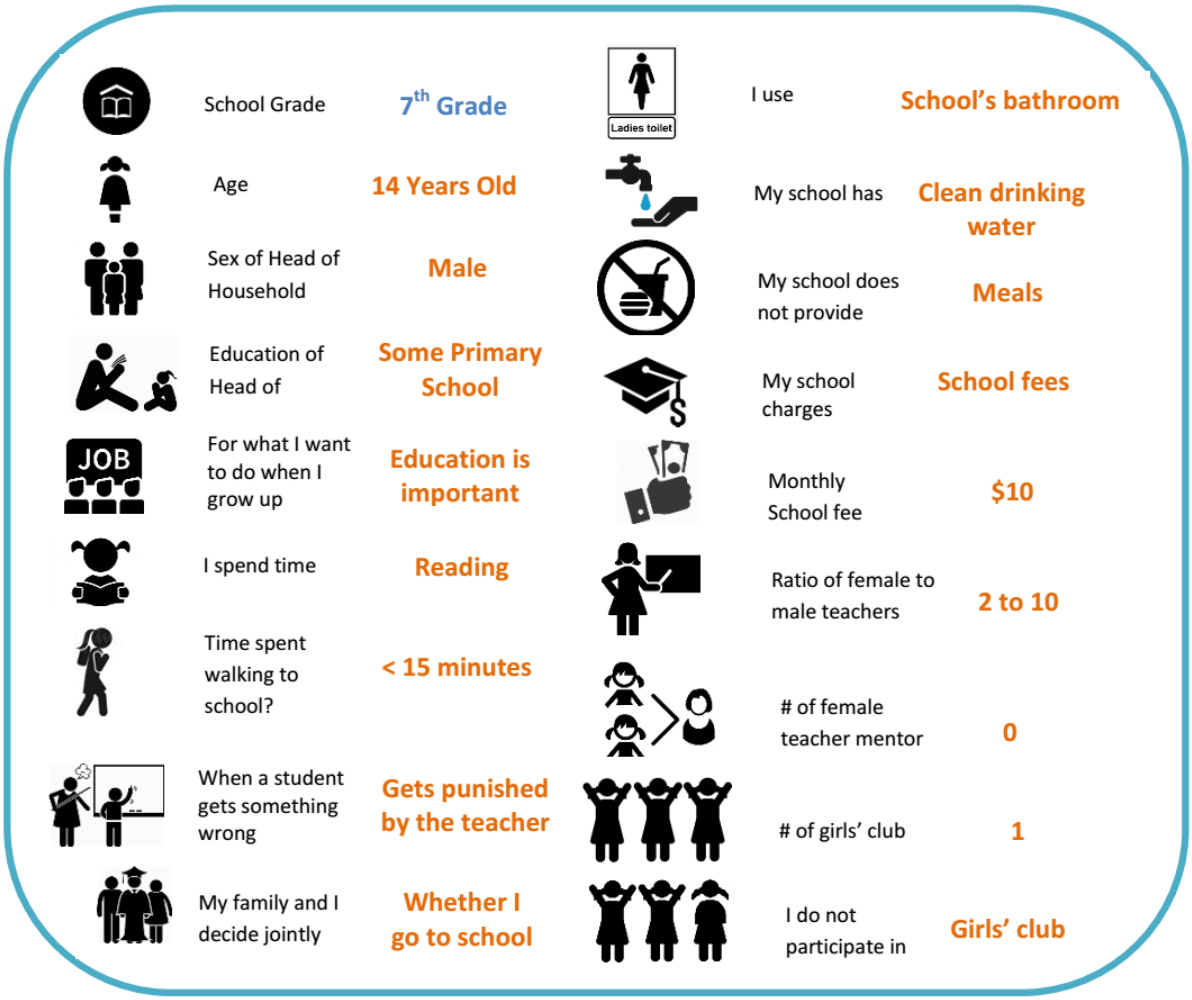
TABLE 4: TYPES OF TARGETED SUPPORT, BY BENEFICIARY GROUP AND MARGINALISATION

Enrolment Status	Location	Grade	Support
Basic Package of Interventions – Marginalised Girls			
In-school	Urban and Rural	G6-F4	• Life skills classes • Remedial classes • English proficiency classes • Teaching and learning materials • Teacher Professional Development
Expanded Package of Interventions – Severely Marginalised Girls			
Each group gets the Basic Package plus:			
In-school	Urban	G6-F4	• Bursary support • Teacher Professional Development
In-school	Rural	G6-F4	• Bursary support
In-school	IDP Camp	G6-G8	• Solar lamps • Sanitary kits • Teacher Professional Development
In-school	Urban & Rural	G2-G5	• Bursary support • School uniforms • Payment of school fees or cash grants • Teacher Professional Development
Out-of-school	Urban	F1-F4	• Bursary support • Community back-to-school campaigns
Altered Package of Interventions – Marginalised Boys			
In-school	Urban & Rural	G6-F4	• Participation in boys' clubs • Life Skills classes • Remedial classes • Teaching and learning materials • Teacher Professional Development

Table 4 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 project's Monitoring, Evaluation, and Learning (MEL) Framework, RI notes 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.

RI distinguishes between marginalised and severely marginalised girls by considering whether girls face additional obstacles to learning and completing school. These obstacles include the death of parents, more extreme poverty than is normal in EGEP-T target communities, girls who are displaced, and being ethnic or clan minorities.

FIGURE 3: A PROTOTYPICAL PRIMARY SCHOOL BENEFICIARY



Calculation of beneficiary numbers

Based on documentation provided by RI, this section assesses the assumptions on which beneficiary number counts are based, and the overall quality of RI's approach to counting beneficiaries. As noted above, RI expects to reach 31,411 direct beneficiaries and 98,232 indirect beneficiaries.

RI's beneficiary numbers are based on official enrolment counts for EGEP-T schools, collected for the 2017-2018 school year.[36] While school records are often unreliable, official enrolment counts are the best possible data sources for calculating beneficiary numbers. Relative to estimates derived from the baseline evaluation, they are more comprehensive because they include all project schools, rather than only sampled schools.

In general, we consider RI's beneficiary counts to be conservative. RI includes only in-school children in their beneficiary counts. Direct beneficiaries are girls in project-targeted grades only. Indirect beneficiaries are all in-school boys, and girls in non-targeted grades. However, RI does not count out-of-school girls as potential beneficiaries, although they are likely to benefit from various back-to-school campaigns and efforts to change community attitudes. Similarly, RI does not count students who are enrolled in non-EGEP-T schools in the same communities, who may similarly benefit from the project's community-level interventions. Both omissions mean that the project's expected reach is moderately understated in their beneficiary calculations.[37]

[36] In previous drafts of this report, beneficiary number calculations were based on 2016/2017 enrolment data. Following the first draft of the report, RI completed the collection of new, updated, enrolment figures for the 2017/2018 academic year; the discussion in this section has been updated to reflect this newly-available data.

[37] We do not recommend that RI updates their beneficiary numbers to include out-of-school girls and girls in non-EGEP-T schools, however. To do so would require making a number of assumptions about the project's reach, and the number of both populations within EGEP-T communities. Rather, relying on the more accurate, if conservative, approach that RI has already taken to this point is preferred.

2. Baseline Evaluation Approach and Methodology

2.1 Evaluation Questions

This baseline evaluation of EGEP-T seeks to satisfy two main goals. The first is to establish baseline levels of core project outcomes and intermediate outcomes, and develop reasonable targets against which the project will be evaluated at the midline and endline. The second is to make evidence-driven recommendations regarding future evaluation design and project implementation, based on analysis of the project's Theory of Change, the relationship between intermediate outcomes and core project outcomes, the correlation between demographic characteristics and baseline outcomes, the prevalence of specific barriers to educational attainment and learning, and other issues where data collected at the baseline can shed valuable light. More specifically, this report addresses each of the following questions:

- What are baseline values of learning, transition, and sustainability in EGEP-T schools and communities? What targets for the midline and endline should be established to measure project impact?
- What are baseline values of intermediate outcomes, such as attendance and teacher quality, in EGEP-T schools and communities? What targets for the midline and endline should be established to measure project impact?
- Who are the project's beneficiaries, and what are their characteristics? How many individuals can be expected to benefit from the project?
- How realistic and justified is the project's Theory of Change? Is there evidence that the ToC's assumptions, especially regarding the relationship between intermediate and core outcomes, is justified?
- What barriers exist to education in the project's context, both in terms of individual-, household-, community-, and societal-level characteristics? To what extent are these barriers associated with variation in learning, transition, and other outcomes at the baseline?

2.2 Core and Intermediate Outcomes Matrix

TABLE 5: MEASUREMENT APPROACHES AND FREQUENCY FOR CORE AND INTERMEDIATE OUTCOMES

Outcome	Measurement Level	Data Collection Method	Rationale	Frequency of Data Collection
Core Project Outcomes				
Literacy in Somali and English	Individual	SeGRA	Assessment designed specifically for Somali context	Per evaluation point
Numeracy	Individual	SeGMA		Per evaluation point
Transition within school (i.e. grade to grade)	Individual	HH survey, FGDs, KIIs		Per evaluation point
Intermediate Outcomes				
Intermediate	Classrooms	Attendance spot	Spot checks provide	Per evaluation

outcome 1: Improved attendance	within school; individual level (cohort girls)	checks; school register for individual girls; HH survey	objective measures of attendance, limited by number of checks; registers are less reliable but more comprehensive; triangulate with HH survey indications of relative attendance as well	point; additional as possible by RI
Intermediate outcome 2: Greater self-esteem and empowerment of girls	Individual	HH survey (or survey in-school for baseline)	Self-esteem is best measured through direct, individual-level questions	Per evaluation point
Intermediate outcome 3: Improvement in teaching quality	Classrooms within school; teachers within schools	Classroom observations, Teacher Survey	Teachers' self-reported teaching practices need to be triangulated by actual observation	Per evaluation point
Intermediate outcome 4: Improvement in school management and institutional governance	School; Teachers' responses within schools	Teacher Survey, Head Teacher Survey, FGDs, KIIs	Head Teachers provide insight into most aspects of school management; teachers provide an alternative source of information as a check on Head Teachers' and CECs' performance	Per evaluation point
Intermediate outcome 5: Positive community attitudinal and behavioural change	Communities	HH Survey, Girls Survey, Teacher Survey, Head Teacher Survey, FGDs, KIIs	HH Survey is primary measure, as it is conducted with community members; triangulate with perceptions of community attitudes among stakeholders (teachers, head teacher) and among children themselves (girls)	Per evaluation point

TABLE 6: MEASUREMENT APPROACHES AND FREQUENCY FOR SUSTAINABILITY OUTCOME INDICATORS

Sustainability Level	Measurement Level	Data Collection Method	Rationale	Frequency of Data Collection
School	School	Head Teacher Survey; Teacher Survey; FGDs; KIIs	Bursary support and training programmes measured via survey responses; supplemented by qualitative data to understand extent of CEC engagement	Per evaluation point
Community	Community and School	HH Survey; Head Teacher	HH awareness of project activities, advocacy campaigns,	Per evaluation point

		Survey; FGDs; KIIs	and male/boys action in support of girls' education are core measures; supplemented by head teachers' knowledge of project activities; complemented by qualitative data to describe activities, advocacy campaigns, etc. undertaken by community members	
System	Schools and System-Level	Head Teacher Survey; FGDs; KIIs	Child protection policies and monitoring visits by MoE measured at school level, supplemented by qualitative data; gender development strategy assessed exclusively via qualitative data	Per evaluation point

2.3 Evaluation Methodology

This section provides an overview of the research design employed at the baseline of the EGEP-T project. Specifically, the pre-post evaluation design implemented for EGEP-T is described in detail, followed by a discussion of data collection tools used, and the sampling design employed.

Overall Evaluation Design

The evaluation design for EGEP-T employs a mixed method approaching, utilising a wide range of quantitative and qualitative data, gathered from a number of different respondent populations. The quantitative aspect of the EGEP-T evaluation employs a slightly-revised pre-post research design. Importantly, the chosen design diverges from the standard difference-in-differences approach, which utilises a set of control schools for inferring impact in intervention schools. Instead, given security concerns in Somalia, EGEP-T will employ a pre-post design, tracking the same cohort of girls and schools over time from baseline through the endline. 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.

In addition, the design diverges from the joint sampling approach utilised by many GEC-T projects. In a joint sampling approach, 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, this assessment uses a hybrid 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 project schools, while households are sampled using a random walk strategy in communities that include project schools.

The EGEP-T evaluation uses a hybrid of these two approaches, employing a split sample at the baseline and a joint sample at the midline and endline. That is, in the baseline evaluation, the learning cohort is recruited from EGEP-T schools, and the household sample comprises a different

set of respondents, selected from households in communities surrounding EGEP-T schools. The household sampling strategy is particularly important because – in this assessment – it establishes the benchmark against which project outcomes for transition in future waves will be measured. The household survey establishes the expected rate of transition (i.e. the rate at which students continue to the next grade level) among the general population of communities where EGEP-T schools operate.[38]

At the midline and endline, efforts to assess the impact of EGEP-T will centre on two key comparisons. The first is longitudinal, calculating changes in key indicators and intermediate outcomes over time, from the baseline to the midline or endline. In the absence of the control or comparison group, these longitudinal comparisons will be subject to concerns regarding secular trends – for instance, learning outcomes are difficult to judge in a strictly longitudinal design, because as respondents complete additional years of school, learning outcomes will naturally improve.

Longitudinal comparisons will be used primarily to assess outcomes which are not benchmarked at the baseline and which are unlikely to exhibit secular positive trends in the absence of the EGEP-T intervention. For instance, longitudinal comparisons will be used to assess changes in school management and governance, community attitudes, and teaching quality, among other intermediate outcomes.

The second type of comparison uses benchmarks established at the baseline for comparison. Outcome benchmarks will be established for learning, attendance and transition outcomes. Benchmarks are critical for drawing inferences about project impact in the absence of a control group. Consider the case of learning outcomes. Older children, who have completed an additional year of schooling, would be expected to perform better on learning assessments, even in the absence of project impact. To estimate the impact of the project on learning outcomes for girls in Grade 8 at the midline, their scores will be compared to the scores of Grade 8 girls obtained at the baseline. Grade 8 girls at the baseline serve as a pseudo-control group, because the project – which has not begun significant activities prior to the baseline – should not have influenced the learning outcomes for Grade 8 girls at the baseline. Any gains that midline Grade 8 girls have made relative to baseline Grade 8 girls can be cautiously attributed to the impact of the project.[39]

The combination of benchmarked and non-benchmarked longitudinal comparisons provides the strongest research design possible, in the absence of a true control group of schools and students.

[38] Per RI's MEL Framework, the midline and endline evaluation will follow-up with the learning cohort and conduct household interviews with a subset of the learning cohort. The household sample from the baseline evaluation will not be re-contacted for household interviews. At the midline and endline, transition rates will be calculated for the learning cohort – established at the baseline – and compared to the benchmark transition rates. Importantly, we recommend a modification of this approach, which we document in the Evaluation Plan provided to RI. This modification is not relevant to baseline data collection and analysis; as a result, we provide our recommendation in the Evaluation Plan instead.

[39] While the danger in assessing learning outcomes without a benchmark is that learning outcomes will be overestimated (the project is given credit for improving learning outcomes merely because children have completed an additional year of schooling that would have occurred in the absence of the program), the opposite problem obtains in the case of attendance and enrolment results. For instance, if attendance rates for a given cohort of girls deteriorate between the baseline and endline, this could reflect a negative impact of the project. But attendance and enrolment rates are lower among older girls even in the absence of the program; as such, it is important to compare "apples to apples" – attendance and enrolment among endline Grade 8 girls compared to baseline Grade 8 girls – to account for the important effect of girls' age on these outcomes.

This baseline study establishes a cohort of students – both boys and girls – to be tracked through the midline and endline evaluations. This sample of students will be compared to benchmarks established at the baseline to estimate the program's impact on learning, attendance, and transition rates.

Baseline Evaluation Design

In the previous section, we described the overarching evaluation design with respect to evaluation through the project's lifecycle. That is, the previous section described how data will be collected and inferences will be drawn regarding project impact from baseline to midline to endline. In this section, we more narrowly describe the design of the baseline evaluation, with its unique set of assessment goals.

The baseline evaluation utilises a mixed-methods approach. The goal of the baseline's design was to lay a strong foundation for future evaluation waves at midline and endline, establish indicator targets, and generate insights that can inform programming. As such, the design focuses both on providing firm, rigorous evidence regarding baseline levels of key project outcomes and intermediate outcomes – such as learning, attendance, etc. – and generating more nuanced insights with respect to challenges the project is likely to encounter, contextual factors that may influence project outcomes, and so forth.

The baseline targeted a diverse range of populations, in an effort to cover a widely ranging set of intermediate and primary outcomes, as well as to allow the triangulation of findings across respondent populations wherever possible. In practice, the range of intermediate outcomes targeted by EGEP-T interventions necessitated data collection from myriad populations – information on school management from head teachers, information on teaching qualifications and practices from teachers, and data on community attitudes from community members, to name just some. The evaluation team took the opportunity provided by these diverse populations to triangulate information on the outcomes as much as possible, improving the strength of inferences in the process. A full list of the respondent groups targeted is provided in Table 7 (populations included in quantitative data collection) and Table 8 (populations included in qualitative data collection) in the following two sections.

The evaluation essentially established four unique cohorts of school-age children who will be re-contacted at later evaluation stages. Note that these cohorts can – and occasionally are – combined in our analysis to draw conclusions; at later stages, comparisons across cohorts may be particularly fruitful for making inferences regarding project impact. However, we describe the four groups as separate to highlight the fact that they are, separately, coherent sets of respondents, and because they were sampled independently of one another.

The first group consists of girls who were enrolled in school and in grades 6-8 or forms 1-2 at the time of the baseline.[40] This group comprises the primary cohort for the purposes of establishing learning and transition outcomes over time. The second group consists of girls enrolled in school and receiving bursary support through EGEP-T at the time of the baseline. Like the "cohort girls", these girls are primarily in grades 6-8 and forms 1-2. Importantly, this grade range does not represent the full extent of EGEP-T's bursary support programme, which often targets severely marginalised girls in lower grades. The evaluation team opted to sample bursary girls from the same grade range as the primary cohort, with the goal of drawing inferences regarding the differential impact of bursaries on

[40] Note that the targeted girls were enrolled in school *at the time of the baseline*, though some girls were out-of-school at the start of the project, a fact which we address in greater detail in Section 4.

learning and transition at later evaluation stages, by comparing improvements among bursary girls to improvements among "cohort girls" not receiving bursary support.

The third group consists of boys enrolled in school and in grades 6-8 or forms 1-2 at the time of the baseline. Like the first two groups, this sample of boys will be tracked over changes in learning, transition, attendance, and other outcomes. The fourth group consists of out-of-school (OOS) girls. While each of the first three groups was sampled at project schools, this fourth group was sampled from randomly-selected households within project communities. This group of OOS girls will be tracked over time, providing insight into the impact of EGEP-T on re-enrolment rates and learning among girls who re-enrol.

Specific beneficiary groups were targeted in several ways. Bursary girls, for instance, were incorporated as an explicit, separate sample, to ensure that information about this beneficiary group was captured with a sufficiently large sample size to draw conclusions in future evaluation waves. Similarly, boys – who are classified as indirect beneficiaries – were sampled explicitly and separately at each school, with the goal of establishing a sample with sufficient power to draw conclusions about project impact among indirect beneficiaries.

Other population subgroups were not explicitly targeted for sampling, but their inclusion drove decisions regarding sample design. For instance, the baseline evaluation did not employ a quota for IDPs, either at the level of individual students within schools or at the level of selecting IDP or IDP-dense schools. However, RI's Monitoring & Evaluation Team verified that IDP schools were included in the random sample in the first stage selection of schools. As we discuss below, IDP schools were underrepresented relative to the full population of EGEP-T schools, though the evaluation employs survey weights to partially correct for this imbalance.

Other beneficiary groups were included in the sample through random selection, but were not explicitly targeted. Specifically, children with disabilities are classified as severely marginalised in EGEP-T's approach and receive additional targeted interventions as a result. However, children with disabilities were not targeted through over-sampling, quotas, or other methods. To the extent that children with disabilities and other severely marginalised girls are included at higher rates among bursary girls – as they should be, in line with EGEP-T's targeting methodology – they should be included in the sample of bursary girls established in this evaluation.

Beyond the establishment of cohorts to be tracked in future evaluation waves, one stated aim of the baseline evaluation is to empirically assess the relationships posited in the project's Theory of Change. In other words, a core goal is to study whether intermediate outcomes targeted by the project are actually related to learning, transition, and sustainability.

Our evaluation takes a wide-ranging approach to this goal. First, we consider the project's Theory of Change from a theoretical perspective, assessing the stated and unstated assumptions that underlie it (see Section 6).

Second, and more importantly, we evaluate the posited relationships between intermediate and end outcomes empirically. The evaluation collected rich data on each of five intermediate outcomes: attendance, teaching quality, self-esteem and empowerment, school management, and community

attitudes. We investigate the relationship between each of these factors and the end outcomes of learning and transition formally, using regression analysis.[41]

To illustrate the approach, consider the relationship between attendance and learning which informs EGEP-T's Theory of Change. The baseline collected data on attendance rates at both the school level, via classroom headcounts, and at the individual level for cohort girls, using school attendance records. We use linear regression models to study whether girls with higher attendance rates earn higher marks on learning assessments, while controlling for a number of factors – such as grade, geography, and demographic characteristics – that might also influence both attendance and learning scores.

We employ a similar approach for each of the intermediate outcomes. In Section 5.6, we study the relationship between intermediate outcomes – among other factors – and learning outcomes. In Section 5.7, we study the same intermediate outcomes and their impact on transition rates.[42] Where possible, we bring to bear qualitative evidence regarding these relationships as well. By assessing the relationship between intermediate and end outcomes, the evaluation provides insight into the potential for EGEP-T to have a transformative effect on education in the communities where it is implemented.

Sampling Methodology

EGEP-T targets a diverse range of intermediate outcomes, in addition to the core GEC-T outcomes of learning, transition, and sustainability. Because many of the intermediate outcomes and sustainability indicators are either difficult to measure or multi-dimensional, the baseline evaluation employed several data collection tools. The goal of employing multiple tools, and sampling multiple populations within each school or community, was two-fold: first, to triangulate findings across diverse populations, whose viewpoints may vary in enlightening ways; second, to triangulate measurement of multi-faceted outcomes across many sub-indicators.

To illustrate triangulation across respondents, consider the issue of measuring community attitudes. The most direct approach is to survey a random sample of community members about their attitudes toward girls' completion of school. However, because respondents may not be truthful about their views of this topic, it is useful to ask other subpopulations about community attitudes. We survey head teachers and teachers about the support for education among particular groups within their communities. We also ask female and male students questions designed to gauge *their perceptions* of community attitudes regarding education.

To illustrate triangulation across indicators, consider the difficulty of measuring the quality of school management. School management is, by definition, multi-dimensional: well-managed schools do many small things that poorly-managed schools do not. For this reason, we designed a series of

[41] Note that we do not study the relationship between intermediate outcomes and sustainability. Sustainability as measured is a composite score – and scores on individual sub-indicators – across all schools. The nature of the sustainability indicators means that it is impossible to assess many of them at the school level (e.g., systemic sub-indicators), which is necessary for evaluating the relationship between intermediate outcomes and sustainability.

[42] It is important to note one difficulty that arises in this analysis, which stems from varying levels at which data was collected. For example, we expect high-quality school management to influence learning and transition rates. However, school management is naturally assessed at the school level – all students in a school are subject to the same high- or low-quality management. In these cases, we still employ regression modeling techniques, because a relationship between school management and learning would still manifest *across* schools. However, we urge caution regarding these analyses, because the effective sample size is greatly reduced due to the clustering of students within schools.

metrics designed to capture small differences across schools that, in sum, distinguish between types of schools. Examples include keeping accurate student records, monitoring teacher classrooms and providing feedback on teaching practices, and putting in place tangible, verifiable policies, in written form.

Data collection targeted multiple populations with a variety of survey instruments and sampling approaches. At the highest level of aggregation, the evaluation sampled schools for inclusion in the sample. A random sample was drawn by RI's Monitoring and Evaluation team from among a sample frame of eligible project schools. Schools were declared eligible for evaluation and included in the sample frame according to a single criterion: whether they included the target grade ranges for the evaluation, specifically grades 6-8 for primary schools and forms 1 and 2 for secondary schools. Some primary schools, for example, do not include all three of the grades 6-8, and were therefore excluded from the sample frame. Notably, all other schools were included, with no exclusions made on the basis of accessibility, conflict, or any other consideration. As such, the sample frame is, to the extent possible, a fair representation of the underlying population of schools in which GEC-T is being implemented.[43]

At each school, team leaders conducted an interview with the head teacher or principal (Head Teacher Survey). For all other samples, the selection of schools constitutes the first stage in a multi-stage sample design. These lower-level samples, including the population targeted, the targeted sample size, and the data collection tools employed, are described in Table 7.

TABLE 7: SAMPLED POPULATIONS AND QUANTITATIVE TOOLS EMPLOYED

Respondents	Sample Population	Quantitative Tools Completed
Learning Cohort Girls	• Grades 6-8, Forms 1-2 • 12 girls per school (max n = 1,680) • Random sample, stratified by grade • 4 girls per grade in primary schools; 6 girls per grade in secondary schools	• Girls School Survey • Learning Assessments
Learning Cohort Boys	• 350 in-school boys, sampled from project schools • Grades 6-8, Forms 1-2 • 3 boys per school (max n = 420) • Random sample, stratified by grade	• Boys School Survey • Learning Assessments
Households and Out-of-School (OOS) Girls	• Households in communities surrounding project schools • 4 per community (max n = 560) • Randomly sampled using random walk approach and varied starting points • Households eligible if they have at least one girl aged 11-18 years	• Adult module (any adult) • Caregiver module • Girl module • Learning Assessments (only OOS girls)

[43] Given that the sample was drawn randomly, without stratification, its composition does deviate from some of the population-level characteristics considered key to evaluating the project. As we discuss in greater detail in the next section, we employ survey weights to correct for differences in the drawn sample of schools and the population of GEC-T intervention schools.

Bursary Girls[44]	• In-school girls, project schools, receiving bursary support • Grades 6-8, Forms 1-2 • Randomly sampled from eligible girls in each school • 3 girls per school (max n = 420)	• Girls School Survey • Learning Assessments
Learning Benchmark Girls	• In-school girls, project schools • Forms 3 and 4 • Randomly sampled from eligible girls in each secondary school • 4 girls sampled from Form 4; 2 girls sampled from Form 3 • Combined with learning assessment pilot data for Form 3 girls	• Learning Assessments
Head Teachers	• Head teachers at all sampled project schools (max n = 140)	• School (Head Teacher) Survey

Teachers & Classrooms	• Varied samples of teachers and classrooms
	  ○ Within each grade, randomly select 1 class for headcount, among English, Somali and math classes
	  ○ 5 teachers, randomly sampled, for teacher survey, stratified by gender (at least one female teacher, if available)
	  ○ 2 teachers, randomly chosen as subsample of teacher survey sample, for classroom observations; stratified by gender (at least one female teacher, if available)
		• Headcount Tool
		• Teacher Survey
		• Classroom Observations

At the outset of data collection, the evaluation team reviewed RI's sample frame of schools, which included comprehensive enrolment figures for each school in the sample. At that time, each school in the sample contained a sufficient number of female and male students to fulfil the sample quotas outlined above. However, given past experience working in Somali schools, it was assumed that field teams would encounter schools with dramatically altered enrolment numbers upon arrival. Other potential barriers include inaccessibility, either due to insecurity or refusals to participate at the school level.

To prepare for these possibilities, a replacement strategy was developed to maintain the overall composition of the sample. In the event that – upon the arrival of field teams on-site – a school did not have 12 enrolled girls in eligible grades (Grades 6-8; Forms 1-2), the entire school was replaced. Forcier proactively drew a set of replacement schools for this eventuality. Unfortunately, the set of available replacement schools was small, because the sample itself includes approximately 80 per cent of all schools in the sample frame. For this reason, Forcier developed two complementary procedures for replacing schools in the sample.

[44] Bursary girls were sampled using two different approaches. At the start of fieldwork, bursary girls were randomly sampled from a list of eligible bursary girls in each school. Midway through fieldwork, RI provided a complete list of girls receiving bursary support in each project school. As a result, at schools sampled early in the fieldwork period, bursary girls were list by field teams and randomly selected; in later schools, Forcier's technical staff drew a random sample from RI's comprehensive sample frame.

The first approach maintained the key characteristics of the school being replaced, and was used whenever possible. Key characteristics were defined as location (in terms of state/zone, i.e. Somaliland, Puntland, Galmudug, Hirshabelle, and Banadir), region within location, urbanity, IDP status, and school level. Forcier selected replacement schools *within* the same category on each characteristic. Consider a school with the following characteristics: a primary school, non-IDP, in an urban area of Awdal, Somaliland. If such a school required replacement, the replacement would be selected from the set of non-sampled schools that have all the same characteristics.

Because of the small number of potential replacements and the large number of subgroups defined by the key characteristics described above – 39 in total – there are many subgroups that have precisely zero available replacements. Schools with these characteristics, if they required replacement, did not have any available replacements. In these cases, Forcier's research staff employed the second, less preferred, replacement approach. In this approach, the number of total characteristics on which replacements were matched was reduced in steps, shown in the table below, by eliminating characteristics one-by-one. Step 0 refers to the ideal replacement scenario, in which all characteristics from the main sample draw are maintained. In Step 1, we eliminated region as a characteristic – though maintaining project location at the state or zonal level – and IDP status. In this case, the example school described above would be replaced by a primary school that is also in an urban area of Somaliland. However, the replacement would not necessarily match the replaced school in terms of IDP status or region within Somaliland.

TABLE 8: STEPS FOR SELECTING REPLACEMENT SCHOOLS

Step Number in Replacement Process	Variables Factored into Replacement
0 (Ideal Replacement Scenario)	State, Region, School Level, Urbanity, IDP Status
1	State, School Level, Urbanity
2	State, Urbanity
3	State

If we were still unable to draw a replacement school within this broadened category, we moved to an even broader level, by eliminating school level as a consideration in selection, such that primary schools can be replaced by secondary schools, and vice versa. Moving to the broadest possible grouping, we eliminated urbanity as a consideration, while maintaining selection that matches only the project location of the school being replaced. This overall strategy was only applied in scenarios in which no replacements using the first approach were available; moreover, we only moved to broader replacement criteria when there was a complete lack of available replacement schools within the lower categories.

SURVEY WEIGHTS

Much of the analysis in this report makes use of sampling weights in line with the sampling methodology outlined in the previous section. Our approach to weighting the data – and the purpose that we pursue via weighting – is, perhaps, unusual. Sample weights can serve a number of purposes. For instance, researchers can apply survey weights to account for differences in each individual's probability of selection into the sample. This method is typically employed in complex multi-stage samples where simple random selection was not used.[45] An alternative purpose is to weight the data to mirror underlying population parameters – typically referred to as post-

[45] Or where sampling using probability proportionate to size was not used.

stratification – such that the sample's gender balance, for example, matches a known gender distribution within the target population.

The weights we employ serve two purposes. The first purpose is to mirror the underlying population of GEC-T schools in terms of *school-level* characteristics. Random selection in the first stage ensures a set of sample schools that mirror the population of GEC-T schools in expectation only, not in every possible sample. Therefore, as shown in Table 9 below, the sample schools do not precisely match the characteristics of the GEC-T population of schools.[46] Table 9 highlights especially significant gaps between the overall GEC-T population and the raw, unweighted, sampled schools in terms of location.

We post-stratify at the school level to match the underlying population characteristics of GEC-T schools. The distribution of school characteristics in the weighted sample of cohort girls is provided in the right-most column of Table 9; the weighted data match the population characteristics more closely in almost all cases than the raw sample. It is worth noting that we were unable to precisely match population characteristics, because some strata which contained just one or two schools in the population were, unsurprisingly, not represented at all in the sample.[47] Weighting cannot correct for this omission, which is a natural consequence of sampling across many strata of interest, which means that we can only weight the sample to approximately match the underlying population.

TABLE 9: POPULATION PARAMETERS, UNWEIGHTED AND WEIGHTED SAMPLE DISTRIBUTIONS

Characteristic	Population Distribution	Raw Sample Distribution	Weighted Sample Distribution
Project Location			
Somaliland	38.43%	34.56%	38.05%
Puntland	38.86%	41.64%	41.05%
Galmudug	9.61%	8.20%	7.76%
Hirshabelle	1.31%	0.75%	0.86%
Banadir	11.79%	14.85%	12.28%
Other Characteristics			
Rural	31.44%	26.60%	28.73%
Urban	68.56%	73.40%	71.27%
IDP Status	12.72%	5.03%	11.64%

The second purpose of our weighting is to – to the extent possible – ensure equal weight is given to all schools in the sample. In other words, we construct weights so that 12 cohort girls in School X are given equal weight in the analysis to 11 cohort girls in School Y. Our preference for equal weight is driven by two considerations. The first is to ensure that learning improvements in different schools are treated fairly – if a single school with many girls included in the learning cohort were to exhibit improvements (or declines) in learning or transition, we would not this single school to drive project-level outcomes. The second is to ensure that the idiosyncratic outcomes of fieldwork do not influence the results. Each school in the sample was allocated an equal number of cohort girls. But not all schools have an equal number of achieved interviews. By re-weighting the data to equally

[46] This is especially true because some schools were excluded from the sample frame, as noted previously, because they lacked the necessary grade levels for the purposes of evaluation.

[47] As one example, consider rural, IDP schools in Somaliland: only one such school exists in the population of GEC-T schools, but it was not included in the sample – therefore, weighting cannot correct for this small imbalance.

weight *schools*, we ensure that a school in which only 10 interviews were completed is given equal aggregate weight to a school with 12 completed interviews.

Importantly, the first purpose outlined above took priority in our approach. That is to say, we calculated weights to match population characteristics first and foremost. Within each post-stratification strata, we ensured that schools were equally weighted.[48]

In the interest of transparency and replicability, it is important to understand why we did not elect to weight the sample to match the underlying population characteristics of *students*, rather than *schools*. Project schools vary significantly in size, from schools with fewer than 50 total students, to those with more than 2000. Weighting observations – which were collected in approximately equal numbers at each school – to mirror the underlying population of schools would require vastly different weights to be applied across schools. Large disparities in weights tend to make parameter estimates unstable and unreliable. Moreover, such weights would imply that learning improvements among a cohort girl in a large school would be considered far more important, statistically, than learning improvements among a cohort girl in a small school, counter to our goal of equally weighting schools. It is also the case that panel attrition will complicate and undermine weighting schemes that weights school unequally. As in past GEC projects, we expect the evaluation to encounter significant attrition at the midline and endline. If schools are weighted unequally at the baseline, the weighting issue will be complicated further by differential attrition over time. Our strong preference is to weight schools equally, to allow for comparable data to be constructed at the midline and endline with straightforward weighting.

Throughout this report, we weight results when multiple observations were selected at the same school or in the same community. That is, we weight household observations to ensure each *community*, as an aggregation of multiple household surveys, is weighted equally. We weight classroom observations in a similar fashion. The sole exception are surveys of head teachers, in which one interview was completed per school – as such, no weights are needed to ensure equal weight across schools.

Qualitative Tools

In addition to extensive quantitative data collection, the evaluation conducted dozens of FGDs and KIIs with a wide range of populations, in an effort to triangulate findings, assess aspects of the project whose evaluation are not conducive to quantitative measurement, and to generate additional insights into barriers to girls' education and other factors that might influence project design and implementation. FGDs were conducted with groups of fathers, mothers, female students, male students, and CEC members, while KIIs were conducted with head teachers, female teachers, male teachers, and an assortment of MOE officials. Table 10 provides a breakdown of qualitative interviews conducted, including their locations.

TABLE 10: QUALITATIVE INTERVIEWS, BY POPULATION AND LOCATION

Type of Interview	Population Interviewed	Locations	Interviews Completed
FGD	Fathers	Somaliland, Puntland, Galmudug, Banadir	9
FGD	Mothers	Somaliland, Puntland, Galmudug, Banadir	8
FGD	Girls	Somaliland, Puntland, Banadir	8

[48] Ensuring equal weight for all schools across the sample is not possible while also mirroring population parameters, which is the reason that weights are equal within strata only.

FGD	Boys	Somaliland, Puntland, Galmudug, Banadir	9
FGD	CEC Members	Somaliland, Puntland, Galmudug, Banadir	8
KII	Head Teacher	Somaliland, Puntland, Galmudug, Banadir	8
KII	Female Teacher	Somaliland, Puntland, Galmudug, Banadir	8
KII	Male Teacher	Somaliland, Puntland, Galmudug, Banadir	10
KII	MOE Official (Child Protection)	Somaliland, Puntland, Banadir	3
KII	MOE Official (Quality Assurance)	Somaliland, Puntland, Banadir	3
KII	MOE Official (Gender Equality)	Somaliland, Puntland, Banadir	3
KII	MOE Official	Galmudug	1

Benchmarking Outcomes

For the purposes of drawing inferences regarding project impact in future evaluation waves, this baseline establishes benchmarks for learning and transition rates. The logic of benchmarking is straightforward: girls in Form 2 who are tested at the baseline will exhibit natural improvement rates over the course of each year, independent of project interventions. To account for counterfactual improvement rates – the rate of improvement that would occur under counterfactual situation in which no EGEP-T interventions occurred – the girls' performance at the midline, in Form 3, needs to be compared to typical performance for a girl in Form 3.

For most grade levels, learning scores can be benchmarked against other respondents in the same sample, who are one or two grades ahead. In other words, girls in Grade 6 at the baseline, when assessed at the midline, can be compared to girls in Grade 7 at the baseline. This approach breaks down in two cases: the first is for students currently in Forms 1 and 2, most of whom will have advanced beyond Form 2 at later evaluation stages – because the cohort samples targeted respondents only up to Form 2, no benchmark is available for these students in future waves. The second case concerns transition rates: because all girls in the learning and transition cohort were enrolled in school at the baseline, we cannot use them to assess transition rates at different grade levels. In each case, benchmark samples were drawn, with sample sizes and methodology as described in the previous section. To clarify the approach to benchmarking, the box below describes the relationship between grade levels of the cohort and their respective learning benchmarks.

In the case of learning, benchmarking is relatively straightforward, with comparisons for Form 1 and Form 2 cohort girls being made to Form 3 and Form 4 girls drawn explicitly for the purpose of benchmarking. In the case of transition, benchmarking is done by age, because many girls in the transition benchmark sample are not enrolled in school. To draw comparisons, then, we assign benchmarks based on age. Importantly, note the fact that no benchmarks are available for cohort girls currently aged 18-20. This fact arises because the transition benchmark sample targeted girls aged 11-18 only.[49] As such, only cohort girls who are 17 or below have suitable benchmark peers at the midline.

[49] The decision to include only girls 11-18 in the transition sample contradicts RI's MEL Framework – which specified an age range of 10-22 years – and stems from a misunderstanding on the part of Forcier's evaluation team at the start of data collection.

Benchmarking for learning

Baseline	Midline (1 year later)	Endline (2 years later)
Project grades		
G6	G7	G8
G7	G8	F1
G8	F1	F2
F1	F2	F3
F2	F3	F4
Benchmark grades		
F3	n/a	n/a
F4	n/a	n/a

Benchmarking for transition

Baseline	Midline (1 year later)	Endline (2 years later)
Project ages		
11	12	13
12	13	14
13	14	15
14	15	16
15	16	17
16	17	18
17	18	No benchmark available
18	No benchmark available	No benchmark available
19	No benchmark available	No benchmark available
20	No benchmark available	No benchmark available
Benchmark ages		
12	n/a	n/a
13	n/a	n/a
14	n/a	n/a
15	n/a	n/a
16	n/a	n/a
17	n/a	n/a
18	n/a	n/a

GESI Standards in the Baseline Evaluation

Part of GEC-T's approach to structural inequalities within societies where GEC-T projects are being implemented is explicit inclusion and assessment of gender and other inequalities during the evaluation process. As noted elsewhere in this report, EGEP-T meets the Gender Equality and Social Inclusion (GESI) minimum standards as a project – the project actively seeks to transform gender inequalities and, at minimum, accommodates and works around other prevalent sources of inequality within Somali society, especially disability.

The project's evaluation also meets the GESI minimum standards. The baseline evaluation collected sufficiently detailed information about respondents and their households to allow disaggregation by a broad set of groupings and potential barriers, including a number of known sources of inequality in

the Somali context. The evaluation collected – or will collect, during the midline and endline – data on the following partial list of demographic and related characteristics important for identifying and studying EGEP-T's role in reducing existing inequalities:

- Marriage and motherhood among girls
- Orphanhood
- Mother tongue, for identifying linguistic minorities
- Disability, including specific types of disability and severity
- Internal displacement and migration status
- Religion, for identifying religious minorities

The evaluation also collected data on barriers that, in addition to simple demography, arise in the Somali educational context, including seasonal migration, the influence of drought, and the impact of internecine conflict.[50]

The evaluation places a heavy emphasis on investigating the role of these barriers in shaping learning and retention outcomes. At the broadest level, the sampling design and targeting of both boys and girls allows the evaluation to directly assess the existing learning and retention gap – and gaps of other types, such as gender-based differences in agency over children's schooling decisions – between boys and girls. Rather than assume that girls' performance on these metrics lags that of boys', the evaluation is able to study it empirically. Beyond this gender gap, the rich data collected enables the identification of population subgroups of girls (and boys) who are at the greatest risk or disadvantage relative to their peers across Somalia. The performance of out-of-school girls can be tracked vis-à-vis in-school girls, and the evaluation can and will disaggregate outcomes, where appropriate, by urbanity, IDP status, disability, drought-impacted, conflict-impacted, and a range of other characteristics.

Both in this baseline report and in the broader evaluation methodology, the influence of inequality-producing and inequality-reinforcing barriers to educational attainment are considered extensively. In many cases, the evaluation has gone an additional step in identifying barriers to education that impact the most disadvantaged girls. As noted above, the evaluation collects detailed information on disability status, type and severity in EGEP-T schools and communities. But the evaluation also attempts to identify the extent to which disability is currently accommodated in EGEP-T schools – and other schools in EGEP-T communities – through the use of follow-up questions to students and parents about the difficulties they face at school, direct observation of how disability is handled in the classroom, and qualitative investigation of if and how school administrators are seeking to target children with disabilities for specific assistance. In many cases, similar approaches have been taken to other characteristics and barriers identified above, with the goal of assessing not just the severity of existing gaps in outcomes between disadvantaged students and their peers, but also the extent to which their schools and communities are actively seeking to alleviate these disadvantages.

[50] Importantly, while a wide range of data on relevant barriers and characteristics was collected at the baseline, the evaluation will collect even richer data at the midline and endline, as household surveys will be conducted with cohort girls at that time.

2.4 Baseline Data Collection

PRIOR TO DATA COLLECTION

Prior to the start of training and data collection, Forcier's technical staff met with the Monitoring and Evaluation team from RI to discuss possible sampling designs and data collection tools that could be employed in the baseline. Based on the set of targeted core outcomes and intermediate outcomes, Forcier proposed a set of data collection tools designed to collect significant information on each outcome, including triangulation of sources and measurement across multiple sub-indicators in many cases. Where appropriate, tools utilised in previous GEC interventions in Somalia (i.e. EGEP and SOMGEP) were adapted extensively to meet the needs of the new evaluation; in other cases, such as the survey of cohort girls at schools, survey instruments were designed from scratch to accommodate the unique requirements of the sampling design. RI's Monitoring and Evaluation team provided multiple rounds of comments on the tool design and proposed within-school and within-community sampling strategies.

To facilitate the midline and endline evaluations, special care was given to tracking cohort girls. Based on Forcier's previous experience tracking GEC beneficiaries in Somalia, each cohort girl was assigned a unique ID which was used consistently for every survey. In addition, Team Leaders provided tracking sheets for each school to the Project Officer which contained each girl's name, her caregiver's name, the name of the head of the household, and up to two phone numbers to reach the girl on. These data were stored on both paper copies that were returned at the end of fieldwork and in a digital database. In addition, notes on how to find each girl or school were provided by Team Leaders.

Piloting of the data collection instruments took place at two times. Prior to Forcier's engagement, RI and CARE jointly developed and piloted learning assessments, as discussed in greater detail below. The household survey was piloted by enumerators on the final day of training. On the intended day of piloting (November 9, 2017), unexpected pre-election rallies took place in and around Hargeisa, where training and piloting was taking place. As a result of the delays, several of the teams were unable to pilot the household survey extensively or at all. A second round of piloting took place on November 12, 2017. In response to feedback during training and in the pilot, certain questions, related to clan and religion, were removed as they garnered very negative results both by enumerators and respondents. In addition, translations were checked and updated in all quantitative tools.

Beyond piloting, the instruments were reviewed extensively by Forcier's technical staff, Forcier's core Somali staff members, and RI's Monitoring and Evaluation team. Particular attention was paid to the nature and accuracy of the scripting, i.e. the skip logic between questions, constraints placed on responses to ensure reasonable answers, and so forth. Additionally, the instruments were translated by Forcier's core Somali staff, and many of the instruments were re-translated for comparison or replacement by RI's staff, to ensure accuracy was maintained in the Somali versions.

Field staff members were a mix of Forcier Consulting's core (full-time) staff members, temporary hires that Forcier had worked with in the past, and temporary hires recruited through Forcier staff members. Six team leaders and two enumerators were Forcier Consulting staff. The rest of the team leaders and enumerators were temporary staff hired for the EGEP-T baseline. All the team leaders had worked with Forcier before, including on previous GEC evaluations. The enumerators, recruited via Forcier staff, lived in the locations of interest. This allowed for better facilitation at the schools, as the staff already lived in those areas and knew the community well.

The enumerators undertook a five-day long training in Hargeisa between November 5, 2017 and November 9, 2017. The first three days included in-depth overviews of each survey, led by Somali Forcier staff. Enumerators used the CATI survey during the overviews of each survey. In addition, the qualitative tools were reviewed with the team leaders as they were the ones in charge of conducting the qualitative tools. The fourth day included practice household and learning assessment tools for the enumerators and a review of the school survey for the team leader. On the fifth day of training, the enumerators spent the morning piloting tools in the field. Because schools were closed, enumerators conducted household surveys and learning assessments, which are the most technically challenging of the surveys. Then they spent the afternoon providing piloting feedback and receiving instructions for fieldwork. The first teams left for fieldwork on November 11, 2017.

DURING DATA COLLECTION

Fieldwork started on November 14, 2017 and concluded on December 20, 2017. Due to elections and permit delays, most teams started at different times. Please see Table 11 for fieldwork dates per project location.

TABLE 11: FIELDWORK START DATES, BY LOCATION

Project Location	Start Date	End Date
Somaliland	November 18, 2017	December 17, 2017
Puntland	November 14, 2017	December 20, 2017
Banadir	November 22, 2017	December 19, 2017
Galmudug/Hirshabelle*	November 14, 2017	December 18, 2017

*One team was in charge of data collection for both Galmudug and Hirshabelle

Qualitative interviews were conducted at the same time as quantitative data collection and at the same schools or communities. If a school needed to be replaced because of too few girls for quantitative data collection, any qualitative interviews were also done at the replacement school.

Enumerators were required to sign a child protection policy at the beginning of training. Enumerators were a mix of females and males to ensure girls can be surveyed by a female and operated in pairs at all times. Enumerators were required to inform Forcier Consulting who would then inform Relief International in the event a child protection concern arose during the data collection. A detailed module during the enumerator training provided explicit instructions never to interview a child alone in a private place. All interviews were conducted within view of another responsible individual.

Forcier staff also accommodated children with disabilities as needed. Enumerators were trained and were expected to treat children with disabilities with respect and sensitivity. Team leaders were provided with large-print copies of the learning assessment, to facilitate participation by children with impaired vision. Enumerators were also instructed to be aware of accessibility concerns for children with mobility impairments by, for instance, selecting only easily-accessible locations for conducting interviews, assessments, and FGDs. Unfortunately, no formal accommodations were made for children with impaired hearing or difficulty communicating.

Enumerators were also kept safe during fieldwork. Forcier made one school replacement as the area was deemed sufficiently insecure at the time to prevent fieldwork. Fieldwork was also halted during an outbreak of violence in Galkayo, and – partially to ensure security – fieldwork did not begin in Somaliland until after the Somaliland elections on November 13, 2017 were completed and the results were finalised.

To ensure the quality of data collection, Team Leaders reviewed ODK surveys every night before submitting them to the Project Officer. They also worked with enumerators if mistakes were identified by either the team leader or Project Officer to improve data collection and fix any mistakes.

On most days of data collection, Forcier's Project Officer reviewed data that had been uploaded the previous day. This process involved downloading the data from Ona's servers, and running a custom-written cleaning script for each data collection tool, which was adapted from Forcier's standard data cleaning and QC approach to meet the unique needs of data collection for EGEP-T. Specifically, the cleaning process verified that target sample sizes had been met at individual schools, checked for inconsistencies in some of the survey questions, and determined whether tracking ID numbers had been used properly to identify cohort girls, cohort boys, etc. A variety of diagnostic checks were also performed: for instance, Forcier staff reviewed the gender distribution of boys and girls included in the household survey, the gender distribution of heads of household, and other variables that could be compared to expected population averages to check consistency. Finally, Forcier's technical staff spent considerable time looking for anomalies or problematic entries that were correlated across enumerators and teams, to check for fraudulent data.

TABLE 12: ACHIEVED SAMPLE SIZES, BY TARGET POPULATION

Population	Instrument(s)	Sample Size (n)
Learning cohort girls	Girls School Survey; Learning Assessments	1,609
Learning cohort boys	Boys School Survey, Learning Assessments	398
Bursary girls	Bursary Girls School Survey; Learning Assessments	428
Head Teachers	Head Teacher Survey	140
Teachers	Teacher Survey	516
Classrooms	Classroom Observation	270
Classrooms	Headcounts	893
Households	Household Survey	559

In total, Forcier completed fieldwork at 140 schools, with achieved sample sizes per instrument and population provided in Table 12.

AFTER DATA COLLECTION

At the conclusion of data collection, Forcier's data collection team performed additional checks for consistency of the data and cleaned the data more extensively to allow for merging of data between datasets, where necessary for analysis. Open-ended survey responses were translated by Forcier's Somali-speaking staff. Qualitative interviews, which were recorded during fieldwork, were transcribed and translated by Somali-speaking staff; translations were then checked by a second, independent Somali-speaking staff member, and corrections were made where necessary.

The quantitative and qualitative data were then provided to Forcier's evaluation team, who was responsible for performing the analysis and composing this report. The evaluation team engaged in additional, extensive data cleaning, checking for duplicate observations, removing implausible responses, and re-coding variables as necessary. After the initial data export from Ona's servers in SPSS (.sav) format, the data were stored and analysed exclusively using Stata.

2.5 Data Collection Challenges

Under the unique circumstances faced by researchers in Somalia, data collection proceeded with minimal challenges. Multiple contested and overlapping jurisdictions between regional governments and the federal government in Mogadishu complicated the granting of permissions, and violent conflict, in Galkayo particularly, contributed to fieldwork delays. Outside of delays, few significant challenges arose.

During the design and pilot testing of the data collection tools, concerns arose regarding at least two sensitive survey questions. A number of enumerators – and some respondents during pilot testing – indicated that they were uncomfortable with questions regarding respondents' clan membership. Clan identity is politicized in much of Somalia, and questions regarding clan membership, especially from a stranger, can cause concern among respondents and even prompt confrontation. A smaller number of enumerators expressed concern regarding questions about a respondent's religion. In response to these concerns, the clan identity question was removed from the survey, while the religious identity question was left in place.

A breakdown of more specific challenges encountered by geographic location is provided below.

PUNTLAND

- In Daawad International Primary School (PL100), the team was not allowed access to grade records beyond the cohort girls because the head teacher claimed the grades were confidential.
- In New Waberi Primary School (PL105) the head teacher was fired the day before the team reached the school and so they were initially delayed in starting as no one knew where the records were kept.
- There was violence in Galkayo midway through fieldwork that delayed all Galkayo teams. When the violence stopped, they were able to continue with their work.
- The Qanlaye region was too dangerous for the team to travel because of ongoing conflicts. The schools in Qanlaye were replaced with schools in Bosasso.

BANADIR AND GALMUDUG

- There was a twelve-day delay in securing the permit for the Mogadishu teams. A Forcier employee had to fly to Mogadishu to meet with the MOE in order to secure the permit.
- Teams were delayed in reaching several schools as the main roads in Mogadishu were shut down for security purposes. The roads eventually opened up and the teams were allowed to continue.
- When the Galmudug team reached Guriel, they found all the schools were closing two days after their arrival. As such, there are no headcounts or classroom observations for three schools in Guriel.

SOMALILAND

- There was a clause in the Somaliland MOE permit that required the teams to travel with a representative of the REO. Negotiating with the REOs and the MOE delayed the start of the Hargeisa teams by five days. The REOs later made more demands, requiring Forcier to provide transportation for them outside of Hargeisa, and requiring extra per diem for the Sahil REO as the Sahil teams were working in two different locations.
- In Arabsiyo Secondary School (SL019), the head teacher was new and did not have any enrollment or grade records. The team was able to collect the rest of the data without delay.

- In Ma'allin Da'uud Primary School (SL022), the head teacher complained about working with ADRA because of apparent poor communication between ADRA and the school.
- In Sheikh Muhumed Warsame Primary School (SL020), the head teacher refused to provide the school survey information beyond the bursary girls, arguing that the school was a private institution. The team was able to collect the rest of the data without delay.
- In Darasalam Primary School (SL018), the head teacher had not shown up to the school in a month and so the team was unable to locate school records.

2.6 Limitations

- **Lack of a control group of schools or students**

The lack of a control or comparison group – of either schools or students – represents the single most important limitation in the evaluation of the EGEP-T project's impact. 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.

As outlined in Section 2.3 above, the design employed relies on two comparisons to draw conclusions about project impact. The first approach employs comparisons across time, comparing progress in the same set of schools and the same cohort of students from the midline and endline to the baseline. For instance, community attitudes toward girls' education will be measured at the endline and compared to attitudes among the same communities at the baseline. Similarly, girls' self-esteem and empowerment will be measured at the endline and compared to the self-esteem and empowerment measures among the same cohort of girls at the baseline.

The weakness of such comparisons stems from the possibility of positive trends over time in the broader project environment. If community attitudes in Somalia are generally becoming more supportive of girls' school completion, the evaluation may attribute impact to the project 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 slow-changing outcomes, such as community attitudes toward girls' education, it may be reasonable to attribute large-scale changes to project impacts. Community attitudes are unlikely to shift dramatically due to broad societal change in just two years; if the endline evaluation documents such a shift, it is likely due to the EGEP-T intervention. Additionally, the magnitude of any changes can be compared to the extent of project activities in particular communities, to determine whether there is a correlation between project activities and the size of changes observed. If communities with the most intense project implementation – especially in outcome-specific realms (i.e. activities meant to promote community attitudinal change) – show the most change in community attitudes, this would suggest that the project has had an effect on community attitudes.[51] Finally, some interventions have a natural control group, in the sense that not every individual is exposed. For instance, EGEP-T messaging campaigns will not reach all members of a

[51] Note, however, that this type of analysis is less useful for outcomes which are prone to quicker change in intervention and non-intervention communities, alike.

community – by comparing attitudes among those who report hearing the message (or those capable of describing the message) to attitudes among those who do not, tentative conclusions can be drawn regarding the impact of messaging on community attitudes.

The second type of comparison employs benchmarks established at the baseline to judge the progress attained over the life of the project. For instance, girls' learning outcomes – which will be assessed at each stage of the evaluation – will be compared to benchmarked learning outcomes from the baseline. To illustrate, consider female students in G6 at the baseline, which will serve as a continual example throughout this discussion: in the midline and endline evaluations, they will, assuming transition, be in G7 and G8, respectively. In order to control for natural improvement in learning outcomes, which would obtain between G6 and G8 even in the absence of the EGEP-T intervention, later evaluation waves will compare this particular grade cohort to scores obtained by G8 girls at the time of the baseline. Baseline scores for G8 girls at the baseline serve as a benchmark: girls reaching G8 at the time of the endline should exceed these scores, if the project has improved learning outcomes.

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 G6 to G8) and program-drive 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 project 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 G8 at the baseline) will not have been affected by these changes at the time of the benchmark's establishment; in contrast, "project girls" (i.e. those in G6 at the baseline) will have been affected from the changes by the time of the endline, when they will be in G8 and being compared to the pre-established benchmark.[52] 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.

An additional source of bias in the case of benchmarked transitions is the possibility that "benchmark girls" and "project girls" are fundamentally different. Because benchmark girls were not assessed at G6, it is not possible to determine whether their G6 learning outcomes were equivalent to project girls at G6. We must assume that their benchmark scores at G8 represent a reasonable expectation for project girls to reach, in the absence of the EGEP-T intervention. If benchmark girls are advantaged or disadvantaged in some way – either due to sampling variation or some more systematic bias – comparing project girls against their benchmark values may not be a fair

[52] 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-project 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.

assessment of the project's impact.[53] A key output of the baseline evaluation will be an assessment of similarity between project girls and their relative benchmark groups, based largely on demographic characteristics, and recommendations for post hoc adjustments (i.e. controlling for specific demographic characteristics that are imbalanced between the two groups) that should be made during analysis at the midline and endline stages.

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. The design limitations noted here should temper conclusions regarding project impact drawn from this evaluation, but should also be used to guide analysis at the midline and endline stages to ensure that the firmest possible conclusions are drawn from the available evidence.

- **Varied assessments across evaluation waves**

The evaluation of EGEP-T, in line with advice from the FM, will employ different learning assessments across each of the three evaluation waves. A significant concern with this approach stems from the possibility that future assessments will be systematically different – either more or less difficult – than the baseline assessments. Even small differences in the difficulty of the assessments from wave to wave exacerbate the limitation concerning the lack of comparison described previously. 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.

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.

Assessments used at the midline and endline will be pilot-tested extensively and calibrated against the baseline assessments, in an attempt to mitigate concerns regarding differences in overall test difficulty from one evaluation period to the next.

- **Timing of the baseline evaluation after the start of project implementation**

[53] While this discussion has focused on bias that may occur due to *possible* differences between benchmark and project girls, one difference between the two groups is known: all benchmark girls are currently enrolled in school at G8. Project girls, who are in G6 currently, will not all be enrolled in school at G8. This difference can be accounted for by restricting comparisons between benchmark girls and project girls that remain enrolled in G8 at the endline. However, even this comparison is problematic – benchmark girls have remained in school until G8 in the absence of a program, while project girls have remained in school while a project encouraging their continued education was in place. This implies that benchmark girls may be systematically different from project girls, because they represent girls of a specific cohort who remained in school despite the lack of a project encouraging their enrolment; in other words, they may be more motivated or have a more encouraging family than the average girl in their cohort, accounting for their continued enrolment until G8. Even a comparison that includes only project girls that remain in school at the endline may be comparing them to a benchmark composed of girls that are systematically different, in a way that would produce bias *against* finding a positive impact of the project on learning outcomes.

The timing of the baseline evaluation complicates the evaluation, though efforts have been made to mitigate this impact. Data collection for the baseline is being collected in November and December, 2017, while project implementation began in May of the same year. To the extent that the project has already begun to have an impact on key indicators, this fact could bias the results of the evaluation. To be concrete, any positive impact the project has already had on the indicators will reduce the scale of the total impact found in the midline and endline evaluations, since the latter would be measured against a baseline that already incorporated some project impacts into its measurements.

Two factors mitigate this effect. The first is the timing of specific aspects of project implementation, which – according to RI's Monitoring and Evaluation staff – have focused on back-to-school campaigning to encourage girls to either enrol in school or stay enrolled during the transition to a new academic year, and bursary support for severely marginalised girls. In the case of back-to-school campaigns, we do not expect the activity to have significantly influenced learning outcomes, community attitudes, or most other indicators at this stage. The primary expected pre-baseline impact from back-to-school campaigns is in enrolment and transition rates, due to the focus on back-to-school campaigns. In the case of bursary support, most girls receiving bursary support are continuing from prior support received during EGEP. This presents problems to the extent that bursary girls were already impacted by at least one intervention at the baseline – the bursary itself, not to mention their exposure to other, previous EGEP activities, which we discuss in greater detail below. However, because bursary girls comprise a unique sample, the pre-baseline impact of on-going bursary support is naturally limited to this sample, and should not impact non-bursary girls or the cohort of boys.

The second mitigating factor seeks to address the fact that the primary expected pre-baseline impact concerns enrolment and transition rates. The household survey – which assesses transition and enrolment rates at the community level – collected information regarding current enrolment as well as information regarding previous enrolment, especially enrolment in the previous academic year. This information will allow the evaluation to estimate enrolment rates in the 2016-2017 academic year, before the start of EGEP-T implementation, and, specifically, enrolment at the conclusion of the 2016-2017 academic year, which is arguably the proper baseline rate of enrolment for evaluation purposes.

Nonetheless, the importance of this limitation should not be underestimated. In addition to making it difficult to attribute early impacts to the project, because they occur before baseline data collection was undertaken, there is also a significant pernicious effect on the establishment of benchmark transition rates. The project collected data from a sample of respondents in EGEP-T communities, whose transition rates from last year to the year of the baseline serve as a benchmark of expected transition rates, *in the absence of project intervention*, among the community.

However, to the extent that the project has had an impact on re-enrolment and transition rates in these same communities, the benchmark transition rate is biased upward *due to project activities*, and does not represent the unbiased, untreated counterfactual needed for comparison. This problem is not strictly limited to project activities that target the community as a whole – such as attitudinal change campaigns – because many of the respondents captured as part of the benchmark transition sample are enrolled in EGEP-T schools. Therefore, they are exposed both to community-level interventions *and* school-level interventions, such as the distribution of food rations as part of drought response efforts, bursary support that has continued since the first phase of EGEP, and other efforts.

To account for the role of project activities in shaping benchmark transition outcomes, the evaluation team has adjusted the transition targets downward for EGEP-T. The transition targets established are lower than those recommended by the FM, but take into account the circumstances outlined above.

- **Evaluating the second phase of a project**

The fact that EGEP-T is the continuation of a previous intervention poses certain challenges for drawing inferences regarding the project's impact. While this report comprises the baseline of EGEP-T, the vast majority of EGEP-T schools have already benefitted from the first phase EGEP intervention. In fact, in the baseline sample of 140 schools, 92.1 per cent were also beneficiaries of the previous EGEP intervention.

The continuation from a prior intervention has two main consequences on the evaluation. First, making improvements in project indicators will be more difficult, because progress has typically already been made from a true baseline, non-intervention state. It is usually safe to assume that, over a project's lifecycle, improvements become more difficult over time: projects typically impact "low-hanging fruit" first – though not necessarily by design – and, in later stages, gains must be made among individuals or communities which were unmoved by the early stages of intervention. This problem is amplified in the case of EGEP-T, where a long-term previous intervention was already implemented.

Second, where project priorities or activities shift from one phase to another, we may expect to see reversion to the mean in some indicators. If interventions undertaken in the first phase improved outcomes but are not continued in the second phase, individuals, schools or communities may backslide partially. It is important that this not be misconstrued as evidence of negative impact by the second phase project.

Both of these consequences suggest a need to temper expectations regarding EGEP-T's impact. Gains are likely to be difficult, and we may observe minor reversals in some indicators that were impacted by the project's first phase. Setting realistic, achievable targets for all indicators will go a long way to alleviating this limitation.

- **Panel attrition through out-migration**

High levels of out-migration pose a threat to the longitudinal panel design of the sample. It is known that school-age girls have a tendency to migrate or emigrate, temporarily or permanently leaving their households and their communities. This migration threatens to remove a significant number of girls from the sample between the start of the study and its end. It also threatens to make some portion of the sample of girls inaccessible during the time of fieldwork if they have migrated temporarily, e.g. on a seasonal basis. If levels of seasonal migration and permanent out-migration are high enough, levels of attrition from the panel may lead to a significant reduction in sample size and thus a reduction in the power of comparisons that can be made using the panel sample of girls.

Even more problematic is the impact such attrition and replacement will have on the comparability of the samples across time. Replacements can be made to the longitudinal sample (from one wave to the next), but cross-sectional comparisons among heterogeneous populations of girls are less valid than comparisons of the same girls over time. This is especially true if attrition from the panel is systematically related to project outcomes. That is, if the likelihood that a girl falls out of the panel over time is correlated with worse learning outcomes, the learning cohort may improve over time as a result of panel attrition, biasing findings drawn from over-time comparisons. Ideally, the midline

and endline evaluation should utilise demographic data gathered at the baseline to assess the characteristics that drive panel attrition, to determine whether girls who fall out of the panel are fundamentally different – in terms of key project outcomes at the baseline – from girls who remain in the panel.

- **Lack of household demographics for the learning/transition cohort**

Because the baseline evaluation does not employ a pure joint-sampling approach, no household surveys are conducted with the girls who constitute the learning/transition cohort. As described in RI's Monitoring, Evaluation and Learning Framework, respondents in the learning and transition cohort will be contacted for household surveys at the midline evaluation. However, for the purposes of analysis in the baseline evaluation report, this limitation means that we will have a relatively circumscribed set of individual-level demographic variables. As a result, analysis of, for instance, baseline learning outcomes according to individual-level characteristics (like migration status of the household, disability status of the girl, etc.) would not be possible.

To mitigate this shortcoming, RI and Forcier designed a brief demographic module to be implemented with each girl selected into the learning and transition cohort at the baseline. This demographic module included questions about the girl themselves: their age, approximate location of birth, disability status, and prior enrolment status. Moreover, it included questions about the girl's households: household size, gender of the head of their household, head of household's occupation and educational attainment, and language spoken at home. These questions required that the evaluation team adapt their wording and response options to make them accessible to relatively young respondents. Responses to these questions allow the evaluation team to analyse baseline learning outcomes according to a limited set of household- and individual-level demographic characteristics.

- **Sample Under-representation of IDPs**

As noted in Section 2.3, the first stage sample draw of EGEP-T schools was a simple random sample from a sample frame consisting of *most* EGEP-T project schools. The only schools excluded from the sample were those without the requisite grades for evaluation (G6-8 for primary schools; Forms 1-2 for secondary schools). By any measure, the extent to which schools were included in the sample frame was impressive – despite Somalia's security challenges, no schools were eliminated from the sample frame due to inaccessibility.

However, those schools that were excluded included a disproportionate number of schools for IDPs. RI's Monitoring & Evaluation team was aware of this problem, because many IDP schools lack grades 7 and 8, and were therefore removed from the sample frame. While IDP schools comprise 12.7 per cent of the population of EGEP-T schools, they comprise just 5.0 per cent of the unweighted sample of schools.

As described in Section 2.3, we post-stratified individual samples to approximate the underlying characteristics of the EGEP-T population of schools with respect to IDP status, urbanity, and project location. While the raw, un-weighted sample of cohort girls includes just 5.0 per cent IDPs, IDPs account for 11.6 per cent of the reweighted sample. Rural schools were also slightly underrepresented in the initial sample draw; by reweighting the data, the share of cohort girls in rural schools rose from 26.6 per cent to 28.7 per cent, more closely approximating the distribution of urbanity in the population, where 31.4 per cent of EGEP-T schools are in rural areas.

- **Truncated Age Range of Transition Benchmark and OOSG Learning Benchmark Samples**

As noted in Section 2.3 (fn. 39), the transition benchmark sample consists of children aged 11-18, who were randomly sampled from EGEP-T project communities. This sample captured 1,250 children in total for the benchmark, which provides an estimate of the transition rate that would obtain in the absence in the absence of project interventions.[54] The transition sample was intended to benchmark transition rates among a broader age range of respondents, however, because the learning cohort currently includes girls up to 18 years of age. At the time of the endline, these girls will be 20 years old. The benchmark transition sample does not include girls of this age; as a result, expected (benchmark) transition rates will need to be estimated on the basis of the available data, in order to provide a valid comparison for the oldest cohort girls at midline and endline.

Due to the overlap between the transition benchmark sample and the sample of OOS girls tested for learning outcomes – OOS girls were recruited for learning assessments during the survey of households, which also constitutes the transition benchmark sample – this same limitation applies to learning scores among OOS girls. That is, the sample of OOS girls who completed learning assessments is truncated at age 18, meaning that future evaluation waves will not have a valid benchmark against which to compare older OOS girls.

- **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.

A number of steps have been taken to mitigate social desirability bias in this baseline evaluation. 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 (household survey), responses from female and male students (Boys and Girls School Surveys), and responses from head teachers (School Head Teacher Survey), in addition to FGDs conducted with community members. 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 community attitudes.[55]

Second, in many cases, we have designed 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*

[54] However, see discussion under "Timing of the baseline evaluation after the start of project implementation" above.

[55] 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.

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."

- **Respondents' Understanding of Key Outcomes**

While many of the outcomes targeted by EGEP-T have straightforward definitions that most respondents will understand, some consist of more complex concepts or include multiple facets. This is a particular problem when community members are asked to assess performance on a complex indicator. For instance, one approach to measuring the quality of school management and governance is to directly ask community members, via the household survey, about the quality of the school's management.

Unfortunately, community members may not understand what high-quality school management consists of, or they may focus their attention on a narrow aspect of management with which they are familiar. To mitigate this limitation, the data collection often include questions designed to assess *particular aspects* of key outcomes directly, rather than asking respondents their opinions about key outcomes. In addition to asking community members about the quality of school management, we observe indicators of a well-managed school directly at the schools themselves. For example, we ask head teachers whether the school keeps a logbook of teacher attendance, and team leaders – who conduct the interviews with head teachers – will then ask to both see the logbook, and assess its completeness and neatness. Similar approaches are taken to verifying that schools post their school mission statements, track expenses and budget information, and so forth. A well-managed school will, in general, have more of these observable characteristics than a school that is poorly managed. Critically, this approach reduces the reliance on respondents' understanding of key outcomes.

- **Estimating attendance – inaccuracy of school record-keeping**

The weakest link in the strategy of triangulating school attendance is the data that will come directly from school 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. The most accurate estimate of attendance will be on the basis of the headcount performed during fieldwork, 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 will also allow for estimates of attendance on the basis of attendance levels reported by caretakers, however these estimates are on an individual basis, and are measured based on a coding of qualitative estimates of attendance (e.g. asking if a given girl attended "most" days), rather than quantitative estimates of exact frequency (which would be unreasonable to request of caretakers). These estimates based on the household survey will be more representative, but much less precise than the other measures.

- **Inconsistencies in school records**

Schools in the sample may have different practices in exam marking and other forms of school records. During training, team leaders and other researchers who had previously participated in EGEP and SOMGEP evaluations in Somalia reported that schools occasionally employ idiosyncratic methods for marking examinations. The most common example – though still infrequent – concerned schools that marked exams out of 50 possible points, as opposed to using a scale from 0

to 100. If these methods are not documented, they would bias any analysis of examination grades gathered from school records.

During training, team leaders – who were responsible for gathering examination, enrolment and attendance data from school records – were instructed to document unusual marking or record-keeping. In practice, we received few reports of such practices; we were also able to verify that most schools marked examinations using a 100-point scale, by reviewing the distribution of exam scores from each school. Nonetheless, the possibility that additional idiosyncrasies influence aspects of school record-keeping urges caution when using school records. Throughout this report, we prioritise the use of data collected first-hand by the evaluation team, whose provenance is better known.

- **Random Sampling of Teachers and Classrooms for Observation**

At the baseline, teachers and classrooms were randomly sampled from within project schools, according to the best criteria available at the time. In the case of teachers selected for classroom observations, the sample targeted specific grades and subjects being taught. However, the sampling strategy did not target teachers who will participate in EGEP-T's teacher training programme or other teacher-focused interventions, partially because these programmes had not yet been fully designed at the time of baseline data collection. As a result, several of the indicators reported regarding teaching quality, for instance, are measured among a wider population of teachers than was necessary. At the midline and endline, the evaluation team should consider how to represent teachers engaged in EGEP-T teacher-focused interventions more fully, while still ensuring a fair comparison – i.e. a similar population – from baseline to midline and endline.

- **Inability to Assess EGEP-T's Teacher Mentoring Programme**

At the outset of baseline data collection, RI and its implementing partners were in the process of designing their teacher mentoring programme, which has been designed to improve teacher quality. The programme involves interventions at a range of levels, from coaching visits to teachers' classrooms to focused training at the regional level. However, because the programme was still being designed at the time of baseline fieldwork, we were unable to design a measurement approach that would capture the extent and success of the programme. As a result, the teacher mentoring programme is assumed to have been inactive at the time of the baseline, and this finding is reflected in the score assigned to its associated sustainability indicator.

- **Gaps in the Assessment of Teaching Quality**

The baseline evaluation's approach to measuring teaching quality sought to gather information primarily regarding teaching practices (learner-centred and gender-sensitive approaches, in particular), while also measuring other aspects of quality, such as teacher attendance and the use of corporal punishment. The primary indicator of teacher quality employed by the project concerns the use of improved teaching practices and – while the baseline employed a number of direct observations of classroom practices – data collected from surveys of teachers was less targeted and, thus, less useful. That is, teachers were not asked to describe their own teaching practices or skills in detail. Moreover, data collected via classroom observation was sufficiently vague that some teachers appear simultaneously to be using high-quality participatory methods *and* utilizing rote repetition. More detailed data, capturing the precise duration teachers use each method during one of the three blocks of observations in each classroom may have yielded even better data in this respect, a factor that should be considered in future evaluation waves.

3. Key Characteristics of Baseline Sample

3.1 Project Beneficiaries

The direct beneficiaries targeted by EGEP-T are marginalised and severely marginalised girls of upper primary and secondary school age at the time of the baseline evaluation. As described above, EGEP-T defines all girls in this age range in project communities as marginalised, based on the particularly severe challenges facing girls in Somalia. Specifically, girls face cultural barriers to education, pressure toward early marriage, and outsized demands for them to perform housework and other domestic duties. This is in addition to broader challenges, including continual conflict, political instability, and intermittent drought, which disrupt households and, thus, girls' educational opportunities. The project also impacts boys in project schools as indirect beneficiaries.

EGEP-T focuses primarily on girls in upper-primary (G6-G8) and secondary school (F1-F4) at the time of the baseline. All girls in these grades will benefit from project interventions. Beyond those grades, other primary-age girls (G2-G5) will be selected for benefits if they are considered severely marginalised. As discussed in Section 1.3, above, severely marginalised girls will benefit from bursary support, the provision of solar lamps and sanitary kits, and other project efforts.

RI defines severe marginalisation on the extent of deprivation and/or challenges faced, relative to the typical girl in Somalia. Girls face additional barriers to education if they are displaced, orphaned, severely impoverished, or members of minority clans or ethnic groups. Additionally, girls currently out-of-school are defined as severely marginalised, owing to their current educational marginalisation. Based on RI's MEL Framework, girls are classified by the extent of their marginalisation through a thorough identification process that involves CEC members, Head Teachers, RI/ADRA/CISP staff, and MoE representatives, who are provided detailed instructions on classification, and who complete a form for each girl assessed, explaining how and why the girl meets the standards for a given category.

Our approach to this section is slightly unusual, owing to the sampling and data collection methodologies the evaluation employs. We assess the extent of marginalisation, where possible, among girls and boys learning cohorts, using data collected for girls and boys using in-school surveys. However, because there is significant overlap between this analysis and that in Sections 3.2 and 3.3, we refer readers interested in a straightforward assessment of marginalisation to those portions of the report. Section 3.2 includes an analysis of disability among four distinct cohorts/sampled assessed at the baseline. Section 3.3 describes educational marginalization in terms of respondent characteristics and specific barriers, focusing on analysis of the girls and boys learning cohorts. The data provided in these sections are necessarily limited by the fact that no household survey was conducted with members of the main girls and boys learning cohorts.

However, the evaluation did collect data that provide further information regarding relative marginalisation. Household surveys were conducted with a random sample of households in each project community (n = 559). Two groups of girls are of particular interest within this sample: girls enrolled in a project school – who were sampled by coincidence – and girls who are not currently enrolled in any schools. From a sample of 559 households, 47.6 per cent (n = 266) households included a girl enrolled in a project school; at the same time, 24.5 per cent (n = 137) households included a cohort-age girl not currently enrolled in any school. These two groups can shed light on the circumstances faced by EGEP-T beneficiary households, because both constitute random samples of their respective subpopulations. That is, sampled girls enrolled in EGEP-T schools are

representative of girls in those schools; sampled OOS girls are, likewise, representative of OOS girls in EGEP-T project communities. Helpfully, households falling into both subsamples completed the more extensive household survey. In this section, we utilise this data to shed light on the circumstances of girls living in EGEP-T communities and attending project schools.

The results of the analysis are provided in Table 13, below. We label girls enrolled in a project school as "in-school girls" to distinguish them from the learning cohort sampled from within project schools – these respondents are enrolled in an EGEP-T school, but are not part of the learning cohort being tracked over time. Table 13 reports unweighted, raw sample proportions, with weighted proportions in parentheses. Finally, note that we focus on survey questions which do not overlap with the considerably larger sample of cohort girls, which are described in Sections 3.2 and 3.3. Therefore, analysis of the extent to which beneficiaries are drought-affected, conflict-affected, or attend IDP schools, for instance, is provided in Section 3.3. Table 13 focuses on areas where insufficient data was collected from the learning cohorts to assess marginalisation.

TABLE 13: CHARACTERISTICS OF EDUCATIONAL MARGINALISATION AMONG GIRLS IN PROJECT COMMUNITIES

Barriers	In-school Girls	OOS Girls
Girl Characteristics:		
Mother	5.3% (5.4%)	8.8% (8.3%)
Orphan	0.4% (0.4%)	2.2% (2.5%)
Single-parent HH due to parent death	14.3% (14.7%)	13.1% (14.1%)
Any physical or cognitive impairment	3.0% (3.1%)	2.9% (2.8%)
Household Characteristics:		
HH migrates seasonally	2.3% (2.0%)	8.8% (9.8%)
At least one child 10-19 has migrated away from HH	20.8% (21.3%)	27.3% (30.9%)
HH does not own a mobile phone	16.2% (15.7%)	14.6% (14.3%)
Unable to meet basic needs without charity	25.2% (26.3%)	28.5% (29.0%)
Gone to sleep many/most nights feeling hungry	10.9% (9.8%)	17.5% (16.3%)
Gone without enough clean water for home use many/most days	20.7% (19.7%)	24.8% (24.2%)
Gone without medicines or medical treatment many/most days	27.8% (27.5%)	36.5% (38.3%)
Gone without cash income many/most days	33.8% (33.5%)	30.7% (31.9%)

The figures presented in Table 13 are instructive for a number of reasons. At the level of individual girls, we see that a moderate share of girls who are mothers or who face physical or cognitive impairments. At the same time, the sample includes very few girls who are completely orphaned, but a significant share – 14.3 per cent among girls enrolled in EGEP-T schools – who have lost a mother or father. Unsurprisingly, the results also provide evidence that a girl's circumstances, especially motherhood and orphanhood, influence enrolment, as girls who are mothers or orphans are more likely to be out-of-school.

The results also suggest that household circumstances may be more common sources of marginalisation for girls than the individual-level characteristics described above. Migration of young boys and girls away from the household is extremely common. And while relatively few beneficiary girls live in households that migrate seasonally, the survey did not adequately capture information about households that have migrated permanently as a result of conflict, economic opportunities, or

other factors.[56] As such, the findings almost certainly drastically understate the extent to which households in EGEP-T communities are migrants. This is suggested, further, by the frequency with which children have migrated away from the household.

EGEP-T communities also have a high rate of marginalisation stemming from financial difficulties. Over one-quarter of households indicate that they are not able to meet their basic needs in the absence of charity. And, when asked to describe their household's level of deprivation over the past year, a large share indicated that they had gone without clean water, medicine or medical treatment, and – especially – cash income – on many or most days over that time period.

Usefully, RI has undertaken a comprehensive data collection effort which identifies the number of girls facing particular challenges (i.e. IDPs, orphans, etc.). This data was not available at the time of the draft baseline report, but RI completed data collection prior to finalisation of this report. This data makes clear the extent of marginalisation. The data suggest that around 1 per cent of girls in EGEP-T schools are identified by school staff as disabled – a fairly low prevalence rate, which we discuss in greater detail below. At the same time, RI's data suggest a much higher rate of orphanhood than the baseline data, with – conservatively – 9.9 per cent of girls enrolled in EGEP-T schools classified as orphans.[57]

As noted above, the data presented in this section is incomplete. A more complete picture of educational marginalisation is provided by combining these results with those from Sections 3.2 and 3.3, below. The midline evaluation will make it possible to provide a more comprehensive assessment of marginalisation – including the extent to which barriers overlap and reinforce one another – among children in EGEP-T communities.

3.2 Achieved Sample

In line with the discussion of sampling methodology and replacements selected in Section 2.3, the achieved sample broadly matched the planned sampling strategy. Each table in this section details the composition of four samples of respondents:[58]

- Learning cohort girls, sampled from project schools
- Learning cohort boys, sampled from project schools
- Bursary girls, sampled from project schools
- OOS girls, sampled from households in project communities

Table 14 describes the achieved sample in terms of geographic distribution; Table 15 disaggregates the samples by age; Table 16 disaggregates the samples by grade, and Table 17 disaggregates the samples by disability status, including only those samples – the girls learning cohort – where questions regarding disability were included. All distributions reported are unweighted.

---

[56] The survey question regarding seasonal migration was, in retrospect, poorly-designed. Seasonal migration is presumably less common in EGEP-T communities – many of which are urban – than permanent or semi-permanent migration. Seasonal migration typically occurs among pastoralist households, while EGEP-T communities more commonly host migrants that have permanently or semi-permanently fled conflict, drought, or are seeking opportunities.

[57] This figure is slightly conservative, because orphanhood data is missing from a few EGEP-T schools.

[58] Note that the learning cohort described in this section is coterminous with the transition cohort. At the same time, the baseline evaluation surveyed a transition benchmark sample, whose purpose was to benchmark transition rates in EGEP-T communities at the start of the project. This sample is not described in this section in detail, though further information is provided in Section 4.3.

Based on the data presented in this section, the sample is not representative of the broader beneficiary population. However, where the sample deviates from the population, this deviation was *by design*. Specifically, the sample targets learning cohort girls and boys in Grade 6 through Form 2, while the project targets students through Form 4. This sampling decision was made with RI, to ensure that cohort girls could be tracked effectively at the endline and that the project's transition rates would be evaluated on valid grounds, since girls leaving secondary school may transition into any number of pathways.

Similarly, the main learning cohort sample does not represent the severely marginalised girls targeted by EGEP-T with bursary support. Rather, a supplemental sample of bursary girls was drawn to ensure coverage of this distinct subpopulation. Moreover, the population of bursary girls receiving interventions from EGEP-T is much broader than that represented in this distinct sample: while many bursary girls are in grades 2-5, the bursary girl sample is limited to those in grades 6-8 and forms 1-2.[59]

For the purposes of establishing baseline learning outcomes – and much of the other analysis conducted in this report – the sample is limited to learning cohort girls as described below, rather than including this additional subsample.

TABLE 14: COMPOSITION OF LEARNING COHORTS, BY PROJECT LOCATION

	Girls Learning Cohort	Girls Bursary Cohort
Somaliland	34.6% (556)	31.8% (136)
Puntland	41.6% (670)	46.3% (198)
Galmudug	8.2% (132)	6.3% (27)
Hirshabelle	0.8% (12)	0.9% (4)
Banadir	14.9% (239)	14.7% (63)
Total	100% (1,609)	100% (428)
	Boys Learning Cohort	**OOS Girls Cohort**
Somaliland	34.7% (138)	24.8% (34)
Puntland	41.0% (163)	46.0% (63)
Galmudug	8.3% (33)	12.4% (17)
Hirshabelle	0.8% (3)	0.0% (0)
Banadir	15.3% (61)	16.8% (23)
Total	100% (398)	100% (137)

TABLE 15: COMPOSITION OF LEARNING COHORTS, BY AGE

	Girls Learning Cohort	Girls Bursary Cohort
Aged <12	3.6% (58)	4.2% (18)
Aged 12-13	26.1% (420)	27.3% (117)
Aged 14-15	42.2% (679)	39.7% (170)
Aged 16-17	25.7% (414)	23.6% (101)
Aged 18+	2.4% (38)	5.1% (22)
Total	100% (1,609)	100% (428)

[59] The evaluation team opted to focus the bursary girls sample on Grades 6-8 and Forms 1-2, grade ranges that match those of the main learning cohort, rather than the overall grade range in which bursary girls fall (grade 3 and up). The logic dictating this approach was that bursary girls could be compared to the main learning cohort over time if they were drawn from the same grade range, allowing inferences to be drawn regarding the impact of bursaries, above and beyond other project interventions. The drawback to this approach is a lack of representation of bursary girls outside these grade ranges, whose numbers are considerable, according to EGEP-T project documents.

	Boys Learning Cohort	OOS Girls Cohort
Aged <12	3.5% (14)	19.7% (27)
Aged 12-13	18.6% (74)	23.3% (32)
Aged 14-15	35.9% (143)	17.5% (24)
Aged 16-17	32.2% (128)	24.8% (34)
Aged 18+	9.8% (39)	14.6% (20)
Total	100% (398)	100% (137)

TABLE 16: COMPOSITION OF LEARNING COHORTS, BY GRADE

	Girls Learning Cohort	Girls Bursary Cohort
Grade 6	27.2% (438)	28.5% (122)
Grade 7	25.9% (416)	25.9% (111)
Grade 8	25.1% (404)	25.9% (111)
Form 1	11.5% (185)	9.4% (40)
Form 2	10.3% (166)	10.3% (44)
Total	100% (1,609)	100% (428)
	Boys Learning Cohort	**OOS Girls Cohort**
Grade 6	25.1% (100)	N/A
Grade 7	27.4% (109)	N/A
Grade 8	26.1% (104)	N/A
Form 1	11.3% (45)	N/A
Form 2	10.1% (40)	N/A
Total	100% (398)	N/A

TABLE 17: COMPOSITION OF LEARNING COHORTS, BY PHYSICAL AND COGNITIVE IMPAIRMENT

	Girls Learning Cohort	Girls Bursary Cohort
Vision impairment	1.3% (21)	No data
Hearing impairment	0.1% (2)	No data
Mobility impairment	0.1% (2)	No data
Cognitive impairment	0.2%(3)	No data
Self-care impairment	0.1% (1)	No data
Communication impairment	0.1% (2)	No data
Any impairment	1.9% (31)	No data
	Boys Learning Cohort	**OOS Girls Cohort**
Vision impairment	No data	1.5% (2)
Hearing impairment	No data	1.5% (2)
Mobility impairment	No data	0.0% (0)
Cognitive impairment	No data	0.0% (0)
Self-care impairment	No data	0.0% (0)
Communication impairment	No data	0.0% (0)
Any impairment	No data	2.9% (4)

Beyond the precise representativeness of the design and achieved sample, the prevalence of disabilities reported in Table 17 is worthy of additional discussion. The evaluation team identified respondents using the reduced set of Washington Group (WG-6) questions:

- Do you have difficulty seeing, even if you are wearing glasses?
- Do you have difficulty hearing, even if you are using a hearing aid?
- Do you have difficulty walking or climbing steps?

- Do you have difficulty remembering things or concentrating?
- Do you have difficulty with self-care, such as washing all over or dressing?
- Using your usual language, do you have difficulty communicating; for example understanding or being understood?

Table 17 contains prevalence rates of each type of impairment among the girls learning cohort and OOS girls identified through the household survey. In the case of the former, WG-6 questions were addressed directly to the girls; in the case of the latter, WG-6 questions were directed to the girl's primary caregiver. Equivalent questions were not included in surveys of bursary girls or the boys learning cohort, though both can and will be assessed with regard to disability at the midline. The evaluation classified a child as being disabled if they indicated that they had either "a lot of difficulty" with a task in the WG-6 questions, or that they were not able to do that task at all.

Our estimates of impairment prevalence are lower than expected, especially given the context in which EGEP-T is being implemented. Due to Somalia's history with conflict, poverty, and lack of access to healthcare, the prevalence of disability (both physical and mental) is higher than the global prevalence and has been estimated to be approximately 20% of the population.[60] This prevalence is much higher than what was seen in the EGEP-T cohort. The difference in prevalence can be attributed to a variety of reasons. First, and most importantly, the SIDA study cited does not appear to have used the Washington Group criteria for defining disability; indeed, it is unclear from the study report what criteria were used. Given the nature of measurement issues surrounding disability, with especially thorny arguments regarding definitions, differences in measurement criteria could completely explain gaps in prevalence. Second, disability questions were asked of the girls directly, rather than their parents; it is possible that the girls either did not define themselves as disabled or they were unaware that they had physical or mental disabilities.[61] Third, it is possible that girls with significant disabilities have systematically lower enrolment rates than average, which would partially explain the low prevalence rate among the EGEP-T cohort relative to the Somali population at large.[62] The cohort being described here includes girls who were enrolled in school at the time of the baseline; if girls with disabilities are less likely to be enrolled, especially in the higher grades that EGEP-T targets – they would be less likely to be included in the EGEP-T cohort, and prevalence rates among the cohort would not match the population at large. This is a natural consequence of the fact that disability prevalence varies by population group, and the EGEP-T

[60] Estimates are provided by the Swedish International Development Cooperation Agency (SIDA. 2014. "Disability Rights in Somalia." Available at https://www.sida.se/globalassets/sida/eng/partners/human-rights-based-approach/disability/rights-of-persons-with-disabilities-somalia.pdf). It is important to note that the source of figures provided by SIDA is unclear and may be based on extremely limited quantitative data. No systematic, nationally-representative surveys regarding health outcomes have been conducted in Somalia in the last 20 years, to the best of our knowledge; as a result, prevalence rates such as these should be treated with significant caution. The recent Population Estimation Survey conducted by UNFPA includes some limited data on disability, but only describes the prevalence of disability relative to individuals who would otherwise be engaged as part of the labour force, which is not comparable to prevalence rates reported for other countries based on demographic and health surveys, national censuses, or other sources. For additional discussion of disability in Somalia, see: Rohwerder, Brigitte. 2018. "Disability in Somalia". Institute of Development Studies report. Available at https://opendocs.ids.ac.uk/opendocs/handle/123456789/13534.

[61] This possibility may stem from two sources: first, the stigma associated with disability in Somalia (see, e.g.: Rohwerder, Brigitte. 2018. "Disability in Somalia". Institute of Development Studies report. Available at https://opendocs.ids.ac.uk/opendocs/handle/123456789/13534.); second, the possibility of simply misunderstanding, among girls included in the sample, about what constitutes a disability.

[62] The descriptive statistics in Table 17 support this interpretation to some degree: the prevalence of impairment is slightly higher among out-of-school girls than among the EGEP-T learning cohort, though the gap is very small.

baseline is not a random sample of the Somali population, but of a very specific sub-population – girls of a specific age, enrolled in EGEP-T schools. Therefore, we should not expect prevalence rates among the cohort to match those among the population at large. Lastly, the high prevalence of disability in Somalia is largely attributable to conflict. Because of this, it is likely that disability is more common in older adults and men who have been involved with or more affected by the conflict than young girls.

3.3 Educational Marginalisation

A number of girl- and household-level characteristics contribute to educational marginalisation. To what extent does the sample of cohort girls include girls with typical forms of marginalisation with respect to education? Table 18 documents a number of key characteristics of cohort girls in the EGEP-T baseline.[63] These characteristics include parental education, gender of head of household, and IDP school status. Because household surveys were not conducted with the cohort girls in the baseline, a number of other potential characteristics indicative of marginalisation – such as orphan status, motherhood among female students, and so forth – were not collected. Table 19 also shows barriers to accessing and completing education in cohort girls – barriers which consist not of a girl's characteristics, but their experiences at school or at home.[64]

In both Tables 18 and 19, several barriers and areas of marginalisation feature prominently. Approximately one-quarter of girls live in households where the head of household has no education, and one-third of girls attend drought-affected schools. A further 15.7 per cent of students attend conflict-affected schools. As one would expect, aspects of marginalisation are overlapping: for instance, 69.6 per cent of girls in IDP schools are also affected by drought in the area and – unsurprisingly – drought and migration are closely related. FGD participants in Bari indicated that "when drought happens, it touches the whole environment…it causes a lot of students [to leave] school to move with their family".[65] Migration among pastoralist families is common as many lost their livestock and were forced to "move to villages from rural areas when droughts affect them".[66] It is also important to note that our results regarding IDP status systematically underestimate the number of IDPs impacted by EGEP-T, because it counts only respondents enrolled in an IDP school, rather than IDPs who are enrolled in non-IDP schools. In practice, this latter group makes up a larger share of all IDP beneficiaries under EGEP-T.

The results regarding language of instruction may be surprising, given the Somali context, where Somali is the dominant mother tongue. All girls in the learning cohort report speaking either Somali (99.8 per cent) or English (0.2 per cent) as their primary language at home. However, a number of schools – mostly, but not exclusively, secondary schools – teach in languages other than Somali. For instance, 40.7 per cent of schools, including 30.3 per cent of primary schools, teach in English.[67] As a

[63] Throughout this section and Section 3.4, we report raw sample proportions and, in parentheses, proportions which employ sampling weights.

[64] All barriers described in Table 19 are drawn from surveys conducted with cohort girls via the Girls School Survey, unless otherwise noted.

[65] FGD with fathers in Bari

[66] FGD with fathers in Maroodi Jeeh; FGD with mothers in Nugal.

[67] According to head teachers, 30.3 per cent of primary schools teach primarily in English, and a further 5.5 per cent teach primarily in Arabic, with the remainder (64.2 per cent) using Somali. Even among schools that serve exclusively primary-age children (as opposed to schools that combine primary and secondary instruction), 22.1 per cent use English as the language of instruction. Among secondary schools, 77.4 per cent use English as the language of instruction, while the remaining 22.6 per cent report using Somali.

result, the 44.1 per cent of girls who are instructed in a language different from their mother tongue are heavily concentrated in secondary schools, with the remainder in primary schools that use English or Arabic as the language of instruction.

As we described in Section 1.1 of this report, the official language policies of the respective ministries of education in Somaliland, Puntland and the Federal Republic of Somalia provide for instruction in Somali during primary school and English during secondary school. In practice, the language of instruction varies from school to school, partially as a function of whether the school is public or private in nature, with public schools much more likely to adhere to the relevant official language policy. We note the specifics of language policy here because differences between mother tongue and language of instruction may not be characteristic of explicit marginalisation in this context, because English-language teaching is widespread at the secondary level. Indeed, students who speak a different mother tongue than their language of instruction are either simply secondary school students enrolled in schools that follow official language policies or they are primary school students enrolled in schools, typically private, that teach in either English or Arabic.[68] As a result, we do not emphasise the importance of "linguistic mismatch" – students who speak a mother tongue different from the language of instruction – as a marker of an individual's marginalisation, though we do note the divergence between official language policy and policy in practice at schools, and the influence this application may have on learning outcomes, which we discuss in slightly more detail in Section 4.2.

TABLE 18: EDUCATIONAL MARGINALISATION CHARACTERISTICS IN COHORT GIRLS

Characteristics	Percentage at Baseline (Weighted Percentage)
Sample Breakdown (Girls)	
Living in female-headed household	44.4% (43.2%)
Attends IDP school	5.0% (11.6%)
Language of instruction different from mother tongue	44.1% (39.0%)
HoH has no education	23.9% (25.6%)
Attends drought-affected school	29.5% (33.4%)
Attends conflict-affected school	15.7% (13.4%)
Sample breakdown (Boys)	
Living in female-headed household	41.2% (39.9%)
Language of instruction different from mother tongue	44.7% (39.0%)
HoH has no education	27.1% (27.8%)
Attends IDP school	4.8% (12.5%)

[68] This contrasts with other contexts where GEC-T is being implemented, such as Sierra Leone or Kenya, where many children speak a wide variety of local languages at home but are taught in Kiswahili or English at school. Under such circumstances, students whose mother tongue is Kiswahili, for instance, are at a marked advantage over students who speak a different mother tongue, and students who speak particularly localised languages or languages of smaller minority groups may be both disadvantaged in the educational system and in the country in more generally. In Somalia, the situation is different: virtually all students speak the same mother tongue, but they attend schools with differing language policies – thus, while language of instruction is undoubtedly an important factor in learning, it is less clearly a marker of marginalisation between students within the same schools. Moreover, in contrast to other contexts, mismatch between mother tongue and language of instruction in Somalia – at least among EGEP-T beneficiaries – does not appear to ever be a symptom of official marginalization of one's ethnic group, because nearly all beneficiaries speak the same mother tongue, and language mismatches stem from school ownership (public versus private) and school level (primary versus secondary), rather than from characteristics of the students themselves.

Attends drought-affected school	29.9% (34.0%)
Attends conflict-affected school	15.8% (13.8%)

Table 19 describes the sample in terms of marginalisation that does not stem directly from a girl's demographic characteristics. For instance, as shown in the top section of Table 19, most girls said they were safe walking to and from, and while they were at, school. Among those who reported feeling unsafe during the journey to school, the vast majority (48.9 per cent) cited the long distance they needed to travel as the main reason why they felt unsafe. Other, less common, responses included the risk of verbal abuse at the hands of other children (11.1 per cent), roadblocks (6.7 per cent), floods and other environmental factors such as heat or rain (8.9 per cent), traffic (4.4 per cent) and others.[69] These findings generally correspond with adult respondents in the baseline evaluation's household survey, among whom 3.1 per cent reported that it was fairly or very unsafe for girls to travel to school in their area.[70]

From the perspective of household support for education, girls overwhelmingly report that they receive support to stay in school. In total, just 2.4 per cent of girls disagreed with the statement "I get support I need from my family to stay in school and perform well."

In school, one in five girls said there were not enough seats, and thirty per cent said there were insufficient books, for everyone in class. The issue of insufficient learning materials was raised often during qualitative interviews. Teachers in Somaliland indicated that students need better access to books that they can read and – in Hargeisa – fathers report that "there are no books" at the local school.[71] Overall, one of the primary barriers to girls' attendance and school completion is the lack of funds for books, which are expensive to buy.[72]

Table 19 also displays the share of girls who report that they do not use various facilities – drinking water, toilets, and the area where children play and socialize – at their schools. Responses to these questions are slightly difficult to interpret, because they could reflect either the absence of such facilities at a girl's school altogether, or her unwillingness to use them for hygienic, modesty, or other reasons.[73]

In practice, responses appear to capture *both* considerations: for instance, 5.8 per cent of girls attend schools in which the head teacher indicated lacking toilet facilities of any kind, and a further 10.8 per cent attend schools that lacked separate toilets for boys and girls. However, even among schools with separate toilets for boys and girls, 24.8 per cent of girls indicate that they do not use the toilet at school. Focusing on all girls who reported not using toilets at school, they indicated that no toilet was available (47.9 per cent), that they were unable to access the toilet (19.8 per cent), and

[69] Interestingly, some girls who cited the long journey as the reason they felt unsafe also report that it takes them less than 30 minutes to reach school on a typical day, suggesting that the long journey may reflect fears over insecurity or other issues on the road, rather than the length of the journey per se. The structure of the question regarding *why* they felt unsafe may be partially responsible for this pattern, as the survey only allowed respondents to select one reason why they felt unsafe during the journey.

[70] The reasoning of adult respondents in EGEP-T communities also mirrored that of girls themselves: the most common arguments regarding the danger of travel for girls focused on traffic (35 per cent of respondents), the distance (30 per cent), open conflict or violence (30 per cent) and a number of less common concerns.

[71] KII with teachers in Somaliland; FGD with fathers in Hargeisa

[72] FGD with fathers in Maroodi

[73] Questions regarding a child's use of drinking water, toilet, and playground facilities were part of the GEC-T standard survey template provided by the FM.

that the toilets were not acceptable for use (14.8 per cent).[74] To provide a more complete picture of infrastructural limitations of EGEP-T schools, Table 19 also reports the share of girls who attend schools lacking clean drinking water and gender-specific toilets, according to reports from head teachers. The diverse reasons cited by girls with regard to toilets are representative of their reasons for electing not to use their school's drinking water and playground facilities as well: in the majority of cases, failure to use these facilities reflects the fact that they are unavailable, but a substantial minority of girls also cite their inability to access the facilities, and the unacceptable nature of the facilities.[75]

In contrast to the resource limitations described above, girls' appraisals of their teachers were more positive. The vast majority of girls describe their teacher as welcoming, though nearly half of girls agree with the statement "I am afraid of my teacher." Many girls clearly view their teachers as both welcoming and intimidating. This is possible because teachers may act in different ways over the course of the day or week and may be strict during lessons but welcoming of students' questions after class. In addition, over one-third of girls said that their teacher treats boys and girls differently. However, this finding should be interpreted with caution – the question does not specify that girls are being treated *worse* than boys, just differently. To the extent that girls and boys face different challenges, benefit from different teaching approaches, or have other differences, we may observe girls and boys being treated differently, but such differences may be entirely appropriate. We also asked girls whether they agree with the statement that girls are treated equally to boys, and 82.8 per cent of respondents agreed. This suggests that girls and boys may be treated differently, but not in a way that is necessarily detrimental or discriminatory toward girls.

TABLE 19: EDUCATIONAL MARGINALISATION BARRIERS OF COHORT GIRLS

Barriers	Percentage at Baseline (Weighted Percentage)
Sample breakdown (Girls)	
Home – community	
Safety:	
Doesn't feel safe travelling to/from school	3.2% (3.6%)
Parental/caregiver support:	
Doesn't get support to stay in school and do well	2.4% (2.4%)
School level	
Attendance:	
Doesn't feel safe at school	2.0% (2.0%)
School facilities:	
Not enough seats for all students	20.1% (21.0%)
Doesn't use drinking water facilities	36.1% (36.9%)
Doesn't use toilet at school	26.1% (28.0%)

[74] A further 17.6 per cent of respondents who reported not using their school's toilets indicated that they did not know why they chose not to use them.

[75] It is unclear whether "inability to access" reflects physical barriers to access stemming from a child's disability, or some other access barrier, such as school staff or students who prevent them from using the facilities. Unfortunately, sufficiently detailed response data was not collected to allow us to distinguish between these types access barriers. It is worth noting, however, that the share of students who report being unable to access, for instance, playground facilities, far outstrips the share of students who report mobility-oriented disabilities, even when using a less stringent standard for defining disabilities, as a student reporting *any* difficulty walking, climbing steps, or engaging in self-care such as washing.

Doesn't use areas where children play/socialise	43.6% (44.4%)
Books are not available for students	29.1% (28.5%)
Computers are not available for students	87.3% (88.0%)
No girl-specific toilet at school (head teacher)	16.7% (16.9%)
Does not have clean drinking water at school (head teacher)	27.8% (27.0%)
Teachers:	
Disagrees (a little/a lot) that teachers make them feel welcome	3.4% (3.4%)
Agrees (a little/a lot) teachers treat boys and girls differently in the classroom	36.1% (39.3%)
Agrees (a little/a lot) teachers treat girls equally to boys	82.8% (83.6%)
Agrees (a little/a lot) teachers often absent from class	24.0% (24.3%)
Agrees (a little/a lot) that they are afraid of their teacher	47.8% (49.4%)
Teacher uses physical punishment (1+ times per week)	25.9% (26.0%)

Additionally, one out of four girls reported that their teacher was often absent from class. This is apparently so common that in one focus group, male students reported other students teaching the class on their own initiative while the teacher was absent.[76] The CEC and Head Teacher monitor attendance but there is little incentive for teachers to attend their classes – many teachers are not paid their salaries on a regular basis, if at all.[77] Teachers are expected to find their own replacements if they are not able to attend class and one teacher reported that if teachers are frequently absent, they are suspended, saying "If the teacher is absent all day long the teacher might face special suspension and lastly, if a teacher is absent from morning classes he might face two suspensions."[78]

Only a quarter of girls and boys reported physical punishment being used in the classroom in the past week (see Table 19). The most common physical punishment mentioned in the qualitative interviews was 'murga' or the practice of making a child stand with his head to his knees and forcing him to hold his ears at the same time.[79] However, even this punishment was only mentioned sporadically in the qualitative interviews, as most interviewees did not specify the type of physical punishment used. Teachers also reported making a student stand in a corner facing the wall and hitting students with sticks.[80] Punishments were commonly used to discipline students for getting into fights, not finishing their homework, or using social media but the murga was typically used for students who had disrupted the classroom.[81] However, in these same qualitative interviews, it was noted that physical punishment was much more likely to be used on boys than girls.

Importantly, the claimed rate at which corporal punishment is used, according to students, may understate the extent to which corporal punishment is common in Somali schools. The results reported in Table 19 describe the share of students who report that their teacher used physical punishment *within the past week* on at least one student in their class. But, while just 25.9 per cent of students report recent physical punishment, 46.0 per cent report that their teacher uses physical

[76] Focus Group Discussion with Male Students. Hargeisa, Somaliland.

[77] Focus Group Discussion with CEC. Awdal, Somaliland.

[78] Key Informant Interview with Female Teacher. Hargeisa, Somaliland.

[79] Key Informant Interview with Male Teacher. Awdal, Somaliland; Key Informant Interview with Female Teacher. Garowe, Puntland.

[80] Key Informant Interview with Female Teacher. Garowe, Puntland.

[81] Focus Group Discussion with Male Students. Hargeisa, Somaliland; Key Informant Interview with Female Teacher. Galkacyo, Puntland; Key Informant Interview with Male Teacher. North Mudug, Puntland; Key Informant Interview with Female Teacher. Garowe, Puntland.

punishment occasionally. Even this figure may be too conservative: according to direct observation of 270 classrooms during the baseline, both boys and girls were observed being disciplined in over half of classes during the observation period.[82]

As the lower panel of Table 18 shows, there is broad similarity between the boys and girls samples, with respect to most characteristics of educational marginalisation.[83] Given that many of these characteristics – such as drought and IDP status – are measured at the school level, this is not surprising. Nonetheless, the results demonstrate that both EGEP-T's direct and indirect beneficiaries tend to be impacted by multiple aspects of marginalisation.

A further limitation to this analysis concerns a lack of data regarding the girls learning cohort. Due to EGEP-T's sampling approach, which drew the learning cohort from within schools – as opposed to at households – household surveys were not conducted with families of the cohort girls. This fact necessarily limits the extent to which we can identify markers of marginalisation, especially at an individual level. While attempts were made to collect critical information directly from girls during interviews at schools, the information they provided was not always reliable, and some questions were either too complex or too sensitive to ask of girls directly.[84] At the midline, household surveys will be conducted with all cohort girls, which will provide additional insight into the extent of marginalisation among this population.

3.4 Intersection between key characteristics and barriers

Table 20 provides a cross-tabulation of barriers and cohort girl characteristics, drawn from the previous sections on educational marginalisation. As the table shows, the demographic characteristics of a girl's head of household (i.e. their education level and gender) do not appear to be systematically related to the existence of barriers at the individual or school level. For instance, girls in female-headed households report similar levels of support for staying in school from the parents/caregivers as the full sample. They are no more likely respondents in the full sample to have teachers who are often absent, or teachers that make them feel unwelcome, although they are significantly more likely to be in a class where the teacher uses corporal punishment.

Barriers interact most significantly with girls' characteristics among those girls attending IDP schools. Girls attending IDP schools are more likely to report feeling unsafe while traveling to school, Focus group discussions with boys in Mogadishu further emphasise security concerns as "it's possible to face danger while they are walking in the streets".[85] The level of security is also variable across

[82] Unfortunately, the reported rates of physical punishment from direct observation appear very high; they also suggest that girls are more often physically punished than boys, which runs counter to the qualitative evidence – where many interviewees noted that teachers are unwilling to physically punish female students. It is possible that enumerators overstated the extent of physical punishment that they observed; even accounting for some degree of systematic bias, however, it appears that rates are higher than those reported by students in Table 19.

[83] Table 19 does not report barriers for the boys cohort, because data on these characteristics were not collected at the baseline from boys. In future evaluation waves, when the boys cohort is subject to household surveys, additional information will be available on barriers and characteristics that might impact their educational attainment.

[84] To illustrate, the evaluation team included a question for cohort girls regarding their migration status. However, in an attempt to simplify the question, its utility was reduced, such that we are not able to determine *when* a girl's family migrated, or if they migrated during her lifetime. Other questions addressed the occupation of the head of household, but may not reliable – younger students, especially, may not be fully aware of their parents' occupation, or how to describe it. Finally, other questions were deemed too sensitive for inclusion, including questions assessing whether a girls' parents are still living.

[85] FGD with Boys in Mogadishu.

districts: in “some [districts], you will find peace and security and in others, danger and bombing”.[86] Girls in IDP schools are also much more likely to report that their teacher uses corporal punishment on a regular or semi-regular basis than girls in non-IDP schools. Among girls in IDP schools, 57.1 per cent report that their teacher uses physical punishment, and 34.6 per cent report that their teacher used physical punishment within the last week. In contrast, just 45.3 per cent of girls in non-IDP schools reported that their teacher ever uses physical punishment, and 25.5 per cent report its use sometime during the last week. Importantly, these rates do not necessarily suggest that girls in IDP schools are physically punished more often than girls in non-IDP schools, because respondents were asked simply whether their teacher uses physical punishment at all, on any student in the class.[87] Girls in IDP schools are *also* more likely to report that they personally received physical punishment from their teacher in the last week, but this gap between IDP and non-IDP schools is much smaller: 17.3 per cent of girls in IDP schools personally received physical punishment within the last week, compared to 14.1 per cent of girls in non-IDP schools. The conclusion these findings suggest is that teachers in IDP schools are much more likely to use physical punishment generally, and that girls enrolled in IDP schools are *slightly* more likely to receive physical punishment at the hands of their teachers than girls in non-IDP schools.

Beyond physical punishment, students who attended IDP schools were also significantly more likely to report that there were not enough seats for all the students in the class. This is confirmed by the qualitative interviews, many of which stated that many people have now joined the school as they have had to leave their homes because of drought or conflict.[88] Most significantly, girls attending IDP schools are also much more likely – 59.3 per cent, compared to 27.4 per cent in the full sample – to live in an area that is drought-affected.

The results in this section are merely suggestive, for two reasons. First, we have only investigated the association between a few demographic characteristics and specific educational barriers. Second, the subsamples employed, especially in the case of girls enrolled in IDP schools, is fairly small.[89] Though tentative, the results highlight the fact that systemic obstacles to education – such as drought and conflict – are likely correlated with other metrics of marginalisation. To the extent that much marginalisation in Somalia stems from financial obstacles that prevent girls from completing a full cycle of education, it should not be surprising that factors that impact household finances (drought, migration, etc.) would be correlated with schools that are under-resourced or performing poorly.

---

[86] FDG with boys in Mogadishu.

[87] This distinction is important, because not all students receive physical punishment and there may be gendered aspects of who is punished physically, though the evidence on this question is mixed. From the qualitative data, some focus group participants suggested that girls were more likely to receive physical punishment than boys (FGD with boys in Sheikh); still, most of the qualitative evidence suggested the opposite – that girls were *less* likely to be punished physically (Focus Group Discussion with Male Students. Hargeisa, Somaliland; Key Informant Interview with Female Teacher. Galkacyo, Puntland; Key Informant Interview with Male Teacher. North Mudug, Puntland). The quantitative data do not provide much additional clarity: enumerators observing classrooms during the baseline were slightly more likely to report that girls were punished physically, compared to boys; however, male respondents to the in-school cohort survey were slightly more likely than female respondents to report that they had personally been physically punished within the last week (15.7 per cent versus 14.2 per cent). While we expect that teachers are somewhat more comfortable physically punishing boys than girls, the data do not allow clear and unambiguous conclusions in this regard.

[88] KII with Female Teacher. Sahil, Somaliland; FGD with Fathers, Maroodi Jeeh, Somaliland; FGD with Fathers, Sheikh, Somaliland.

[89] The set of girls attending an IDP school includes just 81 respondents.

TABLE 20: BARRIERS TO EDUCATION BY CHARACTERISTIC

Barriers:	HOH has no education	HOH is female	Attends IDP school	Full Sample
Parental/caregiver Support:				
Doesn't get support to stay in school and do well	2.6%	2.5%	1.2%	3.1%
School Level				
Doesn't feel safe traveling to school	3.9%	3.4%	4.9%	3.2%
No seats for all students	23.4%	25.9%	38.3%	20.1%
Disagrees teachers make them feel welcome	4.9%	2.8%	2.5%	3.4%
Agrees teachers treat boys and girls differently in the classroom	35.8%	32.0%	34.6%	36.1%
Agrees teachers are often absent from class	26.0%	23.8%	21.0%	24.0%
Teacher uses physical punishment	40.6%	46.2%	56.0%	25.3%
Community Level				
Drought-Affected	37.4%	24.5%	59.3%	27.4%
Conflict-Affected	13.0%	20.0%	14.8%	15.7%

3.5 Appropriateness of project activities to the characteristics and barriers identified

The design of EGEP-T interventions benefitted significantly from lessons learned from the first phase of EGEP programming. The experience of RI and its implementing partners has resulted in a set of interventions that are well-suited to addressing the barriers to education identified during this baseline.

In general, the barriers and characteristics identified above mirror those identified by EGEP-T and its beneficiary mapping. During and after the baseline evaluation, RI conducted an independent mapping of beneficiary characteristics and barriers. This effort included documenting total beneficiary numbers, referenced elsewhere in this report, as well as key characteristics of girls and households that are predictive of educational marginalisation, such as IDP status. This beneficiary mapping, in general, did not measure the same barriers as the EGEP-T baseline. Because the beneficiary mapping consisted of collecting data from school administrators, household variables were not collected and emphasis was primarily on school-level characteristics.

The one indicator the beneficiary mapping and the baseline had in common was enrolment at an IDP school.[90] In the baseline, only 5.0 per cent of girls in the raw, unweighted sample attended an IDP school. Similarly, the beneficiary mapping indicated that 6.0 per cent of girls attended an IDP school.

[90] Other factors, such as the impact of the drought and conflict, were derived from RI's own research, rather than that of the external evaluator. As a result, RI's beneficiary mapping does not provide an independent check on the distribution of these characteristics within the baseline sample, because the distribution within the baseline sample was derived directly from RI's beneficiary mapping.

At the same time, 12.7 per cent of all EGEP-T schools are IDP schools. The disjuncture between the share of IDP schools and the share of students enrolled in those schools reflects the fact that IDP schools tend to be smaller, on average, than non-IDP schools. Nonetheless, the similarity, with regard to IDP status, between the sample and the underlying EGEP-T project population suggests that the sample is capturing an appropriate cross-section of beneficiaries.

Beyond IDP status, many of the most common barriers identified by the evaluation are explicitly targeted by EGEP-T activities. We organise our discussion below in terms of specific "families" of barriers – the resource constraints facing schools, financial constraints facing households, girls' relative comfort at schools, teaching quality and school management, and drought. While EGEP-T has identified additional barriers, we focus our attention on those which the evaluation was able to identify empirically with the data collected.

Previous analysis showed that many schools lack learning and teaching resources necessary for students to succeed. For instance, 20.1 per cent of girls in the main learning cohort attend schools without a sufficient number of desks for all students. Teachers also report a lack of necessary materials – 36.9 per cent state that they do not have the basic supplies necessary for them to teach effectively. Community members also report inadequate teaching and learning resources at schools and the need for more supplies.[91] A number of EGEP-T activities should enable schools to purchase more supplies and otherwise reduce financial constraints on schools – EGEP-T will provide teachers and classrooms with teaching and learning resources, and EGEP-T will engage in capacity-building among CECs. The latter activity should increase the ability of CECs to mobilise resources from among their communities and elsewhere, which qualitative interviews suggest is a goal of CECs, but one which they often have trouble achieving.[92] In addition, EGEP-T is targeting drought-impacted schools with cash grants, and training CECs in grant management, which should ease resource constraints in many of the most disadvantaged schools.

Financial constraints at the household level also figure prominently in the analysis of barriers and educational marginalisation. A wide range of results from the evaluation indicate that community attitudes are broadly supportive of girls' education, in general and in the abstract. There is somewhat less support for girls to complete the full cycle of education but, even here, attitudes themselves are not the primary barrier to school completion. Caregivers, parents, and other community members express widespread support for girls completing even higher levels of education. However, when faced with resource constraints at the household level – common for Somali families – girls' education often takes a backseat to other household needs.[93] As a result, easing the financial burden on households generally, and the burden of educating girls specifically, is critical. Many EGEP-T interventions are geared toward this end, especially providing bursary support,

[91] FGD with mothers, Somaliland; KII with male teacher, Galmudug; FGD with CEC members, Somaliland.

[92] FGD with CEC members, Somaliland; FGD with CEC members, Puntland.

[93] We discuss this distinction – between support, in principle, for girls' education, and support in practice, when faced with resource constraints or difficult decisions about how to spend limited household funds – in Section 5.5. In general, the data reported in Section 5.5 suggest that adults in EGEP-T communities support girls' education in principle and have high aspirations for their daughters to complete their education through secondary school and beyond, but some are also willing to sacrifice their daughters' education when unexpected circumstances arise. For instance, when asked to consider a hypothetical scenario in which their sister was sick, 21.9 per cent of respondents said that they would withdraw their daughter from school – rather than sell household goods – in order to help their sister. In another scenario, in which their daughter has an offer of marriage, 12.6 per cent of respondents would encourage their daughter to get married and *not* continue their schooling.

sanitary kits, solar lamps, and uniforms to severely marginalised girls.[94] By reducing the direct (fees-based) and incidental (uniforms, etc.) cost of girls' school enrolment, EGEP-T should reduce the impact of financial constraints on enrolment rates.

The previous analysis also suggested that discomfort at school may factor into poor educational performance among girls. Although the vast majority of respondents indicated that their teacher made them feel welcome in class, nearly half of respondents simultaneously stated that they are afraid of their teacher. The contradictory nature of the results aside, this finding suggests that many students, especially girls, feel uncomfortable in class. Qualitative data reinforced this view repeatedly, with teachers, school administrators and girls alike stating that more female teachers were needed, to ease discomfort among female students.[95] Specifically, interviewees point out that girls need female role models, and that female students may be hesitant to ask questions of male teachers. EGEP-T activities designed to address this barrier include the provision of psychosocial support, and sponsorship of girls' clubs (and boys' clubs), which should increase girls' sense of community and support at schools. Even more targeted, EGEP-T will also train female teacher mentors, who should, ideally, provide the precise female role models that interviewees identified as lacking.

The analysis above also identified teaching quality and school management as barriers to student achievement, albeit loosely. For instance, around one-quarter of students report that their teacher is often absent for a lesson – a finding which we corroborate using reports from head teachers themselves in Section 5.3. A large proportion of teachers also employ corporal punishment. A number of other indicators of poor teaching quality and school management – not addressed specifically above – are documented extensively in Sections 5.3 and 5.4, where we show that the use of participatory and learner-centred pedagogical techniques is limited, and that many schools and CECs fail to maintain basic records and exercise limited oversight of teachers.

EGEP-T interventions were designed with both teaching quality and school management in mind. For instance, a key activity for EGEP-T is continuing professional development for teachers, focusing on subject content training, and inclusive and gender-responsive teaching techniques. Numerous teachers and CEC members requested additional training for teachers explicitly during qualitative interviews, highlighting the weak qualifications of many teachers in project schools.[96] To the extent that teaching quality is related to the availability and quality of teaching resources, EGEP-T efforts to provide teaching and learning resources to schools should also improve teaching quality. Finally, EGEP-T intends to conduct capacity-building among both CECs and MoE officials, both of which should increase accountability at the school level and improve school management.

One of the largest "unknowns" in EGEP-T's Theory of Change is the role that the drought has and will play in shaping student learning (*Note:* More information can be found in Section 6). Drought

[94] Sanitary kits and solar lamps address both financial and non-financial burdens to attendance. The EGEP endline evaluation documented girls' use of solar lamps to allow them to study. By giving girls the ability to study after dark, girls are somewhat less constrained by the decision of whether to complete their required housework or study, though the housework burden on girls is still disproportionate and impinges on their ability to learn (see, e.g.: FGD with boys, Somaliland; FGD with boys, Galmudug; KII with MoE official, Puntland). Similarly, sanitary kits present a financial obstacle for households, and menstruation is widely seen as both a barrier to girls' attendance and a source of discomfort for them at school (see, e.g.: FGD with mothers, Somaliland; FGD with CEC members, Somaliland; KII with MoE officials, Somaliland).

[95] See, e.g.: FGD with CEC members, Puntland; KII with female teacher, Puntland.

[96] KII with male teacher, Galmudug; FGD with CEC members, Somaliland; KII with MoE official, Puntland.

presents an overarching barrier that is currently impacting student performance and could continue to affect project outcomes over the next two years. Under the best conditions, much of Somalia will remain food-insecure; if the drought continues or intensifies, it will significantly hamper efforts to improve girls' education in the region. Drought exposure is correlated with other barriers, such as IDP status: teachers in drought-affected schools are more likely to indicate that they lack basic teaching supplies and are less likely to have completed schooling beyond secondary school. Most telling is the fact that drought is correlated with lower learning outcomes, as shown in Section 5.4. EGEP-T activities currently target drought-affected schools with an array of support, including psychosocial support for students and teachers, cash grants, and food rations, among other things. As much as possible, these efforts should dampen the impact of the drought on educational outcomes, the effects of which are likely to be transmitted through additional financial constraints on families, pressure to migrate, and a decline in teacher retention and attendance, among other channels.

The project provides a full response to the findings in this section, and in the report more generally, in Annex 13 of the report.

4. Findings – Core Outcomes

4.1 Learning Outcomes

The Girls' Education Challenge targets three outcomes: learning, in terms of numeracy and literacy, and transition from primary school to secondary school and beyond. In this section, we describe in detail the baseline levels of learning in EGEP-T project schools, establishing targets for later evaluation waves, and discussing foundational gaps in learning that were observed during the baseline.

Assessment Design

As with other GEC projects, 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, though this official language policy is not closely followed.[97] Given this complexity – and the fact that the project targets girls in both primary and secondary school – separate assessments were conducted in English and Somali. Importantly, by assessing even secondary school-age students in Somali, the project will be 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.

The project's three learning assessments were designed in conjunction with CARE, who are also implementing a GEC-funded project ('The Somali Girls Education Promotion Programme – Transition') in Somalia. The coordination between RI and CARE allowed for greater expertise to be brought to bear on the design – including Monitoring and Evaluation and technical education staff from each organization – and made piloting the tests more cost-effective. The tests were designed

[97] In the baseline sample of 140 schools, six were taught in Arabic. Moreover, the distinction between school levels is not adhered to closely: 22.1 per cent of the primary schools (exclusively primary, rather than mixed primary/secondary) sampled reported using English as the language of instruction, and 25.9 per cent of the secondary schools sampled reported using Somali.

under guidance from the GEC FM for the development of Secondary Grade Reading and Maths Assessments (SeGRA and SeGMA, respectively), mirroring the structure of the Early Grade Reading and Maths Assessments (EGRA and EGMA). The assessments test the following general skills (full-text versions of the assessments are provided in Annex 7):

NUMERACY

- Subtask 1: Identifying missing numbers
- Subtask 2: Addition (level 1)
- Subtask 3: Subtraction (level 1)
- Subtask 4: Addition (level 2)
- Subtask 5: Subtraction (level 2)
- Subtask 6: Addition and subtraction word problems
- Subtask 7: Multiplication (level 1)
- Subtask 8: Multiplication (level 2)
- Subtask 9: Division (level 1)
- Subtask 10: Division (level 2)
- Subtask 11: Multiplication and division word problems

SOMALI LITERACY

- Subtask 1: Reading fluency or word recognition using high-frequency words
- Subtask 2: Reading comprehension (difficulty: easy)
- Subtask 3: Reading comprehension (difficulty: medium)
- Subtask 4: Reading fluency or word recognition
- Subtask 5: Reading comprehension (difficulty: hard)
- Subtask 6: Writing (fill in missing words)
- Subtask 7: Writing (convert sentence to negative form)
- Subtask 8: Writing (convert sentence to future tense)

ENGLISH LITERACY

- Subtask 1: Letter identification
- Subtask 2: Reading fluency or word recognition
- Subtask 3: Reading comprehension (difficulty: easy)
- Subtask 4: Reading fluency or word recognition
- Subtask 5: Reading comprehension (difficulty: medium)
- Subtask 6: Reading comprehension (difficulty: hard)
- Subtask 7: Writing (fill in missing words)
- Subtask 8: Writing (convert sentence to negative form)
- Subtask 9: Writing (convert sentence to future tense)

Each subtask comprised a set of individual items, ranging from one to ten per subtask. Piloting of the assessments took place in September 2017, with 310 randomly-selected students in grades G5, G7, G8, Form 1 and Form 3. Based on the results of the pilot, RI and CARE jointly revised the assessments. Particular care was taken to prevent ceiling effects, which were common during evaluations in EGEP's first phase (i.e. GEC-1).

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 in that subtask. The aggregate score is the mean of the subtask scores. As an example, a subtask with 5 individual test items, each of which results in a binary correct or incorrect

scoring, would be scored as the percentage of items answered correctly (i.e. four out of five items results in a subtask score of 80 per cent). These subtask-specific scores are then averaged to arrive at a final aggregate score for the assessment. This ensures that subtasks with many questions are not over-weighted in the final score.

Assessment Validity

Based on the data collected in the baseline evaluation, the assessments appear to be well-designed and were generally well-implemented. First, the assessments have strong *prima facie* validity as assessments of learning, because measured outcomes are strongly correlated with grades. Among all three major samples of students who were administered learning assessments – cohort girls, cohort boys, and bursary girls – learning scores increase monotonically (i.e. stepwise) with grade, as shown in the primary results tables in the next section. Table 21 illustrates this trend using an aggregation of both boys' and girls' scores, with a total sample size of 2,073 respondents – at every grade level, scores increase over the previous grade.

TABLE 21: MONOTONICALLY INCREASING LEARNING SCORES, BY GRADE

Grade	Numeracy	Somali Literacy	English Literacy
G6	62.3	67.1	28.2
G7	69.1	77.3	34.7
G8	75.7	80.9	44.3
F1	82.5	86.3	62.6
F2	84.0	89.1	66.5

*Table aggregates learning scores across the main samples of cohort girls and cohort boys

Readers will note that, in Table 21, we report student scores for Somali and English literacy separately. Throughout most of this report, we follow a similar tack.[98] Although a single, unified literacy score would be easier for readers and project staff to digest, discussions between the evaluation team and RI's Monitoring & Evaluation team prompted the separation of English and Somali scores. Two factors drove this decision: first, English and Somali performance are only loosely related from a conceptual perspective. While Somali is the predominant mother tongue – and true literacy is therefore best assessed in Somali – learning English is functionally equivalent to students learning a second or third language. There is no guarantee, or even expectation, that a students' performance across the two subjects should be linked.

Second, the empirical evidence suggests that English and Somali are only loosely correlated with one another. The left panel of Figure 4 plots Somali and English literacy scores among cohort girls and cohort boys (n = 2,007), where the orange fitted curve plots the association between the two scores.[99] The right panel of Figure 4 plots Somali literacy scores against numeracy scores among the same sample of respondents. On one hand, the results in the left panel exhibit a strong correlation between English and Somali literacy. On the other hand, this correlation is actually weaker than that between Somali literacy and numeracy.

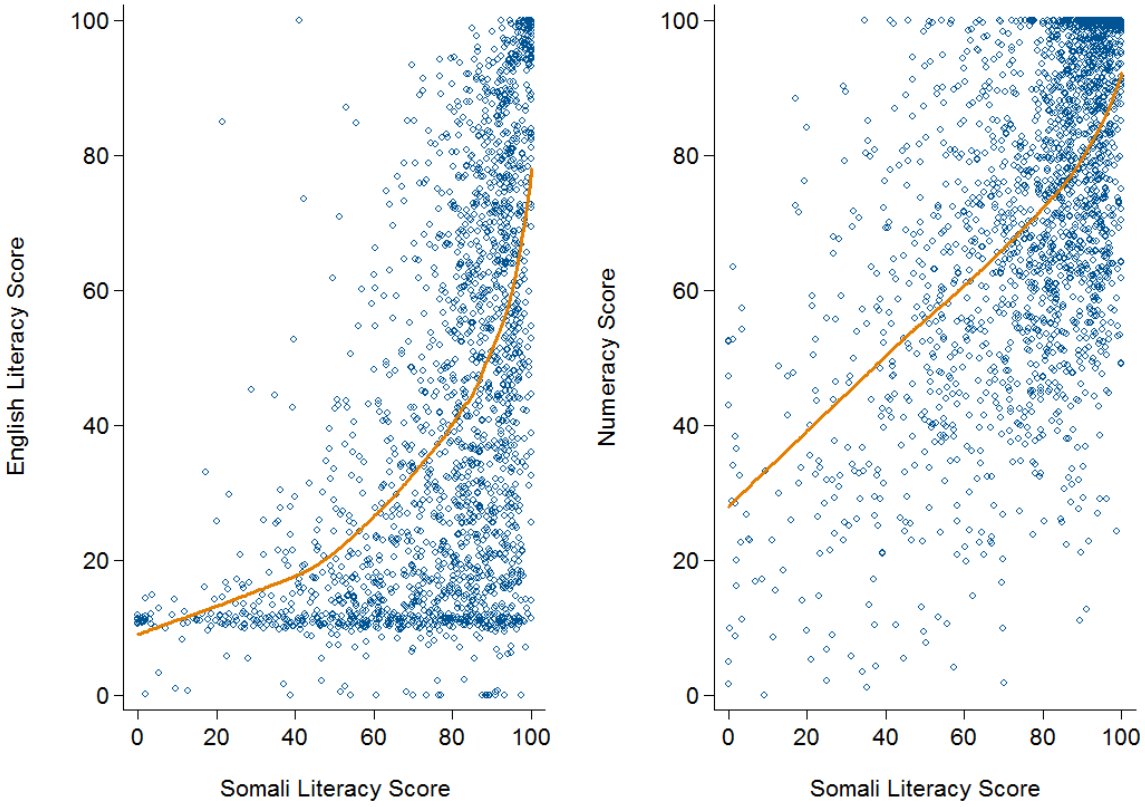
[98] The major exceptions are analyses in which *all* learning assessments, including numeracy, are aggregated into a single learning score. This occurs primarily when we study the correlation between intermediate outcomes and learning.

[99] The orange line in each graph is the locally-weighted smoothing (LOESS) curve, which plots a local regression of the relationship between Somali and English literacy at each level of Somali literacy.

In fact, numeracy exhibits a stronger relationship to both English and Somali literacy than either of these assessments do to one another. The correlation results below illustrate this point:

Numeracy – Somali: 0.56
Numeracy – English: 0.56
Somali – English: 0.51

FIGURE 4: CORRELATION BETWEEN SOMALI LITERACY, ENGLISH LITERACY, AND NUMERACY SCORES

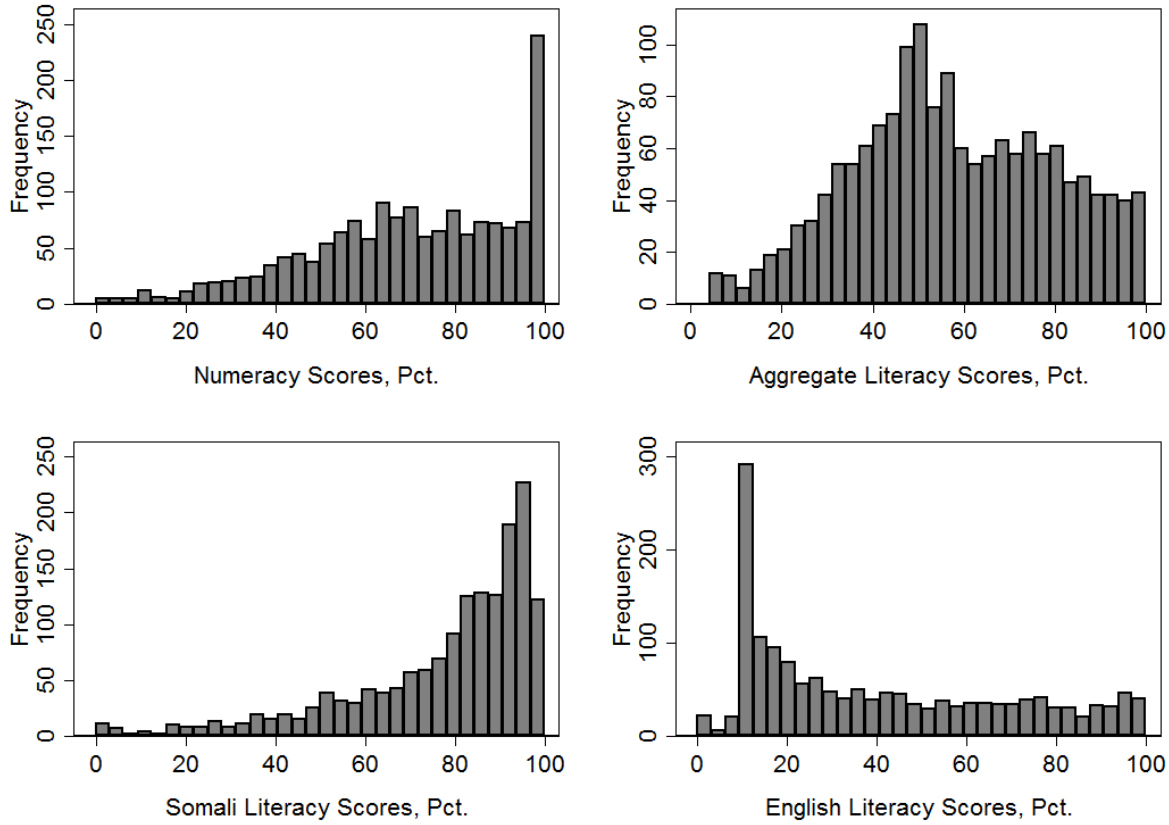


While a student who scores well on their Somali assessment is likely to score well on their English assessment, their relatively weak correlation does not suggest that performance on the two tests is linked by any special relationship. If they were, we would expect their correlation to be higher than that between the literacy and numeracy tests. In short, numeracy is a better predictor of English performance than is Somali, and numeracy is a better predictor of Somali performance than is English.[100]

Because the two literacy assessments are testing fundamentally different skills – literacy in one's mother tongue, versus acquisition of a second language – we recommend treating them separately for the purposes of evaluating EGEP-T's impact on learning outcomes. Where feasible, we analyse English and Somali literacy separately in this baseline evaluation; we also provide benchmarked values for both English and Somali, to facilitate comparisons at the midline and endline.

[100] The correlations reported remain even if we split the sample by school level. Since English is use more intensively at the secondary level, we might expect the correlation between Somali and English performance to be more pronounced at the secondary level. It is not – among both primary and secondary students, numeracy is a better a predictor of both English and Somali performance.

FIGURE 5: DISTRIBUTION OF LEARNING ASSESSMENT SCORES



The single weakness of the assessment design concerns ceiling effects. Figure 5 provides the distribution of scores among the full sample of cohort boys and cohort girls for the numeracy and literacy assessments, and for English and Somali literacy components of the literacy test separately. There is no evidence of ceiling or floor effects in scores for English literacy. While there are a number of very low scores – approximately one-third of the sample scored below 20 per cent – very few respondents scored 0 per cent. Somali literacy scores were much higher, on average, and exhibit mild ceiling effects, with 3.2 per cent of respondents achieving a perfect score. While this is not a significant concern at the baseline evaluation, it indicates that ceiling effects may pose problems in the midline and endline; as student learning increases, there may not be sufficient room for high-level learners at the baseline to exhibit improvement.

The most significant ceiling effects are found in the numeracy assessment (Figure 5, top-left panel). In this assessment, 11.8 per cent of students achieved perfect scores at the baseline, and over 20 per cent achieved scores of 95 per cent or higher. Ceiling effects of this magnitude could limit the ability of evaluators at the midline and endline to uncover positive project impacts in numeracy.[101]

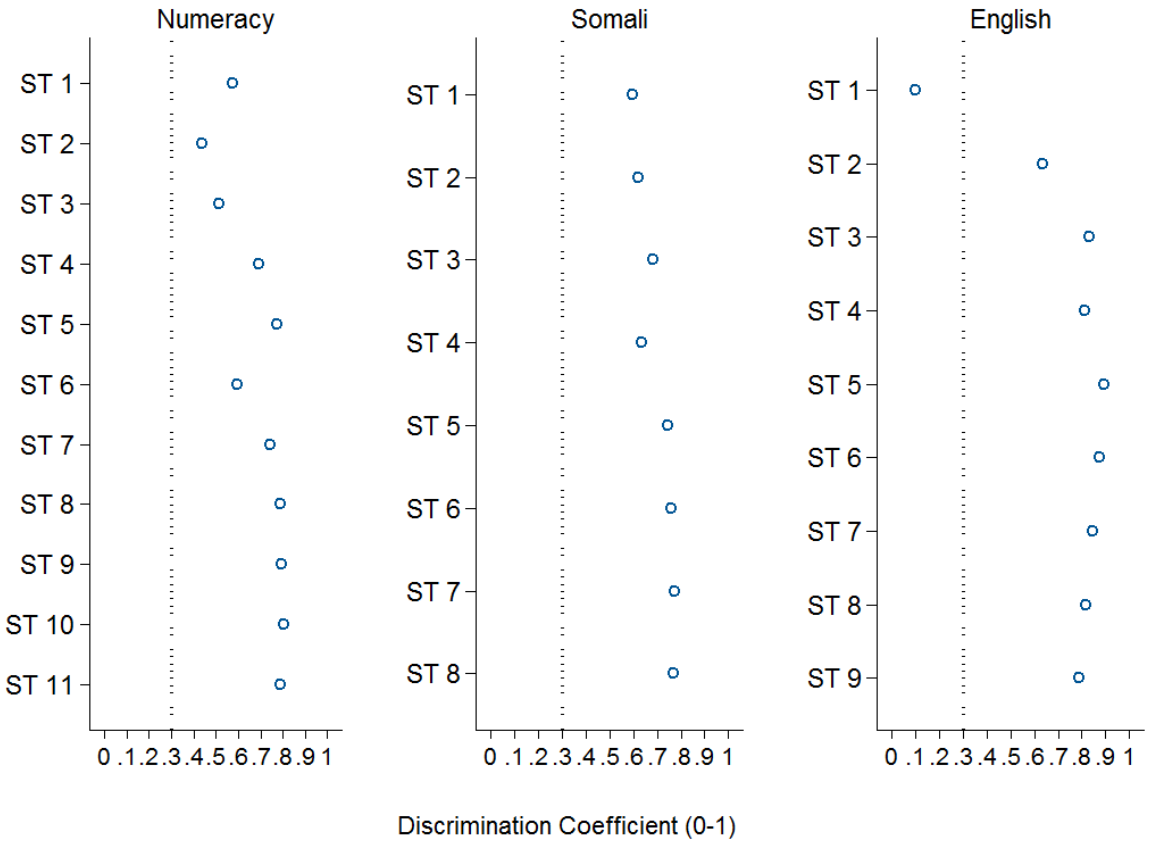
[101] While we are not aware of any standardized method for assessing the effect of ceiling effects on project impacts, the intuition is clear: if we assume an arbitrary 5 per cent "true effect" of the project on numeracy, uniformly distributed across the population, then this effect can only be observed for students who scored at or below 95 per cent at the baseline. In our sample, a 5 per cent "true effect" would translate into an observed 4.15 per cent effect in the sample, once we factor in the right-censoring that occurs among students who scored above 95 per cent at the baseline. By contrast, the potential for right-censoring in the Somali literacy assessment is smaller – a 5 per cent true effect, uniformly distributed across all respondents, would translate into a 4.6 per cent effect in the sample.

Unfortunately, there are no ready-made solutions for ceiling effects; however, we provide some suggestions for the midline and endline evaluation at the conclusion of this report.

Beyond floor and ceiling effects, we also evaluated each assessment in a classic item analysis framework. Specifically, we reviewed individual test items for their ability to discriminate between low- and high-performing students, calculating a discrimination index for each test item. The discrimination index measures the correlation between providing a correct answer on an individual test item and achieving a higher overall score on the exam.[102] The fundamental idea is that a test question which elicits correct answers from otherwise high-performing students and elicits incorrect answers from low-performing students is able to *discriminate* between the two groups – the ultimate purpose of assessment. In Annex 9, we provide the discrimination index for each individual test item in both assessments. Figure 6 summarizes these findings by subtask. A common metric for the type of discrimination index presented in Figure 6 is that an item is effective if its index score is above 0.3. In Annex 9, we show that nearly every individual test item met this standard – only 3 of 112 test items fell below the 0.3 rule-of-thumb.

At the level of subtasks (i.e. groups of individual items, grouped by the type of skill required), Figure 6 shows that most subtasks had very high discrimination levels. Only one subtask fell below the 0.3 threshold, and most items achieved discrimination scores above 0.7.

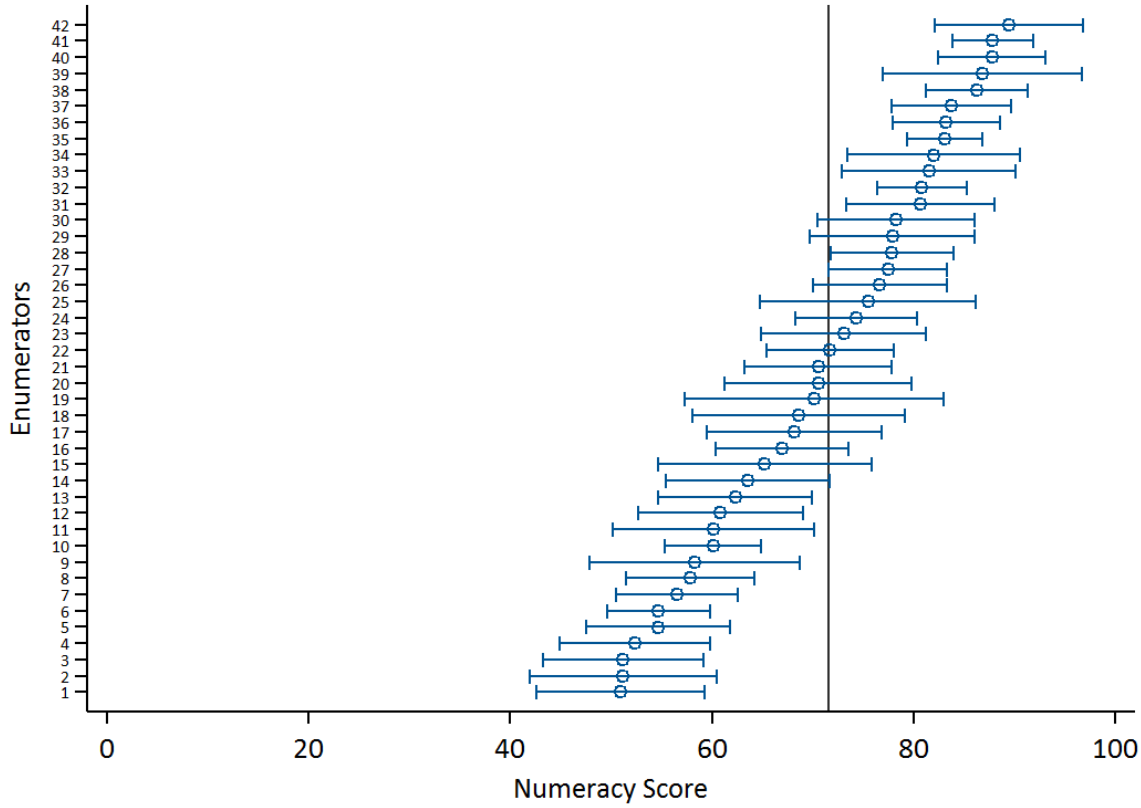
FIGURE 6: SUBTASK DISCRIMINATION SCORES FOR NUMERACY AND LITERACY ASSESSMENTS



[102] See Annex 9 for more details on the calculation of the discrimination index.

In addition to calculating discrimination indices, we performed a number of checks on the quality of the collected data. That is, while the discrimination index focuses on the quality of assessment design, we also checked the quality of data collection efforts. Assessments can be sensitive to the performance of individual enumerators for any number of reasons. For instance, enumerators might not follow instructions precisely, allowing students additional time, or confusing students with unclear instructions. The baseline data provide a strong test for such enumerator effects, because most enumerators completed assessments with 20 or more students. Such a large sample would allow us to detect if students being assessed by a single enumerator consistently under- or over-performed.

FIGURE 7: ENUMERATOR EFFECTS – MEAN NUMERACY SCORES, BY ENUMERATOR



We calculated mean scores for each enumerator on each subject assessment, limiting the sample to those enumerators who assessed at least 20 students.[103] Figure 7 plots each enumerator's mean numeracy scores, along with 95 per cent confidence intervals. We relegate a similar graph of literacy scores by enumerator to Annex 9 because they show qualitatively similar results. As Figure 7 shows, enumerators clearly vary in terms of their students' performance, with mean scores ranging from 51.0 to 89.5 per cent. Importantly, variation by enumerators should be expected, because enumerators were assigned to regions with varying levels of student performance, and scores should be correlated within schools and within districts. However, the 95 per cent confidence intervals show that no single enumerator produces particularly extreme scores. That is, even in the case of the lowest- and highest-scoring enumerators, other enumerators obtained similar average scores. In

[103] This sample includes 42 enumerators, who account for 98.6 per cent of all assessments completed in the baseline.

this analysis, no enumerator stands out, suggesting that no individual enumerators were dramatically influencing student scores.[104]

Learning Outcomes

Aggregate learning outcomes for the three subjects are presented in Tables 22, 23 and 24, below. In each table, we report the mean score of the girls learning cohort by grade, followed by the standard deviation of scores in that grade. We also report the mean score and standard deviation, by grade, for the boys learning cohort. Note that the evaluation design of EGEP-T does not include a control group; therefore, no control group scores are presented. In total, the girls learning cohort consists of 1,609 girls, in Grade 6 through Form 2, across 140 schools; the boys learning cohort includes 398 boys in the same grades and schools. For the establishment of baseline values, we employ sampling weights – as described in the methodological discussion above – to ensure that each school is equally weighted in the analysis.[105]

TABLE 22: MEAN NUMERACY SCORES AND VARIANCES, MAIN GIRLS AND BOYS COHORTS

Grade	Girls Learning Cohort Mean	Girls Learning Cohort Std. Deviation	Boys Learning Cohort Mean	Boys Learning Cohort Std. Deviation
Grade 6	59.50	23.02	69.60	18.70
Grade 7	66.50	23.16	76.04	21.72
Grade 8	72.55	21.83	83.16	15.01
Form 1	79.22	19.44	90.82	10.65
Form 2	81.64	17.56	90.66	11.93
Aggregate	68.85	23.16	79.00	19.04

TABLE 23: MEAN SOMALI LITERACY SCORES AND VARIANCES, MAIN GIRLS AND BOYS COHORTS

Grade	Girls Learning Cohort Mean	Girls Learning Cohort Std. Deviation	Boys Learning Cohort Mean	Boys Learning Cohort Std. Deviation
Grade 6	65.80	25.99	71.44	23.14
Grade 7	76.31	19.66	78.55	20.06
Grade 8	80.14	18.75	81.94	17.21
Form 1	84.93	14.09	90.02	8.55
Form 2	88.36	10.51	89.81	9.28
Aggregate	76.39	21.64	79.71	19.65

TABLE 24: MEAN ENGLISH LITERACY SCORES AND VARIANCES, MAIN GIRLS AND BOYS COHORTS

Grade	Girls Learning Cohort Mean	Girls Learning Cohort Std. Deviation	Boys Learning Cohort Mean	Boys Learning Cohort Std. Deviation
Grade 6	26.27	22.10	34.04	24.80
Grade 7	31.78	23.06	38.93	26.58
Grade 8	41.45	27.30	48.14	27.04
Form 1	58.90	27.51	68.68	24.56
Form 2	63.95	26.99	67.87	22.43

[104] Beyond enumerator effects, we also analysed aggregate school effects, and checked enumerator effects for individual subtasks. None of these checks suggested concern about the quality of data collection.

[105] In analyses that report standard errors, confidence intervals, or p-values for formal hypothesis tests, we generally cluster standard errors at the school-level.

Aggregate	38.60	27.86	45.69	28.34

As expected, learning scores on all three subject assessments increase monotonically with grade level.[106] Of note, girls' learning scores lag behind those of boys in the same grade, across all grade levels and both subjects. In numeracy, girls' scores are approximately ten points lower than boys' scores in each year. A frequent theme that emerges from qualitative interviews with teachers and students is that girls' opportunities and learning performance have increased significantly in recent years and that – in many cases – girls' performance is now equal to that of boys.[107] Interviewees occasionally cite the fact that top performing students in recent years have included female students, as one MoE official noted: "this year, who has the highest mark and second one – both of them are girls".[108]

The stated performance of girls, according to interviewees – performing at the highest level, with girls among the top-performing students overall – runs counter to the *average* performance of girls and boys in EGEP-T schools, where the mean male student outperforms the mean female student. These findings are not necessarily contradictory; indeed, it is unsurprising that some female students are able to compete for the top scores in their respective areas. In fact, the data from the baseline evaluation exhibit a similar pattern: a small minority of both girls and boys achieved perfect scores on both the literacy and numeracy assessments. At the very highest level of achievement, girls are approximately equal to boys in this sample. But convergence at the top end of the score distribution masks divergence across the remainder of the distribution, where boys score over 10 points higher on numeracy and 5.4 points higher on literacy, on average.[109]

Interviewees who cite top students may not be aware of the pernicious, and persistent, gender gap in performance among the bulk of students, because this data is not readily available to community members or even, in many cases, to teachers. In addition, community members and educators who cite top-performing girls may be reflecting the lower expectations placed on girls, such that – when girls surpass standard expectations – their achievements are considered especially noteworthy. This trend, in which adults have lower expectations of girls and are therefore surprised when some girls achieve parity with boys, is natural, given the baseline levels of girls' education in Somalia, and the widely understood fact that girls in the region are underperforming in school. While the performance of girls at the top of the learning distribution is encouraging, the gender gap in overall performance documented in Tables 22, 23 and 24 above highlight the continued need for a focus on girls' education specifically.

[106] The sole exception is found among boys in Forms 1 and 2, where the mean score of older boys was slightly lower than that of younger boys on all three assessments. However, the regression from Form 1 to 2 among boys is sufficiently small that it can be reasonably attributed to sampling variance.

[107] KII with male teacher, Saxiil Somaliland; KII with female teacher, Mogadishu, Banadir; KII with Ministry of Education Quality Assurance official, Galkayo, Puntland; KII with male teacher, Bari.

[108] KII with Ministry of Education Quality Assurance official, Galkayo, Puntland.

[109] Taking the distinction between top-end performance and mean performance further, we can consider the gender gap at different percentiles: at the 95th percentile, the gender gap in total scores – the average of literacy and numeracy, on a 100-point scale – is just 2.9 points. As we move to lower scores, this gap increases – at the 75th and 50th percentiles, the gap is 6.5 points and 10.2 points, respectively. At the 10th percentile, the gap has widened even further, to nearly 20 points. The gender gap is not, in statistical terms, a "location shift", i.e. a uniform shift across the entire distribution; rather, the gender gap reflects a long left tail of female scores in the low end of the distribution.

The scores follow markedly different distributions, as shown in Figure 5 when discussing floor and ceiling effects previously. Aggregate literacy scores are distributed in an approximately normal pattern – see the top-left panel of Figure 5 – while scores on Somali and English components of the literacy assessment are left-skewed and right-skewed, respectively. Differential skewing by language is expected, because the majority of cohort girls are in upper primary school and have not yet shifted to an environment where the language of instruction is English.[110] No floor effects were found in either the aggregate literacy assessment or the English component thereof.

As noted previously, there are significant ceiling effects in the case of the numeracy assessment; in the context of the girls learning cohort, 173 girls (10.8 per cent of the cohort sample) achieved perfect scores, with these high-achieving girls occurring in every grade, but most heavily concentrated among Grade 8, Form 1 and Form 2.

The evaluation also assessed learning outcomes among three additional subpopulations. The first was bursary girls – girls who are being provided bursary support through EGEP-T, and who cover a wide range of grade levels. The evaluation surveyed and assessed 428 such girls, with the intention that they would be recontacted at the midline and endline, allowing evaluators to assess the differential impact of bursary support, while controlling for all other aspects of EGEP-T programming.[111] The second subpopulation was girls in Forms 3 and 4, who were assessed to assemble a benchmark learning level, against which some members of the learning cohort will be assessed. We discuss this group in greater detail later in this section. Finally, the evaluation included a sample of out-of-school girls, randomly sampled from communities surrounding EGEP-T project schools. This sample may be re-contacted in later evaluation waves to assess rates of re-entry into school, and the effects of such re-enrolment on learning outcomes. In the interest of setting baseline learning levels for bursary girls, we report their learning outcomes in Table 25 below. Note that benchmark girls – the second subpopulation mentioned – are omitted, because they are discussed in greater detail elsewhere in this report.

TABLE 25: MEAN ASSESSMENT SCORES, BURSARY GIRLS AND OUT-OF-SCHOOL GIRLS

Grade	Bursary Girls Numeracy Mean	Bursary Girls Somali Literacy Mean	Bursary Girls English Literacy Mean
Grade 6	59.07	67.16	25.33
Grade 7	62.24	67.68	29.24
Grade 8	72.35	80.03	39.23
Form 1	79.26	86.11	54.37
Form 2	77.96	87.65	61.02
Aggregate	67.26	74.58	36.52
Age	**OOS Girls Numeracy Mean**	**OOS Girls Somali Literacy Mean**	**OOS Girls English Literacy Mean**
11	7.21	4.95	11.23
12	4.69	9.65	10.76

[110] To the extent that the outcome of interest is aggregate literacy across both languages, the literacy assessment is well-designed and the distributions reported in this section suggest that there will be minimal problems of floor and ceiling effects going forward. To the extent that the outcome of interest is language-specific literacy, ceiling effects in the Somali component of the test may complicate the analysis and threaten the validity of the results, primarily by dampening the score improvements possible from baseline to endline.

[111] As discussed in Section 2.3, the sample of bursary girls targeted those girls receiving bursary support in grades 6-8 and forms 1-2. However, this does not represent the wider grade range targeted by EGEP-T's bursary intervention, which includes severely marginalised girls from grades as low as grade 2.

13	16.93	17.97	12.96
14	34.95	29.87	24.90
15	18.29	15.49	11.63
16	26.41	18.58	13.83
17	35.54	39.42	20.47
18	31.27	36.23	18.09
Aggregate	20.97	21.04	15.30

Table 25 demonstrates a few findings of interest regarding performance on the learning assessments across the different sample groups. First, performance is universally poor among OOS girls, who score well below their in-school peers at every age and in all three subjects. Even among OOS girls, scores tend to increase with age, though the results in the bottom panel of Table 25 exhibit an unusual peak in performance at age 14, a decline at age 15, and then continued increases beyond among 16- and 17-year old OOS girls. It is possible to over-interpret these patterns, however, as the number of OOS girls is small – just 137 girls in total – and extremely small for individual age groups. There are a mere 10 OOS girls aged 15 years in the sample; as a result, sampling variance is the most likely explanation for the unusual patterns observed among 14-18 year old OOS girls.

In addition, a comparison of bursary girls to their non-bursary cohort girl peers is illuminating. Bursary girls are selected on the basis of particularly severe marginalisation, but they attend the same schools as the main cohort of girls whose scores are reported in Tables 22, 23, and 24. On each assessment, bursary girls underperformed their non-bursary peers, typically by about 1-2 points. When averaged across all three assessment types, the gap between bursary girls and non-bursary cohort girls is 1.8 points, a difference that is statistically significant at the 10 per cent level.[112]

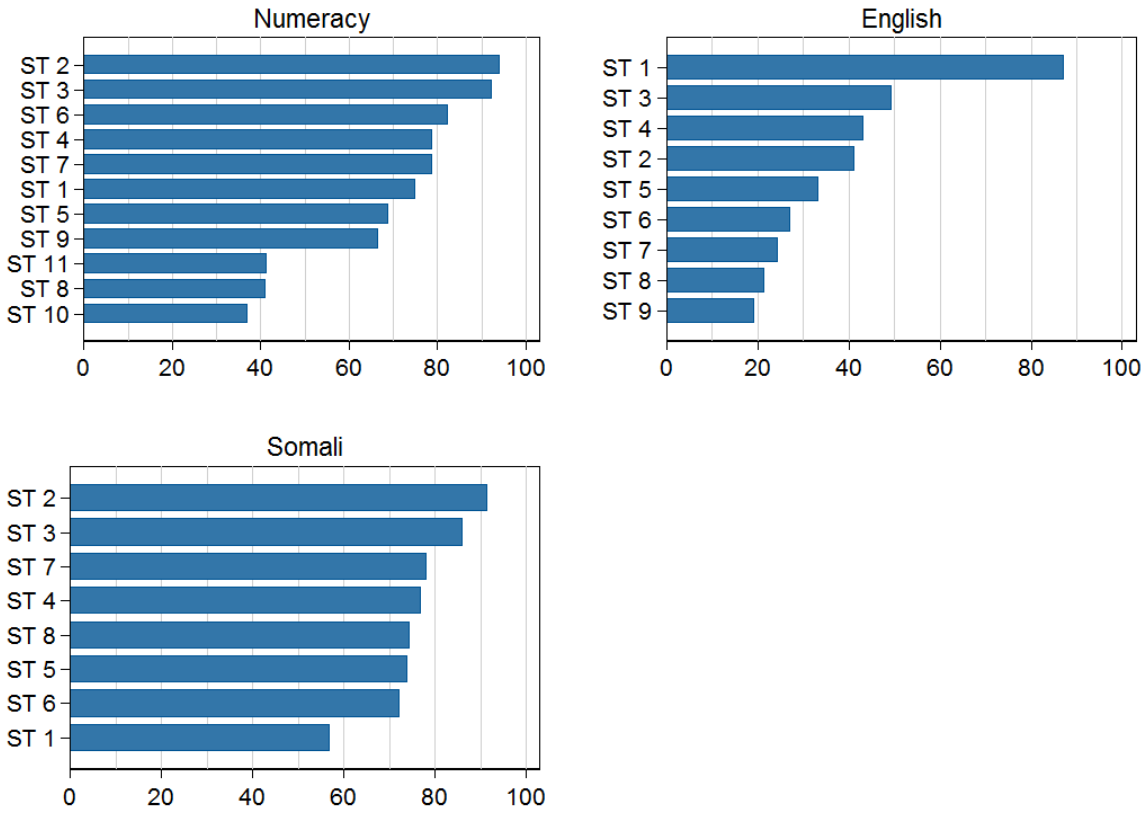
Foundational Gaps

The rich learning data collected at the baseline provide a good opportunity to investigate specific shortcomings in learning in Somalia. By identifying gaps in foundational numeracy and literacy skills, RI may be able to better target their efforts at developing students' skills by, for instance, encouraging focus on these skills during teacher training sessions, or relaying data on specific shortcomings to head teachers.

To start, Figure 8 plots mean scores across subtasks for each of the three learning assessments. The sample described in Figure 8 consists of the primary girls learning cohort, with a sample size of 1,609 girls. Two interesting findings emerge from Figure 8: first, girls perform markedly better on subtask 1 of the English literacy assessment – which assessed letter identification – than on any other subtasks. At the same time, performance on the first subtask of the Somali literacy assessment, which assessed *word* recognition in Somali, was quite low, representing the most difficult subtask of the Somali literacy assessment overall.

[112] P = 0.08. Note that the substantive gap of 1.8 points is the average gap across all three tests. The subject-specific gaps range from 1.59 points in numeracy to 2.09 points in English literacy, with bursary girls underperforming their counterparts on each.

FIGURE 8: MEAN SUBTASK SCORES, GIRLS LEARNING COHORT



In the remainder of this section, we investigate the nature of students' skills in more detail. Tables 26, 27, and 28 report performance of students on individual subtasks of the numeracy, Somali literacy and English literacy assessments, respectively.[113] Because each subtask tests a specific skill, performance on a given subtask is indicative of how well students have mastered that skill. To highlight gaps, students were classified into one of four groups based on the FM's guidance. Non-learners are defined as those students who received a score of 0 per cent on a subtask, indicating that they completely lack the skill tested. Emergent learners scored between 1 and 40 per cent on a subtask. Established learners scored between 41 and 80 per cent on a subtask, indicating a moderate level of proficiency. Finally, proficient learners scored between 81 and 100 per cent, reflecting proficiency or even mastery of a skill.

Before turning to a broader analysis of trends shown in these tables, we discuss findings related to individual subtasks. In general, the patterns of performance across subtasks are consistent with the nature and flow of the assessment, in which subtasks tend to become more difficult as the tests progress. For instance, in the numeracy assessment, there is a strong downward trend in mean scores and in the number of students achieving proficient status over the course of the assessment. This trend is not universal, of course – where the pattern is disrupted, as between subtask 8 and

[113] Note that the analysis presented in this subsection on individual subtasks utilises the full sample of learning assessments conducted in EGEP-T project schools. Namely, the sample includes the learning girls cohort, the learning boys cohort, and bursary girls. It excludes out-of-school girls who completed the learning assessments at their households. The goal of using this wider sample is to strengthen the conclusions that can be drawn regarding individual skills and their deficiencies in project schools.

subtask 9, it appears to be because a more difficult set of multiplication problems were provided before a set of less difficult division problems.[114]

One notable exception is that students performed very well across-the-board on simple addition and subtraction problems (subtasks 2 and 3, respectively), but performed poorly at identifying missing numbers in subtask 1. This suggests that additional attention should be paid to helping students identify missing numbers, especially among primary school students, where performance on this subtask was understandably lowest. At the same time, it suggests that some foundational skills – basic addition and subtraction, especially – are well-covered currently, and learning time could be dedicated to more advanced arithmetic.

As with numeracy, we make note of a prominent outlier in the subtask-level analysis of Somali literacy scores. While students were clustered into the highest proficiency categories in terms of basic and intermediate reading comprehension, relatively few students achieved proficiency in word recognition (subtask 1), which was considered the simplest task on the assessment. It is possible that the structure of subtask 1, which allowed students up to 60 seconds to read a passage in Somali, and assessed their reading speed, renders it incomparable to subtasks 2 and 3, which were more typical binary correct/incorrect test items.[115] Due to differences in scoring approach and response structure, we should be cautious in drawing firm conclusions from comparisons of these subtasks.

Relative to numeracy and Somali literacy, aggregate scores were dramatically lower for English literacy. This was expected both by the evaluation team and the project's Monitoring & Evaluation staff, as English is not the official language of instruction until students reach secondary school. As a result, only a small portion of the students in the sample have been exposed to intensive teaching in English; even those have only been exposed for between approximately 2 and 16 months.[116] This fact accounts for the findings in Table 28, which show very poor performance on all moderate and higher-level English subtasks. To ensure that the inclusion of primary school students did not obscure relevant foundational gaps in English literacy, we also performed this analysis on a restricted sample of only secondary school students. In general, the findings reported in Table 28 hold for the restricted secondary school sample as well.

[114] We do not claim this is a shortcoming of the assessment design. In fact, asking students to perform consecutive, increasingly difficult multiplication exercises before turning to division is a logical approach to assessment design.

[115] As the FM's scoring guidance notes, questions focused on words-per-minute (WPM) reading speeds have structurally different outcome distributions and have been forced to fit an arbitrary 0-100 scale to render them comparable to other subtasks. Students who read faster than 100 WPM are assigned scores of 100, right-censoring the data. It is possible that these differences account for the relatively weaker performance of students on reading comprehension or readingsubtasks (subtasks 1 and 4 on the Somali assessment and subtasks 2 and 4 on the English assessment), relative to other subtasks of comparable difficulty.

[116] Secondary school students in the sample are in Forms 1 and 2, and the assessment was completed around 2-4 months into the school year for most students.

TABLE 26: FOUNDATIONAL GAPS IN NUMERACY - SUBTASK-SPECIFIC NUMERACY OUTCOMES

Categories	Subtask 1 Missing numbers	Subtask 2 Addition (level 1)	Subtask 3 Subtract. (level 1)	Subtask 4 Addition (level 2)	Subtask 5 Subtract. (level 2)	Subtask 6 Word problems	Subtask 7 Multipl. (level 1)	Subtask 8 Multipl. (level 2)	Subtask 9 Division (level 1)	Subtask 10 Division (level 2)	Subtask 11 Word Problems
Non-learner 0%	0.61	1.31	2.23	5.60	11.77	4.43	7.98	40.41	19.71	46.62	50.49
Emergent learner 1%-40%	21.07	1.70	3.24	13.39	15.50	4.24	9.06	18.23	11.90	17.18	0.00
Established learner 41%-80%	25.56	7.44	6.23	24.18	28.73	29.58	25.64	17.46	19.90	13.18	16.06
Proficient learner 81%-100%	52.76	89.54	88.31	56.83	44.00	61.75	57.33	23.90	48.48	23.02	33.45
	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

TABLE 27: FOUNDATIONAL GAPS IN SOMALI LITERACY - SUBTASK-SPECIFIC OUTCOMES

Categories	Subtask 1 **Word recognition**	Subtask 2 **Reading comprehension (level 1)**	Subtask 3 **Reading comprehension (level 2)**	Subtask 4 **Word recognition**	Subtask 5 **Reading comprehension (level 3)**	Subtask 6 **Writing (missing words)**	Subtask 7 **Writing (convert to negative)**	Subtask 8 **Writing (convert tense)**
Non-learner 0%	1.18	1.77	3.47	3.18	9.00	13.87	15.98	18.45
Emergent learner 1%-40%	17.52	1.85	3.61	6.29	7.39	11.75	4.09	6.30
Established learner 41%-80%	70.61	16.18	23.96	33.18	31.62	17.59	9.03	8.41
Proficient learner 81%-100%	10.56	80.20	68.96	56.95	51.99	56.80	70.89	66.84
	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

TABLE 28: FOUNDATIONAL GAPS IN ENGLISH LITERACY - SUBTASK-SPECIFIC OUTCOMES

Categories	Subtask 1 **Letter identification**	Subtask 2 **Word recognition**	Subtask 3 **Reading comprehension (level 1)**	Subtask 4 **Reading comprehension (level 2)**	Subtask 5 **Word recognition**	Subtask 6 **Reading comprehension (level 3)**	Subtask 7 **Writing (missing words)**	Subtask 8 **Writing (convert to negative)**	Subtask 9 **Writing (convert tense)**
Non-learner 0%	5.36	18.86	37.14	36.43	51.97	62.50	64.47	73.32	76.98
Emergent learner 1%-40%	2.58	24.43	7.64	14.09	10.53	6.19	8.13	0.00	0.00
Established learner 41%-80%	6.83	45.93	21.04	23.25	17.97	14.37	13.03	10.29	7.64
Proficient learner 81%-100%	85.23	10.58	34.18	26.04	19.53	16.94	14.36	16.39	15.38
	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Beyond individual subtasks, one curious trend emerged in both the numeracy and English literacy assessments. In both assessments, as subtasks became more difficult, students were increasingly sorted into the two extreme categories – non-learners or proficient learners. As the numeracy results in Table 26 show, on the most difficult four subtasks, very few students fell into the emergent learner category. Rather, the distribution of results became bimodal, with students either scoring zero or scoring above 40 – or even 80 – per cent. In the extreme case of subtask 11, no students fell into the emergent learner category. We observe a similar pattern in English literacy scores, beginning with subtask 4, and becoming more noticeable in subtasks 6 through 9. Again, in the most difficult two subtasks of the English literacy exam, no emergent learners were found, out of a sample of 2,435 students.

To illustrate the bimodal nature of performance on these more difficult subtasks, Figure 9 plots two distributions of numeracy scores. The left panel is the distribution of total numeracy scores in this sample. The right panel is the distribution of total numeracy scores on subtasks 8 through 11. While performance on the overall assessment is not bimodal, performance on the more difficult tasks is.

FIGURE 9: DISTRIBUTION OF SCORES AMONG EASIER AND HARDER NUMERACY SUBTASKS

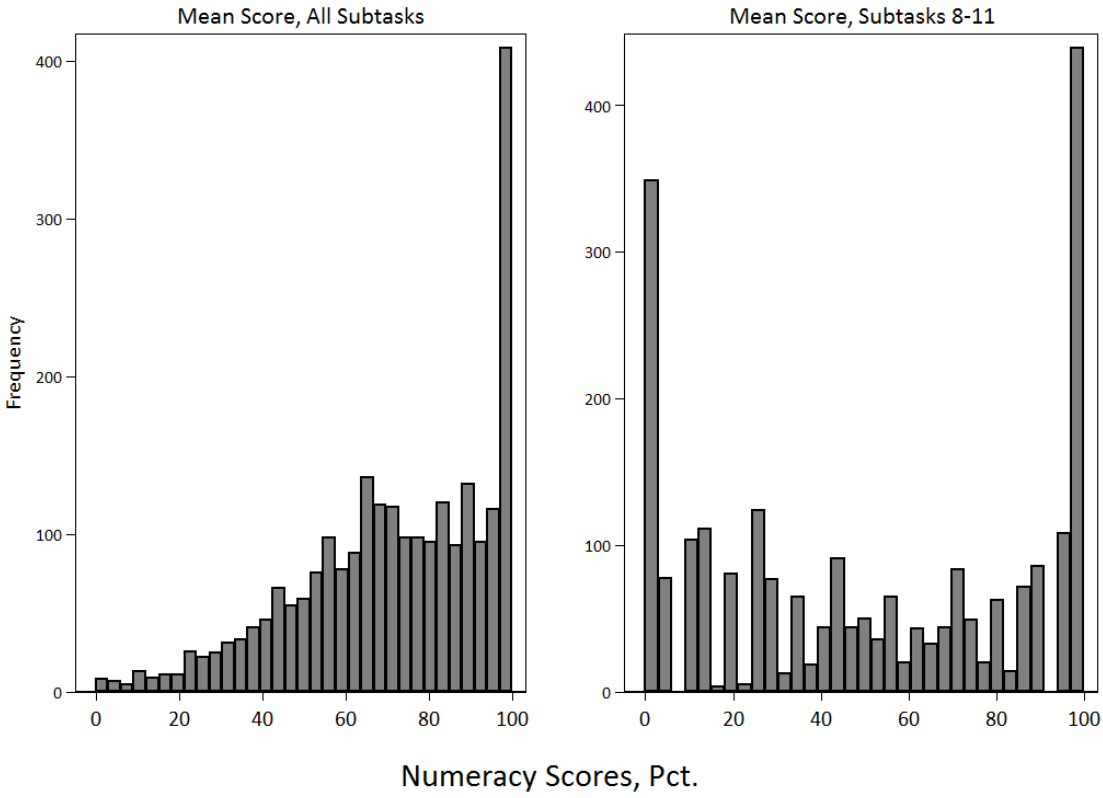


Figure 10 repeats this exercise for the English literacy assessment, but restricting the analysis to students in Forms 1 and 2 because – as outlined above – primary schools' language of instruction is Somali. Again, performance on the overall assessment, in the left panel, shows no sign a sharp bifurcation in student scores. However, performance on the most difficult four subtasks, whose mean aggregate scores are plotted in the right panel, shows a sharp division. On more difficult numeracy and English literacy test items, we find that students sort themselves into one of two

extreme categories – either they are unable to complete any of the test items successfully, or they are able to complete most or all of them. Very few students fall in the middle.[117]

Finally, Figure 11 repeats the analysis for Somali literacy. In contrast to numeracy and English literacy, neither easier nor more difficult Somali literacy tasks exhibit the bimodal distribution seen in the right panels of Figures 8 and 9.

Figures 9 through 11 illustrate a bimodal distribution in scores on more difficult English and numeracy subtasks. However, they do not indicate the full scope of the bifurcation between non-learners and high-achieving students, because they aggregate scores across multiple grade levels. It is possible that all or nearly all of the low-achieving students in such bimodal distributions are younger.

Extending our analysis to investigate the distribution of scores on difficult subtasks by grade level, it appears that, even *within* specific grades, there is a markedly bimodal pattern of scores on more difficult English and numeracy subtasks. For instance, among Form 2 students, 18.4 per cent earned zero points on the four most difficult subtasks of the assessment. At the same time, 20.8 per cent of students achieved a perfect score, while the remaining students were spread thinly across the rest of the score range. This sharp division between non-learners and high-achieving students *within the same grade* is particularly telling, because it shows that some students have fallen behind altogether when it comes to moderate-to-difficult skills in English literacy.

Substantively similar results obtain in the case of numeracy. Across grades, the share of students in the expansive middle of the score distribution – anyone who did not score a zero or a perfect score on the four most difficult numeracy subtasks – stays fairly constant at between 61 and 70 per cent of all students. What shifts as students get older is the share of students who earn zero and perfect scores, respectively, with older students being more likely to achieve perfect scores. However, even in Forms 1 and 2, a non-trivial percentage of students earn zero points on the numeracy assessment's four hardest problems: among Form 2 students, 4.8 per cent of students failed to earn a single point on this set of subtasks.

These findings imply – though they are merely suggestive – that significant gains could be made in student performance by focusing on students who may have low-to-moderate proficiency on easier skills (such as addition and subtraction) but who are unable to complete *any* higher-level tasks, such as division. Specifically, one potential approach would seek to first identify these students systematically within schools by employing a version of these assessments.

[117] Notably, we do not observe similar patterns with respect to Somali literacy.

FIGURE 10: DISTRIBUTION OF SCORES AMONG EASIER AND HARDER ENGLISH LITERACY SUBTASKS

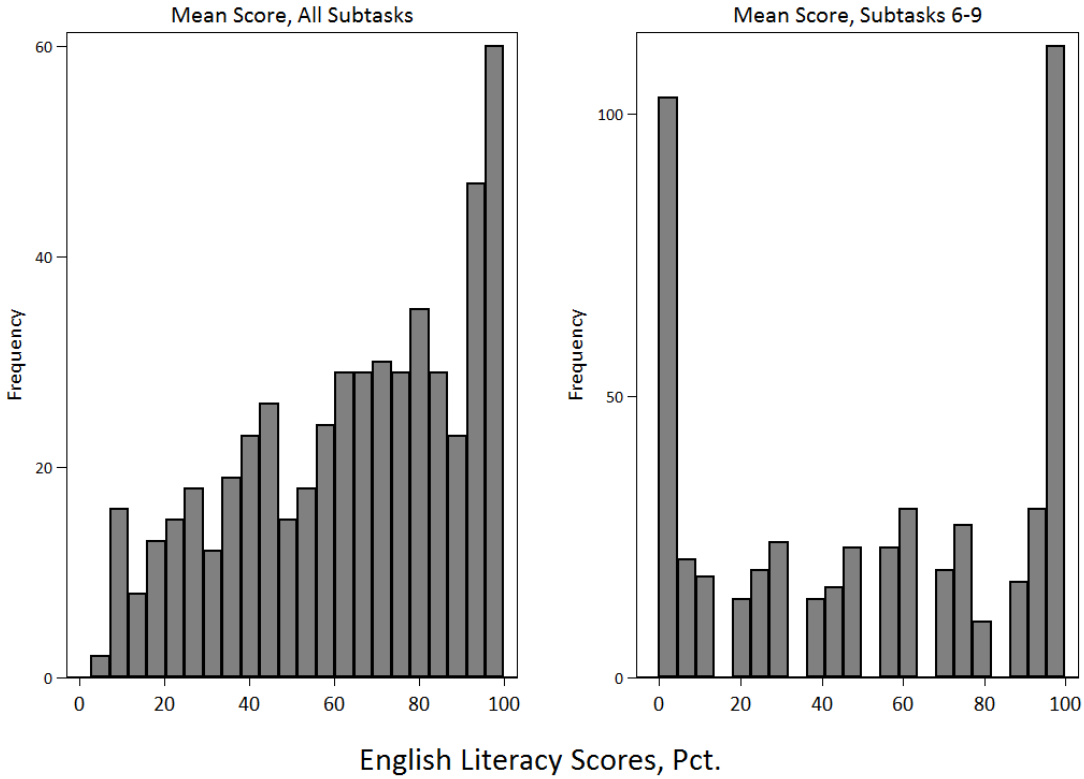
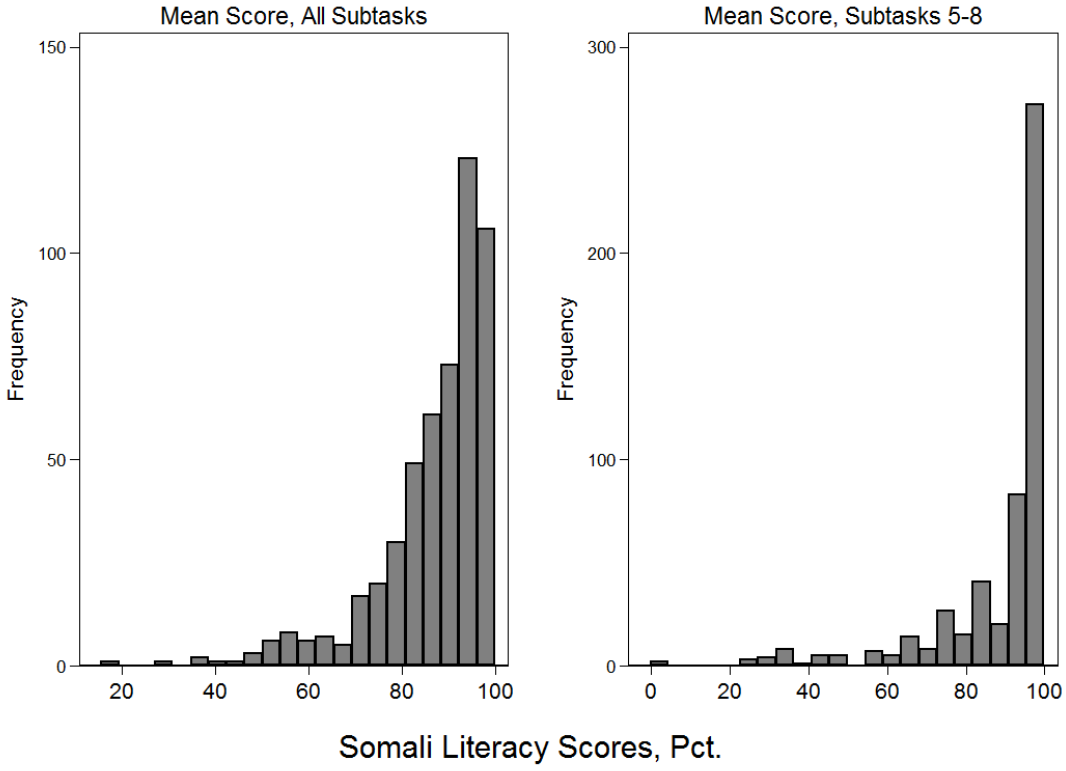


FIGURE 11: DISTRIBUTION OF SCORES AMONG EASIER AND HARDER SOMALI LITERACY SUBTASKS



Next, those students could be targeted for additional assistance with the foundations of higher-level skills, while more proficient students practiced those skills or moved on to harder topics. For instance, targeted students could be given extra instruction – at a conceptual level – in negative sentence construction and in the proper use of tenses, while students who have higher proficiency practiced these skills independently of the teacher's involvement. This approach takes advantage of the fact that, based on the findings above, once students move above a non-learner level on many of these tasks, they are able to acquire moderate levels of proficiency, rather than being stuck in the realm of "emergent learners."[118]

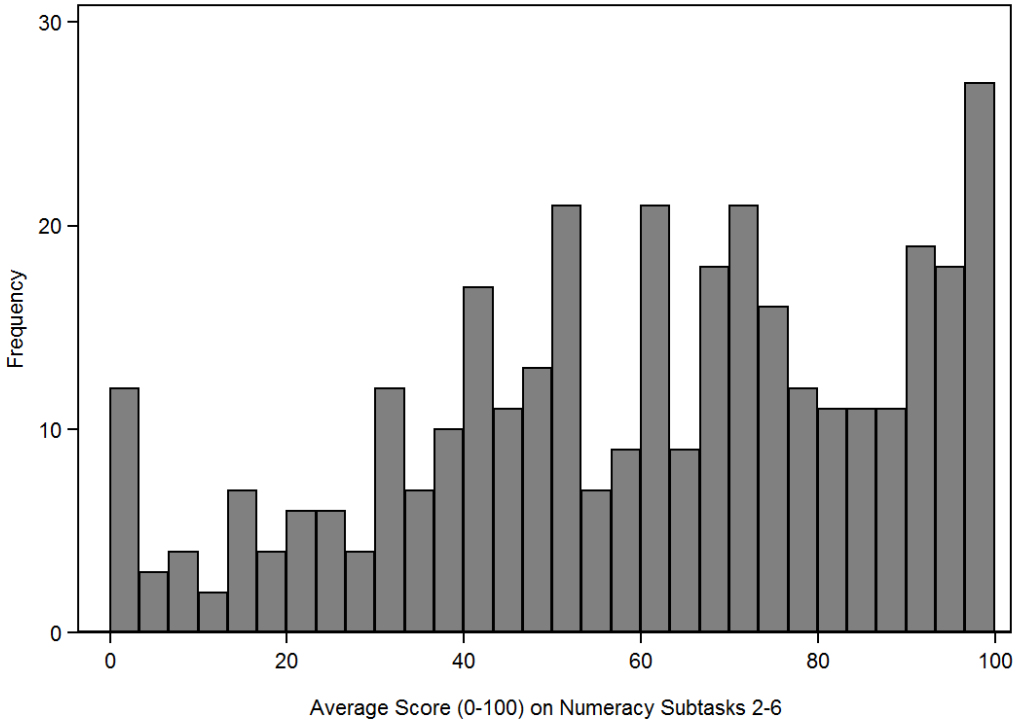
The logic of the approach outlined here is buttressed by the relatively weak correlation observed between easier and more difficult tasks on both the English literacy and numeracy assessments. That is, students' achievement on the easiest numeracy tasks (subtasks 2 through 6) is only moderately predictive of their performance on the harder numeracy tasks defined above. The correlation in students' scores between the easier and harder numeracy subtasks is just 0.57 – while this level of correlation implies a strong relationship, it is less predictive than we might expect in the realm of an assessment that tests similar skills.[119]

To illustrate the extent to which performance on the two portions of the numeracy test are not predictive of one another, Figure 12 plots the easier numeracy scores of students who earned zero points on the harder numeracy subtasks. In other words, we restrict our attention to students who scored zero points, overall, on the harder set of tasks, and plot their average scores on the easier tasks on the same assessment. The starkest finding from Figure 12 is the number of students who achieved a perfect score on the easiest numeracy subtasks but earned no points on the harder tasks. In fact, the bulk of students who earned zero points on the harder tasks actually performed reasonably well on the easier tasks – the mean score on the easier tasks among this group was 60.5 per cent, and the median score was 63.0 per cent. These findings indicate that many students are capable of improving their skill in mathematics – proven by their performance over the first half of the assessment – but lack even the most basic of skills necessary to complete harder tasks, such as multiplication and division. By targeting these students for remedial assistance, the project would be focusing resources on students who need additional help with the foundations, but who are likely to learn quickly once they have a basic grounding in the subject.

[118] One speculative explanation for this phenomenon is that the gap between non-learners and proficient learners on many skills is the understanding and internalization of rules-of-thumb. For instance, forming negative and future tense sentences, or completing division problems, are subject to rules-of-thumb that work in most cases. Once students know these rules and can apply them in one case, they are quickly able to achieve moderate proficiency by applying the rules, only struggling when the rules are occasionally violated or when special circumstances arise and more complicated rules or formulas are needed.

[119] We found a very similar correlation level between scores on the easiest English subtasks (subtasks 1 through 4) and the hardest English subtasks (subtasks 6 through 9). The correlation between students' scores on these two sections was 0.57 among secondary school students, who have the most exposure to English.

FIGURE 12: SCORES ON EASIER NUMERACY SUBTASK, AMONG STUDENTS EARNING A ZERO ON HARDER SUBTASKS



Grade Levels Achieved

This section reports the share of girls achieving each grade level of literacy and numeracy, based on an analysis of the curricula of Somaliland and Puntland. To understand the analysis undertaken and the results, it is important to describe the context of curriculum development in Somalia. 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.[120] 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).

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. Given that Somaliland and Puntland schools are subject to different standards, we attempted to adjudicate between the two as fairly as possible, being slightly conservative with

[120] 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.

respect to the grade at which children should achieve a specific skill.[121] Where skills were not specifically listed in either curriculum, we used our judgment to match them to specifically-enumerated skills in terms of their perceived difficulty.[122] The occasionally subjective nature of our mapping of skills to grades should be borne in mind. Finally, we focus on English literacy and mathematics only, due to the lack of information on Somali learning outcomes.[123]

TABLE 29: GRADE LEVEL STANDARDS FOR MATHEMATICS AND ENGLISH LITERACY

Grade Level Achieved	Mathematics Skills	English Literacy Skills
1	• Number identification up to 99 (portion of subtask 1) • Addition without carrying numbers (portion of subtask 2) • Subtraction without borrowing (subtask 3)	N/A
2	• Number identification up to 999 (portion of subtask 1) • Addition carrying one number (portion of subtask 2) • Addition with 3 digits, carrying up to 1 number (subtask 4) • Subtraction carrying one number (portion of subtask 5) • Addition and subtraction word problems with simple underlying arithmetic (subtask 6) • Multiplication of 1-digit numbers (subtask 7) • Division of 2-digit number by 1-digit number (subtask 9)	• Letter identification (subtask 2)
3	• Subtraction carrying two numbers (portion of subtask 5) • Multiplication of 2-digit numbers (subtask 8) • Word problems with simple multiplication and division (subtask 11)	• Identification of basic words, e.g., classroom objects, foods, animals (subtask 1)
4	• Division of 3-digit number by 2-digit number (subtask 10)	• Reading simple sentences (subtask 3 and portion of subtask 4)
5	N/A	• Reading low-medium difficulty

[121] For instance, if a child in Somaliland was expected to achieve a skill in Grade 3 and a child in Puntland was expected to achieve the same skill in Grade 4, we would classify the skill as being at a grade 4 level.

[122] As an example, neither curriculum available specifies when a child should learn to construct the negative form of a sentence. We consider this skill on par with the difficulty of constructing future tense sentences, which is a skill expected to be developed in Grade 6.

[123] It is also important to note that the available curricula are focused on primary-level outcomes; thus, our analysis is constrained to the achievements of grade levels 1-8.

		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

A full description of the curricula of Puntland and Somaliland, respectively, is provided in Annex 14. Table 29, above, describes the standards developed by the evaluation team for grade level achievement in mathematics and English literacy. In cases where no standard is described, the learning assessments utilised in the evaluation did not include a skill specific to that grade level. To be clear, this means that – based on the imperfect analysis and mapping conducted here – no subtasks on the numeracy assessment tested skills considered necessary for achieving grade 5 or higher numeracy levels. Likewise, no English assessment subtasks were sufficiently difficult to qualify a student as achieving a grade 7 or 8 level of English performance.[124]

In cases in which a subtask is distributed across two grade levels, we distinguish between less and more difficult portions of the subtask, on an item-by-item basis. In order to achieve a given grade level, a student must achieve a score of approximately 80 per cent on subtasks (or portions of a subtask) for that grade, and those for the preceding grades.[125] To illustrate this system, consider a student being assessed for grade 1-level numeracy: they would need to achieve scores of approximately 80 per cent or higher on subtask 3. They would also need to achieve a score of approximately 80 per cent on the grade 1-level portions of subtasks 1 and 2, which test number identification up to 99 and addition that does not require "carrying" numbers. They would *not* need to achieve a passing score on the grade 2-level portions of subtasks 1 and 2, which test number identification from 100 to 999, and addition that requires carrying numbers, respectively. A student being assessed for grade 2-level competency would need to complete each of the subtasks specified for grade 1 *and* those specified for grade 2.

A grade-by-grade breakdown of achievement levels in English and mathematics is reported in Tables 30 and 31. In the top panel of Table 30, we report grade-level English performance for the main girls cohort. To read these tables effectively, note that the columns along the top of the table demarcate the current grade in which the student is enrolled in school. The rows indicate the grade level achieved, ranging from grade 2 to grade 6. Students who did not achieve a grade 2 level of literacy –

[124] The curricular documents reviewed by the evaluation team were limited to primary level. No documents were available from the relevant ministry offices for secondary level. However, even with access to a comprehensive curriculum for secondary level, it is unlikely that the assessments included subtasks with enough difficulty to qualify students for Form 1 or higher levels of achievement, given that the difficulty of the numeracy and English tests appear to be capped at grade 4 and 6, respectively.

[125] Where there are many items in a subtask, we follow the 80 per cent rule. Where there are four items, we allow one wrong answer (75 per cent correct) without disqualifying a student from that grade level.

the lowest level effectively tested by the learning assessment in question – are recorded as having a literacy level "below grade 2".[126] Each cell indicates the share of students in a given grade (column) who achieved the grade level indicated by the row. Concretely, this means that 0.9 per cent of students enrolled in grade 6 achieved a grade 2 level of English literacy, and *an additional* 0.9 per cent of students in grade 6 achieved a grade 3 level.[127]

The top panel of Table 30 describes our findings regarding English performance among the main girls cohort, while the bottom panel reports the same analysis for the boys cohort. Table 31 is structured similarly, providing results for the main girls cohort (top panel) and the boys cohort (bottom panel) for mathematics. In the realm of mathematics, the highest possible grade level achievement is capped at grade 4 because, as shown in the skills mapping above, no numeracy subtasks were sufficiently difficult to justify grade level achievements above grade 4.

TABLE 30: GRADE LEVEL OF ENGLISH ACHIEVED BY COHORT GIRLS AND BOYS, BY GRADE

Grade Level Achieved	Grade 6	Grade 7	Grade 8	Form 1	Form 2
Girls Learning Cohort					
Below grade 2	96.8%	94.9%	88.9%	74.1%	70.5%
2	0.9%	1.0%	1.5%	1.6%	3.0%
3	0.9%	1.7%	1.5%	3.2%	1.2%
4	0.9%	1.4%	3.7%	4.9%	6.0%
5	0.5%	1.0%	2.7%	7.6%	6.0%
6	0.0%	0.0%	1.7%	8.6%	13.3%
Grade Level Achieved	**Grade 6**	**Grade 7**	**Grade 8**	**Form 1**	**Form 2**
Boys Learning Cohort					
Below grade 2	90.0%	90.9%	87.6%	71.2%	72.5%
2	1.0%	0.9%	1.9%	2.2%	2.5%
3	5.0%	1.8%	3.8%	2.2%	5.0%
4	1.0%	1.8%	1.9%	2.2%	5.0%
5	0.0%	2.8%	1.9%	2.2%	0.0%
6	3.0%	1.8%	2.9%	20.0%	15.0%

TABLE 31: GRADE LEVEL OF MATHEMATICS ACHIEVED BY COHORT GIRLS AND BOYS, BY GRADE

Grade Level Achieved	Grade 6	Grade 7	Grade 8	Form 1	Form 2
Girls Learning Cohort					
Below grade 1	38.6%	37.2%	33.7%	23.3%	11.5%

[126] As noted above, the English literacy assessment included subtasks testing skills specifically attributed to grades 2 through 6. Therefore, we omit grade 1, since no subtask is available to distinguish grade 1 performance from grade 2 performance. All students who achieve grade 2 performance can be considered to have achieved grade 1 performance as well. Likewise, we omit grades 7 and 8, because no subtask was included that tested skills specific to grades 7 or 8 but excluded from the lower grades. This is not to say that students in the sample *could not* perform at a seventh grade or higher level. Rather, it means that the assessment did not include sufficiently difficult questions to allow them to demonstrate these skills. This is not a criticism of the assessment, per se – given how few students performed at a sixth grade level, we can be confident that even fewer would exceed this level.

[127] The converse of our example is that 97.7 per cent (100 – 2.3) of students in grade 6 failed to achieve a grade 3 level of English literacy, and that 96.8 per cent (100 – 3.2) of students in the same grade failed to achieve a grade 2 level – or any identifiable level, in this analysis – of English literacy.

1	43.6%	34.4%	29.0%	30.8%	28.3%
2	9.8%	13.5%	12.6%	12.4%	20.5%
3	1.4%	2.9%	2.2%	3.8%	4.8%
4	6.6%	12.0%	22.5%	29.7%	34.9%

Grade Level Achieved	Grade 6	Grade 7	Grade 8	Form 1	Form 2
			Boys Learning Cohort		
Below grade 1	29.0%	28.5%	20.1%	11.1%	15.0%
1	36.0%	33.9%	26.0%	20.0%	10.0%
2	16.0%	11.9%	20.2%	13.3%	17.5%
3	4.0%	1.8%	5.8%	6.7%	2.5%
4	15.0%	23.9%	27.9%	48.9%	55.0%

In line with much of our previous analysis, we find that male students tend to outperform their female counterparts in both subjects, though the gaps in performance vary across grades. Also consistent with the aggregate results regarding mathematics and English literacy reported above, students performed significantly better in mathematics than in English. English literacy is particularly dismal when judged against national curriculum, with remarkably few students achieving performance even remotely similar to their respective grade levels. For instance, just 0.5 per cent of Grade 6 girls in the sample achieve grade 5-level literacy in English. Only 6.6 per cent of these same girls achieved grade 4-level performance in mathematics.

It is difficult to know what conclusions to draw from the results presented here. Certainly, the performance of EGEP-T students at the baseline lags the targets set in the curriculum. But this fact may reflect either lagging performance at EGEP-T schools or unrealistic expectations on the part of ministerial officials and others involved in designing the curriculum. Without knowing the rate at which students in other schools achieve each grade level, it is difficult to adjudicate between these two possibilities.

4.2 Subgroup Analysis of Learning Outcomes

While the analysis of aggregate learning outcomes is essential to establishing overall project impact, this baseline evaluation also serves to test the EGEP-T Theory of Change, and to derive recommendations for project design and implementation. 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. In Section 5.6, below, we systematically extend this analysis to a multivariate framework, highlighting potential determinants of learning outcomes.

Before turning to analysis of specific barriers and demographic subgroups, we provide a more comprehensive breakdown of learning performance by grade, across geographic space. Table 32 reports the mean numeracy, Somali literacy, and English literacy scores for each project location, by grade; Somaliland occupies the top panel in the table, followed by Puntland, Galmudug, and Banadir.[128]

TABLE 32: LEARNING OUTCOMES BY GRADE AND PROJECT LOCATION

Grade	Somaliland	Puntland	Galmudug	Banadir

[128] No results are provided for Hirshabelle, because only 12 cohort girls were sampled in the area.

Numeracy				
Grade 6	48.33	68.91	53.55	69.40
Grade 7	52.71	75.92	73.72	74.24
Grade 8	60.99	78.60	82.10	82.38
Form 1	63.86	85.56	90.45	88.28
Form 2	69.86	86.81	88.72	85.20
Total	55.83	77.02	75.35	77.43
Somali Literacy				
Grade 6	59.50	70.07	56.64	76.50
Grade 7	70.68	79.81	72.48	83.30
Grade 8	74.76	82.33	86.52	85.63
Form 1	81.64	84.32	93.97	88.25
Form 2	81.97	91.27	94.81	87.92
Total	70.19	79.62	78.15	82.95
English Literacy				
Grade 6	15.95	29.71	24.04	49.51
Grade 7	20.78	34.91	37.74	52.25
Grade 8	26.38	46.26	55.47	63.34
Form 1	40.19	65.28	72.49	75.30
Form 2	46.39	70.26	75.23	73.69
Total	24.47	43.68	48.68	58.70

It is important to note the method by which data on demographic characteristics was collected. Because the evaluation employed a split-sampling approach at the baseline, cohort girls were interviewed at their schools, and members of their household were not interviewed. To capture key demographic information, we designed and implemented a brief demographic survey to be administered to each girl. One consequence of this decision is that the set of available demographic variables is more limited than it would be with a full household survey. An additional consequence is that some variables may be measured with error. Adolescent girls may not accurately describe the head of their own households, and may misperceive or misunderstand their parents' respective occupations and education levels. In this discussion, we focus on demographic characteristics in which we have relatively high confidence of accurate answers from even young adolescents.

Further, one should interpret the results presented below with caution, bearing in mind the size of each subsample. In some cases, the subsample on which results are based is exceedingly small, and gaps in learning outcomes that appear substantively important may be statistical noise resulting from sampling variance. In addition, the subgroup scores are not adjusted for other demographic characteristics or predictors of learning outcomes – the scores are raw means, and should be interpreted cautiously as a result.[129]

[129] That is to say, the comparisons drawn in the table do not "control" for other factors that may also determine learning scores. For instance, by comparing the learning scores of girls in rural schools to the full sample, one is simply comparing unadjusted, naïve means. If urbanity is correlated with other determinants of learning scores, this comparison will almost certainly misstate the "effect" of urbanity on learning scores. We encourage interested readers to review the multivariate analysis of learning in Section 5.6.

We disaggregate learning outcomes of cohort girls at the baseline along five primary axes, with results reported in Table 33. First, we consider girls from households headed by a female. Contrary to standard assumptions regarding the marginalisation of female-headed households, respondents from such households marginally outperform the sample mean on all three subject assessments.[130]

Second, we consider the educational attainment of the girl's head of household, because educated parents may be more likely to value education for their children, better able to assist them with their schoolwork, and may have more resources available, such as money for school fees, to support their children's education. We find strong evidence that girls from households in which the household head is less educated perform worse on learning assessments. In the case of heads of household who completed some primary school, girls score between 4.0 and 5.6 points below the sample mean on all three assessments. This effect is more pronounced for heads of household with no education, where learning outcomes fall between 6.0 and 10.5 points lower than the sample mean.[131]

Third, we report learning outcomes among children with self-reported physical and cognitive impairments. Importantly, while the baseline evaluation collected data on a standard range of impairments – impaired vision, hearing, mobility, cognition, communication and impaired ability to engage in self-care – we report results only for girls reporting visual impairment and for girls reporting any impairment from among the broader set. In practice, the sample contained very few girls with any non-visual impairment; for instance, we interviewed only two girls with impaired hearing and two girls with impaired mobility. As a result of this small sample size, we do not report results for these subgroups.[132]

TABLE 33: LEARNING SCORES AMONG DISADVANTAGED SUBGROUPS IN MAIN GIRLS COHORT

Subgroup	Mean Numeracy Score	Mean Somali Literacy Score	Mean English Literacy Score	Subgroup size (n)
Main Girls Cohort				
All girls in cohort	68.85	76.39	38.60	1,609
Living in female headed household	71.97	78.25	43.07	715
HoH – some primary educ.	65.02	70.98	32.27	131
HoH – Quranic educ. only	72.67	79.96	41.71	319
HoH – no education	61.38	71.12	29.23	385
Vision impairment	48.31	69.85	32.27	21
Any impairment	53.38	70.98	31.14	30

[130] It is important to note that our measurement strategy, with regard to information on the heads of girls' households, was somewhat unusual. As we describe above, household surveys were not conducted with members of the girls learning cohort, nor with the boys learning cohort or bursary girls. Without interviewing a child's parents or caregiver, we must rely on children to provide demographic information about their own households. In some cases, such as data on their parents' occupations, this proved so error-prone that we do not report the data in this report. For the purposes of identifying children from female-headed households, we asked children to think of the person who is in charge of their household – the person who makes important decisions, such as how to spend the family's money – and identify whether they are a man or woman. If respondents were unsure, enumerators asked them who was "in charge of the household"; this decision was made because the reference to financial decisions in the original question tended to confuse respondents, as women are often tasked with making decisions about how to spend money within households, even when they are not the heads of their respective households.

[131] As Table 33 shows, girls from households in which the household head completed only Quranic education actually outperform the sample mean.

[132] In practice, the data indicate that 1.3 per cent of girls in EGEP-T schools have impaired vision, and 1.9 per cent suffer from one of the impairments captured in the survey instrument.

Serious illness	65.28	75.06	33.00	148
Conflict-affected school	80.44	78.62	36.40	252
IDP school	70.29	75.72	36.47	81
Drought-affected school	64.08	73.00	31.92	441
Rural	58.51	70.23	26.13	428
Old for grade	60.38	69.08	28.33	224

Girls self-reporting physical or cognitive impairments score significantly below the sample mean on all three assessments, with particularly dramatic differences observed in the case of numeracy. The available qualitative evidence does not provide significant additional information regarding the impact of physical and cognitive impairments on learning outcomes, though one group of students noted that some of their classmates have difficulty seeing the blackboard at the front of the class.[133] According to our sample of teachers, most students with visual impairments sit near the front of the class; however, this is not universally true – 5.3 per cent of teachers indicated that students with visual impairments do *not* sit near the front of the class. Furthermore, accommodation of students with impaired hearing is even less consistent.[134] Given the prevalence of visual impairment within project schools – visual impairment comprised 70 per cent of all documented physical and cognitive impairments – efforts to improve accommodations for physical and cognitive impairments should focus on accommodations for students with impaired vision.

Fourth, we consider the environment in which the girl attends school. Just over 15 per cent of girls in the learning cohort attend schools that are considered conflict-affected by RI's internal analysis, and additional girls attend school in IDP camps or other unfavourable environments. Surprisingly, we find no evidence that conflict-affected and IDP schools underperform as a whole, as students in these schools perform on par with students in non-conflict and non-IDP schools, respectively. Both of these findings should be interpreted with some degree of caution, however, given the relatively small sample of conflict-affected and IDP schools in question, and characteristics of the schools in question.

Considering conflict-affected schools first, it is important to note that schools denoted as such as concentrated almost exclusively in Puntland. To illustrate, 19 of the 21 conflict-affected schools in the sample were in Puntland, and conflict-affected schools make up 31.7 per cent of all Puntland schools sampled. When we limit our analysis of the relationship between conflict and learning outcomes to Puntland alone – in order to ensure that differences across locations are not spuriously driving any relationship between conflict and learning – we find that girls in conflict-affected schools score higher in terms of numeracy, slightly lower in terms of Somali literacy, and much lower in terms of English literacy.[135] In the aggregate, girls in Puntland's conflict-affected schools perform marginally worse than those in non-conflict schools, with an aggregate score – averaging numeracy, Somali literacy, and English literacy – 5.1 points lower than their counterparts in non-conflict schools. This gap shrinks, but does not entirely disappear, when we employ multivariate regression models.[136] In broader models, which we report in Figure 21 in Section 5.6, conflict does not predict

[133] FGD with boys in Awdal, Somaliland.

[134] When asked whether children with impaired hearing sit near the front of the class, 7.9 per cent of teachers indicated that they disagreed a little or disagreed a lot.

[135] Specifically, girls in conflict-affected schools score 4.6 points higher on numeracy, 2.3 points lower on Somali literacy, and 17.6 points lower on English literacy.

[136] We do not report the full results of these models here. However, when limiting the sample to Puntland and incorporating controls for school level (primary versus secondary school), drought, and the grade of individual

learning outcomes, though that analysis expands the sample to all project locations, rather than Puntland alone. Overall, then, we find mixed evidence regarding the impact of conflict on learning outcomes: while it appears to have a statistically significant effect on learning outcomes in the primary conflict-affected area of Somalia, the effect is smaller than that of other factors, including drought.

Turning to IDP schools, our initial analysis suggested that students in these schools performed on par with students in non-IDP schools, as shown in Table 33. IDP schools are not concentrated in a single location – they appear in Somaliland, Puntland and Galmudug – though, as we note in the methodological limitations in Section 2.6, IDP schools were systematically underrepresented in the sample, such that there are only seven IDP schools in the data. Given the small sample size, findings regarding learning outcomes should be interpreted cautiously. More importantly, our data do not fully capture the number of IDPs in the sample, many of whom we cannot identify on the basis of data from the baseline. For the purpose of identifying IDP schools, we rely on RI's classification of schools; however, this classification ignores IDPs who are enrolled in non-IDP schools. Based on RI's monitoring data, almost all EGEP-T schools include at least a handful of IDPs and, in some non-IDP schools, they make up a substantial share of the student body.[137] Unfortunately, our data do not allow us to identify individual IDPs in non-IDP schools, which limits the quality of our analysis regarding the relationship between IDP status and learning outcomes.[138]

As we discuss elsewhere in this report, Somalia is currently suffering a severe drought, which has impacted large swathes of all five project locations: Banadir, Galmudug, Hirshabelle, Puntland, and Somaliland. According to RI's internal analysis of their specific project sites at the time of writing this report, 29 of the 140 sampled schools are "severely impacted" by the drought. Table 34 shows that girls in drought-affected schools underperform the sample mean on all three assessments. At the same time, girls in rural schools underperform even more dramatically, with numeracy scores 11.4 percentage points below the sample average and over 15 points below the mean performance of girls in urban schools. Because urbanity and drought impact are negatively correlated – 72.4 per cent of rural schools are drought-affected, compared to just 15.6 per cent of urban schools – it is difficult to determine which factor is influencing learning outcomes in this bivariate analysis.

Finally, because the ages of girls enrolled in the target grades vary so significantly, we also assess the performance of girls who are relatively older for their respective grade levels. Students who are older than others in their class face unique challenges, including social stigma and shame: students who have not been promoted in previous years may feel ashamed of their performance and lack

---

students, we find that girls in conflict-affected schools score 2.6 points lower in the aggregate than those in non-conflict schools.

[137] In total, RI estimates that they will directly benefit 2,197 IDPs enrolled in IDP schools, and a further 3,322 IDPs in non-enrolled schools. These figures underline the limitations of our analysis of the relationship between IDP status and project outcomes, because our analysis focuses exclusively on the former group, to the exclusion of the latter.

[138] While RI's monitoring data indicate the number of IDPs at all or most of their schools, they do not identify *which* girls are IDPs, and we are unable to link IDP status of each girl to their individual learning outcomes. The evaluation did not collect reliable data on IDP status of individual girls at the baseline, because no household survey was conducted with members of the girls learning cohort; while girls were asked about their family's recent migration status, classifying them as IDPs or non-IDPs on this basis is too unreliable to be used as the basis for analysis. In future evaluation waves, when household surveys are conducted with all members of the girls learning cohort, additional analysis of the relationship between IDP status and learning outcomes will be possible.

confidence in their ability to learn.[139] We categorize girls as "old for their grade" if they are 15 years or older in Grade 6, 16 years or older in Grade 7, and so on. As the bottom row of Table 34 shows, these girls also systematically underperform relative to the sample mean. Such a result could occur if girls who are old for their grade were held back due to poor learning outcomes, but it could also occur if girls temporarily dropped out of school and have consequently fallen behind one or more years. In either case, these girls may be suitable targets for additional help, because they are a readily-identifiable group with systematically weaker performance.

Tables 34 and 35 replicate the analysis of cohort girls, conducted above, in the case of cohort boys and bursary girls. Surprisingly, many of the barriers relevant to girls' learning outcomes do not appear to impact outcomes among boys: for instance, girls in households headed by an individual with no education scored markedly lower on all three assessments, while boys in this subgroup actually appear to perform slightly better than the full sample, overall. Likewise, boys who are old for their grade do not seem to underperform the full sample, though girls in the same circumstances did underperform.

TABLE 34: LEARNING SCORES AMONG DISADVANTAGED SUBGROUPS IN BOYS COHORT

Subgroup	Mean Numeracy Score	Mean Somali Literacy Score	Mean English Literacy Score	Subgroup size (n)
Boys Cohort				
All boys in cohort	79.00	79.71	45.69	398
Somaliland	71.83	74.04	33.99	138
Puntland	82.74	81.67	48.26	163
Galmudug	86.72	85.36	56.31	33
Banadir	82.38	85.98	66.16	61
Living in female headed household	82.17	80.76	47.67	164
HoH – some primary educ.	79.26	79.77	45.26	32
HoH – Quranic educ. only	81.17	82.49	48.82	93
HoH – no education	78.01	78.18	37.43	108
Vision impairment	No data collected			
Any impairment	No data collected			
Serious illness	No data collected			
Conflict-affected school	84.86	84.28	38.12	63
IDP school	72.77	73.99	39.85	19
Drought-affected school	78.99	79.88	40.13	114
Rural	73.22	75.21	32.75	105
Old for grade	75.52	77.57	43.43	117

TABLE 35: LEARNING SCORES AMONG DISADVANTAGED SUBGROUPS IN BURSARY GIRLS COHORT

Subgroup	Mean Numeracy Score	Mean Somali Literacy Score	Mean English Literacy Score	Subgroup size (n)
Bursary Girls Cohort[140]				

[139] According to one teacher, students who fail end-of-year examinations often drop out of school, and this problem is particularly pronounced among students who are already older than their peers (KII with male teacher, Garbadadar, Somaliland).

[140] The sample used in this analysis differs from the complete bursary girls sample, because only a portion of bursary girls who completed learning assessments also completed surveys in which demographic details were collected. In total 428 bursary girls were sampled and completed learning assessments, but just 318 completed surveys, which

All girls in cohort	70.23	76.61	39.40	318
Somaliland	53.28	64.01	23.91	136
Puntland	77.85	81.60	42.48	198
Galmudug	71.90	81.61	43.30	27
Banadir	75.85	81.36	54.76	63
Living in female headed household	70.75	77.21	39.03	186
HoH – some primary educ.	66.16	70.29	40.85	31
HoH – Quranic educ. only	73.84	81.07	47.03	71
HoH – no education	64.46	71.86	28.62	98
Vision impairment	No data collected			
Any impairment	No data collected			
Serious illness	No data collected			
Conflict-affected school	80.16	80.78	35.87	74
IDP school	71.37	80.83	30.96	21
Drought-affected school	65.79	73.14	31.96	112
Rural	57.98	65.38	25.40	108
Old for grade	60.51	71.64	24.73	43

The results among bursary girls are more consistent with those of the main set of cohort girls. Bursary girls from households in which the head of household did not complete any education performed markedly worse than the sample average, as did bursary girls who were classified as "old for their grade" and those in rural schools.

Moving beyond demographic categories, we also report the learning outcomes of cohort girls, categorized by their exposure to a number of hypothesized barriers to learning. For instance, while the education of one's parents or one's age identify potentially marginalised groups, a number of other, often less tangible and more malleable, barriers to learning exist. As in the discussion above, we report the mean learning assessment scores for cohort girls across several thematic barriers in Table 36, below. Data on a number of barriers included in this analysis was collected independently of girls' survey responses. Specifically, data on school-level outcomes, such as infrastructure, teacher pay, and female teaching staff, among others, was collected from school head teachers. Where data was collected from head teachers, the question text is denoted by the letters "HT" in parentheses.[141] It is also important to note differences in data quality and – occasionally – relevance to the Somali or EGEP-T contexts across different measures, which sometimes motivates us to focus on one or two key indicators within a group.[142]

The first thematic group concerns school infrastructure, which can affect how comfortable it is for girls to attend school, and how much additional burden is placed on students to collect water or find external washrooms to use. The factors in this section represent both girl-level and school-level

---

were implemented after the first 1-2 weeks of fieldwork were already finished. In these results, we limit the sample to the 318 bursary girls who completed surveys.

[141] All other data was collected from cohort girls during surveys at their schools.

[142] As an example of a question where we are concerned about data quality, consider students who report that they are "unable to move around the school easily." Over one-quarter of students indicated that they are unable to move easily around their school, despite relatively low rates of self-reported disability prevalence. It is unclear whether this question assesses the physical mobility of a student, or other barriers to free and easy movement, such as school staff telling them that they are not allowed to be in parts of the school, or concerns about their own safety. As a result, we focus on subgroups with a more straightforward interpretation, such as girls who attend schools without gender-specific toilets, a common and contextually-relevant barrier to girls' education.

barriers. That is, some barriers are specific to individual girls (e.g., a girl who does not use the toilet at school), while others are a function of the school itself (e.g., a school does not have a toilet reserved specifically for girls' use). Across each of these barriers, we observe a marginal decrease in learning outcomes, with the largest differences occurring for girls who find it difficult to move around their school and girls who do not use a toilet at school. These findings are consistent with the qualitative evidence, as several respondents highlighted the importance of female sanitation kits and female-specific toilets for improving attendance rates among girls.[143]

The second thematic group also focuses on girls' schools, but moves beyond available facilities to focus on other resources, such as the availability of textbooks, desks for students, and the consistency with which teachers are paid. Among factors in this group, we find a consistent null effect; in other words, there is no systematic difference between girls who report lacking learning materials and girls who do not, nor does the consistency of teachers' pay appear to influence learning outcomes in this analysis.[144]

TABLE 36: BARRIERS TO LEARNING AND LEARNING OUTCOMES

Subgroup	Mean Numeracy	Mean Somali Literacy	Mean English Literacy	Subgroup size (n)
All girls	68.85	76.39	38.60	1,609
School Infrastructure				
Difficult to move around school	61.13	74.13	33.89	414
Doesn't use drinking water facilities	66.46	76.58	37.83	580
Doesn't use toilet at school	62.47	74.36	33.76	420
Doesn't use areas where children play	64.00	75.28	37.27	702
No toilet specifically for girls at school (HT)	64.47	73.76	35.75	304
No clean drinking water at school (HT)	67.81	76.78	38.79	391
No electricity at school (HT)	66.48	74.97	35.21	894
School Resources				
School does not provide feeding programme (HT)	73.24	79.55	44.63	1020
School does not use textbooks (HT)	67.53	75.42	39.72	405
Students share textbooks (HT)	71.25	78.32	40.06	674
Student does not have learning materials	69.44	77.48	39.30	468
Teachers are paid very inconsistently[145] (HT)	70.26	75.97	39.88	217
Not enough seats for children at school	65.67	75.27	38.11	323

[143] KII with female teacher, Maroodi Jeeh, Somaliland; KII with male teacher, Saxiil, Somaliland.

[144] Strangely, schools that participate in a school feeding programme appear to outperform those that do not. Most schools that provide meals to students are primary schools, but this pattern holds even when we restrict the analysis to primary schools. One possible explanation is that schools in higher-poverty areas are more likely to participate in school feeding programmes, and also suffer from worse learning outcomes as a result of more extensive poverty.

[145] Head teachers were asked to assess how consistently teachers in their school were paid on time, on a 4-point scale (very inconsistently, somewhat inconsistently, somewhat consistently, and very consistently).

Teaching Quality and Environment[146]				
Disagrees teachers make them feel welcome	61.48	69.21	32.99	55
Agrees that they are afraid of teacher	64.79	73.94	35.10	769
Agrees teachers treat boys and girls differently in the classroom	66.88	74.69	36.26	580
Agrees teachers often absent from class	65.78	74.44	35.12	386
Teacher absent more than once in last two weeks (HT)	60.57	71.70	29.97	222
Fewer than 5 hours of instruction per day (HT)	65.29	71.06	32.02	340
Agrees teachers punish students for wrong answers	67.67	75.42	38.10	936
Gender Equity				
School has no full-time female teachers (HT)	77.17	83.76	46.87	278
School has no female teachers at all (HT)	79.03	85.32	51.08	219
School does not have a female mentor (HT)	69.60	76.76	39.28	922
Girl is unaware of female mentor	71.26	77.99	38.45	557
Disagrees boys encourage girls to participate	70.37	79.44	44.64	259
Disagrees boys encourage girls to continue schooling	70.18	79.02	43.32	272
Perceives gap in community support for girls' education[147]	67.07	78.63	37.99	233
Other Barriers				
Agrees she has no choice in schooling decisions[148]	67.63	77.10	36.70	1079
Over one hour travel time to school	56.20	65.94	27.97	61
Feels unsafe on way to school	61.77	68.76	28.31	51
Feels unsafe at school	57.13	72.16	27.86	32
Perceives both parents as disengaged[149]	69.83	73.53	37.33	119

[146] Agree/disagree questions in this section utilised a four-point scale: agree a lot, agree a little, disagree a little, and disagree a lot. Respondents are coded as agreeing if they respond that they agree a little or a lot, and as disagreeing if they respond tha they disagree a little or a lot.

[147] See footnote 106 for complete explication of this metric.

[148] This question asked respondents to indicate their agreement with the statement "I cannot choose to attend or stay in school – I must accept whatever happens", which we consider a measure of empowerment with regard to schooling decisions. Girls who indicated strong ("a lot") or partial ("a little" agreement with this statement are included in this subgroup.

[149] Respondents were asked to imagine a scenario in which their parents were asked to attend a meeting with their teacher. We asked respondents how likely it was that their mother and father, separately, would attend the meeting. Respondents who indicated that both parents were somewhat or very unlikely to attend were classified as perceiving their parents to be disengaged.

Teacher quality constitutes the third thematic area. Improving the quality of teachers was a core goal of EGEP in GEC-1, and continues to be a focus of EGEP-T, which will instruct teachers on various aspects of pedagogy, as well as subject-specific training. The need for additional teacher training was occasionally noted by qualitative interviewees, especially teachers themselves and CEC members.[150] As Table 36 shows, we observe consistent differences between girls exposed to poor teaching practices, high absenteeism among teachers, and teachers about which the girls have negative perceptions (i.e. girls who report being afraid of their teacher or who indicate their teacher does not make them feel welcome). The correlation between learning outcomes and absenteeism is particularly striking – girls who report that their teacher missed two or more lessons in the previous two weeks score between 5.2 and 10.2 points lower on each of the three assessments. Schools that report having fewer than five hours of instruction per day also show signs of worse student outcomes, though the difference is less stark than with teacher absenteeism.

Within this third theme, we also note an additional, nuanced, predictor of learning outcomes: language of instruction. As noted in Section 3.3 previously, the official language of instruction for primary schools, Somali, matches the language the vast majority of children speak at home. However, we have also documented that a number of schools – especially private schools – do not conform to official language of instruction policies. Indeed, 22.1 per cent of primary schools use English instead of Somali as their language of instruction, and 25.9 per cent of secondary schools use Somali instead of English as their language of instruction. A small share of schools use Arabic instead. These contextual differences between schools can influence learning outcomes, especially in terms of language-specific literacy. Girls in grades 6-8 in primary schools that use English outperform equivalent girls in primary schools that use Somali on all learning outcomes; however, the performance gap is most notable in English literacy, where they score 19.9 points higher than equivalent girls in Somali-language primary schools. This finding fits with our expectations that girls in schools that use English as the language of instruction should perform especially well in terms of English literacy, especially compared to girls who have merely studied English as a subject, rather than using it daily for most or all educational tasks.

The fourth theme focused on gender equity at school, particularly the presence of female teachers, and a female teacher-mentor. This section also analyses the performance of girls who report that boys do not encourage their educational endeavors, and girls who believe that their community would favour a boy over a girl when it comes to funding their education.[151] Surprisingly, we find no systematic correlation between perceptions of gender bias in their communities, or discouragement from boys in their school, and learning outcomes. Further, we find that girls enrolled in schools that lack female teaching staff outperform the sample mean, though it is possible that this result is an artefact of the types of schools that have all-male staffs. The relationship between schools the lack

[150] KII with male teacher, Bari.

[151] The latter outcome is derived from a hypothetical scenario posed to each girl. Respondents were asked to imagine that a boy and a girl were both admitted to a university, and that their families were seeking community support to raise funds for them to attend. Respondents were asked, for each of the hypothetical students, how likely they were to successfully raise funds from their community. We define a girl as perceiving a gap in support for girls' versus boys' education in their community if they report that the boys' family would be "very likely" or "somewhat likely" to succeed, while the girls' family would be "very unlikely", or if they report that the boys' family would be "very likely" to succeed, while the girls' family would be "somewhat unlikely". Each difference constitutes a two-level gap on a four-point likert scale, implying a fairly significant difference in expectations for an individual respondent.

female teaching staff and higher learning scores is consistent even in regression models that control for grade level, school level, and urbanity.[152] However, we find similar results for boys' learning scores – boys in schools that lack female teachers also perform significantly better than the typical male student. Given that there is little reason to suspect that the presence of female teaching staff would influence male performance, we view the equivalent finding with regard to girls as suspect.

The qualitative evidence also appears to contradict the idea that schools with female teachers would produce worse-performing female students. Indeed, a number of teachers, CEC members and MoE officials emphasized the need for additional female teachers because of their impact on girls' learning.[153] One respondent focused on the role female teachers can play in making girls more comfortable to ask questions in class, noting that girls are often too shy or lack the confidence to ask or answer questions of male teachers.[154] As noted, it is possible that the *type* of school that entirely lacks female teachers may be fundamentally different from those that hire female teachers, but this speculative hypothesis should be studied in more detail.

The bottom panel of Table 36 considers a number of other barriers that do not fit into the thematic areas described above. Most notably, we find that learning outcomes are significantly lower among girls that travel longer distances to school, who report feeling unsafe on their way to school and, especially, among girls who report feeling unsafe *at* school.[155] The results regarding perceived safety and travel time are based on a particularly small number of observations, however; at the extreme, just 32 respondents (out of 1,609) reported that they feel unsafe at school. The small subsample size calls for caution in drawing conclusions regarding the effect of perceived safety or travel time on learning outcomes.

One prominent barrier that we do not study systematically is the role of housework in girls' educational underperformance. The baseline did not collect useful quantitative data on girls' relative burden of housework or chores. However, the role of housework was arguably the most common theme in qualitative interviews when interviewees were asked to discuss barriers to girls' education and why girls and boys perform differently at school.[156] Interviewees indicated that chores prevent girls from doing their homework, even though many interviewees seemed to believe that girls are more motivated to study than boys. For instance, after completing their chores, girls often study to the extent possible, but face time constraints not faced by boys. Girls' housework burden also impacts their attendance: interviewees saw football and other leisure activities as the primary

[152] In a linear model controlling for these factors, the aggregate learning score – the average of numeracy, Somali literacy and English literacy – for girls in schools without female teaching staff was 7.3 points higher than that of girls in schools with at least one female teacher.

[153] KII with male teacher, Saxiil, Somaliland; KII with MoE QA official, Galkayo, Puntland; FGD with CEC members, Mogadishu; KII with male teacher, Garbadadar, Somaliland.

[154] FGD with CEC members, Mogadishu.

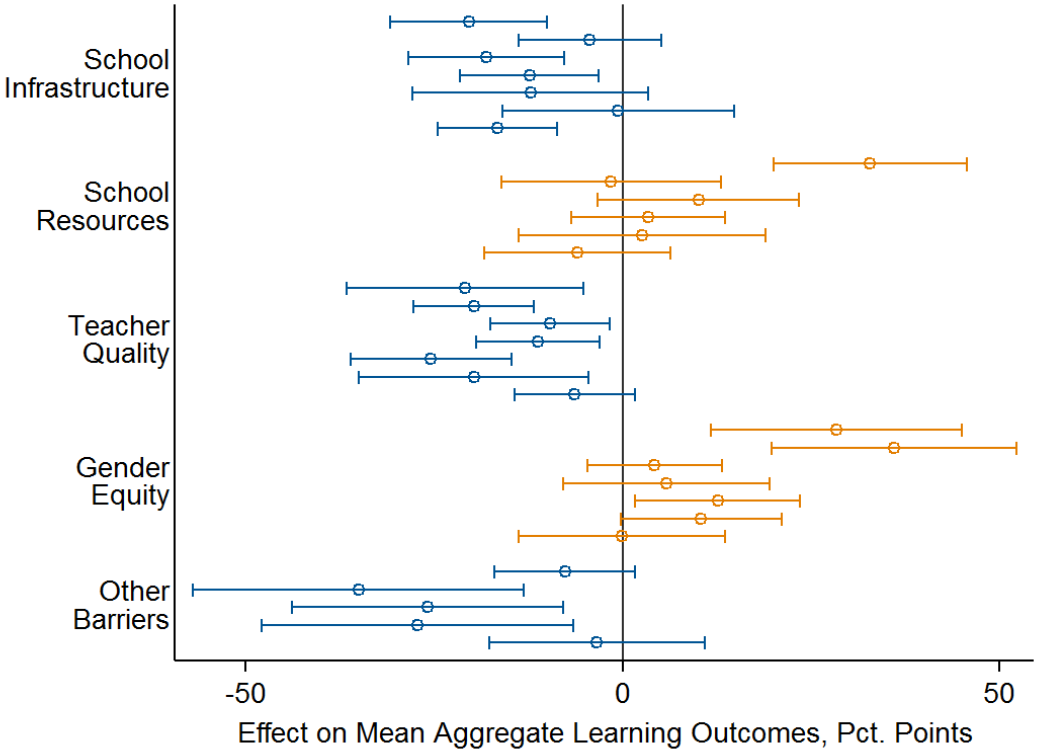
[155] Students' perceptions that their parents are disengaged (i.e. neither parent is likely or very likely to attend a parent-teacher meeting arranged by their teacher) and students' lack of autonomy over schooling decisions are not correlated with learning outcomes.

[156] KII with female teacher, Hargeisa, Somaliland; KII with male teacher, Galgadud; FGD with CEC members, Mogadishu, Banadir; KII with female teacher, Mogadishu, Banadir; KII with male teacher, Bari; KII with male teacher, Garbadadar, Somaliland.

reason why boys were late for school, but noted that girls are often late due to the need to cook breakfast or engage in other housework before school.[157]

Due to the large number of barriers analysed in Table 36, it can be difficult to discern patterns that are meaningful versus those that are merely statistical noise. To summarize the findings regarding barriers to learning outcomes, we perform a secondary analysis and graph the results in Figure 13. In this analysis, we perform a two-sample t-test comparing mean aggregate (numeracy, Somali literacy, and English literacy combined) learning outcomes for girls in each subgroup studied in Table 36, relative to their counterparts outside the subgroup. To illustrate, we compare girls at schools that lack clean drinking water to girls whose schools have clean drinking water. We then plot the *difference* in mean learning outcomes between these two groups. Figure 13 plots these differences for each barrier identified above, grouped by thematic area.[158]

FIGURE 13: RELATIONSHIP BETWEEN BARRIERS AND AGGREGATE LEARNING OUTCOMES



Three findings emerge from Figure 13, which may have been obscured in the previous discussion. First, deficits in school infrastructure appear to be systematically related to worse learning outcomes; while not all of the barriers are statistically significant at the 5 per cent level, the consistent finding across all barriers is indicative of a widespread association. Second, lower teacher

[157] For evidence regarding boys and leisure activities, see: KII with male teacher, Galgadud; KII with female teacher, Hargeisa, Somaliland. For discuss of girls' housework burden before school, see: KII with male teacher, Galgadud.

[158] The figure plots the mean difference and the 95 per cent confidence interval for that difference. A straightforward way to determine whether a barrier is a statistically significant predictor of learning outcomes is to check whether its confidence interval crosses the vertical line at zero. If the confidence interval does not cross the vertical line, the correlation between the barrier and aggregate learning outcomes is statistically significant at the 5 per cent level.

quality is highly correlated with worse outcomes, of varying magnitudes.[159] Finally, the bottom section of Figure 13 shows three "other barriers" with large negative relationships with learning outcomes – the three barriers are safety at school, safety on the way to school, and the distance travelled to school, as noted above.

It is important to temper conclusions drawn from this analysis. While the results in this section often document patterns of marginalisation among particular subgroups, or highlight potential barriers to girls' learning, we cannot make strong claims regarding these barriers in terms of their association with learning outcomes. In a bivariate framework, we are unable to control for a number of confounding factors that may explain correlation – or lack thereof – between particular barriers and learning outcomes. We engage in more systematic multivariate analysis in Section 5.6, with the aim of drawing firmer conclusions about the individual-, school-, and community-level correlates of learning.

Predicting Low-Performing Learners

Thus far, our analysis has focused on average learning outcomes across subgroups. However, a focus on mean outcomes can obscure important patterns in data. If a given subgroup has unremarkable mean scores – similar to the rest of the sample, for instance – this does not necessarily mean that it is not an important factor in driving particularly low levels of achievement.

To evaluate the determinants of learning in greater detail, we developed a method for classifying students whose performance on all three learning assessments raise special concerns. Our goal was to identify students who systematically underperformed their peers on all three assessments, suggesting particular difficulties in learning.

We begin by defining the lowest tercile of assessment performance for each subject, disaggregated by grade level and gender. The lowest tercile includes individuals who fall below the 33rd percentile. We define the lowest tercile for girls in grade 6 as those girls in grade 6 who score below the 33rd percentile amongst all girls in the same grade.[160] We create similar terciles for numeracy, English literacy, and Somali literacy, for each grade-gender combination (i.e. girls in grade 6, boys in grade 6, girls in grade 7, etc.).

We identify students who fall into the lowest tercile of performance on *all three assessments*. We consider these students a particularly interesting subpopulation – while students who perform poorly in English may compensate by excelling in mathematics, the students we identify are consistently underperforming their peers, across all three examinations. Such students are less likely to respond to simple remedial courses, because they have fallen behind in more than one subject, presenting an especially large burden for any efforts to catch up.

Our interest is in determining which types of students fall into this low-achieving category. We study a large sample of students, comprising the main girls cohort, the boys cohort, and the bursary girls cohort (n = 2,345), where sufficient data is available for each respondent. In total, 13.6 per cent of

[159] Again, these barriers are identical to those studied in Table 36, but presented in a different fashion. These barriers include, in order: disagree that my teacher makes me feel welcome; agree that I am afraid of my teacher; agree that teachers treat boys and girls differently in the classroom; agree that teachers are often absent from class; a measure of teacher absenteeism collected from head teachers; attends school in which head teacher reports that a typical school day includes fewer than five hours of instructional time; and agree that teachers punish students for wrong answers.

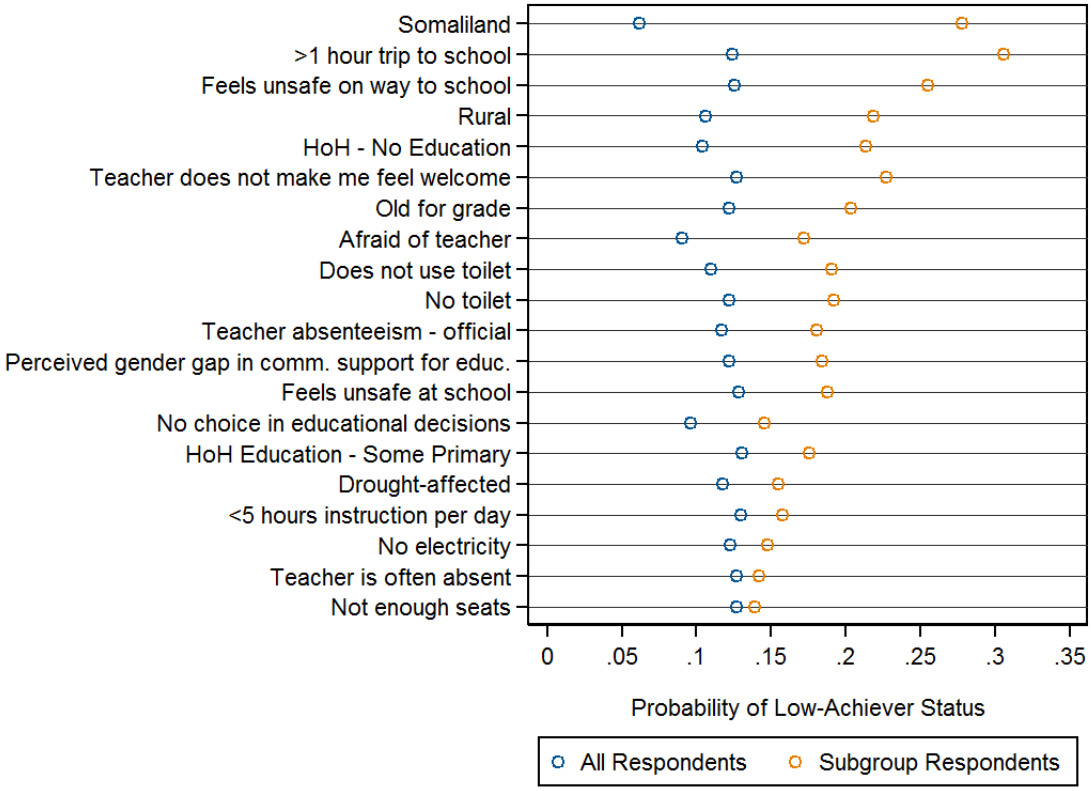
[160] We define terciles *within* grades because, otherwise, the set of low-performers would be dominated by students in the lowest grades.

students are classified as low-achievers, because they fall into the lowest tercile on all three assessments. We call these students, for convenience, "low-achievers."

We study the probability that a student will be a low-achiever, on the basis of their demographic characteristics, their school's resources and quality, and other barriers identified in the analysis that preceded this section. For each barrier, we calculate the probability that a student *not* facing the barrier will be a low-achiever; we then calculate the probability that student facing the barrier will fall into the low-achiever category. By focusing on probability in this way, we are able to immediately see which factors exert substantively strong effects on learning outcomes at the lower-end of achievement.[161]

In total, we identified 34 potential predictors for investigation. For the sake of brevity, we only report results for the 20 predictors that increased the probability of a student being a low-achiever. The results of this analysis are presented in Figure 14. The blue dots represent the probability of low-achiever status among individuals *not* in the subgroup in question. For instance, the top blue dot provides the probability of being a low-achiever if a student does not live in Somaliland, while the orange dot provides the same probability if a student *does* live in Somaliland.

FIGURE 14: PREDICTORS OF LOW-ACHIEVING STUDENTS



Our interest is in those predictors with large gaps between the two probabilities. The barriers in Figure 14 are sorted such that the largest gaps are found at the top of the graph. The single-best

[161] As with our previous findings in this section, it is important to emphasize that this analysis is bivariate in nature, meaning that the analysis does not control for mitigating factors or omitted variables that might simultaneously influence learning outcomes and be correlated to the barrier being assessed. We provide a more rigorous analysis of factors that predict learning outcomes in Section 5.6.

predictor of a student acquiring low-achiever status is residence in Somaliland. Students in other locations have a probability of just 0.06 – or a likelihood of 6 per cent – of being low-achievers; among students in Somaliland, that probability jumps to 0.28.

The case of Somaliland is straightforward, because Somaliland systematically underperforms other locations in terms of average learning outcomes as well. Other findings are more revealing, though. Among female students in schools that lack a toilet specifically dedicated to girls, average performance is lower than that for female students in schools with such a toilet; however, the difference in *mean* performance is relatively modest – just 4.4 points for numeracy, and just under 3 points for English literacy. At the same time, female students in the same schools are significantly more likely to fall into the low-achieving category, with the likelihood jumping from 12.2 per cent to 19.2 per cent when we compare schools with a toilet to those without. A 57.4 per cent increase in probability is substantively large, by any standard.

The results in Figure 14 also serve to further emphasise the role of several barriers that have stood out across multiple analyses. Students in rural schools, for example, are prone to underachievement, as are students in households where adult household members are, themselves, minimally educated.

Learning Benchmarks

Before turning to the project's second core outcome, we establish learning outcome benchmarks, against which midline and endline results will be compared. The learning cohort selected at the baseline included girls from Grade 6 through Form 2. Girls in the learning cohort, as in past GEC evaluations, will be tracked over time, and recontacted at the midline and endline. Because most girls in the cohort will have advanced two grades during the project's lifecycle, it is important that their learning outcomes – rather than being compared to their own scores at the baseline – be compared to grade-appropriate benchmarks.

To establish benchmarks for currently lower-grade girls, EGEP-T uses the scores of upper-grade girls at the baseline. For currently upper-grade girls, we drew a supplemental benchmark sample of Form 3 and Form 4, whose scores serve as the benchmark for girls in Forms 1 and 2 at the baseline. Table 37 clarifies the relationship of the learning cohort to their benchmarks.

TABLE 37: BENCHMARK GROUPS FOR ASSESSMENT AT MIDLINE AND ENDLINE

Age/Grade at Baseline	Benchmark Group at Midline	Benchmark Group at Endline
Grade 6	Grade 7 (Cohort girls)	Grade 8 (Cohort girls)
Grade 7	Grade 8 (Cohort girls)	Form 1 (Cohort girls)
Grade 8	Form 1 (Cohort girls)	Form 2 (Cohort girls)
Form 1	Form 2 (Cohort girls)	Form 3 (Benchmark Sample)
Form 2	Form 3 (Benchmark Sample)	Form 4 (Benchmark Sample)

The benchmark sample was collected from sampled secondary schools. Field teams identified eligible Form 3 and Form 4 girls, and randomly selected up to six girls, stratified by grade. Specifically, two Form 3 girls and four Form 4 girls were selected, where possible. Form 4 girls were overrepresented because the evaluation team and RI planned to use Form 3 respondents to the learning assessment pilot – conducted jointly by RI and CARE – to supplement the benchmark sample reported here. However, a review of the pilot data showed dramatically lower scores among Form 3 respondents in the pilot than in the benchmark sample drawn exclusively from RI schools.

For this reason, we recommend using the benchmark sample exclusively – despite the smaller sample of Form 3 girls that this entails – for comparisons at midline and endline.[162]

In practice, field teams were not able to meet this sample size target in each school; in total, learning assessments were conducted with 56 Form 3 girls, and 108 Form 4 girls. The benchmark values established for the midline and endline, respectively, are provided in Table 38, with benchmarks for numeracy, Somali literacy, and English literacy provided in the top, middle, and bottom panels of the table. Table 39 provides equivalent benchmarks for out-of-school girls identified at the baseline. Note that benchmark learning outcomes for out-of-school girls were only collected for those girls aged 11 to 18 years. As we discuss in Section 2.6, above, this shortcoming stems from the fact that girls outside this age range were mistakenly excluded from the transition and out-of-school girl samples. As a result, no benchmarks are available for older out-of-school girls who may be recruited for assessment in the midline and endline evaluations.

TABLE 38: LEARNING BENCHMARKS AT MIDLINE AND ENDLINE FOR MAIN GIRLS COHORT

Grade at Baseline	Score at Baseline	Midline Benchmark	Endline Benchmark
Numeracy			
Grade 6	59.50	66.50 (G7)	72.55 (G8)
Grade 7	66.50	72.55 (G8)	79.22 (F1)
Grade 8	72.55	79.22 (F1)	81.64 (F2)
Form 1	79.22	81.64 (F2)	85.97 (F3)
Form 2	81.64	85.97(F3)	90.06 (F4)
Somali Literacy			
Grade 6	65.80	76.31 (G7)	80.14 (G8)
Grade 7	76.31	80.14 (G8)	84.93 (F1)
Grade 8	80.14	84.93 (F1)	88.36 (F2)
Form 1	84.93	88.36 (F2)	91.20 (F3)
Form 2	88.36	91.20 (F3)	91.26 (F4)
English Literacy			
Grade 6	26.27	31.78 (G7)	41.45 (G8)
Grade 7	31.78	41.45 (G8)	58.90 (F1)
Grade 8	41.45	58.90 (F1)	63.95 (F2)
Form 1	58.90	63.95 (F2)	70.48 (F3)
Form 2	63.95	70.48 (F3)	78.26 (F4)

[162] To be concrete, Form 3 girls in the benchmark sample (n=56) averaged literacy and numeracy scores of 80.2 per cent and 86.0 per cent. In contrast, Form 3 girls in the pilot sample (n=56) averaged 66.8 per cent and 74.2 per cent. Importantly, if the benchmark and pilot samples were aggregated, Form 2 girls would currently exceed their midline benchmark in both literacy and numeracy. The data collected for the benchmark sample, because it is representative of EGEP-T secondary schools, and matches a logical progression of scores from Forms 2 through 4, provides a more reliable benchmark against which to measure midline and endline learning performance.

TABLE 39: LEARNING BENCHMARKS AT MIDLINE AND ENDLINE FOR OUT-OF-SCHOOL GIRLS

Age at Baseline	Score at Baseline	Midline Benchmark	Endline Benchmark
Numeracy			
Age 11	11.8	4.6 (Age 12)	12.6 (Age 13)
Age 12	4.6	12.6 (Age 13)	30.1 (Age 14)
Age 13	12.6	30.1 (Age 14)	18.3 (Age 15)
Age 14	30.1	18.3 (Age 15)	26.4 (Age 16)
Age 15	18.3	26.4 (Age 16)	34.6 (Age 17)
Age 16	26.4	34.6 (Age 17)	34.8 (Age 18)
Age 17	34.6	34.8 (Age 18)	No benchmark
Age 18	34.8	No benchmark	No benchmark
Somali Literacy			
Age 11	10.3	9.2 (Age 12)	13.6 (Age 13)
Age 12	9.2	13.6 (Age 13)	23.1 (Age 14)
Age 13	13.6	23.1 (Age 14)	15.5 (Age 15)
Age 14	23.1	15.5 (Age 15)	18.6 (Age 16)
Age 15	15.5	18.6 (Age 16)	40.4 (Age 17)
Age 16	18.6	40.4 (Age 17)	37.8 (Age 18)
Age 17	40.4	37.8 (Age 18)	No benchmark
Age 18	37.8	No benchmark	No benchmark
English Literacy			
Age 11	13.4	10.8 (Age 12)	12.7 (Age 13)
Age 12	10.8	12.7 (Age 13)	18.5 (Age 14)
Age 13	12.7	18.5 (Age 14)	11.6 (Age 15)
Age 14	18.5	11.6 (Age 15)	13.8 (Age 16)
Age 15	11.6	13.8 (Age 16)	21.2 (Age 17)
Age 16	13.8	21.2 (Age 17)	20.1 (Age 18)
Age 17	21.2	20.1 (Age 18)	No benchmark
Age 18	20.1	No benchmark	No benchmark

4.3 Transition Outcome

The second core outcome for all GEC-T programming is successful transition. GEC-T defines successful transition largely based on the progression of upper-primary age students into secondary school or alternative, productive pathways, such as skills-training programmes. EGEP-T does not include support for alternative training or education programmes and is focusing particularly on encouraging transition through the school grades, including from primary to secondary school. As per the GEC MEL guidelines, successful transition will be measured on the basis of yearly progression across all targeted grades.

Table 40 describes successful transition outcomes for three 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 evaluation of EGEP-T is following girls who are enrolled in Form 2 at the highest at the baseline, these same girls will be enrolled in Form 4 at the endline, assuming successful transition. As a result, there is no need to define transition rates from secondary school into other forms of education, such as university, or employment for the

purposes of evaluation. Finally, girls currently out-of-school are considered to have successfully transitioned if they enrol or re-enrol in school.

TABLE 40: DEFINING A "SUCCESSFUL TRANSITION", BY GRADE LEVEL

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

Benchmarking Transition Outcome

The goal of the baseline evaluation, with regard to transition, was to establish benchmark values within EGEP-T communities for different age cohorts. The baseline evaluation employed a split sampling approach, in which cohort girls participating in learning assessments were selected from within project schools. As described in detail in the methodological discussion above, this approach means that all learning cohort girls, by definition, successfully transitioned in the previous year.

To establish a baseline expected level of transition, therefore, the evaluation randomly sampled households in EGEP-T communities, and recorded transition outcomes over the previous year for all children in the household aged 11-18 years.[163] In the midline and endline evaluations, transition rates among cohort girls will be assessed against these benchmark values.

The top panel of Table 41 reports both aggregate and disaggregated transition rates for girls aged 11-18. Respondents are disaggregated by age and classified by the transition pathway they followed. Consistent with the definition of transition above, the first two pathways listed (in-school progression and re-enrolment in school) are considered success, while the remaining three pathways are considered unsuccessful transition.

[163] As previously discussed in Section 2.3, the age range of children included in this transition benchmark sample was mistakenly truncated. Per RI's MEL Framework, the appropriate age range included children up to age 22, while the sample drawn targets only those up to 18. Given that some members of the in-school cohort will be 21 years old at the time of the endline, future evaluation waves will need to account for the mismatch in age ranges between the cohort and the transition benchmark sample.

TABLE 41: BASELINE TRANSITION RATES AND PATHWAYS, BY AGE (FEMALE SAMPLE)

Benchmark Group – Female Only							
		Benchmark Transition Pathways					Transition Rates
Age	Sample Size (#)	In-school progression	Re-enrolled in School	Held back a grade	Drops out	Remains out of school	Successful transition rate per age (%)
11	135	89	14	7	7	18	76.3
12	116	82	8	8	3	15	77.6
13	103	72	3	9	4	15	72.8
14	101	76	4	4	7	10	79.2
15	100	75	4	3	5	13	79.0
16	92	63	2	6	4	17	70.7
17	94	61	1	3	5	24	66.0
18	77	47	1	3	8	18	62.3
Overall	818	565	37	46	48	130	72.7*
Benchmark Group – Male Only							
		Benchmark Transition Pathways					Transition Rates
Age	Sample Size (#)	In-school progression	Re-enrolled in School	Held back a grade	Drops out	Remains out of school	Successful transition rate per age (%)
11	63	47	5	6	1	4	82.5
12	64	49	2	3	2	8	79.7
13	73	58	5	2	1	7	86.3
14	61	53	0	3	1	4	86.9
15	51	41	0	4	3	3	80.4
16	49	39	1	4	3	2	81.6
17	37	29	0	1	3	4	78.4
18	34	21	0	3	2	8	61.8
Overall	432	337	13	26	16	40	80.2*

*The overall transition rate accounts for sampling weights designed to match underlying population parameters, ensure that each community is equally weighted in the sample and that households with multiple children are not over-weighted.

The sample size available for analysis includes 818 girls. Overall, 73.8 per cent of the sample successfully transitioned between last year and this year, with the vast majority of those successes involved grade-to-grade progression. Total drop-outs in the sample were 48, and total re-enrolments were 37. As the right-most column shows, transition rates are steady at between 72.8 and 79.2 per cent for girls between the ages of 11 and 15. Beyond the age of 15, however, transition rates begin to decline, falling to a low of 62.3 per cent at age 18.[164]

Before turning to results by age group and – in the next section – transition rates among demographic and other subgroups, it is worth highlighting the relatively high transition rates observed, in the aggregate, in Table 41. Given the barriers to girls' school completion, year-on-year transition of approximately 73 per cent among girls aged 11-18 is higher than might have been expected.

[164] Note that our analysis excludes girls who were enrolled in Form 4 last year, because it is unclear what successful transition for these girls would consist of.

We believe there are three primary reasons accounting for these higher-than-expected transition rates. First, while this report constitutes the baseline evaluation of EGEP-T, the project is being implemented in communities where the previous EGEP project also took place. The first phase of EGEP involved attempts to improve community support for girls' education, and remove financial obstacles to school completion, among a wide range of other interventions that should have prompted higher transition rates. It is, therefore, not altogether surprising that transition rates are higher than expected in the sampled communities. Second, data collection for the baseline evaluation took place after the start of EGEP-T programming, which included efforts to promote enrolment at the start of the school year. To the extent that these efforts prompted either re-enrolment or continued enrolment year-on-year, the reported transition rate may be overstated.[165]

Third, the transition sample is beset by issues of self-selection that may produce unnaturally high transition rates. The transition sample was drawn from a random sample of households in EGEP-T communities; at each household, the evaluation team completely enumerated all members of the household aged 11 to 18 years, and recorded sufficient information about their current and past schooling to calculate transition rates.[166] Critically, the evaluation only enumerated *current* members of the household, excluding girls and boys who were age-eligible but who no longer lived in the household. Children aged 11-18 years that have left the household may do so for a number of reasons, including pursuing employment, getting married, and attending school elsewhere. But, in most cases, we expect transition rates to be lower among this group of children than among those who remain in their households.[167] We discuss this issue in greater detail in our discussion of migration and transition in Section 4.4, below; here, we simply note that the extent of out-migration from sampled households suggests that true transition rates – if they were to take account of all children, rather than just those who currently live in the sampled households – could be significantly lower than those reported in Table 41.

Because the age range of students enrolled in each grade can vary widely in the Somali context, the ages studied in Table 41 do not map neatly to specific grades. As a result, disaggregating the findings separately by grade can provide additional insight regarding when girls tend to drop out of school. As noted above, drop-out rates for girls are relatively steady from age 11-15, but drop precipitously after that point. When we disaggregate the analysis by grade, we find a somewhat lower transition rate between upper primary and lower secondary (from G8 to F1) in the sample than for the grades surrounding it. The transition rate for girls from G7 to G8 was 88.2 per cent, while the transition rate from G8 to F1 was 81.8 per cent, indicating that the transition from upper primary to lower secondary is a small barrier – vis-à-vis more general in-school progression from grade to grade – but not one that causes a sharp decline in transition.

The bottom panel of Table 41 replicates this analysis for boys aged 11-18. The sample size of boys included in the transition benchmark is significantly lower, at just 432 total boys aged 11-18, which suggests caution regarding the interpretation of age-specific transition rates, given the especially small sample sizes for these subgroups.[168] At nearly every age, transition rates for boys are markedly

[165] We discuss the possible bias produced by this issue, tentatively, in Section 4.5

[166] It is for this reason that the transition sample covered 559 households but included data on 818 eligible girls.

[167] In practice, current enrolment rates among girls who migrated out of their households in the last year are just 26.0 per cent, and just 7.3 per cent among girls who migrated in order to get married.

[168] It is not clear why the evaluation team found fewer boys (n = 432) than girls (n = 818) in the target age range. Sampling variance is an unlikely explanation for the gender gap, because field teams included *all* eligible children in each household in the benchmark sample, rather than randomly sampling from among them. Differential rates of out-migration from the household cannot explain the gap either, because the evaluation team documented more

higher than those for girls, with gaps ranging from 2 to 15 points in most cases. Interestingly, the age at which boys' transition rates start to decline noticeably comes later than for girls: while girls' transition begin declining at age 16, boys' rates do not begin to decline in earnest until age 18 – a finding that is consistent with extensive qualitative evidence, and the expectations that shaped project design, that girls lag boys most markedly in terms of completing a full cycle of education.

For completeness, we also report benchmark transition rates for a combined sample of boys and girls aged 11-18 in Table 42. Transition rates in this sample are slightly higher, owing to the higher aggregate transition rate for boys – 76.8 per cent among boys and 72.7 per cent among girls.

TABLE 42: BASELINE TRANSITION RATES AND PATHWAYS, BY AGE (MALE AND FEMALE)

Benchmark Group – Female & Male							
		Benchmark Transition Pathways					Transition Rates
Age	Sample Size (#)	In-school progression	Re-enrolled in School	Held back a grade	Drops out	Remains out of school	Successful transition rate per age (%)
11	198	136	19	13	8	22	78.3
12	180	131	10	11	5	23	78.3
13	176	130	8	11	5	22	78.4
14	162	129	4	7	8	14	82.1
15	151	116	4	7	8	16	79.5
16	141	102	3	10	7	19	74.5
17	131	90	1	4	8	28	69.5
18	111	68	1	6	10	26	62.2
Overall	1,250	902	50	73	65	170	73.4*

*The overall transition rate accounts for sampling weights designed to match underlying population parameters, ensure that each community is equally weighted within the sample and that households with multiple eligible girls and boys are not over-weighted.

4.4 Sub-Group Analysis of Transition Outcomes

In this section, we provide additional analysis of baseline transition rates in EGEP-T communities. The analysis focuses largely on geographic and demographic differences in transition rates, in an effort to shed light on structural barriers to transition that could be targeted by project interventions.

Table 43 reports transition rates across a variety of subgroups, along with the sample size of eligible girls included in each subgroup.[169] Several major trends are evident from the results. First, there are significant differences across project locations, with girls in Somaliland and Banadir much more likely to transiton than those in Puntland and, especially, Galmudug. Despite the small sample size, the differences in transition rates across locations are statistically significant at the 5 per cent level.

aggregate girls (153) than boys (97) migrating away from the households in the sample. While lacking an explanation for the gender gap, we note that it is consistent with Forcier's previous experiences in Somalia and other post-conflict settings (e.g., Sudan and South Sudan), where a persistent gender gap of approximately 5-8 per cent between the number of male and female household members is often observed.

[169] Sampling weights were utilized in each calculation in this section to ensure a valid comparison between subgroup results and the overall results in the top row of the table.

Second, transition rates are not markedly lower in rural and drought-affected communities. Indeed, they are higher, on average, in both settings, which runs counter to expectations regarding the impact of drought especially.[170] With regard to drought, one explanation for the higher-than-expected transition rate concerns EGEP-T's interventions to this point in the project lifecycle. RI and its partners have been working in the affected communities for several years and – prior to baseline data collection – EGEP-T was already implementing an extensive set of drought-response efforts. This fact may help explain the higher-than-expected transition rates in both rural and drought-affected communities, since all communities classified as drought-affected in our data are also rural.[171] On the other hand, transition rates in conflict-affected communities are significantly lower, though it is difficult to separate any effect of conflict from the more generally lower rates of transition in Puntland, as 88.9 per cent of conflict-affected households are drawn from Puntland.[172]

The role that drought plays in the project's operating context is a key issue; while it is not limited strictly to the effect of drought on transition and dropout rates, this is an area where drought is expected to have major effects. We dedicate additional analysis to the role of drought in Sections 5.6 and 6. One point which is worth noting here is the idea that drought has conflicting effects on transition and dropout rates depending on the local context. That is, rural communities faced with drought are likely to see a large number of dropouts, as children or entire families move to urban areas to escape the drought's worst effects. This was a common theme in the qualitative data, where participants indicated that entire schools might close due to drought, where drought might hamper livelihoods to a degree that parents cannot afford school fees, and where migration to urban areas disrupts children's schooling.[173] But qualitative interviewees also emphasised the inverse of this problem in urban areas: in response to drought, children arrive in urban areas and area schools are often *more crowded* as a result.[174] These narratives highlight the complex relationship between drought and enrolment rates, which are likely to confound simple analysis of transition rates in EGEP-T, owing to the mixed urban/rural composition of the project's schools.

Third, girls in households with demographic characteristics that are traditionally considered challenging for the achieving strong female educational outcomes are no less likely to transition than the typical girl in the sample. For instance, girls in female-headed households are marginally more

[170] Households are classified as urban/rural, drought-affected, and conflict-affected on the basis of RI's analysis and data collection.

[171] It is also possible that self-selection explains part of this phenomenon. In rural communities that have been affected by drought, there is likely to be a significant amount of out-migration, with families moving to urban areas to escape the effects of the drought. If households with the thinnest available safety net are the first to migrate, this may explain the relatively high transition rates *among the households that remain behind in rural areas*. That is, if the least resilient households have left rural areas, it would explain our higher-than-expected transition rates among those households that remain.

[172] Similarly, it is possible that the low transition rates for urban respondents are driven partially by the fact that 85.1 per cent of Puntland respondents and 100 per cent of Galmudug respondents are urban, and both locations have lower transition rates. However, even when we restrict the analysis to Puntland alone, rural respondents have dramatically higher transition rates (86.5 versus 68.2 per cent), suggesting that some other factor explains the urban-rural gap in transition rates. This is an issue that we address in greater detail in the regression analysis provided in Section 5.6, where we are able to control for multiple competing explanatory variables.

[173] KII with MOE Quality Assurance official, Galkayo, Puntland; FGD with mothers, Galgadud; KII with female teacher, Maroodi Jeeh, Somaliland; FGD with CEC members, Hargeisa, Somaliland.

[174] FGD with CEC members, Bari, Puntland; KII with female teacher, Lafa-ruug, Somaliland; KII with male teacher, Garbadadar, Somaliland.

likely to transition, and girls with a non-literate caregiver are approximately as likely to transition as the sample average.

TABLE 43: TRANSITION RATES AMONG GEOGRAPHIC AND DEMOGRAPHIC SUBGROUPS

Subgroup	Successful Transition Rate	Subgroup size (n)
All girls	72.7%	818
Somaliland	79.9%	284
Puntland	68.8%	376
Galmudug	50.6%	53
Banadir	78.2%	101
Urban	71.2%	616
Rural	81.4%	202
Drought-Affected	76.9%	210
Conflict-Affected	67.8%	121
Female HoH	74.8%	507
HoH No Formal Education	71.0%	456
Non-literate caregiver	73.9%	227
Family migrates seasonally	51.6%	32
Family migrated seasonally within last 6 months	48.3%	28
Family has lived in community one year or less	58.5%	41
Family does not own a mobile phone	75.4%	141
Family does not own a smartphone	67.2%	494
Family is unable to meet basic needs without charity	70.9%	195
Family has gone to sleep at night feeling hungry many or most days	67.3%	105
Family would withdraw girl from school – hypothetical scenario	60.3%	175
Family would encourage marriage rather than schooling	48.8%	95

*The subgroup transition rates account for sampling weights designed to match underlying population parameters, ensure that each community is equally weighted within the sample, and that households with multiple eligible girls and boys are not over-weighted.

However, other demographic characteristics *do* present a significant barrier to transition—namely, a propensity of the household to migrate on a seasonal basis. Among the very small subset of girls residing in households that report occasional seasonal migration, transition rates are approximately 20 percentage points lower than the sample average.

While both migration and newcomer status within a community are associated with dramatically lower transition rates, more general household economic barriers are only associated with mild drops. Mobile phone ownership exhibits a mixed relationship: girls in households that do not own a mobile phone actually have marginally higher transition rates (75.4 per cent) than the sample average, though children in households without a smartphone appear to be disadvantaged. More telling is relationship between household food security and transition – among households where family members have gone to sleep hungry many or most days over the past 12 months, transition rates are 67.3 per cent.[175]

[175] Caution should be exercised when interpreting this finding, primarily because of well-known problems with recall over such long (12 month) time periods. At the same time, to the extent that respondents are unable to accurately

Finally, our last subgroup focuses on households in which caregivers express lukewarm support for girls' education. Specifically, girls' caregivers were presented with a hypothetical scenario in which their sister is ill and needs financial assistance with her medical bills. We asked caregivers to indicate what they would do if given the choice between two unpalatable options for helping their sister: selling some of their household assets (household goods or an animal) or withdrawing their daughter from school to save on her school fees. Girls in households with a caregiver who indicates that they would withdraw their daughter from school – 21.2 per cent of the sample – are significantly less likely to be enrolled in school. This finding suggests that, in addition to challenges stemming from geography and demography, the attitudes of adults in one's household have a powerful effect on a girl's educational opportunities and outcomes.[176]

Small subgroup sample sizes may drive some of the results presented here, of course, because subgroups often contain fewer than 50 female respondents. In cases where subgroups were particularly small, we repeated our analysis using the full sample of girls *and* boys, to determine whether a larger sample size altered the findings. In general, there is no evidence that including boys in the sample would dramatically change the conclusions for the small subgroups reported. For instance, transition rates among this slightly larger sample are still very low in Galmudug, and in households that migrate seasonally transition rates still hover around 50 per cent.[177]

In other cases, small subgroup sample sizes do not appear to be driving the results, but differences in transition rates between boys and girls are clear. Specifically, households in which caregiver responses indicate relatively weak support for girls' completion of education show divergence between girls' and boys' transition rates, as Table 44 shows. In households where caregivers indicated that they would withdraw their daughter from school when faced with a difficult financial choice, transition rates are just 60.3 per cent among girls, but 80.9 per cent among boys. A less stark, but still telling, gap is found between boys and girls in households where a caregiver indicated support for a girl getting married over completing her schooling. These gaps, in our view, reflect the importance of household-level attitudes on enrolment and transition decisions, and the fact that some girls face attitudinal barriers within their own households that boys do not.

TABLE 44: TRANSITION RATES, AMONG HOUSEHOLDS WITH WEAK SUPPORT FOR GIRLS' EDUCATIONAL COMPLETION

Subgroup	Transition Rate – Girls	Transition Rate – Boys
Family would withdraw girl from school – hypothetical scenario	60.3%	80.9%
Family would encourage marriage rather than schooling	48.8%	58.0%

recall the extent of their household's hunger over the past 12 months, we expect this error to be random in nature, rather than biasing our finding regarding transition rates in either a positive or negative direction.

[176] We chose to focus our analysis regarding household attitudes on these hypothetical scenarios because, as we show in Section 5.5 below, attitudes toward girls' education often appear to be very supportive in principle, but less supportive in practice. While the vast majority of community members sampled reported that they hope their daughters will complete a university education, many fewer were willing to see their daughters pass up marriage opportunities to stay in school. This dichotomy is often reflected in the qualitative data, where respondents would express strong support for their daughers to continue schooling through the highest levels, qualified by the statement that they would do so "if they could afford it" (see, e.g., FGD with mothers in Nugal, Puntland).

[177] In Galmudug, transition rates among boys and girls combined are just 48.8 per cent.

The findings in this section are consistent with other quantitative evidence presented later in this report. In some cases, the findings are also consistent with the views of qualitative interviewees, though this is far from universal. For a more detailed quantitative analysis of the predictors of transition, see Section 5.6, where we employ linear regression to study transition at the individual level. In that analysis, most of the bivariate findings from this section are confirmed.

In contrast to the findings above, where drought-affected communities appear to have higher transition rates than non-drought communities, the qualitative evidence emphasises the importance of drought in transition. According to interviewees, drought fuels movement, especially from rural areas into urban centres like Hargeisa or Mogadishu, which makes re-enrolment less likely.[178] This fact also has consequences for the schools in urban centres, who can both be overwhelmed by an influx of new potential students, and feel it necessary to help those students re-enrol despite a frequent inability to pay school fees.[179] Even among individuals who do not migrate, enrolment rates are affected, as the drought impacts the economic environment and reduces their ability to pay school fees.[180]

Interviewees also mentioned the role of conflict in reducing transition rates, though this was a less common theme than that of early marriage or drought. Conflict, like drought, causes migration of entire households and of individual children, which can reduce the ability of families to pay for schooling, and increase obstacles to re-enrolment in their new home.[181] At the same time, conflict is relative, and many areas of Somalia are more peaceful now than they have been in the past – several interviewees noted that civil conflict used to reduce enrolment and impact education, but does not at this time.[182] Levels of conflict vary dramatically across geographic space and – in some cases, conflict arises for a month or two and then fizzles. This may mean that interviewees understate the importance of conflict – conflict may have subsided, but if children have been forced to stay home from school and have not re-enrolled, the damage from conflict has already been done and is long-lasting.

A notable shortcoming of the analysis of transition outcomes is the evaluation's inability to measure many of the most common reasons cited for girls dropping out of school. A key example of this problem is the role that household chores play in preventing girls from adequately pursuing educational opportunities. Girls are responsible for an outsized share of household chores, and qualitative interviewees often noted this as a fundamental barrier to their education: in particular, teachers and students highlighted the fact that chores can make a girl late for school, or encourage absences; chores also cut into study time and prevent girls from performing as well in school and on final exams.[183] These issues can be exacerbated by a family's circumstances; for instance, one group of boys participating in an FGD indicated that, if a parent gets sick, "the girl will remain at home to take care of them because the boys don't do any chores at home".[184] Despite this importance, the

[178] FGD with CEC members, Puntland; KII with female teacher, Somaliland; FGD with mothers, Banadir; FGD with fathers, Puntland.

[179] KII with male teacher, Somaliland.

[180] FGD with fathers, Bari, Puntland; FGD with fathers, Sahil, Somaliland.

[181] KII with male teacher, Puntland; FGD with fathers, Bari, Puntland.

[182] FGD with fathers, Mogadishu, Banadir.

[183] KII with female teacher, Hargeisa, Somaliland; KII with male teacher, Galgadud; FGD with mothers, Galgadud; FGD with boys, Sheikh, Somaliland.

[184] FGD with boys, Sheikh, Somaliland.

evaluation did not measure individual girls' chore burdens, which would have allowed us to assess the relationship between chores and transition outcomes.

Migration of Children and Transition

In addition to household attitudes, migration status is among the strongest predictors of low transition rates documented in Table 43. The importance of migration in the Somali context is not demonstrated by the data in Table 43, however, as relatively few households in the sample migrate seasonally or have migrated within the past year.[185] Household migration is not the only way in which migration affects girls' educational completion; in fact, many girls and boys migrate alone, without the rest of their households. These situations are likely more important sources of dropouts.

The data collected from households shed some light on this issue. Across 559 households, 250 boys and girls aged 11-18 have left their households in the past 12 months. The reasons for leaving are not all migration-related, as this figure includes deceased children, and those who have left to get married, who may not have migrated away from their community. Across the same set of households, 56 boys and 100 girls in this age group have unambiguously migrated away from their households. The most common reason for migration among boys is to enter the labour force, at 44.6 per cent; the most common reason among girls is to get married, at 41.0 per cent.

Enrolment rates among this population are extremely low, as might be expected. For girls who have migrated, overall enrolment is just 26.0 per cent, and just 7.3 per cent among girls who migrated in order to get married. Among boys who have migrated, enrolment rates are just 35.7 per cent, and particularly low amongst those who migrated to get married, to enter the labour force, or to live with another family member.

The extent of children's out-migration is important to note: somewhat conservatively, we estimate 156 children migrated away from their household within the last year, representing an emigration rate of around 10.6 per cent among children 11-18 years in target communities.[186] In other words, among all children aged 11-18, 10.6 per cent leave their households *and* migrate away from their communities *per year*.[187] Such high rates present significant obstacles to maintaining enrolment and encouraging transition. When children migrate to get married or enter the labour force, it is extremely unlikely that they will re-enrol in their new community. But, even among those who migrate to join the household of a family member, re-enrolment rates are very low – just 31.9 per cent across genders combined. The act of migrating itself may be disruptive, regardless of the reason; moreover, households who send their children to live with family members may be facing particularly difficult financial or other circumstances, making it less likely that they can afford schooling costs for their children.

The importance of migration by children, specifically, is reinforced by the qualitative evidence. Many interviewees highlighted the role that early marriage plays in preventing girls from completing a full

[185] To be clear, we mean this statement relatively. Seasonal migration, even within our sample, is much higher than in most contexts outside Somalia. However, we believe that our sample underrepresents the extent of migration in Somalia at large.

[186] The only children included in this count are those who left their households in the last year, are not deceased, and who now live in either another village or another country. In practice, this approach may underestimate the extent of out-migration, because some children who left their households are not listed as either living in another village or another country, and we exclude them from our count.

[187] Among boys, we conservatively estimate the out-migration rate to be 11.1 per cent per year, while we estimate a rate of 10.4 per cent among girls.

cycle of education.[188] But early marriage is especially problematic for girls, both because they tend to get married at younger ages, and because convention dictates that newlywed girls and women move to join their husband's household. As such, marriage and migration typically occur together and reinforce one another, with migration throwing up additional hurdles – such as finding a new school and ensuring they are enrolled at the same learning level as at their previous school – for newlywed girls who wish to continue their education.

Migration of children may help explain a puzzling finding from the previous section, which suggested that drought-affected communities had higher transition rates than others. In practice, migration of children may result in biased estimates of transition rates, causing higher rates in drought-affected communities and lower rates in non-drought communities. There are two mechanisms by which this might occur. First, children in drought-affected communities may be more likely to leave their households, either to get married, enter the labour force, live with a family member, or for other reasons. If this is the case, these children are not included in estimates of transition rates, because they have left the households being surveyed. To the extent that drought and the out-migration of children are correlated, this may inflate the transition rates of drought-affected communities. Second, children who migrate as a result of drought typically go to urban centres unaffected by drought. To the extent that these children join the households of family members but do not enrol, this would deflate the transition rates of non-drought communities. In both cases, drought is still partially or wholly responsible for the failure to transition, but the communities in which the student is counted varies in a manner that could bias our subgroup estimates of transition rates. More generally, the extent of migration by children is an important factor to consider when establishing targets for transition, as we do in the next section.

4.5 Cohort Tracking and Targets for Transition

The baseline evaluation has two distinct transition samples. The first is the sample of learning cohort girls and boys. Transition rates for these respondents were not described in detail in this report, because – as noted above – they were sampled from within intervention schools. As a result, their transition rates are, by definition, 100 per cent. However, this group, in combination with OOS girls sampled through household surveys at the baseline, will comprise the evaluation sample for transition rates at the midline and endline, assessing year-on-year transition relative to their in-school status at the baseline.

The second sample is a transition benchmark sample, described in Section 4.3. It is this set of girls, randomly drawn from communities in which EGEP-T schools operate, that will guide the establishment of target transition rates, against which the cohort girls will be measured in future evaluation waves.

The baseline transition rate in the benchmark sample among girls was 72.7 per cent, where successful transition is defined as grade-level progression or re-enrolment in school. This represents a moderately-high transition rate; as a result, we recommend a conservative approach to setting targets for future evaluation points, as dramatic improvements above 72.7 per cent are unlikely.

Furthermore, as discussed in the previous section, it is likely that the benchmark transition rates overstate the frequency of successful transition within project communities, because children who had migrated away from home were not included in the benchmark calculations. Since a share of the

[188] KII with male teacher, Somaliland; KII with male teacher, Puntland; FGD with fathers, Somaliland; KII with female teacher, Galmudug; KII with MoE official, Banadir; KII with male teacher, Puntland; FGD with boys, Somaliland.

cohort girls are likely to migrate away from home and fail to transition as a result, the comparison against the benchmark may need to be adjusted at the midline and endline.[189]

Finally, these targets may need to be revised in the future, depending on the extent and severity of any drought conditions over the life of the project. Our recommendations are based on benchmarks derived over the last year, when the drought was extensive but not at its most severe. Depending on the evolution of drought conditions going forward, these targets may be too ambitious.

TABLE 45: TRANSITION OUTCOME TARGETS, BY EVALUATION POINT

Target Type	Baseline Rate	Year 2	Year 3
Target suggested in Outcome Worksheet	72.7%	79.7%	86.7%
Evaluator-Proposed Target	72.7%	76.2%	78.7%

Our recommendations for transition targets are provided in Table 45. The first row provides targets in line with the FM's guidance, based on the baseline level – year-on-year improvements of 7 percentage points. In contrast, we recommend a less stringent standard, which is graduated over time: a 3.5-point increase in the first year, followed by a 2.5-point increase in the final year of the project. This approach reflects the fact that there are diminishing marginal returns to project activities, and that improvements become increasingly difficult as transition rates climb toward 100 per cent.

The lower transition targets suggested in Table 45 are justified for two primary reasons. First, when the evaluation assesses transition rates from the baseline to the midline, it will only partially capture the transition impacts the project had prior to baseline data collection. As noted in the limitations section of this report, baseline data collection began after EGEP-T project activities were already underway. As a result, some members of the primary girls cohort – whose transition rates will be assessed in future evaluation waves – were impacted by project activities, and may have re-enrolled or transitioned from one grade to another as a result of those activities.[190] While the baseline sought to collect data on previous years' enrolment for all girls, it is likely that some degree of project impact was and will be missed.

Second, and more importantly, the benchmark transition rate reported in Table 45 as the baseline (72.7 per cent) likely overstates the true benchmark transition rate. The benchmark in question was established by randomly sampling households in EGEP-T communities, fully enumerating all girls and boys in the household aged 11-18, and classifying them as having successful transitioned or not since

[189] Also, when evaluating transition rates the midline, the evaluation team will need to account for the fact that some members of the learning cohort were not enrolled at the project's start, but enrolled before the baseline was conducted. The evaluation team collected data regarding cohort members' enrolment status in the previous school year to facilitate proper comparisons at the midline.

[190] Using data from the baseline, we can tentatively quantify the extent of bias that is caused by excluding positive transition outcomes that occurred after the start of EGEP-T but prior to baseline data collection. Specifically, for the transition benchmark sample, the baseline evaluation collected data on enrolment in the prior school year, which we can use to determine how many children re-enrolled in school before the baseline was conducted. Out of 232 girls and boys who were out of school at the end of the previous year, 24.6 per cent re-enrolled *this* year. In future evaluations, it will be possible to compare this re-enrolment rate with those obtained at the midline and endline; if re-enrolment prior to the baseline is similar to that immediately prior to the midline and endline, it will suggest that the timing of baseline data collection produced an evaluation that missed pre-baseline impact of EGEP-T interventions.

the previous year. However, the children in this sample had already been exposed to EGEP-T project activities, ranging from community awareness campaigns or back-to-school campaigns, to drought relief food distributions, to bursary support. Importantly, many girls and boys enrolled in EGEP-T project schools were included in this transition sample, though those selected into the main cohort were excluded from the benchmark sample. Nonetheless, this means that members of the benchmark sample are not a true benchmark – the calculated benchmark does not measure the transition rate that would have obtained in the absence of EGEP-T interventions, because the sampled respondents were exposed to such interventions prior to data collection.

The bias incorporated into the benchmark transition sample justifies a less stringent target for transition in the midline and endline, because baseline transition rates are almost certainly starting from a lower point than 72.7 per cent, as implied in Table 45.

4.6 Sustainability Outcome

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.

One of the sustainability indicators of the EGEP-T project, this measure focuses on changing community attitudes toward girls' education through awareness-raising activities. Note that this indicator consists of two distinct components: exposure to project awareness-raising activities, and community attitudes toward girls' school completion. At midline and endline, the goal of the evaluation will be to assess whether changes in community attitudes are correlated with exposure to project awareness-raising activities. In contrast, it is not possible to measure attitudinal change at baseline. As a result, the discussion in this section focuses on establishing three outcomes at the baseline: the extent of exposure to project awareness-raising activities, current community attitudes toward girls' education, and the correlation between exposure and attitudes at an individual level. For the purposes of establishing a baseline level of this indicator, we focus on current community attitudes, though we note that there is significant overlap between the analysis of community attitudes provided here with the analysis provided in Section 5.5, where we discussion the intermediate outcome of community attitudes and behaviour.

Relief International has several concurrent efforts to raise awareness and change opinions within the communities where RI works. Past research – and findings documented in this report – has shown that families can be resistant to their girls going to school because of the financial costs and the decrease in time the girl has to help around the house. The overall goal of the EGEP-T project is to increase girls' educational performance and the share of girls who complete a full cycle of education. However, without changing attitudes, it is unlikely that girls would be allowed to continue their educations after the EGEP-T project ends. These community awareness-raising efforts not only help change the minds of parents whose girls are in EGEP-T schools but also help change the minds of parents of girls at other schools. These activities thus have a broader impact outside of the EGEP-T schools.

Exposure to Awareness-Raising Efforts

The current coverage of EGEP-T awareness-raising activities was measured in two ways. First, randomly-selected households in EGEP-T project communities were surveyed regarding their exposure to awareness-raising activities. Specifically, adult community members were asked

whether they were aware of door-to-door visits regarding the importance of girls' education or back-to-school campaigns encouraging girls' enrolment. Importantly, the household sample was a random sample within project communities, meaning that many non-beneficiary households (i.e. without children enrolled in an EGEP-T school) were included; as such, the sample is a good indication of overall community attitudes, rather than attitudes exclusively among beneficiary households. Second, similar questions were asked of teachers in project schools, as teachers may recall such events to a greater degree than parents.[191]

Based on surveys with community members and teachers, it is clear that awareness-raising efforts are already underway and have relatively high penetration within communities. As shown in Table 46, 46.0 per cent of community members report being aware of door-to-door efforts that occurred over the previous year, and 47.0 per cent report being aware of other types of back-to-school campaigns in the same period.

As expected, teachers are generally more aware of these efforts, with approximately 77 per cent aware of either type of campaign among the full sample of teachers at EGEP-T schools.

TABLE 46: AWARENESS ACROSS PROJECT LOCATIONS

Event	Overall	Somaliland	Puntland	Galmudug	Banadir
Community Members Sample					
Door-to-door campaigns	46.0%	47.1%	54.3%	40.9%	23.8%*
Back-to-school campaigns	47.0%	50.3%	56.7%	30.0%	22.6%*
Teachers Sample					
Door-to-door campaigns	77.3%	78.9%	76.6%	62.7%	79.6%
Back-to-school campaigns	77.2%	76.0%	83.0%	50.6%	70.1%

While the overall results suggest that nearly half of households have heard of girls' education events, project location seems to influence this. In Banadir, only around one in four households had heard of such events.[192] Despite people in Banadir being less likely to hear about these events, there was no difference between teachers in Banadir hearing about these events compared to teachers in other project locations. It may be that there is less penetration of such events into communities in Banadir compared to other project locations.

Beyond project location, there were no significant differences between households or teachers in urban and rural areas hearing about such events but teachers were again much more likely to know about these events taking place than households within the community. Drought had no significant influence on whether households had heard of these events while drought had a significant impact on only whether teachers had heard about back-to-school campaigns, not the other two types of events.

It is important to note that the data collected during the evaluation did not distinguish between awareness-raising efforts of RI and its partners, on one hand, and other NGOs or organisations working in the same communities, on the other. In other words, some respondents reporting

[191] A significant shortcoming of this analysis concerns the range of campaigns that were included in the survey. Respondents were asked about their familiarity with just two types of awareness-raising campaigns, which does not reflect the full range of activities that EGEP-T plans to engage in. At the midline and endline, more attention should be given to the survey design to ensure better representation of different campaign types.

[192] Similar rates were found in Hirshabelle. However, the sample size in Hirshabelle – where only one school was included in the evaluation sample – is sufficiently small that it is excluded from the disaggregated results.

exposure to awareness-raising efforts may have been exposed to actions taken by other organisations. While this shortcoming in measurement may influence the results to some degree, we do not expect to find *other* widespread awareness-raising efforts in most of the communities where EGEP-T is being implemented.[193]

Community Attitudes

Community attitudes present the crux of this indicator, as awareness-raising efforts are expected to influence community attitudes over the project's life. However, as noted above, there is significant overlap between community attitudes discussed for this indicator, the following sustainability indicator, and the more general intermediate outcome focused on community attitudes and behaviour. For a fuller analysis of community attitudes, we refer readers to Section 5.5 of this report. Here we focus on a select set of indicators concerning community attitudes, which are used to score this sustainability indicator.

In each household, the caregiver of a girl was asked a series of questions designed to gauge their support for educating their daughter, specifically, and girls, more generally. The first three measures we employ, based on the survey questions below, are straightforward:

- What level of schooling would you like [girl] to achieve?
- To what extent do you agree that, even when funds are limited, it is worth investing in [girl's] education?
- To what extent do you agree that a girl is just as likely to use her education as a boy?

The findings from these questions, disaggregated by project location, are provided in Table 47. In the first row, we focus on the share of respondents who indicate that they would like their girl to complete university. While this is obviously an aspirational question in an environment where so few girls complete university, we focus on this high level of aspiration because so few caregivers selected lower levels of education. Similarly, in the second and third rows, we focus on the share of respondents who "strongly agreed" with the two statements in question, because – in general – agreement was very high across the board.

TABLE 47: CAREGIVER ATTITUDES TOWARD GIRLS' EDUCATION

Attitudinal Indicator	Overall	Somaliland	Puntland	Galmudug	Banadir
Wants girl to complete university	89.0%	89.7%	87.7%	85.1%	91.5%
Strongly agree it is worth investing in girls' education	77.1%	85.7%	75.1%	78.4%	53.6%
Strongly agree girls use education as much as boys	77.8%	85.9%	71.6%	78.4%	70.2%

The results in Table 47 suggest that caregivers have high aspirations for their girls, and that the vast majority of them feel strongly about the value of educating their girls. Even here, however, notable differences emerge across project locations. Support for girls' education appears to be highest in Somaliland, which has the highest rates of support on all three indicators. Less clear-cut is the lowest-performing location, though this appears to be Banadir, where just 53.6 per cent of

[193] The evaluation team and RI have discussed the need to better measure awareness-raising efforts in future waves of the evaluation. This includes specifying the organization that is behind these campaigns, as well as distinguishing between different types of campaigns and different modes of delivery.

caregivers strongly agreed with the statement that it is worth investing in their daughters' education even when finances are limited.

As we discuss in greater detail in Section 5.5, the findings above represent – in our view – caregivers' support for girls' education *in the abstract* or in principle only. We are not surprised that the majority of caregivers profess support for girls' education in these questions because, in the absence of hard choices and resource constraints, we are confident that most Somali families would prefer that their daughters attend and complete school. Considering this, a better test of community attitudes should examine support for education when more difficult decisions must be made about household resource allocation.

To assess community support for education *in practice*, we asked caregivers to indicate how they would react to two hypothetical situations. In the first, respondents were asked to imagine that their sister was sick and needed financial assistance with her medical bills. In this scenario, respondents were asked to indicate whether they would sell household goods or an animal to help their sister, or whether they would withdraw their daughter from school. This scenario presents the precise type of difficult choice with which many Somali families are faced on a regular basis. A second scenario asked respondents to imagine that their daughter was 15 years old and had a marriage proposal. Respondents were asked to indicate whether they would encourage their daughter to get married, continue their schooling, or both.

TABLE 48: CAREGIVER REACTIONS TO HYPOTHETICAL SCENARIOS REGARDING GIRLS' EDUCATION

Indicator	Overall	Somaliland	Puntland	Galmudug	Banadir
Would withdraw daughter from school to help sister	22.2%	22.0%	24.1%	38.1%	6.0%
Would encourage daughter to get married	29.0%	18.8%	31.3%	47.8%	44.1%

In Table 48, we report the results of these indicators. In the first case, we analyse the share of respondents who state that they would withdraw their daughter from school to help their sister.[194] Overall, 22.2 per cent of caregivers surveyed stated that they would withdraw their daughter from school rather than sell household goods. Contrary to the results presented previously, respondents in Banadir were the *most* supportive of girls' education in this hypothetical scenario, with just 6.0 per cent reporting that they would withdraw their daughter from school.

Findings from the second scenario were similar, though our finding regarding Banadir flips from scenario to scenario. In this case, we calculate the share of respondents who express support for their daughter to either accept the proposal, or accept the proposal *and* continue schooling. We consider the latter an indication of weaker support for schooling – relative to a preference that their daughter simply reject or postpone the proposal to continue schooling – because the number of girls who get married and continue schooling is extremely low. Therefore, encouraging a girl to get married and continue schooling is unlikely to result in them continuing their studies.

[194] A non-trivial number of respondents were not able or willing to choose between the two options; these respondents are included in the denominator, although inability to choose between the two options might also be an indicator of weaker support for girls' education, at least relative to those who unequivocally state that they would sell household goods instead.

Discussion

There are a number of barriers to community attitudinal change regarding girls' education. Aside from cultural beliefs, awareness-raising efforts are only effective if they have strong coverage throughout the community. At this stage in the project's life, there are indications that awareness-raising efforts have already penetrated large sections of the project communities, with nearly half of all respondents – including, crucially, not just those in beneficiary households, who are more likely to hear about EGEP-T activities – having heard of awareness-raising campaigns in the last year. And, while it is possible that some of these awareness-raising activities were sponsored by other organisations or by the government – or took place under the previous project phase, EGEP, if respondents misunderstood the time period in question – the penetration rates are still promising.

Nonetheless, there is significant room for improvement at the level of community attitudes. While the majority of caregivers express support for girls' education in theory, there are two reasons to be sceptical regarding figures based on such abstract notions of support. First, support in the abstract is divorced from the realities that Somali families face when actually making decisions about whether to enrol their daughters in school – realities rooted in resource constraints and other barriers to education. Second, support among caregivers is almost certainly higher than that among the general adult population of project communities, because caregivers are more likely to be concerned about the best interests of the children for whom they care. The fact that support is relatively high for education in principle, but that a significant minority of caregivers would discourage schooling or remove their daughters from schooling under plausible circumstances suggests that community attitudes require significant additional progress.

Indicator Grade: 2 - 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

This sustainability indicator, focusing on the actions taken by men and boys, was developed to understand whether male support, specifically in the form of tangible actions, for girls' education increase during the EGEP-T project. This is important because male support will be critical to sustaining the gains in girls' education and graduation made by EGEP-T. As the primary decision-makers in many households, men often decide if their daughters will attend school and for how long. In addition, boys at school can make the learning experience for girls pleasant or not and can help them learn or not. Thus, it is imperative that boys also are encouraging and respectful to their female peers in the classroom.

Because tangible actions taken to support girls' education can take a wide variety of forms, this evaluation attempted to triangulate this indicator by both assessing a number of different specific actions boys and men could take, and by collecting data from multiple different populations of respondents. For actions by boys, we collected data from both teachers, the main girls learning cohort, and cohort boys. Teachers were asked whether boys provide encouragement to girls in the classroom. Girls in the learning cohort were asked a similar question. In addition, girls in the learning cohort were presented with a hypothetical scenario in which they asked their brother for help completing chores, so that they could dedicate additional time to studying for examinations. Girls were asked to judge whether they thought their brother would be willing to help them. Finally, boys were also presented with the inverse of the hypothetical scenario presented to girls: if asked by their

sister to help them with household chores so that their sister could have additional study time before an examination.

For actions by men, we collected data from teachers and head teachers, presenting them with a question regarding fathers' attendance at a hypothetical parent-teacher meeting. While this question measures the perceptions of teachers and head teachers regarding fathers' interest in and support of education for their daughters, we feel it is indicative of one tangible action engaged fathers would take in this realm. Finally, we also asked members of randomly-selected households to assess the attitudes of men in their community toward girls' education, and the actions they take in support of their daughters' education.

Supportive Actions from Men and Fathers

As we discuss elsewhere in this report, the baseline evaluation documented broadly positive attitudes toward girls' education among all community members. However, action resulting from these positive attitudes is not always seen in the data. Fathers, despite overwhelmingly supporting girls' education, do not involve themselves with going to actual meetings or helping their daughters learn in a substantial way. Similarly, boys may support girls' education but that support may not translate into tangible actions.

In this section, we focus on concrete actions that men and fathers might take in support of girls' education. Naturally, this section does not provide a comprehensive assessment of all the different ways men and fathers can support girls' education; rather, it focuses on easily observable, measurable ways that can be standardized across locations and time.

TABLE 49: FATHERS' EXPECTED ATTENDANCE AT PARENT-TEACHER MEETINGS

Survey	Overall	Somaliland	Puntland	Galmudug	Banadir
Father would attend meeting with daughters' teacher					
According to head teacher	28.6%	31.9%	16.7%	18.2%	61.9%
According to teachers	33.1%	30.4%	26.9%	50.1%	50.0%

Our first approach concerns fathers only. It measures how likely fathers are to attend parent-teacher meetings with their daughter's teacher. Unfortunately, only limited data was collected from fathers at the baseline; as such, much of our analysis relies on data collected regarding the perceptions of *others* about fathers' actions or willingness to take action. In this case, we surveyed head teachers and teachers, separately, and presented them with a hypothetical scenario in which a girl was not doing well in school. In the hypothetical, respondents were asked which, if any, parents would be more likely to attend a parent-teacher meeting they scheduled. As Table 49 shows, 28.6 per cent of head teachers believe that the imaginary girl's father would attend the meeting, while 33.1 per cent of teachers believe the same. Note that this figure does not indicate that mothers would not also attend; rather, the metric is whether fathers would attend with or without the girl's mother. In general, teachers appear sceptical of the idea that fathers would attend meetings regarding their daughters' education.

Importantly, the rates presented in Table 49 may reflect minimal support among fathers for education in general, rather than minimal support for girls' education specifically. When head teachers were asked whether fathers would attend a similar parent-teacher meeting regarding their

son, just 29.3 per cent believed that they would, a rate only marginally higher than that found when the question focused on girls.[195]

A broader view of male views regarding girls' education is provided by respondents to the household survey. This population is representative of the overall EGEP-T communities, not just households whose children are direct beneficiaries, via enrolment, of EGEP-T schools. As such, data collected from respondents to the household survey provide the most comprehensive snapshot of male attitudes and their willingness to take action in support of girls' education.

Importantly, the population sampled within households consists of caregivers of girls aged 11-18 years. These caregivers were asked to assess the support of men in their community – and, later, men in their household – for girls' education. The top panel of Table 50 reports findings focused on caregiver perceptions of male attitudes in their communities. For instance, caregivers were asked whether they agree or disagree with the following statement: "men in this community support girls to complete schooling before marriage." We consider the share of respondents who agree a lot or a little with that statement, and report the overall percentage and the percentages across project locations in the top row of Table 50.

TABLE 50: MEN'S SUPPORT FOR GIRLS' EDUCATION ACROSS PROJECT LOCATIONS

Survey	Overall	Somaliland	Puntland	Galmudug	Banadir
Men in this community					
Support girls to complete schooling before marriage	75.2%	85.7%	66.4%	80.1%	64.6%
Support girls to continue schooling after marriage	66.8%	76.0%	66.9%	56.3%	46.9%
Help their daughters with their schoolwork	73.8%	83.6%	68.2%	66.5%	66.7%
Would hire house-help to give daughter time to study	73.4%	81.5%	68.6%	59.7%	73.8%
Men in this household					
Prefer girls to complete schooling before marriage	84.0%	90.3%	80.3%	91.4%	69.1%
Prefer girls to continue schooling after marriage	75.7%	84.0%	73.1%	70.6%	60.0%
Help their daughters with their schoolwork	81.2%	87.6%	80.5%	79.3%	63.8%

Overall, perceived support for the completion of girls' education is relatively low. It is important to remember that the reported figures include those who agree either a lot or a little; this means that 24.8 per cent of respondents, overall, disagree with the notion that men in their community support girls to complete schooling before marriage. As with many of the other metrics of male support for girls' education, we see dramatic variation across project locations, with Banadir consistently underperforming the sample average. In general, respondents in Somaliland view male support for girls' education as strongest.

[195] Unfortunately, similar data was not collected from teachers, who were only asked the hypothetical question regarding a meeting with parents of a female student.

Even more telling, perhaps, are the results in the bottom panel of Table 50. The same set of caregivers was asked to assess the attitudes of men in their own household regarding girls' education. Here we would expect more positive responses, partially due to social desirability bias – respondents may not want to admit that their household members are not wholly supportive of girls' education – and partially because most individuals have more positive perceptions of themselves and their families than they do of the broader community.

Surprisingly, we still document relatively low levels of perceived support. While 84.0 per cent of respondents say that men in their household prefer girls to complete schooling before getting married, this means that a full 16.0 per cent disagree. Given how innocuous the statement is, this is surprising; we say that it is innocuous because most people would feel free to say that they support girls completing their schooling in this hypothetical situation, where no costs or trade-offs are presented. Even more startling is the lukewarm support found in Banadir, where just 60.0 per cent of respondents say men in their household support girls continuing their schooling after getting married.

Qualitative interviews showed mixed support from men about girls attending school. Some participants said that support varied within their own community with some men supporting their daughters and others not supporting their daughters' education.[196] Some people even reported that fathers supported their daughters more than mothers as mothers wanted their daughters to stay home and assist with chores.[197] Another CEC member reported that fathers had visited the school to see how their daughters were doing and to encourage them in their schooling.[198]

Supportive Actions from Boys

To what extent do boys and men in EGEP-T communities differ with regard to tangible actions taken to support girls' education? Surveys completed with both teachers and girls at EGEP-T project schools asked respondents a series of questions regarding boys' support for girls' education.[199] First, at a relatively basic level, both groups were asked whether boys encourage classroom participation by girls. Second, both groups were asked whether boys encourage girls to continue their education.

These metrics of boys' support for girls' education capture an important aspect of any support – emotional or social support for the value of girls' education. As we discuss elsewhere, some community members perceive girls' education to have little value, partially because girls tend to get married and stay at home. But if their male peers encourage them to participate and, especially, to continue their education, girls will see additional value in schooling, especially since those male peers may later become potential marriage partners.

The top section of Table 51 reports the share of girls who agree (a lot or a little) that boys encourage them to participate in class. Moving horizontally across the table, we can see that 76.9 per cent of girls, overall, agree with this statement, with the highest rates found in Somaliland and the lowest rates – by far – found in Banadir. The row below it reports similar findings, but drawn from the survey of teachers, who generally report similar levels of agreement with the statement as girls.

---

[196] Focus Group Discussion with CEC. Sahil, Somaliland.

[197] Focus Group Discussion with CEC. Hargeisa, Somaliland.

[198] Focus Group Discussion with CEC. Galkacyo, Puntland.

[199] The sample of girls referenced in this section is the main learning cohort of girls (n = 1,609). The teachers referenced here are a randomly sampled set of teachers in EGEP-T schools (n = 516); in other places within this subsection, we refer to head teachers, who constitute a distinct sample from that of teachers.

Importantly, teachers do not show the same scale of location-to-location swings that are found in responses from girls – for instance, teachers in Galmudug and Banadir are no less likely to disagree with the statement than teachers in Somaliland and Puntland, a finding that contrasts sharply with that among girls.

TABLE 51: BOYS' ACTIONS IN SUPPORT OF FEMALE EDUCATION, BY PROJECT LOCATION AND SETTING

Survey	Overall	Somaliland	Puntland	Galmudug	Banadir	Urban	Rural
Boys Encourage Girls to Participate in Class							
According to girls	76.9%	85.1%	78.2%	68.4%	52.0%	72.6%	87.4%
According to teachers	82.5%	80.4%	85.0%	79.5%	85.8%	82.6%	82.3%
Boys Encourage Girls to Continue Education							
According to girls	76.2%	84.6%	78.1%	67.8%	49.3%	71.8%	87.0%
According to teachers	79.5%	78.8%	81.8%	72.3%	81.7%	79.2%	80.2%
Boys help their sister with chores often or sometimes							
According to girls	67.9%	70.5%	70.1%	71.0%	54.4%	65.3%	74.6%
According to boys	80.7%	78.6%	82.8%	91.8%	65.2%	78.7%	86.8%
Boys would help their sister, if asked, before exam							
According to girls	72.9%	61.2%	85.2%	79.4%	68.8%	74.3%	69.4%
According to boys	86.1%	69.8%	90.5%	95.9%	100%	86.6%	84.6%

While encouraging classroom participation is valuable, EGEP-T's attitudinal change efforts are focused especially on encouraging community members to support girls through the complete cycle of education. In other words, the goal is not necessarily to encourage support for girls' education in general, but to foster support for the idea that girls should complete their education through secondary school and beyond. The next metric of boys' support for girls' education is directly related to this issue, as the girls and teachers were asked whether boys encourage girls to continue their schooling. The findings here are similar to those regarding participation in class, with significantly less agreement found among girls in Galmudug and Banadir, relative to those in other areas. Likewise, girls in rural areas believe that boys are more supportive of their continued education than those in urban areas.

An additional metric focuses on tangible ways in which boys can support girls in their studies. Boys and girls enrolled in project schools were asked questions about helping girls with chores when they needed to complete homework or study for exams. The evaluation approached this question in two ways. First, we asked both groups of respondents how often boys help their sisters with chores.[200] Among girls, we report the share of girls who say that their brother helps them with chores "often" or "sometimes." Among boys, we report the share who say that they help their sister with the same frequency. Unsurprisingly, boys perceive themselves to be more helpful than their sisters perceive them to be. But the findings across locations exhibit similar patterns to those regarding boys' encouragement of girls' education – lower in urban areas and, especially, Banadir.

Our second question regarding chores took this approach further, and asked respondents about help with chores immediately prior to an examination. For girls, we asked them to judge whether their

[200] Note, these questions were only asked of boys who indicate they have at least one sister and girls who indicate they have at least one brother.

brother would help them with their chores in order to give them more time to study before an exam; as before, the inverse question was posed to boys, asking them whether they would help their sister. Overall, 72.9 per cent of girls believe that their brother would be somewhat or very likely to help them with their chores under these circumstances, while 86.1 per cent of boys state that they would be very somewhat or very likely to help their sisters.

Relative to the quantitative findings reported above, qualitative data emphasised both positive and negative ways in which boys influence girls' education. On one hand, interviewees generally report that boys are supportive of girls in the classroom, helping them with schoolwork when necessary.[201] At least one interviewee described boys bothering girls in class, but most interviewees adamantly stated that girls were at no risk from boys in the school.[202] Some teachers also commented that their religion created an environment where boys always respected girls in the classroom because it would be against Islam to not respect them.[203]

On the other hand, concerns regarding the safety and comfort of girls in class remain, and concerns specifically about the influence of boys' actions are not entirely without merit. To be clear, many of the comments which inform this opinion are not specific to EGEP-T schools, but reflect concern about the actions of boys in Somali communities more generally. One teacher reported that there used to be a fear among parents that, if they sent their daughter to school, she would be sexually abused by boys or men.[204] The same interviewee also reported that boys sometimes bother girls in class by, for instance, taking the girls' books.[205] More commonly interviewees expressed concern about the safety of girls traveling to school, and the problems that harassment from boys along the road can pose.[206] Given that this indicator focuses on the actions of boys in EGEP-T communities, writ large, it is worth noting that the actions of some boys in the community – though likely not those enrolled in EGEP-T schools – pose barriers to girls' education.

Likewise, boys and girls say that boys would be willing to help their sisters with chores but in the qualitative interviews, boys seemed more resistant to participating in what they consider "women's chores."[207] It seems more common for girls to go to their brothers for help with their homework than for help with chores.[208]

Discussion

Positive attitudes towards girls' education are common among men in EGEP-T communities. However, like attitudes toward girls' education among the broader population, it is arguably easy to affirm support for something in theory, when it is costless to the respondent. Therefore, a more stringent measure of support is the extent to which men (and boys) take tangible actions to support girls' education and girls completing a full cycle of schooling. Fathers, despite overwhelmingly

[201] FGD with boys, Nugal, Puntland; FGD with girls, Sahil, Somaliland; FGD with boys, Mogadishu, Banadir; FGD with fathers, Sahil, Somaliland; KII with Female Teacher, Hargeisa, Somaliland.

[202] FGD with Boys, Sahil, Somaliland; FGD with CEC members, Mogadishu, Banadir; FGD with CEC members, Dhusamareeb, Galmudug.

[203] KII with Male Teacher, Mogadishu, Banadir.

[204] KII with male teacher, Garbadadar, Somaliland.

[205] Ibid.

[206] FGD with mothers, Galkayo, Galmudug; FGD with girls, Maroodi Jeeh, Somaliland; FGD with mothers, Galgadud, Galmudug.

[207] FGD with Boys, Sahil, Somaliland.

[208] FGD with Boys, Hargeisa, Somaliland.

supporting girls' education, do not involve themselves with going to actual meetings or helping their daughters learn in a substantial way. Boys may support girls' education but their actions are not always actually supportive.

Boys and men are crucial groups to get on board with supporting girls' education. Their parental involvement or their support in the classroom would help girls both stay in school and learn better while they are there. When boys support girls' schooling, girls' may feel less pressure to drop out and get married, or may feel supported to continue their education after marriage. Likewise, when men become involved in girls' education, it sends a signal to their daughters that their education is important to their families. It appears currently that boys and men overwhelmingly support girls' education. Boys were likely to say they would help with their sisters' chores and most households reported men in the community supporting girls' education. However, it seems that the qualitative data does not always support this conclusion as some of the boys in focus groups seemed resistant to helping with 'female' chores and teachers expressed concern about boys harassing the girls in the classroom. In addition, fathers may show support for girls' education but they are not involved enough in their children's education.

It is important for EGEP-T to be a success to both get fathers involved in their daughters' education and to make sure boys are creating safe environments for girls to learn in the classroom. EGEP-T should both work on encouraging men to be involved with their daughters' education and make sure boys are helping their sisters with chores and convincing both men and boys of the importance of girls' education. Because there were not large differences between project locations or urban/rural areas, EGEP-T should make sure they have an aspect of this type of programming in all of their project locations. Without making sure boys and men are on board with girls' education, it will be challenging to ensure that girls learn while they are in the classroom and are allowed to continue their education through secondary school.

Indicator Grade: 1 – Latent

Sustainability Indicator 3 – Community Leaders' Advocacy

Percentage of EGEP target communities where community leaders are leading campaigns and advocacy events

The third sustainability indicator for EGEP-T focuses on the role of community leaders, and their particular role in influencing local attitudes. This indicator was developed to understand how involved community leaders are in encouraging girls' education and retention in schools. By encouraging community leaders to lead campaigns and advocacy events promoting girls' education, it will help create a sustainable environment, at the community level, for girls to continue their schooling.

The majority of the data reviewed for this indicator come from qualitative interviews with community members, including mothers, fathers, and CECs. In addition, the findings are triangulated using quantitative data where possible – among households sampled from EGEP-T communities, respondents were asked whether they had observed community leaders leading events in support of girls' education. To be precise, the question asked of household members was "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." The results in this indicator focus on responses indicating awareness of "events where community leaders gave

speeches encouraging girls' education."[209] While this type of event is only one of several in which a community leader could participate, it is also a common type of event, and likely the one which community members will most readily recall.

Quantitative Findings

Of the 559 households surveyed in the EGEP-T baseline, only 48.6 per cent had heard of an event where a community leader led an awareness-raising event for girls' education. Hearing of these events was most common in Puntland where 61 per cent of respondents reported hearing about an event. In Galmudug, and Banadir, very few people had heard of these events with only 17.0 per cent, and 29.8 per cent of people recalling these events, respectively. There was a marginal four point difference in rural versus urban areas.

It is important to remember that household recollections are not foolproof. This is suggested by the fact that households in the same communities often disagree with one another about the occurrence of such an event – as such, this metric should be seen as an indicator of the frequency and prominence of such events, but not a precise measure of how many communities have witnessed them.

Moreover, one should be cautious interpreting some of the findings in Table 51, above, where disaggregation by project location results in relatively small samples in, especially, Galmudug (n = 44). It is possible that the poor performance of Galmudug, especially, stems from sampling variance, given the small sample size. To alleviate concern regarding small location-specific samples, we utilise similar data collected from teachers about their recollection of events led by community leaders advocating for girls' education. To the extent that the results mirror those found among community members, it would instill additional confidence in the results. The bottom panel of Table 52 reports results among teachers in project schools.

TABLE 52: SHARE OF RESPONDENTS WHO RECALL COMMUNITY LEADERS' LEADING EVENTS SUPPORTING GIRLS' EDUCATION

Subgroup	Share of Respondents
Household Respondents	
Overall	48.6%
Somaliland	46.7%
Puntland	62.9%
Galmudug	17.0%
Banadir	29.8%
Rural	51.3%
Urban	47.4%
Teacher Respondents	
Overall	73.7%
Somaliland	70.8%
Puntland	81.9%
Galmudug	64.4%
Banadir	66.5%

[209] Importantly, other sustainability sub-indicators focus on questions regarding back-to-school campaigns and other awareness-raising events. In contrast, our focus in this section is exclusively on public events where community leaders explicitly lent support to the cause of girls' education.

Rural	72.8%
Urban	74.2%

Not surprisingly, it was far more common for teachers to recall such events, likely because they are more deeply engaged with the topic of education in their communities. Overall, 73.7 per cent of teachers surveyed reported that they recalled a community event in support of girls' education, led by a community leader, during the previous year. Teachers in Puntland were the most likely to have heard of an event with 81.9 per cent saying that they had heard of such an event. Teachers in Galmudug had the lowest reporting rate with only 64.4 per cent saying that they had heard of a community leader-led event for girls' education. As with the broader sample of community members, there was not a large urban-rural divide, with a minimum 1.4 point difference between teachers in urban and teachers in rural areas.

Across both teachers and households, respondents in Galmudug, and Banadir were far less likely to report hearing about these events than respondents in other locations, and these areas should be a focus of EGEP-T interventions targeted at encouraging community leaders' involvement in awareness-raising efforts. These findings are consistent with the fact that caregiver respondents in Banadir and Galmudug rated religious leaders as much less supportive of girls' education than respondents in Puntland and Somaliland. This could be because of the context in Banadir, especially, where rates of violence driven by extremism are higher than in other regions. It may be that community leaders in Banadir either do not support girls' education or they are concerned about the consequences of speaking out in favour of girls' education.

Qualitative Findings

In line with the quantitative findings reported above, a wide range of interviewees participating in FGDs and KIIs recalled community events – led by community leaders – encouraging girls' education. For instance, in one focus group discussion with boys in Somaliland, they reported seeing or hearing community leaders talk about girls' education on the television and radio.[210] Many of these awareness-raising events seemed to be conducted or led by those leaders already affiliated with the school in some way, such as principals, CECs, or school committees. Several interviewees in Puntland and Somaliland said that girls' education was often promoted during school ceremonies where children and parents are likely to attend.[211] In an FGD with CEC members in Galkayo, one respondent said that religious leaders were very enthusiastic about girls' education, as it made them better mothers.[212] Likewise, in Banadir, a female teacher in a key informant interview stated that religious leaders supported the school.[213] And, in Bari, a male teacher participating in a KII stated that religious leaders had come to the school and made speeches about education, particularly regarding the importance of educating children with disabilities.[214]

Importantly, many interviewees described community events that did not appear to be affiliated directly with EGEP-T, RI, or its implementing partners. Of course, in most cases it is not possible to determine, from interview transcripts, who organised or sponsored an event, so some of the events reported by interviewees likely resulted directly from EGEP-T activities. But others clearly were not

[210] FGD with boys, Maroodi Jeeh, Somaliland.

[211] KII with female teacher, South Galkayo, Galmudug; FGD with boys, Maroodi Jeeh, Somaliland.

[212] FGD with CEC, Galkayo, Puntland.

[213] KII with Female Teacher, Mogadishu, Banadir.

[214] KII with Male Teacher, Bari, Puntland.

sponsored by EGEP-T, such as a women's organisation in Galmudug that encourages girls' education.[215] This is a potentially important point, because the goal of EGEP-T is *not* to simply hold community events; rather, the goal is to build sufficient support for girls' education about community leaders that they hold such events themselves. As such, while we cannot clearly attribute the cause of most of the events described by interviewees, we find it encouraging that some events appear to be occurring without any direct link to EGEP-T itself, suggesting some degree of local motivation for raising awareness and encouraging girls' education.

While several interviewees described events that were directly linked to schools or the education sector – such as pro-education given at graduation ceremonies – these are not the only types of events possible. Several qualitative interviews suggested that religious leaders are advocates of girls' education; quantitative findings also confirm the perception among community members that religious leaders support girls' education.[216] However, most of the qualitative interviews were insufficiently specific about the nature of support from religious leaders – thus, it is not clear whether these leaders were hosting events to promote girls' education, or just generally encouraging girls' education in a less public or specific manner. In the former case, such events would obviously be directly relevant to this indicator. But, even in the latter case, the fact that religious leaders are seen as supportive suggests two things: first, that religious leaders are not decrying girls' education – and are likely lending support to it in some fashion; and, second, that religious leaders are good potential partners who, with encouragement, would likely lead such events, even if they were not already doing so.[217]

Discussion

Community leaders and their influence should not be understated. Community leaders are often a trusted source of information and of value systems. If they promote girls' education, then it follows that community members may be more open to changing their minds about girls' education and its importance. Changing people's opinions is often one of the most important ways to ensure a project's sustainability because the community will continue the desired actions after the project has ended. Using community leaders to influence the community on girls' education is one of the best ways to change opinions and create sustainability. By achieving this sustainable outcome, transition outcomes will be more achievable. Girls will be encouraged to stay in school and do well in school as the community starts to support her.

Looking forward to the midline and endline, additional information needs to be collected on this indicator. The limited data on community leaders and their activities reduced our confidence in the results, and limited the extent to which results could be triangulated across sources and data types. There should be both specific questions in relevant quantitative surveys and qualitative interviews. While we now have baseline data on the percentages of people who recall community leaders leading events in support of girls' education, there should be more questions relevant to who the

[215] KII with Female Teacher, South Galkayo, Galmudug.

[216] For example, 77.9 per cent of caregiver respondents report that religious leaders in their community support girls' education, and 75.8 per cent report that religious leaders support girls completing their education before getting married. Broadly positive – though slightly less so than among caregivers – results obtain among teachers, 63.4 per cent of whom state that religious leaders are either somewhat or very supportive of girls' education. At the same time, 21.3 per cent of teachers polled state that religious leaders are very unsupportive, which suggests that opinions among this important group are sharply divided.

[217] That is to say, if religious leaders support girls' education but are not publicly advocating for it in a prominent way, they will be easier to convince to participate in awareness-raising activities than those who might oppose or be ambivalent toward girls' education.

community leader is, what kind of event it is, and others that would help to pinpoint the type of events that are happening. Additional data of this kind would allow EGEP-T to understand what kinds of events are most common, and what events might be best at encouraging girls' education.

Secondly, in Galmudug, Hirshabelle, and Banadir, it does not appear that these events are particularly common. As noted above, this likely stems from the context in which communities live: conflict could affect leaders' willingness and ability to speak out in favour of girls' education; and that same conflict has likely reduced the scope and frequency of *any* community events, given the fact that there are few safe spaces in which to hold such events. These regions should be a primary focus for EGEP-T in campaigns with community leaders for girls' education.

Currently many awareness-raising events are being held during school ceremonies or with girls who are already in school. To the extent that EGEP-T's focus is on improved learning and encouraging successful transition among girls currently enrolled, this approach is reasonable. But, to the extent that EGEP-T wishes to reach out-of-school girls in a broader way, they may need to make efforts to expand these events beyond school-related activities or places. Regardless, events held at schools are likely critical to building strong support among enrolled students and their families to continue schooling – reinforcing the importance of education, and promoting a view among enrolled students that their efforts are valuable.

EGEP-T should continue to work with religious leaders in this arena. Most religious leaders are perceived by members of their community to be supportive of girls' education, and encouraging them to speak out more vocally could be a particularly useful approach to encouraging enrolment and discourage dropouts. At the same time, the number of religious leaders who are not supportive of girls' education – according to the perceptions of community members – is surprising; to the extent that EGEP-T can carefully identify willing and sympathetic partners, they will have more success in encouraging religious leaders' participation. At the same time, those who are unsupportive should, perhaps, not be entirely neglected, as convincing them of the value of girls' education could pay significant dividends within their communities.

Indicator Grade: 2 – Emerging

Sustainability Indicator 4 – CEC Bursary Support

Percentage of CECs providing bursary support to marginalised girls in EGEP target schools

For many children in Somalia, fees charged for attendance at school – and/or additional costs of education, such as uniforms and supplies – present a significant barrier to enrolment. According to head teachers, the most common reason for students to drop out of school is an inability to pay school fees.[218] As such, it would be helpful for local institutions to provide financial aid to marginalised girls so they may attend school.

To measure baseline bursary support to girls, we focus on the sample of head teachers, who were asked whether the CEC currently supports girls with bursary support for their school fees. Head teachers were also asked to report the number of girls being supported, either partially or completely, by their CEC.

Among head teachers at 139 schools, 30.7 per cent report that their CEC provides bursary support to at least one female student. The number of girls supported varied dramatically, from a single

[218] For girls, 25.5 per cent of head teachers indicate that an inability to pay school fees is the most common reason for girls to drop out in a typical year.

student to as many as 142 in larger schools. Among the 38 schools reporting a specific number of girls supported, the average share of female students supported was 10.0 per cent. That is, among schools in which the CEC supports at least one girl, the average school had CEC support for 10.7 per cent of their female students. Across all schools, head teachers report that a total of 1,146 girls are currently being supported by their CEC.

As shown in Table 53, the share of CECs providing bursary support varies markedly by location, with the highest rates in Banadir, where 76.2 per cent of CECs are supporting at least one girl with her school fees. At the same time, among schools where bursary support is provided by CECs, CECs in Banadir tend to support a smaller share of their total female students, at 6.3 per cent. This fact may reflect the tendency for EGEP-T schools in Banadir to be somewhat larger than average.

TABLE 53: CEC BURSARY SUPPORT, AMONG DROUGHT-AFFECTED AND RURAL/URBAN COMMUNITIES

Grade at Baseline	Share of CECs Providing Bursary Support	Mean share of girls supported by bursaries, among CECs providing support
Somaliland	14.9%	23.1%
Puntland	30.0%	9.3%
Galmudug	18.2%	41.1%
Banadir	76.2%	6.3%
Drought-Affected	15.8%	15.8%
Not Drought-Affected	36.6%	9.3%
Rural	19.4%	10.7%
Urban	34.6%	9.9%

The finding regarding Somaliland is worth highlighting specifically, because Somaliland has formally abolished school fees at the primary level. Given this fact, it is surprising that 14.9 per cent of school CECs in Somaliland are providing bursary support to girls, especially since all sampled schools in Somaliland were described by head teachers as public. In actual practice, however, head teachers at a majority of primary schools in Somaliland report that their schools charge fees to students or their families.[219] Those Somaliland schools that report providing bursary support through their CECs all report that they charge school fees.

Overall, the level of bursary support does not appear to vary by school level; in both primary and secondary schools, the share of CECs providing support is similar. Across locations, schools in Banadir enjoy the greatest support from CECs in terms of bursary support, followed by Puntland. In Galmudug and Somaliland, only 18.2 and 14.9 per cent of schools report bursary support from their CECs.

Taking it one step further, we compared the performance of CEC bursary support under two conditions that can be expected to influence the financial resources of CECs: drought, and urbanicity. We expect that CECs in rural areas will be less likely to provide bursary support to girls; we also expect that the drought has negatively impacted CECs' ability to provide bursary support. As shown in Table 53, schools in drought-affected areas are much less likely to report bursary support from their CECs. Similarly, rural schools enjoy less bursary support. The difference between urban and

[219] Of primary schools sampled in Somaliland, 69.2 per cent (n = 27/39) report that they charge school fees.

rural schools reported in Table 53 actually underestimates the gap between the two, as urban schools also report higher numbers of girls receiving bursary support, on average.[220]

HOW DOES THIS AFFECT SUSTAINABILITY AND WHAT ARE THE BARRIERS?

Interview responses from CEC members and teachers emphasize the need for more funds in order to boost the quality of education (e.g. teacher salary, better sanitation, more facilities, books, and teaching equipment.[221] However, funds are also needed by households and students to cover school fees. When teachers were asked what was needed to be done in order to sustain girl's education, their response was "to pay fees [for] girls [who] are not able to pay school fees".[222] Other teachers also brought up the issue with the government, and how the "MoE should cover school fees…if the girls [are] not able to pay".[223] In a social environment where parents may not see the value of their daughters continuing and completing school, school tuition requirements may further increase the opportunity cost incurred by parents sending their daughters to school-even when tuition is free, the opportunity cost of school attendance is less time spent running errands for parents or helping with chores at home. In an interview with a local teacher, she stated that "when the mother doesn't have school fee, then she tells her daughter to stay at home, and wash the dishes, and work at home".[224] To remove this barrier, CECs and government institutions may need to prioritise funding more than awareness. Interviews indicate that even when parents are cognizant of the importance of education, the lack of financial aid still keeps children out of classrooms.[225]

Indicator Grade: 2 – Emerging

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

Community Education Committees are composed of local parents, community leaders, and school teachers that have come together to improve the quality of schools through raising awareness and providing financial and non-financial support to marginalised girls. As such, their operations and contributions should be strongly associated with general academic success of schools in their communities. To gauge the percentage of CECs in EGEP target schools that are providing match funding for school improvement initiatives, we focus primarily on quantitative data collected from school head teachers. Head teachers were asked whether their school had a functioning CEC, and whether their schools receive cash contributions from their CEC. Because many CECs contribute to schools in other ways, we also asked whether CECs were providing non-cash contributions to schools. These non-cash contributions typically took the form of donated time and labour by, for instance, painting classrooms, erecting a fence, maintaining school furniture, or managing school feeding programmes.

---

[220] In schools that report at least one girl receiving bursary support from their CEC, rural schools report that 19.5 girls, on average, receive support, compared to 30.5 girls in urban schools. These figures should be interpreted with caution, as only four rural schools report bursaries being provided.

[221] FGD with CEC members, Mogadishu, Banadir.

[222] KII with Teacher, South Galkayo.

[223] KII with Teacher, Bari, Puntland.

[224] KII with Teacher, Somaliland.

[225] FGDs with girls, Sahil, Somaliland.

Table 54 reports the share of schools receiving financial and non-financial support from their CECs. First, it is important to note that 12 of the head teachers interviewed indicate that they do not have a functioning CEC associated with their school.[226] Naturally, if a CEC is not functioning, it cannot provide either financial or non-financial support to schools; as such, we assume that schools without a functioning CEC are not receiving support of this kind. Overall, just 28.6 per cent of schools report receiving financial support from their CECs, while 33.6 per cent report non-financial contributions. Importantly, cash and non-cash contributions are not simply substituted for one another: 12.9 per cent of schools report receiving *both* types of contributions, while 50.7 per cent report that they have not received either kind of assistance.

TABLE 54: CEC ACTIVITY AND FINANCIAL SUPPORT

Response	Functioning CEC	Financial Support from CEC	Non-Financial Support from CEC
Yes	91.4% (128)	28.6% (40)	33.6% (47)
No	8.6% (12)	71.4% (100)	66.4% (93)
Total	100%	100%	100%

Disaggregating the results geographically, we find that there are significant differences in the performance of CECs, on this metric, across project locations. Figure 15 reports results by project location. In each of the four project locations, a minority of schools reported receiving cash contributions. Contributions of this kind were more common in Banadir and Galmudug, though the geographic differences were more pronounced with regard to non-cash contributions. When considering non-cash contributions, CECs in Galmudug performed best, with 72.7 of schools in the area reporting non-cash contributions by their CECs. Puntland schools were the lowest-performing on both metrics.

Focus groups with CEC members in Galmudug indicate that CECs have contributed money that was used for a variety of large projects, including constructing a school gate, "minor things like painting school", and "repairing billboards".[227] Other uses of CEC funds helped students cover their tuition fees.[228] CEC members also mentioned that parents have taken a more active role in improving the quality of education by, for instance, supporting feeding programs in local schools. Meanwhile, CEC members in Banadir reported that they were in the middle of a campaign which aims to raise funds for the construction of separate areas where female students can dress.[229]

This indicator of sustainability concerned CECs as a whole, and their ability to provide material support to their schools. Head teachers, though their capacity was not assessed specifically, are a critical component of this indicator, because head teachers sit on the CECs as members. The baseline evaluation focused on CECs as a whole, because EGEP-T's interventions focus on improving

[226] Enumerators did not define what it meant for a CEC to function during the interview. Rather, head teachers were asked whether their school currently had a "functioning Community Education Committee", and judged whether their CEC fit with their own definition of "functioning." This assessment, while subjective, reflects a meaningful judgment from an informed individual – head teachers are typically members of the CEC, and would certainly be aware of its activities, or lack thereof – that can be used to draw conclusions, at the least, about relative activity levels. Notably, a further four head teachers reported that, while their school had a functioning CEC, it had not met for over 12 months.

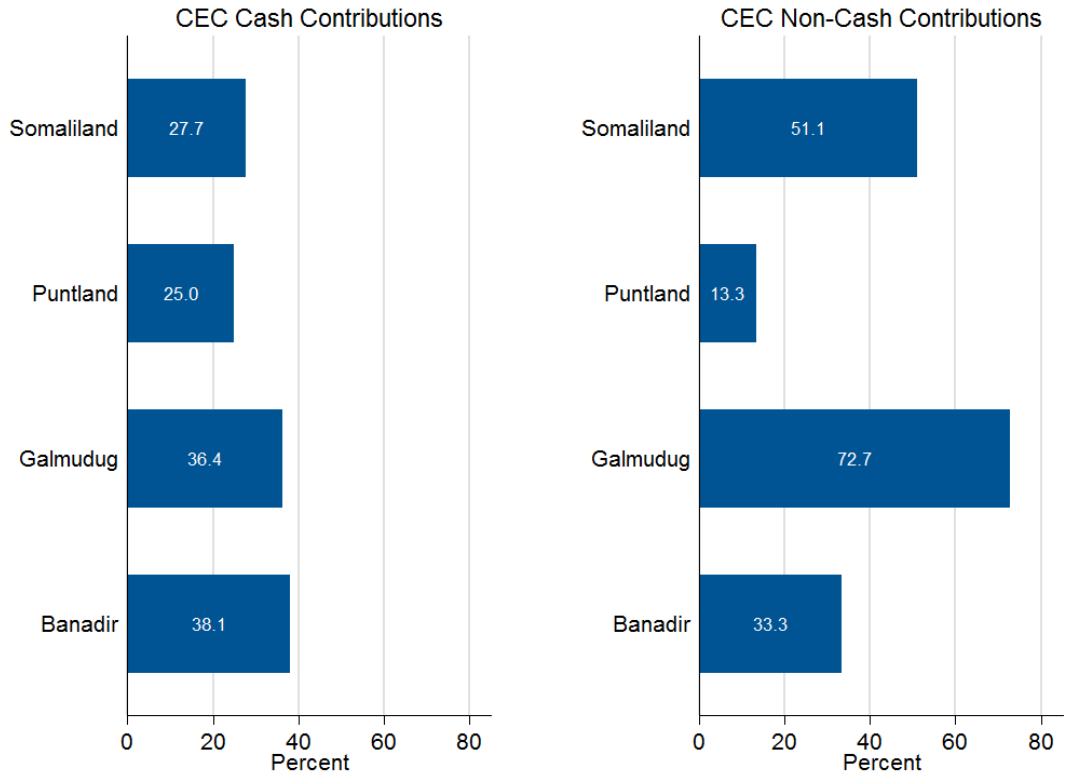
[227] FGDs with CEC member, Dhusamareeb, Galmudug.

[228] FGDs with CEC members, Dhusamareeb, Galmudug.

[229] FGDs with CEC members, Mogadishu, Banadir.

governance at this level, rather than at the level of head teachers, where no specific, targeted interventions are planned at this time. However, the evaluation did capture rich data on the capacities of head teachers; for instance, teachers were asked to indicate how often their schools had full-staff meetings, whether their head teacher or principal regularly observes their classroom, whether they regularly provide performance evaluations to teachers, and whether they receive tangible, written feedback following these observations and evaluations. Across most schools, teachers report consistent staff meetings: 89.9 per cent of teachers state that their school holds staff meetings, and 79.7 per cent report that a staff meeting was held within the last month. Head teachers' oversight of teachers is also fairly regularized, with 96.0 per cent of teachers reporting that their head teacher observes their classroom, and 88.8 per cent indicating that their classroom was last observed within the past month. These findings suggest that, although EGEP-T does not specifically target head teachers for capacity-building, head teachers are generally exercising oversight over teachers in their schools, and have the ability to contribute meaningfully to the project's sustainability as members of local CECs.

FIGURE 15: CEC CASH AND NON-CASH CONTRIBUTIONS, BY PROJECT LOCATION



WHY IS IT IMPORTANT FOR SUSTAINABILITY?

Based on interviews with teachers and CECs, CECs and NGOs are the largest advocates for improving the quality of schools. CEC members report that they receive help from the government; however, it comes only intermittently. As one CEC member described it, the government helps support the school by paying teacher salaries, but they only pay them once every three months, on average.[230] Other interviewees have described a range of inputs that the government provides to their schools,

[230] FGDs with CEC members, Mogadishu, Banadir.

but a theme of infrequent, and potentially ad hoc, support comes across in the qualitative interviews.[231] An interview with a teacher in Garbadadar, Somaliland reported that the government only "raised awareness, nothing else" when it came to helping marginalised children.[232]

As we discuss in detail in Section 5.5, below, CEC contributions pay teacher salaries, keep students safe, and provide higher quality education. CEC members in Mogadishu report that a major issue is the timely payment of salaries to local teachers. One of the interviewees indicated that it is not uncommon to find a teacher who has not received their salary for two months.[233] This finding comports with more systematic data, described in Section 5.5 below, showing that many teachers are not paid consistently from month to month. Furthermore, 76 per cent of teachers surveyed indicated that they have used personal funds to purchase school supplies needed for teaching their class. Increased financial support for CECs could help them overcome many of these problems, as they are aware of the particular problems faced in their schools. may improve the timeliness of salary payments to teachers.

Community Level Score: 2 – Emerging

Sustainability Indicator 6 – Teacher Continuous Professional Development Programmes

Percentage of schools with teachers peer-mentoring programme in operation (TBC depending on approach for teacher training)

An important aspect of EGEP-T programming concerns the training and mentoring of teachers. Building on the experience and lessons learned from EGEP, RI is implementing a teacher training programme focused on continuous professional development (CPD), with an emphasis on face-to-face and in-classroom support through the use of trained teaching coaches. Based on feedback received from the endline study of EGEP, sustainability has taken centre stage in the design of EGEP-T's peer-mentoring programme. For example, the programme will utilise training coaches based in Teacher Training Institutes, engage the respective MoEs in quality monitoring, and incrementally transfer ownership of the approach to Head Teachers over the life of the project.

Unfortunately, it was not possible for the baseline evaluation to fully measure this indicator, as the project's approach to teacher training and mentoring was still under development at the time of baseline data collection. As a result, it was not possible to develop a measurement approach around it. Relief International have informed Forcier that there is currently no teacher peer mentoring programme operating in Somalia; as a result, this indicator should naturally receive the lowest sustainability score at the baseline, *Latent*.[234] Beyond scoring this indicator, the information below provides a brief overview of EGEP-T's teacher-mentoring programme model. Progress against the plan will be measured at midline.

---

[231] FGD with MoE official, Galkayo, Puntland; FGD with CEC members, Hargeisa, Somaliland; FGD with CEC members, Galkayo, Puntland.

[232] KII with Teacher, Garbadadar, Somaliland.

[233] FGDs with CEC members, Mogadishu, Banadir.

[234] Note that it was not possible for Forcier to independently verify the lack of a teacher peer-mentoring programme at the time of the baseline, due to the timing of the programme's development relative to baseline data collection.

Teachers

Two teachers (one male and one female where possible) from each school will be selected for Continuous Professional Development. At primary level, the teachers will be those who are engaged in upper primary classes, from Grade 5 to Grade 8, and who have a specialisation in English and mathematics. At secondary level, mathematics and English teachers will be selected. The selection process will be led by the Head Teacher and Regional Education Officer, with guidance from the Key Resource Persons, including staff members from the implementation agencies. Where possible, and to aid sustainability, priority will be given to qualified teachers; however, where appropriate, the opportunity will be widened to community or non-qualified teachers. Teachers will have access to the following CPD inputs:

- Face-to-face training at regional level: 2 days per year – initial focus on inclusive pedagogy, numeracy and remedial education
- Face-to-face training at cluster level: 4 days per year – content to respond to needs identified during the classroom support – may differ from cluster to cluster
- In-classroom support through coach visits: a minimum of 1 day per month
- Interactive digital platform focusing on English as a Second Language and numeracy, as well as reference materials from the CPD training

Coaches

Teacher Training Institutes, in partnership with EGEP-T agencies, will be responsible for the recruitment and management of coaches. Coaches will be hired through a competitive process and will be required to have a background in teacher training, and where possible, classroom teaching. Each coach will be responsible for the schools in his/her allocated cluster. Coaches undergo annual training of trainers on coaching skills, numeracy, literacy, inclusive and gender-sensitive pedagogy, remedial education and English as a Second Language and will have access to on-going professional support from the Key Resource Person (KRP) team.

During the initial start-up period, coaches and KRPs will work closely with District and Regional Education Officers to establish effective relationships and ways of working with schools and teachers. Coaches will then visit each school at least once a month throughout the academic year to provide on-the-job support to selected teachers. They will spend a whole day in one school per month to observe the teaching techniques of targeted teachers and the impact on students' learning. Debriefing sessions with teachers will be held following observations where coaches will discuss strengths and gaps and identify action points to support skills development. This will also inform cluster and regional face-to-face training.

Coaches will participate in hub meetings once a month. These will be at cluster level and will be facilitated by KRPs. These meetings will provide a platform for reflection on successes and challenges, peer learning and joint monthly planning for visits and trainings. They will also be an opportunity for any course correction and for KRPs to identify skills gaps and discuss the responding training/ support inputs.

Key Resource Persons

Each state or zone has a team of Key Resource Persons (KRPs) – a technical representative from the partner TTI, MoE and implementing agency. KRPs are ultimately responsible for the training of coaches and ensuring that their work is of a sufficiently high standard. They will lead the different TOT sessions, support the coaches during the delivery of the face-to-face teacher training modules and undertake regular monitoring of coaches' work. They will be expected to respond to any issues

and skills gaps, liaising with the Technical Steering Committee to design and implement appropriate responses. KRPs will meet as a zonal team quarterly and as a consortium team on an annual basis. They will report quarterly to the Technical Steering Committee and will identify themes for in-depth learning and further research.

Technical Steering Committee

The TSC comprises central EGEP-T technical persons, including RI's Education Technical Lead, the Consortium Education Director and the Education Programme Managers from each implementing agency. The TSC will have overall oversight and quality assurance of the EGEP-T CPD approach. This will be ensured through quarterly review and planning meetings, regular field visits and communication and coordination with the KRP team.

Indicator Score: 1 – Latent

Sustainability Indicator 7 – 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.

As mentioned in earlier sections of this report, if girls' education is to be sustainable post-project, the respective Ministries of Education in each location will need to take a more prominent role in funding, management, and oversight of schools. This increased role includes, but is clearly not limited to, a need to play a more active role in gender issues within the educational system. One broader systemic goal of EGEP-T is to assist the relevant ministries in the development and implementation of a gender strategy. EGEP-T has specifically targeted the federal level Ministry of Education and the Galmudug Gender Unit for this intervention, because – based on their experience working with the ministries – neither has an existing gender strategy in place.

To assess the current status of this indicator, the evaluation included KIIs with MOE officials specifically situated within gender units, and surveyed head teachers regarding their level of engagement with their respective Ministry officials on issues of gender. The information gleaned from these sources is limited, but suggestive regarding the current policy environment around gender in the various ministries.

At the federal level, there is some evidence that a specific gender education policy exists. According to an MOE official in Banadir, the Ministry of Education is currently developing a strategic plan to guide most aspects of its work; more importantly, they indicate that the department has drafted a gender education policy.[235] This policy, according to the interviewee, includes efforts to promote the hiring of female teachers, especially in those schools that currently lack them.[236] The ministry is also engaged in other efforts to promote girls' education specifically, including the appointment of "Girl Ambassadors," who are educated girls that will visit schools and serve as role models for others.[237] Even if the federal MOE has a gender education policy in place, it is important to distinguish this from a strategy document, which lays out specific steps needed to achieve the policy's goals. Based on interviews with MOE officials at the federal level, there is no clear evidence that such a strategy currently exists.

[235] KII with MOE official, Mogadishu, Banadir.

[236] Ibid.

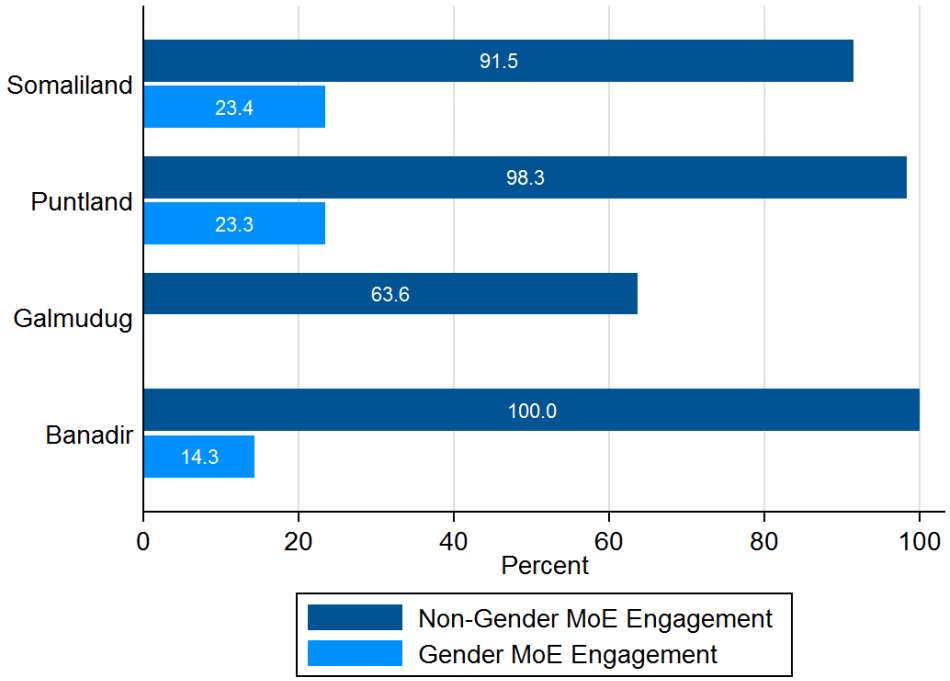
[237] Ibid.

Evidence from the other respective ministries paints a broadly similar picture. An MOE official in Galmudug – who was not specifically stationed within Galmudug's Gender Unit – reported that the ministry makes a variety of efforts to promote girls' education, especially via awareness-raising campaigns.[238] But no formal gender policy or strategy was mentioned.[239]

In Puntland and Somaliland, interviews were conducted with members of the areas' respective gender units. Again, neither interviewee indicated that their units or ministries had a tangible gender policy or a gender strategy in place.[240] However, it is important to note that interviewees may not have been asked about gender strategy documents in a sufficiently direct way. It is possible that such documents exist, but that the interviewees understood the questions to be seeking information about what efforts, even ad hoc, have been made to promote girls' education, as opposed to whether there exists a specific document that has been developed by their units.

Despite the relatively limited information collected regarding explicit policies and strategies within the ministries, we believe it is safe to conclude that the federal-level MOE and the Galmudug gender units do not have an explicit *strategy* document in place for promoting girls' education and, certainly, that implementation of any existing gender policy is sporadic. Surveys and KIIs with head teachers and CEC members appear to confirm this notion – CEC members describe a wide range of efforts made by the ministries to support their schools, but there does not appear to be a systematic approach being applied to all schools in an area.[241]

FIGURE 16: HEAD TEACHERS' ENGAGEMENT WITH MOE OFFICIALS



[238] KII with MOE official, Galmudug.

[239] Ibid.

[240] KII with MOE official, Bari, Puntland; KII with MOE official, Maroodi Jeeh, Somaliland.

[241] See, e.g.: KII with female teacher, Garowe, Puntland.

Among head teachers, most (93.6 per cent) report some form of engagement with the MOE over the past year. But these engagements are concentrated in areas of engagement that are not explicitly about gender policy or efforts to promote girls' education: engagement with regional education offices and examination boards lead the list. In fact, of the types of engagement included in the survey, interaction with officials from the gender department is the *least* common, with just 20.0 per cent of head teachers reporting such interaction over the previous year.

Interaction with gender department officials is also lowest among the two areas specifically targeted by EGEP-T for assistance with the development and implementation of a gender strategy. Figure 16 plots the share of head teachers in each zone who have had some form of interaction with non-gender-focused MOE officials in the last 12 months (dark blue bar), and the share who have interacted with officials from their MOE's gender department over the same time period (lighter blue bar). As Figure 16 shows, none of the 11 head teachers interviewed in Galmudug report interaction with a gender department official, and just 14.3 per cent of those interviewed in Banadir report such interaction. Compared to the much higher rates of interaction with other MOE officials – regional education officers, examination board members, and officials from departments charged with developing curricula – the lack of engagement with gender-focused officials is clear, and suggests that any gender policies that the MOEs have promulgated are not being widely and consistently implemented.

Indicator Score: 1 – Latent

Sustainability Indicator 8 – Child Protection Mechanisms

Percentage of Head Teachers and relevant ministry that report the established child protection mechanism is being actively used.

Effective child protection at the level of either a school or an education system requires organisations to have in place a complete system of policies and procedures to protect students. Schools should have an explicit child protection policy, and teachers and staff at the schools should be able to identify and refer cases to authorities. In addition, schools should have procedures in place that allow children to safely report cases, and which indicate how records should be kept, and how to follow up on cases. To assess the baseline quality of child protection systems in place in project schools, a number of metrics were employed, reflecting the range of procedures that must be in place for effective child protection. The primary metrics are:

- Does the school have an official child protection policy? Can the head teacher produce a copy of this policy?
- Does the school maintain a record of all cases related to child protection?
- Does the school have an official procedure for following up on child protection cases?
- Does the school have an established focal point for child protection cases, who is responsible for following up on such cases?
- Do new teachers receive any induction training on child protection?
- When administrative personnel, guards, and cooks are hired, are they required to complete induction training on child protection?
- Are teachers in this school required to agree to a Code of Conduct?
- Are students aware of how and to whom they could report child protection issues stemming from mistreatment or abuse by a teacher?

School-Level Policies

Do schools have official policies in place for protecting children? To what extent are standard procedures, such as training staff on child protection and requiring teachers to sign a Code of Conduct, followed by schools? Overall, the results from the baseline indicate that most schools lack fundamental components of child protection.

Only about one-third of head teachers reported that their school has an official child protection policy in place, and even fewer – just one-quarter – were able to produce a copy of the policy for field researchers to review. Other aspects of child protection were more mixed: most schools require teachers to agree to a Code of Conduct when they are hired; among schools with an official Code of Conduct in place, our field researchers report that over half (54.7 per cent) included a specific section dealing with child protection issues.[242] On the other hand, relatively few schools maintain case records for child protection issues, which may limit the types of actions that can be taken against perpetrators of abuse.

Worryingly, relatively few teachers are required to complete child protection training upon being hired. While some teachers may have received training of this kind during their university or post-secondary training programmes, arguably, new teachers should still undergo refresher courses to emphasise their responsibilities and to familiarise them with existing procedures at their new school. Other staff members are more often required to complete induction training of this kind, though rates are still low: among schools employing a watchman and/or administrative staff, 42.9 per cent and 50.0 per cent report that these individuals undergo induction training, respectively.

TABLE 55: SCHOOL CHILD PROTECTION POLICIES AND PROCEDURES

Indicator	Share of Schools
Full Sample	
Child protection policy	36.4%
Child protection policy available for viewing	25.0%
Maintains case records	15.0%
Procedure for following up on cases	32.9%
Focal point for cases	24.3%
School has Code of Conduct	63.6%
Teachers sign Code of Conduct when hired	47.1%
Induction training on child protection for teachers	34.3%
Somaliland	
Child protection policy	23.4%
Child protection policy available for viewing	12.8%
Maintains case records	8.5%
Procedure for following up on cases	19.2%
Focal point for cases	19.2%
School has Code of Conduct	44.7%
Teachers sign Code of Conduct when hired	34.0%
Induction training on child protection for teachers	12.8%
Puntland	

[242] Field researchers asked to review schools' Code of Conduct documents, and verified the existence of a child protection section personally.

Child protection policy	45.0%
Child protection policy available for viewing	36.7%
Maintains case records	13.3%
Procedure for following up on cases	30.0%
Focal point for cases	25.0%
School has Code of Conduct	66.7%
Teachers sign Code of Conduct when hired	58.3%
Induction training on child protection for teachers	50.0%
Galmudug	
Child protection policy	27.3%
Child protection policy available for viewing	27.3%
Maintains case records	45.5%
Procedure for following up on cases	81.8%
Focal point for cases	36.4%
School has Code of Conduct	100%
Teachers sign Code of Conduct when hired	90.9%
Induction training on child protection for teachers	45.5%
Banadir	
Child protection policy	47.6%
Child protection policy available for viewing	19.0%
Maintains case records	19.1%
Procedure for following up on cases	42.9%
Focal point for cases	23.8%
School has Code of Conduct	81.0%
Teachers sign Code of Conduct when hired	23.8%
Induction training on child protection for teachers	28.6%

Table 55 also disaggregates these findings by project location, to provide insight into the areas with the weakest child protection policies and implementation. Across the four areas studied, schools in Somaliland appear to systematically underperform with respect to the promulgation and implementation of specific child protection policies. Just 12.8 per cent of Somaliland schools have a child protection policy that could be viewed by a member of the evaluation team, half the rate in the full sample, and significantly fewer Somaliland schools have a Code of Conduct in place. In fact, across the eight indicators reported in Table 55, Somaliland is the worst-performing location on seven.

Existing mechanisms for reporting child protection issues at most schools are problematic, and could discourage reporting by making it uncomfortable for students. In other cases, students may simply be unaware of who they should report such issues to. For instance, when asked to whom students should report child protection cases, head teachers overwhelmingly indicated that cases should be reported to themselves. In 79.3 per cent of schools, head teachers report that cases should be reported to a man, which can reduce the likelihood of female students, especially, reporting instances of abuse.

Even more worrying is the fact that most female students do not believe there is an adult they could talk to at the school about abuse by a teacher. Girls in the learning cohort were presented with a hypothetical scenario in which one of their friends was being mistreated by a teacher. Respondents

were asked whether there was an adult they could recommend their friend talk to at the school, and just 62.0 per cent indicated that there was.[243]

Students' willingness and ability to report mistreatment by a teacher appear to track the implementation of child protection policies across locations outlined above. That is, girls in Somaliland were least likely to indicate that there was an adult at the school to whom they could report such mistreatment, while rates were significantly higher in Puntland, Banadir, and, especially, Galmudug.[244] Girls in rural schools are almost much less likely – 47.0 per cent versus 68.0 per cent – than those in urban schools to feel that they can report mistreatment. However, it is difficult to separate this finding from that of project location, because the majority of the sampled rural schools are also located in Somaliland, which itself had lower rates than the other project locations.

School Level Grade: 1 – Latent

Sustainability Indicator 9 – MoE Monitoring Visits

Percentage of EGEP-T schools receiving follow-up monitoring visits from MoE officials, including Gender Focal Points, District/Regional Education Officers/RES

A critical component of an effective educational system is two-way interaction between schools, at the local level, and relevant ministries at the district and national levels. This interaction should include schools transmitting their needs to administrators at higher levels within the system, as well as monitoring of schools by administrators. Monitoring by MOE officials can ensure accountability of schools for their students' performance, as well as for keeping accurate records of student enrolment, grades, and the proper administration of examinations, among other things.

The frequency of MOE visitations is important, as is the cooperative relationship between MOE officials and all other relevant stakeholders. But, more than the simple fact of a visit from MOE official, the role of follow-up visits, in which MOE officials provide feedback from their visits back to the schools, is especially critical. If MOE officials are collecting data from schools and CECs, but do not provide follow-up information and guidance, it is unlikely that local stakeholders can improve their performance. In order to triangulate the quality and frequency of MOE monitoring efforts, we assembled a number of indicators:

- The number of visits by MOE or other education-related officials in the past year to monitor facilities, teaching quality or teacher attendance.[245]
- Has there been an increase in monitoring visits from the year before?
- Do MOE officials provide feedback to schools after monitoring visits?
- Has the head teacher engaged with MOE officials in other ways, other than monitoring visits?

---

[243] Reinforcing the previous argument regarding the importance of gender when it comes to reporting cases, female students showed a strong bias in favour of reporting cases to female staff members or teachers. Despite the relative scarcity of female teachers, more students indicated that they would direct their friend to a female teacher than a male teacher (head teachers excluded), and more students indicated that they would direct their friend to a female non-teaching staff member than an equivalent male staff member.

[244] Just 44.6 per cent of girls in Somaliland would be willing to report mistreatment to an adult at school, compared to 68.8, 68.7, and 96.0 per cent of girls in Puntland, Banadir, and Galmudug, respectively.

[245] The survey instrument used to measure this indicator specifically lists Regional Education Officers, District Education Officers, Gender Focal Points, Regional Supervisors, Minister for Education, or Deputy Ministers for Education as examples of possible visitors.

Each of these indicators was measured by reported visits and interactions with MoE officials by head teachers at sampled schools.

TABLE 56: MOE MONITORING VISITS, BY PROJECT LOCATION

Indicator	Overall	Somaliland	Puntland	Galmudug	Banadir
Visit in last year	80.0%	87.2%	85.0%	72.7%	57.1%
Increase in monitoring since last year	62.5%	70.7%	64.7%	12.5%	58.3%

The primary results regarding MOE interaction with schools, disaggregated by project location, is provided in Table 56. Most schools, both in the aggregate and in each project location, report having received a visit by an MOE or education-related official within the last year. However, most schools report relatively few visits: 61.4 per cent of head teachers report two or fewer visits over the previous year. Optimistically, most schools report progress with respect to the number of visits: 62.5 per cent say that there are more visits this year than the year prior.

On both indicators reported above, schools in Banadir underperform those in other areas.[246] Head teachers in Banadir are least likely to indicate that a monitoring visit happened in the last year; they also report the fewest number of visits, on average – only 9.5 per cent of schools in Banadir received more than two visits last year. This figure is particularly striking, given the fact that all sampled schools in Banadir are located in Mogadishu, meaning that visits by ministry officials should, in theory, be less complicated logistically and less burdensome financially.

Among all the schools that received a visit from MOE officials in the past year, only 38.4 per cent received feedback or reports about monitoring visits.[247] When this data is disaggregated by project location, only Galmudug had a majority of schools receiving feedback from MOE officials post-visit. Among the sampled schools that did receive MOE feedback or reports, head teachers reported receiving encouragement, mostly in the form of verbal and written compliments regarding their progress and lesson plans.

TABLE 57: MOE VISITS AND POST-VISIT FOLLOW-UP, BY PROJECT LOCATION

Project Location	Share Schools Visited in Last Year	Share of Schools Receiving Feedback in Last Year
Overall	80.0%	30.7%
Somaliland	87.2%	42.6%
Puntland	85.0%	25.0%
Galmudug	72.7%	36.4%
Banadir	57.1%	19.1%

The activity levels of MOE officials – both in terms of initial visits and the provision of feedback to schools based on those visits – are disaggregated by project location in Table 57. Across locations, it is clear that visits from MOE officials are least common in Banadir, where just 57.1 per cent of schools report having received a visit in the previous year. This result is especially surprising, because

[246] The relatively small number of total schools in Galmudug means that results from this area should be interpreted with particular caution.

[247] Among all schools, just 30.7 per cent both received a visit and a follow-up visit from MOE officials.

EGEP-T schools in Banadir are exclusively located in Mogadishu, where visits would presumably be less costly in terms of time and resources. Given how many EGEP-T schools in Somaliland are in rural areas (57.5 per cent), we would expect fewer visits to occur in that area; instead, head teachers in Somaliland report the highest rates of MOE visitation.

The right column of the table focuses on the feedback that MOE officials provide to head teachers following a visit. While visits in which MOE officials engage in monitoring are important, they are not very effective in the absence of timely and specific feedback provided to head teachers afterward. Yet, as noted above, just 38.4 per cent of schools reporting a visit indicated that they received feedback afterward, and just 30.7 per cent of all schools in the sample report receiving both a visit and follow-up feedback.

Feedback rates, like visitation rates, are lowest in Banadir, where just 19.1 per cent of all schools receive a visit and follow-up feedback. As with visitation in general, Somaliland outperforms the other project locations: however, even in Somaliland, just 42.6 per cent of head teachers report that they have received visits from MOE officials *and* received feedback following those visits.

The low rate of follow-up across all locations suggests that interaction with MOE officials is not an iterative process. Rather, MOE officials visit schools and then depart with the information they gathered, without using those visits to guide schools' efforts to improve their performance or offering suggestions on areas where improvement is needed. Without feedback following a visit, head teachers, CECs and school administrators are left to wonder about their performance relative to other schools, and left to make improvements to their school without coherent guidance that is consistent across schools.

System Level Score: 1 - Latent

Aggregate Sustainability Findings

Although EGEP-T only began implementation in earnest approximately 2-3 months before the start of the fieldwork for this report, there are a number of positive signs regarding the potential sustainability of the project. Table 58 outlines the baseline values on each of nine sustainability indicators, and an aggregate baseline sustainability score. On a 0-4 scale, we currently rate the project's sustainability at 1.45 overall.

TABLE 58: SUSTAINABILITY OUTCOMES MATRIX – BASELINE VALUES

	Community	School	System
Indicator 1:	Awareness-raising and attitudinal change Score: 2	CEC bursary support for girls Score: 2	Gender Development Strategy Score: 1
Indicator 2:	Male support for girls' education Score: 1	CEC financial support for schools Score: 2	Child protection mechanisms Score: 1
Indicator 3:	Community leaders' advocacy for education Score: 2	Teacher mentoring programmes Score: 1	MoE Monitoring Visits Score: 1
Baseline Sustainability Score (0-4)	Mean Score: 1.67	Mean Score: 1.67	Mean Score: 1.0
Overall Sustainability Score (0-4, average of the three level scores)		1.45	

While some of the community-level and school-level sustainability outcomes are generally positive for such an early stage in the project life-cycle, system-level indicators are less encouraging. As discussed in detail above, the current state of gender policy development and implementation at the ministry level is disorganized and characterized by ad hoc interventions without a unifying theme or agenda. There is also little evidence of a systematic approach to the development of child protection mechanisms, which contributes to the limited uptake of such systems at the school level. Importantly, while our analysis documented the promulgation of child protection policies at the school level, the extent to which they were implemented and institutionalized was limited.

At the system level, even output-focused indicators, such as the performance of school monitoring visits – have limited uptake thus far. While the baseline data document many visits by MOE officials to individual schools, follow-up and feedback is rarely provided to schools, undermining the value of such visits. Unfortunately, these types of systematic efforts require buy-in from stakeholders that EGEP-T's implementing partners may have difficulty influencing. In short, the areas of sustainability where EGEP-T is likely to have the most difficulty making an impact are those which are currently lagging.

The other primary cause for concern with regard to sustainability lies in the area of community-level *actions* to support girls' education. For instance, while men stated strong support for girls' education (Community Indicator #2), they were unlikely to take tangible action that demonstrated this support. The other action-focused sustainability indicator at this level, advocacy efforts by community leaders, showed more promising results, as the qualitative and quantitative evidence both supported the view that community leaders were actively leading events aimed at encouraging girls' education. However, this indicator would benefit from collecting richer and more specific data at the midline, so our findings thus far should be interpreted with cautious optimism.

Changes needed for sustainability [to be completed by project]

TABLE 59: CHANGES NEEDED FOR SUSTAINABILITY (PROJECT COMPLETES)

School	
Change: What change should happen by the end of the implementation period?	A more girl and child-friendly learning environment
Activities: What activities are aimed at this change?	EGEP-T will provide capacity development at the institutional level. Child Protection monitoring tools have been developed in GEC 1 and in GEC-T staff from the respective ministries responsible for child protection will be trained in Child Protection and in use of the tools. EGEP-T will engage 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. The project will redesign the CEC training manual to ensure it is relevant to the required roles and responsibilities of the CECs with regards to Child Protection. School Child Protection Policies will be developed, school staff orientated in Child Protection and schools monitored to ensure a

	child friendly version of the policy is displayed and is being adhered to. Girls and boys will receive training in child protection as part of the Girls' Club activities and girls will be encouraged to raise general child protection concerns (as opposed to specific cases) in meetings with the CECs. Girls' Clubs have already started to do this in GEC-1 and this progress will be built on. Relief International will continue to engage with the Somalia Child Protection Working Group and the Protection Cluster, and will exchange learning with group members. Tools developed through the programme will be shared with the group, as well as lessons in best practice regarding child protection practice in schools. Therefore it is expected that other actors will replicate systems and the impact will be spread to schools beyond the EGEP target schools, and sustained beyond the project lifetime.
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, ministry officials, community members and girls themselves.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Child protection services are limited in Somalia. Referral mechanisms must rely solely on services provided by other NGOs working in the area.
Community	
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 training will build on the progress made in GEC 1. Training content will be designed to meet the capacity gaps identified through the course of GEC 1. It will focus particularly on the following areas: Longer term school development planning; mobilisation of funds and in-kind contributions to support plans, community-led initiatives to support the most severely marginalised girls, tracking teacher attendance, Disaster Risk Reduction at the school level, Child Protection principles and referral mechanisms, and strengthened administration systems. CEC members will play a particularly important role in the following four areas: i) Developing and implementing plans for school improvement; ii)Supporting the most marginalised girls; iii) Following-up with any girls potentially dropping out of school; and iv) Strengthening child protection initiatives in the school. Training and ongoing follow-up with the CECs will focus on enabling them 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.
Stakeholders: Who are the relevant stakeholders?	EGEP aims for CECs to play a more effective role in school governance and management. To achieve this, it is important that linkages are strengthened 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.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Capacity development approaches are often targeted at individuals meaning that the knowledge leaves the institution when the staff member does. The continuing drought has led to a high level of migration amongst CEC members as with other community members. The project is therefore providing refresher training covering similar material as done in GEC-1 but also adding a focus of resilience.
Systems	
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 will provide capacity-building opportunities to the MOE across all zones, particularly at the decentralised level, to aid the ministries in taking increased ownership and management of the target schools and wider education system. A focus on strengthening links between the schools themselves and the Regional Education Officers is an important step in the process. EGEP will continue to train and strengthen the Gender Focal Points. Monitoring tools developed by EGEP are now being used and this will continue in EGEP-T, critically with the addition of a feedback process with the schools. An important strategy employed by EGEP is to establish and strengthen bottom-up accountability mechanisms that enable communities to increasingly hold those in authority to account.
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.
Factors: What factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	There is generally very high turnover of MoE staff. EGEP will therefore concentrate on establishing tools and processes at the institutional level, to ensure that knowledge and practice does not disappear when individuals do.

5. Findings – Intermediate Outcomes

This section of the report presents findings related to each of EGEP-T's five selected intermediate outcomes. For each intermediate outcome, we present findings related to the primary indicator targeted by EGEP-T, and construct targets to be achieved by the midline and endline. In addition, the evaluation team – in consultation with RI's Monitoring & Evaluation team – developed a set of ancillary indicators during the design of the baseline evaluation. For each intermediate outcome, we first highlight results related to the primary indicator, and then report results for these ancillary outcomes, where appropriate.

5.1 Improved girls' attendance at primary and secondary school

Indicator: Improvement in marginalised girls' attendance in schools throughout the life of the project

The first of EGEP-T's intermediate outcomes is improving attendance among marginalised girls. Attendance is a natural and vital intermediate outcome, because it is one of the primary mechanisms through which EGEP-T project outputs are translated into improved learning and transition outcomes.

Three aspects of EGEP-T programming are expected to influence student attendance. First, EGEP-T will provide support to marginalised and severely marginalised girls, primarily through the provision of bursaries, school uniforms, solar lamps, and female sanitary kits, and the payment of examination fees. Second, EGEP-T will promote activities within schools that provide peer support for girls – such as training female teacher-mentors. By providing stronger peer support and confidence-building activities, girls will be more comfortable in school; by providing remedial courses, girls who have fallen behind will be given the opportunity to catch up – encouraging better attendance. Third, EGEP-T will seek to improve the performance of teachers through teacher training including, but not limited to, gender-responsive pedagogy. Intuitively, by improving the quality of teaching and the ability of teachers to identify with and respond to girls' needs, girls' attendance should improve.

By improving attendance, EGEP-T may naturally improve learning outcomes. A consistent predictor of academic performance in many contexts is the frequency with which students attend school. As we show in Section 5.6 of this report, attendance and learning outcomes are correlated among the baseline sample of both female and male students. Given the tight theoretical connection between attendance and learning – more time in a classroom provides more opportunities for learning and practicing key skills – we expect improving attendance rates to have positive knock-on effects on learning.

We also expect improved attendance to translate into better transition outcomes. One argument for this connection is that the decision to drop out of schooling is not typically a binary choice. Rather, students often start by missing school but intending to continue. After falling behind, they may become discouraged and drop out; this is especially true if they fail to pass the examinations required for their promotion to the next grade. Alternatively, if students begin occasionally missing school, it is a much smaller leap to dropping out than for a student who attends regularly. As a result, we expect that improved attendance will also contribute to increased transition rates.

Measurement and the Selection of Indicators

In line with guidance from the FM, RI has prioritised measuring attendance rates through the use of classroom registers or school records, triangulating this data through the household survey and via

classroom headcounts. While the use of school registers would be preferable under many circumstances, the evaluation team strongly recommends revising the measurement strategy to prioritise classroom head counts. As we describe at length in the discussion of measurement issues below, school attendance records are of poor quality both theoretically and empirically. The data collected during the baseline suggest that preference should be given to classroom headcounts collected by enumerators at the evaluation points, rather than relying on school records.

Primary Indicator: Girl-Specific Attendance Records

While the baseline evaluation 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 was collected is somewhat smaller than the overall sample of cohort girls. The girls cohort includes 1,609 respondents, while the sample we analyse in this section includes 1,190 girls.[248] The available sample is even smaller when we analyse the correlates of attendance, because we are occasionally missing data on other variables. Nonetheless, this individual-level sample allows us analyse the relationship between attendance and attributes of individual girls and their families, as well as their individual perceptions of their teachers and school.

Overall, we find that attendance among the sampled learning cohort girls is higher than that documented using classroom headcounts in the previous section. Across 1,190 cohort girls, mean attendance rates were 93.7 per cent, compared to a girls' attendance rate of 83.8 per cent in physical headcounts performed by enumerators. This difference could stem from at least two sources.

First, the sample of girls studied here differs dramatically from that included in the classroom headcount. Headcounts included 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. The sample analysed in this section includes only cohort girls, who were sampled from among girls *in attendance at the time of data collection*; as such, the cohort girls sample is biased toward girls who attend school 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.

Second, schools with extremely poor or nonexistent attendance records do not appear in our analysis of individual-level attendance, because attendance data was not available for individual girls.[249] In contrast, classroom headcounts included any classroom with plausible enrolment figures.

[248] The sample size discrepancy stems from the fact that attendance records were not available or complete for all girls in the cohort sample. Where records were incomplete, unavailable, or obviously incorrect (i.e. including implausible values), we recoded girls' attendance to be missing.

[249] This fact has repercussions for the midline and endline evaluation, as some learning cohort girls may have attendance records at the midline who did not appear in the baseline (or vice versa). It will be necessary, at the midline, to match attendance records for girls across the two evaluation waves, and ensure that analysis of changes in attendance over time include only those girls with records in both periods. Of course, to the extent that schools improve their recordkeeping and additional girls are included in their records over time, this provides an interesting

Underlying differences in the sample, such as these, explain why we observe higher attendance rates among cohort girls than in classroom headcounts. They also suggest that care will be needed at the midline and endline, both to ensure that the comparisons being made from one period to the next are comparing identical populations, and to provide primary focus on headcounts over individual-level attendance records.[250]

To shed light on the factors that influence attendance, we employ linear regression models. Our models take the following functional form

$$y_i = B_0 + \gamma_i + \delta_i + \mu_i$$

where y_i is the attendance rate for girl *I*, measured from 0 to 100. In our first model, we include a vector of basic demographic control variables, denoted by γ_i, which include project location, urbanicity, grade level, and the characteristics of a girls' head-of-household, among others. In every subsequent model, we include this set of demographic controls, and incorporate additional predictors of attendance rates, denoted by δ_i, which we describe in greater detail below. Finally, we incorporate an error term, μ_i, and we cluster standard errors at the school level to account for intra-school correlation in the error term.

We estimate four regression models, whose results we summarize in Figure 17 below. As noted above, our first model includes only demographic variables. The top-left panel in Figure 17 reports results from this model, by plotting the beta coefficients and 95 per cent confidence intervals for individual independent variables, as well as p-values for each coefficient along the right-hand margin. The vertical line at zero represents a null effect – for variables whose 95 per cent confidence interval crosses the vertical line, the correlation between the variable and attendance rates is not statistically significant.

Two main findings emerge from the demographics-only model. First, the differences in cohort girl attendance rates across project locations are generally small. Girls in Somaliland, Puntland, Hirshabelle and Banadir – the omitted reference category in our regressions – have attendance rates that are statistically indistinguishable from one another. The exception to this finding is Galmudug – girls in Galmudug have attendance rates 3.3 percentage points lower than the rest of the sample, holding all else equal.

Second, there is a monotonic relationship between grade level and attendance rates. Although the correlation between grade level and attendance is not statistically significant in this model, the upward trend from Grade 6 (the omitted reference category) through Form 2 is clear from the results in Figure 17. Notably, each of these findings is consistent across models: even when we control for additional factors in more saturated models, the findings regarding Galmudug and grade level remain.

Our second model, whose results we report in the top-right panel of Figure 17 – incorporates school-level factors that might influence girls' attendance, such as a lack of clean drinking water, or an

---

indicator of improved school management, though this is more relevant to an analysis of school management, rather than attendance itself.

[250] This is not to suggest that such individual-level attendance records are without merit. They are useful for analyzing the predictors or correlates of attendance at an individual level, though not among a truly random sample of girls. Moreover, 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.

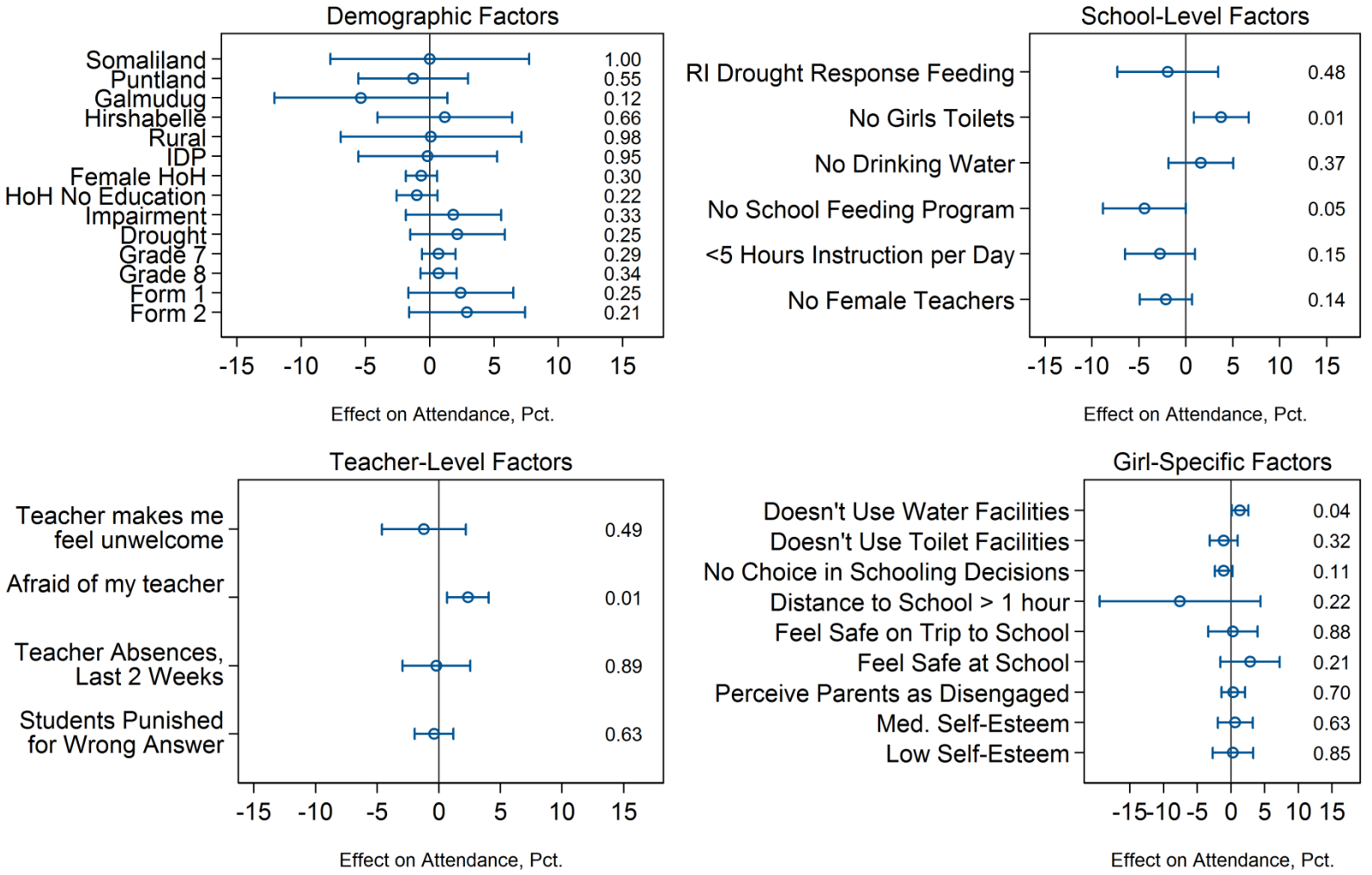
entirely-male teaching staff. While still controlling for a broad range of "demographic" factors, we also find that characteristics of a girls' school are associated with attendance rates. Counter-intuitively, girls at schools that do not provide girls with a separate toilet have higher attendance rates. This finding is unexpected from both a theoretical and empirical perspective. Theoretically, we see little reason why girls' attendance would be higher at schools that lack a toilet: the qualitative evidence suggests that girls' attendance is negatively affected by a lack of toilets, with girls leaving school to find a suitable toilet for use. And, even if this were not a significant problem for girls, we would at least expect better-equipped schools to otherwise have higher attendance, all else equal. Empirically, the results are unexpected because they run counter to the findings documented using classroom headcounts, where schools that lacked female-specific toilets had noticeably lower attendance rates. We do not have a compelling explanation for this discrepancy, though it may stem from the differences, noted above, between analysis based on classroom headcounts and analysis based on individual-level attendance records.

Another notable finding from the second regression model is the impact of school feeding programmes. To assess the role of school feeding programme, we employed data collected from head teachers, who were asked whether their schools participate in a feeding programme. In total, 35.1 per cent of schools report that they participate in such a programme. Note that these feeding programmes were not necessarily provided by EGEP-T – rather, we study whether any feeding programme, as reported by the head teacher, impacts student attendance. Separately, we study whether EGEP-T drought response activities that include a feeding component impact attendance rates.[251]

In schools that lack a general feeding programme, attendance rates are 4.5 points lower, while controlling for all other factors. In addition, we find suggestive evidence that other school characteristics reduce student attendance. Based on data collected from head teachers, we found that a surprising number of schools have either abbreviated school-days, or otherwise reduced instructional time. Across the sample, 21.2 per cent of head teachers report that their schools teach for fewer than five hours per day. When we analyse the attendance rates of girls in these schools, we find that their attendance rates are 2.9 percentage points lower than other girls, all else equal. Likewise, we find that schools without a single female teacher have attendance rates that are 2.2 points lower than otherwise similarly schools. These two findings are merely suggestive, as they are not statistically significant at conventional levels ($p = 0.12$ and $p = 0.13$, respectively); nonetheless, given the consistently of the findings and the fact that they occur in the theoretically-expected direction, they provide tentative evidence of a relationship.

[251] Evaluating the impact of RI's drought response activities in this way is not, in our view, a fair assessment, because it drought response is highly correlated with drought itself. Although there are good theoretical reasons to believe that drought response efforts incorporating rations for students would improve attendance rates, we do not think it likely that the impact would completely overcome the effect of drought.

FIGURE 17: REGRESSION RESULTS – RELATIONSHIP BETWEEN DEMOGRAPHIC, SCHOOL-LEVEL, TEACHER-LEVEL, AND GIRL-LEVEL FACTORS AND ATTENDANCE



In a third regression model, we remove school-level characteristics included in the previous model, and incorporate characteristics of a girl's teacher, as described by the girl herself. For instance, we include a measure of teacher absenteeism, as reported by the girl, and a variable indicating whether a girl strongly agrees with the statement "I am afraid of my teacher."[252] As shown in the bottom-left panel of Figure 17, the only teacher-level factor in our model that is significantly associated with attendance is a girl's fear of their teacher. However, the direction of the effect is unexpected – girls who report being afraid of their teacher have higher attendance rates, contrary to the findings from our analysis of aggregate headcount data, where schools with more girls who report being afraid of their teacher had lower overall attendance rates.

Our last regression model moves beyond school- and teacher-level characteristics, focusing instead on characteristics of girls and their families. As with all models reported in this section, we control for the demographic factors that were included in model 1. The results of this final model are reported in the bottom-right panel of Figure 17. In general, girls' attendance does not appear to be affected systematically by their willingness to use school facilities, the distance between their home and school, or their perceptions of safety at school.[253] We also do not find any evidence that the level of engagement of their parents influences attendance rates.[254]

We include a measure of girls' self-esteem in this model, to assess whether girls with greater self-esteem are more likely to attend school regularly. These measures are derived from a set of questions related to self-esteem and empowerment, which were asked to each girl in the learning cohort. We provide greater detail in Section 5.2, which focuses on self-esteem and empowerment as an intermediate outcome; here we note that we combined data from 23 distinct questions, to form an index ranging in value from 8 to 56. Based on their index score, we split respondents into three categories, loosely comprising girls with high, medium, and low self-esteem.[255] The omitted category in our regression is high self-esteem, meaning that the effects of medium and low self-esteem are measured *relative* to girls with high self-esteem. As the results in Figure 17 show, girls in the medium and low self-esteem categories have almost precisely the same attendance rates as girls with higher self-esteem. Surprisingly, even when we model the relationship between self-esteem and attendance in a non-linear – and even entirely non-parametric – fashion, there does not appear to be a systematic relationship between the two.

[252] Teacher absenteeism is the number of days their teacher has missed in the last two weeks, as reported by girls. Like students' fear of their teacher, we measure the extent to which students do not feel welcomed by their teachers on a 4-point likert scale, with girls who strongly or slightly disagree that their teacher "makes them feel welcome" coded as a "1" and other girls coded as a "0."

[253] In reality, we find that girls who are unwilling to use the drinking water facilities at school attend school more often ($p = 0.04$), while girls who are unwilling to use their school's toilets are somewhat less likely to attend school ($p = 0.32$). The contradictory nature – and generally small substantive sizes – suggests that infrastructure quality, at least in terms of drinking water and toilets, are not strongly predictive of attendance rates among girls in the sample.

[254] We define parental engagement based on student responses to a question about parent-teacher meetings. If students report that *both* of their parents are either very unlikely or somewhat unlikely to attend a meeting arranged with their teachers, we code a girl as perceiving their parents to be disengaged from their education. If at least one parent is somewhat or very likely to attend, the girl is coded in the opposite direction.

[255] In our primary analysis of self-esteem, we divide the index into five categories. Due to the smaller sample size studied here, we collapsed the two top and two bottom categories, to create three self-esteem groups.

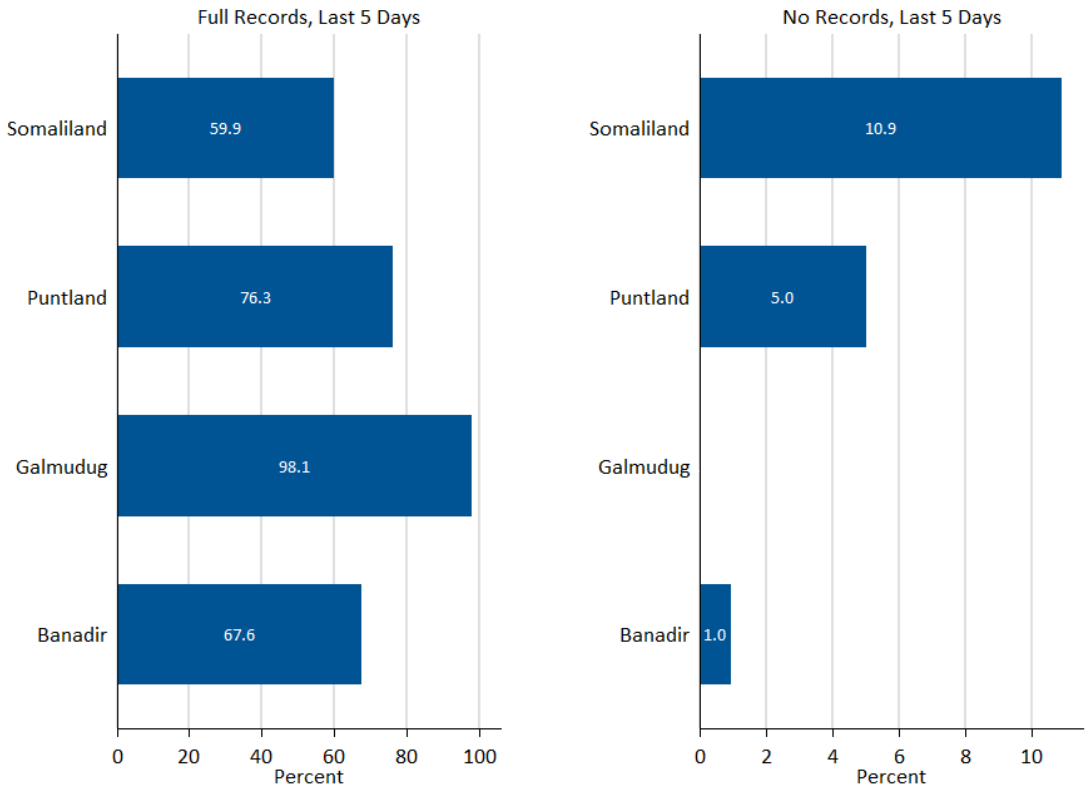
Measurement Error in Measuring Attendance

Beyond the girl-specific attendance records described and analysed in the previous section, the baseline evaluation also collected data by conducting headcounts in randomly-selected classrooms in each EGEP-T school. These headcounts were conducted by Forcier researchers on the day of fieldwork at a given school. Data collection teams randomly selected classes to visit from among mathematics, English and Somali classes at a given school. One class was sampled from each grade level in the school. In total, the sample consists of headcounts conducted in 893 classrooms.

To perform the headcounts, enumerators collected enrolment and attendance information from the teachers and head teachers for the selected classroom. Specifically, they recorded the number of children enrolled in the class (i.e. the maximum possible number of attending students) and the number of children that were marked present on the attendance roll each of the two previous days. Finally, the enumerator performed an independent headcount of students in the classroom at the time of data collection. The primary indicator for attendance is the attendance rate documented through classroom headcounts. That is, the number of students counted present by the enumerator, divided by the number of students enrolled in the class.

It is important to note concerns regarding data quality in school attendance records. Past GEC evaluations, both in Somalia and elsewhere, have provided significant evidence that attendance records are poorly-kept in many schools, and are unreliable as a metric of progress over time. Indeed, RI's MEL Framework noted the poor quality of attendance records observed during GEC-1, and explicitly designed their MEL strategy around collecting alternative measures of attendance for purposes of triangulation.

FIGURE 18: AVAILABILITY OF ATTENDANCE RECORDS, BY PROJECT LOCATION



The data from the baseline underline this concern. Schools sampled at the baseline often had verifiably incomplete attendance records. As part of the classroom headcounts described above, enumerators collected data on the frequency with which attendance was taken in each class over the past five days. Overall, only 71.3 per cent of classes had full attendance records available over the previous five days. As shown in the left panel of Figure 18, this rate varied systematically by project location; only schools in Galmudug consistently collected attendance information. This problem goes beyond missing a single day in a week, however; 6.1 per cent of classrooms sampled had not collected *any* attendance information over the previous week. Again, the results vary significantly by project location, with schools in Somaliland the most likely to completely lack records.

Furthermore, even when teachers recorded attendance information, they did not always do so accurately. When we compare the official attendance records on the day of data collection to headcounts conducted in the same classrooms, we find that attendance is misreported in many classrooms' official counts. Figure 19 plots official attendance counts on the y-axis against physical counts performed by enumerators on the x-axis. Each point represents one classroom. In cases where the counts agree, the points should lie on the diagonal line bisecting the graph. Points that fall off the line indicate discrepancies between the teachers' count and our own.

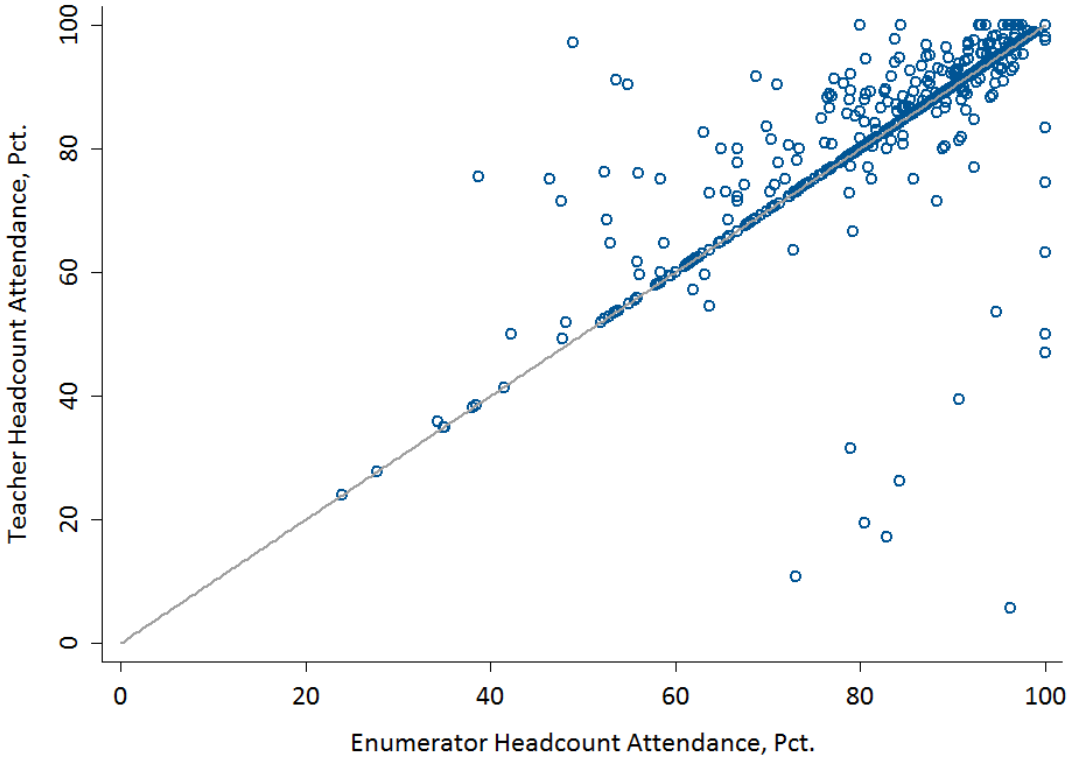
Two aspects of Figure 19 are worth noting. First, we limited this analysis to classrooms in which teachers reported plausible levels of attendance. If no attendance information was available for the day, or if the records indicated implausible attendance levels – e.g., more than 100 per cent, or attendance of more students than could be present in the classroom – we removed these classes from the analysis presented in Figure 19. As such, the results actually understate the number and magnitude of discrepancies in the broader data.

Second, the number of discrepancies is alarming – 28.6 per cent of this subset of classrooms had counts that disagreed with the physical count performed by an enumerator. It is possible that our count differs from those of teachers for several reasons that do not indicate systematic problems with the official records *or* with the physical headcounts performed by enumerators. For instance, if students arrived after attendance was taken, left before our headcount, or were away from the classroom for some other reason, the two counts would differ. However, in 28.1 per cent of cases, our count differs from those of teachers by more than one student; in 16.2 per cent of cases, our counts differ by more than five students.

In addition, classroom registers are occasionally marked by noting how many students are absent, rather than the number who are present. This could explain those points in the far lower-right corner of the graph, where the number of students present recorded by enumerators are similar to the inverse of the number of students present recorded by teachers (i.e. 80 per cent attendance according to enumerators, 20 per cent missing according to teachers). However, the fact that most discrepancies fall above the diagonal line means that most discrepancies involve teachers recording higher attendance rates than enumerators – discrepancies of this kind cannot be explained by the possibility that teachers marked the register according to the number of students absent. More generally, even explanations for small discrepancies, such as students leaving class early, or coming late, cannot account for larger differences documented in Figure 19.

Our findings regarding the quality of records are consistent with those reported in Section 5.4, focusing on school management, below. In that section, we show that schools often lack essential records, and that the records they keep are often incomplete, poorly organized, or both.

FIGURE 19: CORRELATION BETWEEN ENUMERATOR HEADCOUNTS AND OFFICIAL ATTENDANCE REGISTERS



Unfortunately, official attendance records have one significant advantage over classroom headcounts: the availability of individual-level data. Our headcount data assess overall attendance in a given classroom, and provide a high-quality measure of overall attendance rates at each school. However, they do not allow us to track the attendance of a specific girl or boy in our sample cohort. To the extent that we want to assess the relationship between attendance and learning outcomes, either in the baseline or in future evaluation waves, using individual-level data is extremely helpful.

For these reasons, we report and analyse both official, individual-level, attendance records, and aggregate headcounts collected by field teams. We suggest that evaluations at the midline and endline measure progress against both indicators. At the conclusion of this report, we also provide recommendations for improving attendance data collection in future evaluations, and triangulating findings across multiple data sources.

Results from Classroom Headcounts

In this section, we report baseline attendance levels drawn from physical classroom headcounts performed by Forcier researchers. Owing to the measurement concerns raised above, we describe these alternative metrics of attendance in detail. We also strongly recommend that classroom headcounts be used to establish baseline attendance rates and set project targets going forward; in line with these recommendations, we set targets in this section using classroom headcounts as the primary outcome of interest, with girl-specific attendance records fulfilling the role of ancillary indicator.

TABLE 60: CLASSROOM ATTENDANCE RATES, BY GEOGRAPHIC AND SCHOOL-LEVEL FACTORS

Subgroup	Girls Attendance	Boys Attendance	Total Attendance
Overall	83.3%	83.8%	83.7%
Somaliland	80.1%	79.6%	79.8%
Puntland	85.7%	86.9%	86.7%
Galmudug	79.2%	80.3%	79.6%
Banadir	87.5%	88.2%	87.6%
Primary schools	83.1%	84.2%	83.8%
Secondary schools	84.3%	81.8%	83.2%
Grade 1	81.6%	86.0%	83.5%
Grade 2	81.2%	84.0%	83.2%
Grade 3	88.1%	82.6%	86.1%
Grade 4	84.0%	85.9%	84.5%
Grade 5	82.4%	83.2%	83.4%
Grade 6	84.2%	84.8%	84.3%
Grade 7	81.0%	82.6%	81.5%
Grade 8	82.8%	84.3%	84.1%
Form 1	86.0%	84.5%	85.6%
Form 2	80.5%	76.3%	77.3%
Form 3	85.5%	84.5%	84.7%
Form 4	83.7%	82.1%	84.2%
Rural schools	79.2%	82.2%	80.7%
Urban schools	84.9%	84.4%	84.8%
Drought-affected schools	78.1%	81.9%	80.2%
Has school feeding program	80.0%	81.6%	81.0%
IDP schools	83.1%	82.0%	82.9%
Conflict-affected schools	86.2%	85.7%	86.1%
Many girls feel journey to school is unsafe	81.4%	84.1%	83.2%
Many girls feel school is unsafe	86.7%	88.1%	87.2%
Has girls' toilets	84.7%	84.5%	84.7%
Does not have girls' toilets	79.4%	82.3%	80.9%
Many girls do not use toilets at school	78.7%	76.2%	78.1%
Many girls feel unwelcome by their teacher	85.2%	86.3%	85.8%
Many girls feel afraid of their teacher	79.9%	82.2%	80.9%

Overall, attendance rates are moderate in the sampled schools, with attendance in mathematics, Somali and English classes averaging 84.2 per cent across the 838-class sample. Table 60 disaggregates classroom attendance rates by a number of geographic, demographic and other school-level correlates of attendance. Across project locations, attendance rates are lowest in Somaliland and Galmudug, with 79.8 and 82.0 per cent overall attendance, respectively.[256]

[256] We do not report results for Hirshabelle, because the subsample in Hirshabelle includes only one school.

Consistent with expectations regarding the impact of geography on attendance rates, we find that rural schools have lower attendance rates (80.9 per cent versus 85.5 per cent) than urban schools. Similarly, we find lower attendance rates in drought-affected schools.

School infrastructure appears to have an outsized impact on attendance rates. Schools which specifically provide toilets for girls have slightly higher overall attendance rates, at 84.9 per cent, compared to the sample average of 84.2 per cent. However, these results understate the true difference, because most schools provide toilets for girls. When we compare schools that provide girls' toilets to those that do not, we find that attendance rates in the former are 84.9 per cent, while attendance rates in the latter are just 80.9 per cent. Importantly, this gap is even more pronounced in the case of girls' attendance – attendance rates among female students are 5.3 points higher in schools that provide toilets than those that do not.

Beyond geography and school infrastructure, we also analysed the relationship between attendance and less-tangible school characteristics. Using data collected from the survey of cohort girls, we classified schools according to five additional criteria:

- At least 10 per cent of girls report the journey to school is unsafe (31 schools)
- At least 10 per cent of girls report they feel unsafe at school (20 schools)
- At least 50 per cent of girls report not using toilets at school (32 schools)[257]
- Schools with a mean "teacher welcoming" score under 2.5 on a 0-3 scale (22 schools)
- Schools with a mean "afraid of teacher" score above 2.25 on a 0-3 scale (15 schools)

These variables have mixed relationships with attendance, as shown in Table 60 above. Where more girls report that the journey to school is unsafe, attendance rates are 3.9 points lower than the sample average, and 4.8 points lower than schools where fewer girls describe the journey as unsafe.[258] On the other hand, attendance rates are higher in schools where more girls report the school itself is unsafe, a finding that is particularly counter-intuitive.[259] Finally, the strongest correlation documented above is the relationship between girls' toilet use and attendance. In schools where more than 50 per cent of surveyed girls report that they do not use toilets at school, attendance rates among both boys and girls are significantly lower than schools where this is not true – 7.2 points lower for girls and 10.4 points lower for boys.[260]

One important caveat regarding this analysis concerns the influence of our rules for including classrooms in the sample for analysis. While headcounts were performed in 893 classrooms, we removed classrooms from the analysis for multiple reasons. In classrooms in which enrolment records were not kept up-to-date, we cannot calculate attendance percentages accurately. Likewise, headcounts that documented attendance over 100 per cent were discarded.[261]

---

[257] This category includes girls at schools in which no toilet is available, but also includes those with girl-specific toilets that a majority of girls report being unwilling to use.

[258] Note that these results – though substantively large – are not statistically significant when we account for intra-school clustering of classrooms.

[259] Girls' relationships with their teachers exhibit similarly counter-intuitive results – in schools where more girls report being afraid of their teacher, attendance rates are marginally lower among both boys and girls, but the opposite effect is found in schools where fewer girls report that their teacher makes them feel welcome.

[260] For the difference in girls' attendance, $p = .05$; for the difference in boys' attendance, $p = .001$.

[261] In addition, we removed classrooms with documented enrolment and/or attendance by one sex but not the other. Unfortunately, the data do not allow us to determine whether these classrooms are all-boys or all-girls; as a

We also removed classrooms in which headcounts documented total attendance of over 150 students, or class records showed over 150 students were enrolled. Decisions regarding such cut-offs are arbitrary; in practice, we cannot imagine a situation in which a classroom actually has 150 students enrolled or present.[262] It is possible that enumerators documenting enrolments or attendance of more than 150 students in a single classroom were recording information for multiple classes in the same grade or subject. Unfortunately, our data do not allow us to determine the exact cause of these outliers.

Practically, we do not expect such classrooms – those with high enrolment and headcount attendance figures – to bias our analysis systematically. Because we removed schools with implausibly high attendance *rates*, the remaining classrooms may have high raw attendance numbers, but their rates are not abnormally high, given their similarly high enrolment numbers. Nonetheless, to ensure that classrooms with unusually large enrolment and attendance numbers are not impacting the results presented above, we repeated our analysis after removing a further 47 classroom observations. In this subsample, total attendance rates are 83.8 per cent, just 0.4 points lower than the primary results reported above. For both boys and girls, attendance rates in the subsample are a mere 0.2 points lower than in the full sample used above.

Beyond the quantitative results from either school attendance records or the evaluation's classroom headcounts, qualitative interviews provided additional insight into the reasons that girls (and boys) were absent from school. Some interviewees suggested an attitude that boys are often more absent than girls even though the latter have more responsibility outside of school.[263] However, people also commented that girls are frequently absent for whole weeks during menstruation.[264] Some teachers also said that students are occasionally kept late at their religious schools (which start in the early morning) and may miss school.[265] Lastly, girls stated that sometimes chores would prevent them from going to school that day.[266]

Targets for Midline and Endline

In preparation for future evaluations, it is uncertain how EGEP-T's performance should, with respect to attendance, be judged. Before establishing targets for the midline and endline, it is important to note that attendance rates have natural ceiling effects – no project will be able to increase attendance rates above – to cite an admittedly arbitrary threshold – 98 per cent, as even the most motivated students in the best schools are occasionally sick or miss school for other reasons. Given the relatively high attendance rates observed among cohort girls, especially, there may be limited room for improvement.

Table 61 suggests midline and endline targets for three attendance rates: girls' attendance and boys' attendance derived from enumerator headcounts, and attendance rates for cohort girls, derived from official school records. In each case, there are suggested targeted improvements of between 1.0 and 2.0 per cent from wave to wave. We suggest lower improvement targets (1.0 per cent per

---

result, we employ a conservative approach and remove them, under the assumption that data is missing for one of the two groups.

[262] Anecdotal field reports from the data collection teams suggest classrooms with more than 50 students are extremely rare and that classrooms with more than 100 students do not exist.

[263] Key Informant Interview with Female Teacher. Maroodi Jeeh, Somaliland.

[264] Ibid.

[265] Key Informant Interview with Male Teacher. Galgadud, Galmudug.

[266] Focus Group Discussion with Female Students. Galgadud, Galmudug.

wave) for cohort girls' attendance rates, as this indicator is the most likely to bump up against ceiling effects.[267] We also suggest slightly lower improvement targets (1.6 per cent versus 2.0 per cent) for boys' attendance versus girls' attendance, owing to the fact that EGEP-T – although it should benefit boys as well – focuses predominately on girls' school performance. Table 61 also provides targets for specific project locations, and according to urbanity, conflict-affected status, and IDP status; however, for these subgroup targets, we focus on attendance rates based on headcounts exclusively, as they appear to be a more realistic indicator of baseline attendance rates, without bias from the sampling method employed for selecting cohort girls. In general, we recommend a focus on targets derived from headcounts at the time of the midline and endline.

TABLE 61: MIDLINE AND ENDLINE ATTENDANCE TARGETS

Indicator	Baseline Level	Midline Target	Endline Target
Overall Targets			
Girls' attendance rates, headcounts	83.3%	85.3%	87.3%
Boys' attendance rates, headcounts	83.8%	85.4%	87%
Cohort girls' attendance rates	93.7%	94.7%	95.7%
Location-Specific Targets			
Somaliland			
Girls' attendance rates, headcounts	80.1%	82.1%	84.1%
Boys' attendance rates, headcounts	79.6%	81.2%	82.8%
Puntland			
Girls' attendance rates, headcounts	85.7%	87.7%	89.7%
Boys' attendance rates, headcounts	86.9%	88.5%	90.1%
Galmudug			
Girls' attendance rates, headcounts	79.2%	81.2%	83.2%
Boys' attendance rates, headcounts	80.3%	81.9%	83.5%
Banadir			
Girls' attendance rates, headcounts	87.5%	89.5%	91.5%
Boys' attendance rates, headcounts	83.8%	85.4%	87%
Subgroup Targets			
Urban			
Girls' attendance rates, headcounts	84.9%	86.9%	88.9%
Boys' attendance rates, headcounts	84.4%	86%	87.6%
Rural			
Girls' attendance rates, headcounts	79.2%	81.2%	83.2%
Boys' attendance rates, headcounts	82.2%	83.8%	85.4%
IDP			
Girls' attendance rates, headcounts	83.1%	85.1%	87.1%
Boys' attendance rates, headcounts	82%	83.6%	85.2%
Non-IDP			
Girls' attendance rates, headcounts	83.4%	85.4%	87.4%

[267] The high baseline rates for cohort girls' attendance, as derived from official school records, are surprising. As noted previously, this is likely due to the fact that cohort girls were sampled from among girls in attendance on the day of baseline data collection, biasing the sample toward girls who attend school regularly.

Boys' attendance rates, headcounts	84%	85.6%	87.2%
Conflict-affected			
Girls' attendance rates, headcounts	85.3%	87.3%	89.3%
Boys' attendance rates, headcounts	84.9%	86.5%	88.1%
Non-Conflict			
Girls' attendance rates, headcounts	83.1%	85.1%	87.1%
Boys' attendance rates, headcounts	83.6%	85.2%	86.8%

5.2 Girls feel more empowered with greater self confidence

Indicator: Increase in self-esteem and self-confidence of marginalised girl

One of the intermediate outcomes developed by EGEP-T was greater self-esteem and empowerment of girls. RI, through EGEP-T, is involved in several areas of programming that they expect will raise the self-esteem and agency of girls they work with. According to the EGEP-T Theory of Change, girls clubs, female teacher mentors, improvement of teacher quality, and bursary support to vulnerable girls are expected to improve the self-esteem of girls in EGEP-T schools. Improvement in the empowerment and self-esteem of girls in EGEP-T schools is expected to encourage them to ask questions during lessons, and have the confidence to pursue their dreams and continue their education. Empowering girls is likely to improve both their learning and their likelihood of a successful transition.

As the multidimensional concept of self-esteem presents measurement difficulties, the evaluation team developed a number of indicators related to self-esteem and empowerment. The primary indicator, selected for targeting by RI, seeks to assess the self-confidence of girls in taking on leadership positions. The baseline measured self-confidence of this kind by asking respondents to agree or disagree with the statement "When I have the opportunity, I can organize my peers or friends to do an activity." The indicator measures the share of girls who strongly agree with the statement as a measure of self-confidence.

Two secondary indicators were developed to measure self-esteem and empowerment. The first was a broader index of self-confidence focusing on a range of non-leadership areas of self-confidence and empowerment. The index closely mirrored the self-esteem and empowerment module utilised in past GEC projects. Specifically, the index aggregated self-assessments across 18 individual items. The second focused on the agency girls enjoy with regard to schooling, marriage, work, and time use.

Each of the indicators described in this section were collected through surveys with girls in the learning cohort, who comprise a random sample of approximately 12 girls per school from grade 6 through form 2.[268] In addition, a condensed version of the self-esteem and agency module was given to three boys at every school to allow for comparisons between boys' and girls' self-esteem. A similarly condensed version, focused on girls' agency, was also given to girls sampled as part of the household survey.

Primary Indicator – Girls' Leadership

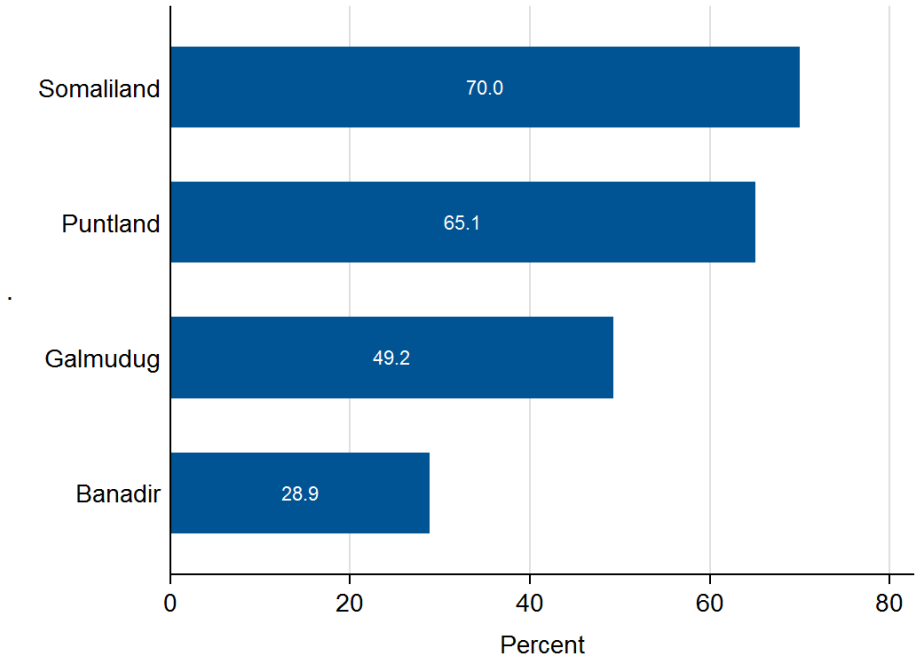
The evaluation's primary indicator of girls' self-esteem and empowerment focuses on the willingness of girls to take on leadership positions. The evaluation specifically focused on leadership within a

[268] As part of the Girls School Survey (see Section 2.3 for description).

girls' peer group, assessing whether girls felt they were able to lead their friends to accomplish a goal or participate in an activity. Respondents selected for inclusion in the main girls cohort were asked whether they agree or disagree with the statement "When I have the opportunity, I can organize my peers or friends to do an activity."

Overall, the vast majority of girls agreed with the statement, though the strength of their agreement varied. In total, 61.7 per cent of respondents agreed strongly that they are able to lead their friends or peers in this way, while another 24.6 per cent agreed less strongly. Virtually no respondents, just 0.6 per cent, disagreed strongly.

FIGURE 20: SHARE OF GIRLS EXPRESSING CONFIDENCE IN THEIR ABILITY TO ORGANISE THEIR PEERS



Rates of agreement were strongly correlated with girl and community characteristics, including drought – at the community level – and disability status – at the individual level. The starkest variation, however, concerned geography, as shown in Figure 20: the share of girls strongly agreeing with the notion that they can organise their friends effectively ranged from a high of 70 per cent in Somaliland to a low of just 28.9 per cent in Banadir. Even this wide swing does not fully account for the geographic variation, as 91.7 per cent of respondents in Hirshabelle expressed strong agreement; however we exclude Hirshabelle from the primary analysis, due to the small sample size of available respondents (n = 12).

Less dramatic variation occurred in other areas. Contrary to expectations, respondents attending IDP schools and in drought-affected communities expressed greater self-confidence on this metric than their counterparts in non-IDP schools and non-drought communities. Similarly, more girls in rural schools than urban schools – 65 per cent versus 60.4 per cent – report strong agreement with the statement in question, and disagreement rates were lower in rural areas as well.

TABLE 62: BASELINE OUTCOME AND FUTURE TARGETS FOR SELF-ESTEEM INDICATOR

Subgroup	Baseline Outcome	Midline Target	Endline Target
Overall	61.5%	65.0%	68.5%
Somaliland	70.3%	73.3%	76.3%
Puntland	64.6%	67.6%	70.6%
Galmudug	53.7%	57.7%	61.7%
Banadir	28.6%	33.6%	38.6%

Midline and endline targets for girls' self-esteem are provided in Table 62. Note that the targets are location-specific, owing to the vast gulf in girls' self-reported self-esteem across areas. In the areas achieving a higher score at the baseline, we recommend a targeted gain of 3 percentage points per evaluation wave; in the lower-performing areas of Galmudug and Banadir, we recommend targeted gains of 4 points and 5 points per wave, respectively. The location-specific targets reflect both the greater need for self-confidence gains in these areas, and the possibility that gains may be easier if one is starting from a lower baseline level of self-esteem.

Despite girls being quite willing to take on leadership positions, teachers often reported girls as very shy in the classroom.[269] One CEC member said that girls felt shy when the teacher was a man and another CEC said girls felt shy if they didn't have the newest school uniforms. A teacher said that boys are often quick to speak up and to participate in class but girls were often reticent to speak up.[270] It is important to note that the teacher who said this was female and shows that female teachers will not always be the solution to girls feeling comfortable in speaking up in class.

Self-Esteem Index

In order to capture the multidimensional nature of self-esteem, girls were asked to assess themselves with regard to 18 statements that indicate varied aspects of self-esteem and self-confidence. The 18 items are described in Table 63. Each item was measured in the self-esteem and agency module given to cohort girls (n = 1609). An index was created to summarize the self-esteem questions. Each self-esteem question was ranked 0 – 4 with an answer of 0 being the highest self-esteem answer possible and 4 being the lowest self-esteem answer possible. Scores across all questions were then summed for each girl. The index was divided into five categories: high self-esteem, moderately high self-esteem, neither high nor low self-esteem, moderately low self-esteem, and low self-esteem. The categories were created using a maximum score indicator. For example, the maximum score for moderately low self-esteem would be 54 because there were 18 questions and 3 was the moderately low self-esteem score for each question. The minimum score was the previous score's maximum score indicator plus one.

Using these indexed results to summarize the self-esteem module questions, it was found that 0.4% of girls had high self-esteem and 24.7 per cent of girls had moderately high self-esteem. Table 64 shows the self-esteem distribution across the five project locations and in urban and rural areas. Table 64 also shows the distribution across secondary and primary schools, IDP and other types of schools.

[269] Focus Group Discussion with CEC. Awdal, Somaliland; Focus Group Discussion with CEC. Hargeisa, Somaliland.

[270] Key Informant Interview with Female Teacher. Garowe, Puntland.

TABLE 63: LIST OF QUESTIONS USED TO MEASURE SELF-ESTEEM AND AGENCY INDICATORS

Survey	Question Text	Agency or Self-Esteem
Girls/Boys	I am able to do things as well as my friends	Self-Esteem
Girls	I want to use the skills I've learned during my education	Self-Esteem
Girls/Boys	I get nervous when I have to speak in front of an adult	Self-Esteem
Girls/Boys	I get nervous when I have to speak in front of a group of people my age	Self-Esteem
Girls/Boys	I feel confident answering questions when I'm in a group of people	Self-Esteem
Girls	I can stay focused on a goal despite things getting in the way	Self-Esteem
Girls/Boys	I can put a plan in place and stick with it	Self-Esteem
Girls	I recognize when choices I make today can affect my life in the future	Self-Esteem
Girls/Boys	I try to find ways to improve my life	Self-Esteem
Girls	I can describe my thoughts to others when I speak	Self-Esteem
Girls	If someone does not understand me I try to find a different way of saying what is on my mind	Self-Esteem
Girls	When others talk I pay attention to their body language, gestures and facial expressions	Self-Esteem
Girls/Boys	I can work well in a group with other people	Self-Esteem
Girls/Boys	When I have the opportunity, I can organize my peers or friends to do an activity	Self-Esteem
Girls	I often feel lonely	Self-Esteem
Girls/Boys	I ask an adult if I don't understand something	Self-Esteem
Girls/Boys	When I succeed at a task it is because I worked hard	Self-Esteem
Girls/Boys	If I succeed at a task it is because I am lucky	Self-Esteem
Girls/Boys/HH	Whether or not you will go to school	Agency
Girls/Boys/HH	When/ at what age you will get married	Agency
Girls/Boys/HH	What type of work you will do	Agency
Girls/Boys/HH	How you spend your free time	Agency
Girls/Boys/HH	How often you spend time with your friends	Agency

Geographical differences in self-esteem are apparent with girls in Somaliland reporting much higher self-esteem than girls in other project locations. Girls in Banadir also had significantly lower self-esteem than those in other locations. In all locations, most girls had neither high nor low self-esteem providing a large opportunity for RI to improve those figures in the following years of the project. In addition, girls in private schools were found to have lower self-esteem than girls in public school; however, the effect was not statistically significant. Self-esteem did not differ significantly between primary and secondary girls or between girls in IDP schools and non-IDP schools. Certain questions related to self-esteem scored lower than others. Even though most girls had moderate or high self-esteem, most simultaneously considered themselves lucky when scoring well on a test and more than half of the girls responded that they get nervous when they have to read in front of others. Some teachers recognized the connection between self-esteem and staying in school. As one teacher put it, "First of all they must be given motivation and encouragement and tell them that girls can learn and compete with boys. They must be told that they are equal to boys because an

educated girl is an educated community. And if girls are educated then they will be useful to their nation."[271]

One notable school-level factor which appears correlated with self-esteem is conflict. As the table shows, fewer girls in conflict-affected schools scored in the high or moderately high category of self-esteem than in the overall sample. Because conflict is highly correlated with geographic region – the vast majority of conflict-affected schools in the sample were in Puntland – our joint analysis of self-esteem and agency in the next section considers differences between conflict-affected and non-conflict schools within the same overall geographic locations.

TABLE 64: COHORT GIRLS' SELF-ESTEEM, ACROSS DEMOGRAPHIC AND GEOGRAPHIC GROUPS

Self-Esteem	High	Mod. High	Medium	Mod. Low	Low
Overall	0.4%	24.7%	68.5%	6.4%	0.1%
Project Location					
Somaliland	0.7%	29.2%	65.1%	4.8%	0.2%
Puntland	0.2%	24.4%	68.8%	6.6%	0%
Galmudug	0%	33.3%	66.7%	0%	0%
Banadir	0%	8.7%	78.1%	13.2%	0%
School-Level Factors					
Urban	0.4%	25.4%	67.0%	7.2%	0%
Rural	0.2%	23.0%	72.5%	4.3%	0%
Drought-Affected	0.1%	23.0%	72.7%	4.2%	0%
Conflict-Affected	0.0%	13.9%	76.1%	10.0%	0%
Non-Drought	0.4%	25.1%	66.7%	7.7%	0%
Type of School					
Primary	0%	22.8%	70.5%	6.6%	0%
Secondary	1.5%	32.7%	60.3%	5.4%	0%
IDP	0%	28.6%	63.2%	8.2%	0%
Non-IDP	0.4%	24.2%	69.2%	6.1%	0%
Public	0.4%	24.0%	69.0%	5.6%	0%
Private	0.4%	24.0%	68.9%	6.7%	0%
Demographic Factors					
Under 12 years	0%	12.1%	61.2%	24.8%	1.9%
12 years and over	0.4%	25.2%	68.8%	5.7%	0%
Girls with disabilities	0%	24.1%	69.5%	6.5%	0%
Girls without disabilities	0.4%	24.8%	68.4%	6.3%	0%

Girls' Agency

In addition to self-esteem, the baseline also measured agency of girls with regard to decisions about their education, marriage, future work, friends, and use of their free time. For each area, girls were asked who exercises control over their decisions – the girl herself, her family, or her and her family jointly. Many girls viewed decisions in these five areas as being the purview of themselves alone, or

[271] Key Informant Interview with Female Teacher. Mogadishu, Banadir.

them and their family jointly. For instance, 87 per cent of girls reported that they made decisions either by themselves or jointly with their family on where they would work in the future and when they would spend time with their friends. The lowest agency measure was the decision to continue education with only 77 per cent of girls saying that they made the decision to continue their education themselves or jointly with their family. Table 65 shows the percentage of girls who said they had all or partial decision-making power across each project location and in urban and rural areas. Girls often felt they had limited power to control whether they continued their education, particularly if they got married.[272] Marriage was raised as the predominant deciding factor in whether a girl continued her education in qualitative interviews.[273] Many girls also reported their parents discouraging them from getting married so that they could continue their education.[274]

There were no significant differences between project locations in measuring agency of girls. Reported agency of girls was high in almost every category and did not show much variation between subjects. There were also significant differences in reported agency in school and work between rural and urban girls, with rural girls reporting higher agency in those categories.

TABLE 65: AGENCY OF COHORT GIRLS ACROSS PROJECT LOCATIONS

Agency	Overall	Somaliland	Puntland	Galmudug	Banadir	Urban	Rural
School	77.1%	83.6%	67.1%	82.9%	86.0%	74.0%	84.6%
Marriage	81.5%	87.1%	73.0%	84.7%	90.6%	81.2%	82.1%
Work	87.9%	92.3%	83.1%	87.8%	89.7%	86.5%	91.6%
Time	79.4%	86.5%	70.7%	84.9%	81.8%	78.2%	82.3%
Friends	87.2%	88.0%	83.1%	95.0%	93.3%	87.8%	85.8%

Beyond project location and urbanity, we also disaggregated agency scores by school level (primary versus secondary), IDP status, and by ownership or control over the school (public versus private). While the differences across most of these categories were not statistically or substantively significant, the difference across types of schools was notable. Girls who attended private schools were shown to have both relatively lower self-esteem, and lower agency over their decision-making. However, the self-esteem and agency reported by girls in private schools are still high in an absolute sense.

Results were also disaggregated by age and disability, measured by those who had significant difficulty seeing, hearing, and walking. There were no observable differences in agency among girls with disabilities and girls without disabilities. Likewise, girls under and over the age of 12 had similar reports of agency.

Drought did not appear to be a determinant of self-esteem or agency. Importantly, girls affected by drought showed higher agency in deciding whether to continue their education. This may be because they now live with family members who do not make those decisions for them or because they have had to become more independent because of the drought.

Finally, to understand the role of conflict in girls' self-esteem and agency, the results were disaggregated according to a metric of conflict status provided by RI's Monitoring & Evaluation team. While RI rates conflict on a 4-point scale of severity, the relatively small sample size of conflict-

[272] Focus Group Discussion with CEC. Awdal, Somaliland.

[273] Focus Group Discussion with Female Students. Galgadud, Galmudug.

[274] Focus Group Discussion with Male Students. Nugal, Puntland.

affected girls motivated us to collapse the scale to a binary indicator, distinguishing respondents who live in conflict-affected communities, of any severity, and those who do not.

TABLE 66: SELF-ESTEEM AND AGENCY, BY CONFLICT STATUS AND LOCATION

	Puntland		Galmudug		Banadir	
	Conflict	Non-Conflict	Conflict	Non-Conflict	Conflict	Non-Conflict
			Self-Esteem			
High	0%	0.2%	0%	0.6%	0%	0%
Mod. High	11.4%	29.9%	50%	27.4%	33.3%	7.5%
Average	77.7%	65.1%	50%	69.6%	66.7%	78.6%
Mod. Low	10.9%	4.8%	0%	2.4%	0%	13.9%
Low	0%	0%	0%	0%	0%	0%
			Agency			
School	50.2%	74.2%	91.7%	82.3%	91.7%	85.7%
Marriage	52.1%	81.8%	83.3%	84.8%	100%	90.1%
Work	56.8%	94.3%	100%	86.9%	100%	89.2%
Time	47.2%	80.7%	91.7%	84.4%	100%	80.9%
Friends	60.1%	92.9%	100%	94.6%	100%	92.9%

As conflict is highly correlated with project location – with almost all conflict-affected schools being found in Puntland – we disaggregated self-esteem and agency by conflict status within locations. The most notable observations occur in Puntland – these observations are also the most reliable, since Puntland includes the largest sample of conflict-affected girls and a larger overall sample of girls than either Galmudug or Banadir. Among conflict-affected girls in Puntland, self-esteem is markedly lower than among girls in communities unaffected by conflict: just 11.4 per cent of conflict-affected girls have a moderately high or high self-esteem, compared to 30.1 per cent among unaffected girls.

Similar results are seen in the bottom panel of Table 66, where only 50 per cent of Puntland girls affected by conflict say that they have agency over whether they attend school, whereas 74 per cent of Puntland girls not affected by conflict say that they have agency over whether they go to school. The results show that conflict is a heavy driver of lower agency and self-esteem. While conflict is outside the scope of the EGEP-T project, it should be remembered during programming to focus on empowering girls, particularly girls in Puntland, who live in conflict-affected communities.

Boys also took a shortened version of the self-esteem assessment. The goal of data collection among boys was two-fold: first, to determine whether boys and girls have markedly different levels of self-esteem at the outset of the project, which could guide EGEP-T programming. Second, boys in EGEP-T schools are considered indirect beneficiaries and will benefit from a number of school-wide interventions including the institution and support of boys' clubs. These interventions may have an impact on self-esteem among boys, which can be determined by comparisons from baseline to midline and endline.

The study helped to determine if girls had lower self-esteem than boys in EGEP-T schools. However, the results were almost identical, without large differences boys' and girls' self-esteem. In addition, girls in four households surrounding each school were asked the same agency questions as the cohort girls in order to understand whether self-esteem among enrolled girls already differed significantly from OOSG. Table 67 shows the calculated index scores of boys and girls on the smaller subset of questions asked. The measures of agency between girls in EGEP-T schools and girls in the communities surrounding EGEP-T schools are also shown.

TABLE 67: COMPARISONS OF SELF-ESTEEM AND AGENCY TO COHORT GIRLS

Self-Esteem	Boys	Girls	Agency	EGEP-T Girls	Community Girls
High	7.5%	7.1%	School	77.1%	77.2%
Mod. High	48.3%	46.0%	Marriage	81.5%	76.4%
Average	43.9%	46.4%	Work	87.9%	80.6%
Mod. Low	0.4%	0.5%	Time	79.4%	74.4%
Low	0%	0%	Friends	87.2%	79.2%

Interpretation and Discussion

The indicators, self-esteem and agency, were measured through a module – previously used in GEC evaluations – given to cohort girls in the EGEP-T schools. They were fit for purpose as they were composed of validated questions that successfully measured self-esteem and agency. The questions were also successfully used in GEC-1 to measure self-esteem and agency. In addition, the two indicators were logical being able to measure the intermediate outcome of greater self-esteem and empowerment in girls.

Based on the baseline data, there are already high rates of self-esteem and agency among the EGEP-T cohort girls. There was no noticeable difference between the agency of community girls and cohort girls but this established a baseline to hopefully see improvement in agency among the cohort girls. Agency surveys should continue with community girls each year to understand the impact of EGEP-T programs.

By raising girls' self-esteem, they will feel more confident in school and in their ability to learn. Lower self-esteem may be linked to poorer performance in school, so encouraging girls' self-esteem – in addition to encouraging their learning – may be crucial to their ability to successfully learn in school. In addition, agency in deciding whether a girl should go to school was often the lowest measurement of any category of agency (school, marriage, work, time, and friends). Girls even stated that they had more agency in deciding who they married than whether they went to school, which may be surprising given the Somali context. Improving agency in school-related decision-making is important as it is the lowest-scoring agency category and because school-related agency has the most direction connection to the EGEP-T goals of learning and transition.

Given this, there are two learning points for EGEP-T. The first is that empowerment programs should focus on the girls with the lowest 'school' agency scores. These tended to be girls in IDP schools, and girls with disabilities. These girls, and the schools they are in, should be the ones that EGEP-T targets with female mentorship and other empowerment programs. Secondly, as girls have less agency in this area, the community, and parents specifically, should be targeted for positive girls education messaging. These programs are underway and should continue, especially in the target areas listed previously. In this way, even if empowerment programs are less successful in empowering girls to improve their agency in deciding to go to school, the parents may be more willing to send their girls to school anyway. With this two-pronged approach, it may be more likely that girls stay in school with improved self-esteem and agency.

5.3 Improved Teaching Quality

Indicator: Percentage of trained teachers demonstrating improved teaching practices

Indicator Development

Improvement in teaching quality is an additional intermediate outcome for the EGEP-T project. Teaching quality is poor across Somalia. Teaching quality affects both a girl's learning and her transition to secondary school and post-secondary school. A girl will likely learn more when the teacher employs high quality teaching methods. Effective teaching may also be an incentive to the girl to stay in school as she feels like she is learning and that school is worth her time and money. However, the opposite is true too – a girl may be more inclined to drop out of school if the teacher is often absent or ill-prepared. Similarly, she will not learn as much if the teacher fails to engage both boys and girls.

One primary and four ancillary indicators were selected to examine teaching quality. The primary project indicator is a measurement of teaching approach generally, with a focus on participatory and student-centred teaching methods. The indicator was chosen to understand how teachers are trying to educate students in the classroom and where there is room for improvement. During the classroom observations, researchers noted when they saw different teaching styles (for example: group work, activities, or lecture). Cohort girls and boys also were asked about how they viewed their teacher's teaching style and whether their teachers ever exhibited certain 'good' teaching behaviours.

The ancillary indicators were developed at the start of the evaluation to provide a nuanced picture of teaching quality in EGEP-T schools. Importantly, some aspects of teaching quality analysed in this section are not being targeted by the project; for instance, EGEP-T does not expect to exert significant impact on teacher absenteeism, as it is a function of systemic shortcomings with respect to teacher pay. For this reason, our focus for setting targets is on the project's primary indicator – teaching approach – which the project will target most directly with its interventions. The ancillary indicators described below provide a more detailed picture of teaching quality, but should not generally be used for setting targets for the midline and endline.

The first ancillary indicatory assesses gender-sensitive teaching and gender equity. This indicator was measured in four different surveys and was selected because inequality may lead to girls learning less and dropping out. The classroom observation was done with two classrooms in each school and asked team leaders to observe whether teachers differed in their approach to girls and boys including how they provided feedback, who they directed questions to, and whether boys and girls had similar access to materials and desks.[275] The boys and girls surveys also asked the students whether they thought there was gender equity in the classroom. Lastly the teacher survey measured teachers' attitudes towards gender equity and their views on the behaviours of boys and girls, with the assumption that such views could influence the ways they teach.

[275] The teachers selected for observation were almost universally teaching English, mathematics, or Somali. In total, 31.5 per cent of those observed taught English, 35.6 per cent taught mathematics, and 29.6 per cent taught Somali. Just 3.3 per cent taught other subjects, including biology and religion. Teachers were drawn from a wide range of grades, but these observations were concentrated among grades targeted explicitly by EGEP-T interventions: 84.5 per cent of classroom observations were conducted with classes from Grade 6 through Form 4. At times, observations were conducted with younger classes, though most teachers teach multiple grade levels.

A second ancillary indicator, preparation, was selected to understand how much effort teachers were putting into their lessons. 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 asked whether they thought their teacher was prepared

The third ancillary indicator, punishment, was measured in three different surveys and was selected due to its importance for child protection and allowing children, including girls, the space to grow and learn within school. Cohort girls and boys were asked about whether their teacher used physical punishment and how often.

The fourth, and final, ancillary indicator focuses on teacher absenteeism, measuring how often the teachers were actually in the classroom teaching. This was measured in three different surveys. Cohort girls and boys were asked about how often their teacher was absent for a whole day or part of a lesson. The school survey also included questions for the head teacher about how many days each individual teacher had missed in the past two weeks.

Primary Findings – Teaching Approach

As might be expected, teaching approaches varied from classroom to classroom, although several common trends emerged. Across the board, teachers tended to "gather participation"; that is, teachers tended to go around the room, calling on a wide variety of students, and – often – calling on students who were not actively participating. The majority of teachers also lectured, according to cohort boys and girls.

Teaching approaches were assessed primarily by enumerators who observed classrooms at EGEP-T intervention schools. In total 270 classrooms were observed across 140 schools, with each observation lasting approximately 45 minutes. An additional source of information on teaching approach came from surveys of female students, who were asked to assess the extent to which their teachers encouraged participation, and the frequency with which their teachers lectured during class time.

Results drawn from these two sources are reported in Table 68, below. According to female students, the vast majority of teachers across all locations encourage participation on a regular basis. At the same time, students report that they spend a significant portion of their instructional time repeating what their teacher says. This disjuncture appears contradictory. However, as we discuss below, we view teachers as engaging in a menu of activities during a class period, and that menu can, and often does, include both significant periods of rote repetition as well as encouragement of active participation.[276]

The frequency with which teachers use repetition is confirmed, to some degree, by the data collected during classroom observations. Across observations of 270 teachers, enumerators indicated that 59.3 per cent of teachers used a substantial portion of class time for rote repetition during at least one of the three blocks of 15 minute observations conducted. At the same time,

[276] The precise nature of the questions posed to students may also drive a portion of this counter-intuitive finding. Students were asked the frequency with which their teacher encourages participation, but were provided no guidance on what constitutes a level of encouragement that would justify a response of "often" or "sometimes". As a result, students were left to make subjective assessments about whether their teachers' encouragement was very frequent or not very frequent; for many students, even limited encouragement may have justified a strongly positive response, if they have few expectations around best practices for teaching. As a result, students may view their teacher as being quite encouraging by the standards they believe are reasonable, while still using rote repetition during much of the class period.

relatively few (14.4 per cent) teachers used repetition during *all three blocks*. This finding suggests that most teachers use repetition as part of their teaching approach – perhaps even pre-dominantly – but that they also use other methods. To illustrate, consider the set of teachers who were recorded as using a significant amount of rote observation during exactly one observation block: among these 53 teachers, 59.3 per cent of them also actively sought out the participation of disengaged students at some point during the class period.[277] While this rate could be improved dramatically, this finding highlights the fact that teachers engage in a number of activities within a single class period. It also suggests the importance of more nuanced measurement regarding the *share* of each class consumed by different activities going forward.

Overall, the rates at which teachers lectured (i.e. students repeated the words of their teacher), were relatively high, though much lower than the rate indicated by students themselves. Rates also varied, though not markedly, across locations, as shown in the bottom panel of Table 68 – lecturing of this kind was least common in Galmudug, where only 52.9 per cent of teachers lectured extensively during at least one of three observation periods.

TABLE 68: TEACHING APPROACHES ACROSS PROJECT LOCATIONS

Teaching Approach	Overall	Somaliland	Puntland	Galmudug	Banadir
Outcomes Based on Student Surveys					
Teacher often or sometimes encourages participation	95.3%	92.9%	97.0%	99.4%	94.3%
Agree: We spent much of the day repeating what my teacher says	86.0%	87.3%	83.8%	95.1%	84.3%
Outcomes Based on Classroom Observations					
Teacher lectured*	59.3%	55.1%	66.1%	52.9%	52.5%
Teacher used student-centred activities**	49.6%	40.8%	49.6%	52.9%	70.0%

*Classrooms were observed in blocks. Teachers were recorded as lecturing if students spent most of their time repeating the teacher's words aloud during at least one of three 15-minute blocks of observations in each classroom. **Teachers were classified as using student-centred activities if enumerators recorded that they used such activities or games at least once during any of the three blocks of observations.

Lecturing was also more common in secondary schools than primary schools, according to the classroom observations.[278] Using activities to get the students engaged seems like a less popular method with only 49.6 per cent of teachers in the classroom observation attempting to use activities to engage the students. This was most common in Banadir where over half the teachers attempted to use some sort of activity for engagement.[279] Activities and participation were also more common

[277] This finding is not limited to the distinction between repetition and participation. Among teachers who engaged in rote repetition during at least one 15 minute block of observations, 60.6 per cent also had students complete group work, 75.6 per cent had students teach each other, and 67.5 per cent used student-centred activities or games at some point during the overall period of classroom observation.

[278] Note that classroom observation data was not collected in Hirshabelle, because school examinations had already started by the time fieldwork began in the area.

[279] Note that the rates of lecturing and student-centred activities documented in the bottom panel of Table 66 do not sum to 100 per cent. This is because some teachers did not spend a significant portion of their time lecturing, but also did not engage in student-centred activities, such as students teaching one another or working in groups. In short, the two approaches do not constitute an either-or distinction.

in IDP schools.[280] In qualitative interviews, teachers introduced several different student-centred activities they have used in their classrooms. Some teachers had students give presentations on their work.[281] Other teachers said they brought in props for their lessons and let the students play with them. The example one teacher described was that they brought leafs and twigs in to teach students about nature.[282] Group discussions were also used to involve the students in the classroom.[283]

A participatory-based approach may be a recent development, at least in Puntland, where one female teacher said they had used the lecture method for many years but recently adopted a system where students were supposed to engage in the class.[284] Similarly a teacher in Puntland said they have started to do group work in order to identify students weaker in specific subjects.[285] The same teacher commented that a lack of teacher training was one of the largest challenges.[286] A Somaliland teacher said she had started to connect lessons to previous and future lessons – one of the key behaviours measured in the classroom observation.[287]

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 (see Table 69 for a complete list). An index was created of these fifteen behaviours by assigning positive behaviours a 0 for not observed and 1 for observed and negative behaviours a 0 for observed and 1 for not observed. As noted previously, teachers were observed for three separate 15-minute blocks; if a behaviour is observed during *any one of the three blocks*, the behaviour is coded as having occurred. To illustrate, consider a teacher who has students complete group work during the first block of observations and is observed reprimanding a student who offered an incorrect answer during the second block – the teacher would be coded as having both approaches observed, even though the approaches only occurred during one the three blocks. The index ranged from 0 to 16 points, with higher scores representing the most positive possible mix of teaching approaches.

One further note of clarification is necessary with regard to two specific approaches documented in Table 69: students copying from the board, and students repeating their teacher's words aloud. To be counted as observed, these behaviours had to be observed for the *majority of the observation.*[288] For example, if an observation lasted 20 minutes, the team leaders would indicate that a behaviour had been observed if it had been observed for more than ten minutes. Thus behaviours were only

[280] It is important to note that enumerators were not given specific criteria for identifying "student-centred activities or games"; as a result, these findings may reflect differences in definitions across enumerators. However, the overall rate of student-centred activity use generally accords with the rates at which enumerators classified students on more specific measures. For instance, rates of group work are similar to those of the more vague "student-centred activities", with 47.8 per cent of teachers employing group work in at least one block of observations, and the highest and lowest rates being observed in Banadir and Galmudug, respectively. As with student-centred activities, group work rates were also higher in IDP schools.

[281] Key Informant Interview with Female Teacher. Garowe, Puntland.

[282] Key Informant Interview with Female Teacher. Maroodi Jeeh, Somaliland.

[283] Key Informant Interview with Female Teacher. South Galkacyo, Galmudug.

[284] KII with Female Teacher, Garowe, Puntland.

[285] KII with Male Teacher, North Mudug, Puntland.

[286] Ibid.

[287] KII with Female Teacher, Hargeisa, Somaliland.

[288] In both cases, the data collection instrument instructed enumerators to record if students spent *most of the time* copying from the board or repeating their teacher's words aloud. This standard was not applied to other approaches included in the index.

counted if they made up a large part of the observation and not if they were only performed for a small amount of time. These were then aggregated into an index with three categories – poor quality, average quality, and good quality.

TABLE 69: OBSERVATIONS USED TO COMPILE TEACHER QUALITY INDEX

Type of Behaviour	Question	Observed at least once during an observation
Positive	Is the teacher confident in their presentation of the material?	97.4%
Negative	Students spent most of the time copying from the board.	85.2%
Negative	Students spent most of the time repeating teacher's words aloud.	59.3%
Positive	Teacher used student-centred activities or games.	49.6%
Positive	Teacher allowed students to instruct each other (e.g. come to board to demonstrate something, or explain to classmate).	65.2 %
Positive	Teacher asked open-ended question (requires more than simple answer) that encourages thinking.	84.8%
Positive	After a student gave an incorrect answer, did the teacher explain the concept in a new way?	87.0%
Positive	After correcting a student who gave an incorrect answer, did the teacher verify the student understood the question now?	81.5%
Negative	If a student gives an incorrect answer, are they reprimanded (verbally or physically)?	31.1%
Positive	Teacher called on or actively tried to involve a student who was not participating.	87.0%
Positive	Students worked together in groups.	47.8%
Positive	The teacher summarized and clearly stated a key concept or takeaway point from the lesson.	85.9%
Positive	The teacher referred back to previous lessons, relating this lesson to previous lessons.	75.6%
Positive	The teacher stopped the lesson and invited questions from students.	78.5%
Positive	The teacher employs a variety of explanations that differ in difficulty for the diverse learners in the classroom.	78.9%

It is important to note that the contents of the 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 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.

Results for the teacher quality index are provided in Table 70. The bottom panel of the table establishes targets for the midline and endline on this indicator. Importantly, for the purposes of defining targets, we focus on the share of teachers who score in the "good" range. From a practical standpoint, the targets have been adjusted across locations to reflect the higher overall quality of teachers in Galmudug and Banadir at the baseline. Using the current scoring approach, there is little room for improvement in these two locations, and the established targets reflect this fact. Prior to

the start of the midline evaluation, the evaluation team will reassess the scoring approach utilised and will – in consultation with RI and the FM – adjust the approach such that baseline scores are lower and more realistic targets can be established for future evaluation waves.

TABLE 70: TEACHER QUALITY INDEX ACROSS PROJECT LOCATIONS AND SCHOOL LOCATION

Quality	Overall	Somaliland	Puntland	Galmudug	Banadir	Urban	Rural
Poor	4.1%	8.2%	2.6%	0.0%	0.0%	2.0%	9.6%
Average	21.1%	31.6%	20.0%	11.8%	2.5%	15.7%	35.6%
Good	74.8%	60.2%	77.4%	88.2%	97.5%	82.2%	54.8%
			Targets				
Midline	77.8%	64.2%	80.4%	90.2%	98.5%		
Endline	80.8%	68.2%	83.4%	92.2%	99.5%		

Absenteeism

Absenteeism was measured using two different approaches. The first approach relied on reports and ratings of absenteeism by teachers from in-school girls. Girls were asked to assess their teacher's attendance using three survey questions, as outlined below:

- My teachers are often absent from class (agree or disagree)
- Over the last two weeks, how many times did this occur: my teacher was absent for a lesson.
- Over the last two weeks, how many times did this occur: my teacher was gone from the classroom for 30 minutes or more, and no other teacher was present

Notably, boys enrolled in EGEP-T project schools were asked a similar set of questions, with similar results. As the results were similar across the two samples, we focus on the much larger respondent pool of girls in this analysis.

The results drawn from surveys with cohort girls are reported in the top panel of Table 71. According to the girls surveyed, the typical teacher across the sample was absent for 0.54 lessons over the previous two weeks. The rate was much higher in Somaliland than other project locations, as respondents in the boys survey reported their teacher missing 0.81 lessons per two week period. Among boys, reports of absenteeism in Somaliland are even higher, at almost a full lesson missed per two week period.

Teacher absenteeism was mentioned as a problem in several qualitative interviews. Male students stated that they had taken to teaching their own class when the teacher was absent because the teacher was absent on such a frequent basis. [289] Teachers are often not paid on a regular basis and this may be why they do not attend class on a regular basis. [290] Some teachers reported punishment if they were gone too long. These punishments included suspension but it is not known how widely these punishments are enforced. No punishment for absenteeism was mentioned, such that teachers who are chronically absent are still present in the classroom (i.e. not suspended). [291]

[289] Focus Group Discussion with Male Students. Hargeisa, Somaliland.

[290] Focus Group Discussion with CEC. Awdal, Somaliland.

[291] Key Informant Interview with Female Teacher. Hargeisa, Somaliland.

TABLE 71: TEACHER ABSENTEEISM ACROSS PROJECT LOCATIONS

Absent - Survey	Overall	Somaliland	Puntland	Galmudug	Banadir
Reports from Cohort Girls					
Times teacher missed lesson, last two weeks	0.54	0.81	0.38	0.24	0.43
Times teacher left class for 30+ minutes, last two weeks	0.85	0.90	0.65	0.19	1.75
Agree: My teachers are often absent	25.2%	34.1%	19.6%	13.8%	22.8%
Reports from Head Teachers					
Avg. Full Days Absent in Previous 2 Weeks	1.35	0.7	1.7	0.4	1.7

Respondents were also asked to assess their teachers' attendance in a more informal manner by indicating whether they agree or disagree with the notion that "my teacher is often absent." We classify children as agreeing with this statement if the agree a little or a lot. In total, 25.2 per cent of respondents agree that their teacher is often absent; as with the results previously – in which absences from lessons were counted over a two-week period – girls in Somaliland report that their teachers are absent much more often than do girls in other project locations.

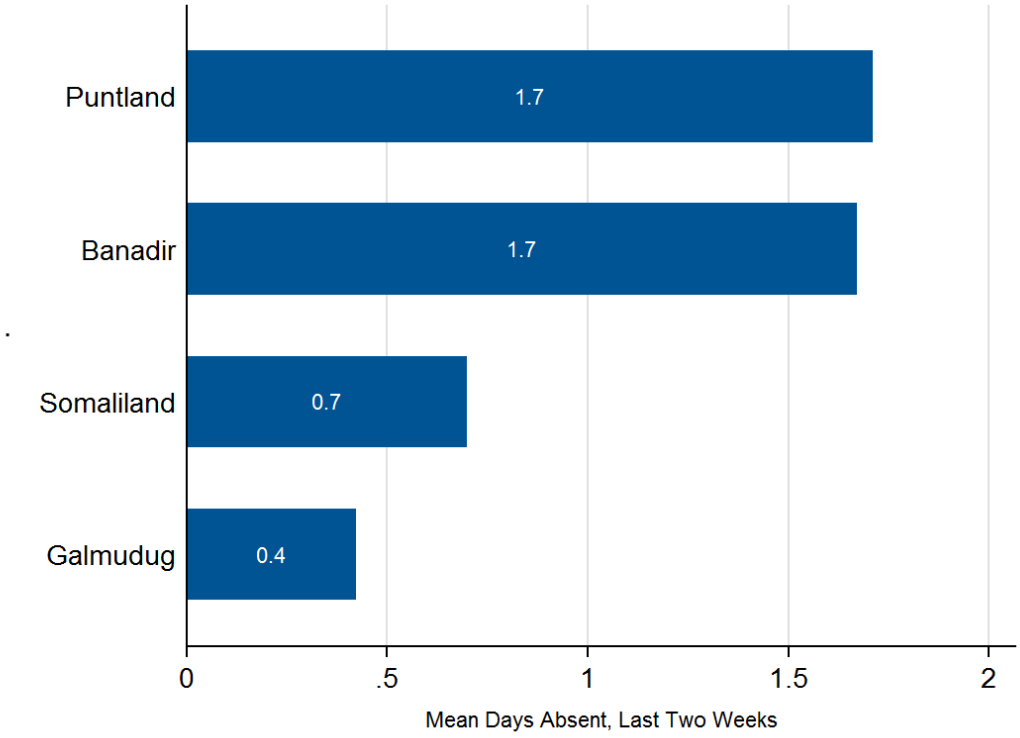
The bottom panel of Table 71 moves away from reports by students, and focuses on reports of absenteeism collected from head teachers. At the time of the baseline, head teachers were surveyed regarding the operation of their school. As part of the process of sampling teachers for participation in other data collection efforts, enumerators collected information on absences for a random sample of teachers in each school. For each teacher, head teachers were asked to report the number of full and partial days that teacher had missed over the previous two weeks. In the bottom panel of Table 71, we report the mean number of full days missed by teachers, by project location.

Overall, head teachers document higher rates of absenteeism than students. However, this discrepancy may stem from the sampling approach, as girls may be referring to any number of teachers from whom they receive lessons, while head teachers are referring to a specific set of teachers, who are drawn from teachers of specific subjects. More concerning is the extent to which head teacher reports fail to match with reports from cohort girls with respect to variation across locations. According to girls, absenteeism is highest among teachers in Somaliland; but, according to head teachers, Somaliland has lower teacher absenteeism rates than both Banadir and Puntland.

Absenteeism, according to the reports of head teachers, was lower in primary schools than in secondary schools, with teachers missing an average of 1.17 days in primary schools and 2.4 days in secondary schools. Schools with mixed primary and secondary grades had much lower absenteeism rates, though this is likely due to the small sample of teachers drawn from such schools.[292]

[292] Interestingly, absenteeism by this metric is lower in secondary schools in Somaliland specifically. As RI noted in comments on this report, this fact may stem from the fact that school fees are paid for secondary school, but not primary school, in Somaliland. As a result, teachers may be less likely to receive consistent salaries in Somaliland primary schools. While we cannot test this hypothesis directly, it is consistent with the simple fact, noted above, that absenteeism is lower in Somaliland secondary schools than primary schools, but the opposite pattern obtains in the other project locations. At the same time, it is worth noting that a number of Somaliland primary schools report charging school fees; thus, the extent to which fees are not charged at the primary level is not entirely clear.

FIGURE 21: TEACHER ABSENTEEISM IN PREVIOUS TWO WEEKS BASED ON HEAD TEACHER REPORTS



Preparation

Overall, most teachers were able to start their class on time and appeared to be prepared for class with 92.3 per cent of teachers starting class on time and 66.0 per cent having a lesson planned. According classroom observations done by trained researchers, preparedness was least common in Puntland where only 88.3 per cent of classes started on time and only 57.9 per cent of teachers had a lesson plan. However, student perceptions of preparedness were lowest in Somaliland. Preparedness was not defined in the student surveys and their perceptions most likely are more encompassing than the start time and lesson plan preparation that the researchers observed in the classroom. There were not large differences between urban and rural or primary and secondary schools. The boys and girls survey data seems to mirror the data recorded in the classroom observation as generally supporting the idea that teachers are well prepared for their lessons (refer to Tables 72 and 73 for disaggregated data).

Qualitative interviews suggest that teachers' preparation depended on their teacher training. One female teacher stated, "Some teachers had trainings and know how to teach and some of them have not had a training. For those who had trainings, they prepare and plan their lesson on a monthly basis and sometimes they plan mid-term lessons or the academic year lessons by following guidelines and setting up a goal that they can achieve for themselves." A male student focus group reported that the teachers in their school put effort into lessons and are also willing to sit with students after class to make sure they understand the material.[293] One focus group also pointed out that teachers were not willing to put in a lot of work into their classes when they were not paid on time or, in some cases, at all.[294]

[293] Focus Group Discussion with Male Students. Hargeisa, Somaliland.

[294] Focus Group Discussion with CEC. Hargeisa, Somaliland.

TABLE 72: PREPARATION ACROSS PROJECT LOCATIONS

Preparation - Survey	Overall	Somaliland	Puntland	Galmudug	Banadir
Outcomes Based on Classroom Observation					
Lesson Started On Time	92.3%	94.6%	88.3%	93.9%	97.5%
Teacher had Lesson Plan	66.0%	71.6%	57.9%	93.9%	61.5%
Outcomes Based on Student Survey Responses					
Agree: My teacher is prepared for class each day	84.8%	75.4%	89.1%	99.4%	89.6%

TABLE 73: PREPARATION ACROSS SCHOOL LOCATION, IDP STATUS, SCHOOL LEVEL, AND SCHOOL TYPE

Preparation - Survey	Urban	Rural	IDP	Non-IDP	Primary	Second.
Outcomes Based on Classroom Observation						
Lesson Started On Time	91.0%	96.2%	100%	92.2%	91.7%	95.2%
Teacher had Lesson Plan	64.7%	69.5%	66.7%	66.0%	66.1%	65.6%
Outcomes Based on Student Survey Responses						
Agree: My teacher is prepared for class each day	86.1%	81.9%	85.4%	84.8%	84.1%	87.8%

Gender-Sensitive Teaching

Our approach to measuring gender-sensitive teaching emphasises gender equity. While the goal of EGEP-T programming is to encourage gender-sensitive teaching, in which teachers are likely to treat male and female students differently depending on their backgrounds, measurement of gender-sensitivity in teaching is challenging. When teachers treat students differently, this fact may reflect positive discrimination meant to help female students overcome barriers to their education; but it may also reflect implicit bias or another form of gender inequity. Unfortunately, without a clear understanding of the specific context and a detailed observation of teachers over time, it may be difficult to determine whether a teacher's approach is gender-sensitive, as opposed to gender-equal.

The approach we take focuses, to the extent possible, on equal treatment where such treatment is most in line with gender-sensitivity. We investigate the treatment of gender within classrooms using data derived from four distinct survey questions:

- For female students: Girls are treated equally to boys (agree/disagree)
- For male students: Girls are treated equally to boys (agree/disagree)
- For teachers: Who should be prepared for a professional career (boys or girls more, or equally)
- Classroom observation: Girls and boys have equal access to learning materials

These questions, though focused on gender equity, are reasonable proxies for gender-sensitivity on the part of teachers. For instance, even if teachers treat students differently in an effort to help girls perform better in school, it is generally undesirable for female (or male) students to feel that they are being treated differently.

The majority of girls reported that their teacher asked the same amount (91 per cent) and same difficulty of questions (89.9 per cent) to both boys and girls in the classroom. However, only 61 per cent of girls said they felt their teacher treated boys and girls equally in the classroom. As noted

above, these findings need to be interpreted cautiously, because they reflect a distinction between gender-equal and gender-sensitive teaching.

Although many students perceive their teachers to treat students differently based on gender, the qualitative data suggest that these differences often take the form of "positive discrimination". Quantitative surveys – at least as implemented during the baseline – do not provide sufficient detail to differentiate between these explanations, but qualitative data, where respondents were able to explain the nature of gendered differences in treatment, shed greater light on the issue. For instance, boys participating in FGDs occasionally indicated that they were treated better in school than girls.[295] In contrast, teachers tended to assert that they treated children the same, regardless of gender or – in some cases – actively encouraged female students to participate.[296] In at least one case, a teacher said, because girls are shy, they would make sure to select girls to answer questions as they would not answer independent of encouragement.[297]

To the extent that quantitative data is available to test differences in teachers' treatment of students by gender, it tends to suggest that girls are treated better. In one example, enumerators recorded the number of times girls and boys, respectively, were called on to answer questions in class. Across all classroom observations, girls were asked marginally more questions than boys, in line with the reports of teachers who say that they proactively call on girls to compensate for their shyness.[298] Ultimately, more inquiry needs to be done to into whether differential treatment exists and, if it does, whether it is beneficial towards girls (i.e. providing more support to girls).

Beyond the perceptions of students, teachers were also surveyed regarding their attitudes toward male and female students. At least 20 per cent of teachers consistently showed attitudes that were more favourable towards boys in each question. See Tables 74 and 75 for disaggregated results.

TABLE 74: GENDER EQUITY ACROSS PROJECT LOCATIONS

Gender Equity - Survey	Overall	Somaliland	Puntland	Galmudug	Banadir
Girls perceive equal treatment	61.1%	47.7%	66.8%	85.1%	65.4%
Boys perceive equal treatment	56.7%	45.3%	62.7%	84.9%	49.1%
Teacher: Girls and boys should be equally prepared for a professional career	69.6%	61.7%	70.8%	77.1%	74.9%
Classroom observation: equal access to learning materials[299]	88.2%	86.8%	86.3%	93.9%	94.6%

[295] Focus Group with Boys. Maroodi Jeeh, Somaliland

[296] Key Informant Interview with Female Teacher. Garowe, Puntland. Key Informant Interview with Male Teacher. North Mudug, Puntland.

[297] Key Informant Interview with Female Teacher. Hargeisa, Somaliland.

[298] Similarly, when teachers were asked who is assigned chores in their classroom, 49.2 per cent indicated that boys and girls are assigned chores equally. Of those reporting a difference between girls and boys, a higher share (23.8 per cent versus 19.5 per cent) of teachers assigned additional chores to boys than to girls.

[299] Classroom Observation Question "Girls and boys have EQUAL access to desks, learning materials, etc. (e.g. same amount of sharing of books, desks)."

TABLE 75: GENDER EQUITY ACROSS SCHOOL LOCATION, IDP STATUS, SCHOOL LEVEL, AND SCHOOL TYPE

Gender Equity – Survey	Urban	Rural	IDP	Non-IDP	Primary	Second.	Public	Private
Equality of Treatment - Girls	65.4%	50.5%	65.4%	60.5%	60.6%	62.9%	56.4%	73.7%
Equality of Treatment – Boys	58.6%	51.4%	56.1%	66.7%	56.5%	57.5%	52.6%	63.6%
Equality of Beliefs – Teacher	71.3%	61.5%	N/A	N/A	68.9%	71.8%	N/A	N/A
Equal Access – Class Obs.	85.8%	92.4%	100%	87.6%	90.1%	79.8%	86.8%	N/A

Punishment

To evaluate the extent to which teachers punish students, especially through the use of corporal punishment, the baseline incorporated questions focused on punishment into surveys with students and into classroom observation tools. Our focus in this section is on the responses provided by female and male students, respectively, when asked the following three questions:

- How do the teachers punish students? Use of physical punishment
- How do the teachers punish students? Use of detention
- Think about the past week at school, or the last week you were in school. In that week, did you see a teacher use physical punishment on other students?[300]

Beyond these three questions, the evaluation collected additional data. For instance, students were asked whether teachers had used corporal punishment *on them* in the past week. In addition, enumerators observing classrooms were asked to observe whether students were disciplined physically during class. We opt to focus on the data sources outlined previously, as we feel these responses are less subject to social desirability bias, and similar concerns. Students may be hesitant to admit that they were personally punished by their teacher – since it might imply that they had done something wrong – but should be more candid when describing teachers' actions with regard to other students. Likewise, the use of physical punishment during classroom observations may be muted if teachers think that they will be judged by the researcher for doing so.[301]

The top panel of Table 76 provides the rates at which female students report the use of corporal punishment and detention within their schools. Overall, 46.3 per cent of girls report that their teacher uses physical punishments in their classrooms. A slight majority of girls, 53.5 per cent, reported that their teacher used shouting as a form of punishment. Only 36.0 per cent of girls said their teacher used detention as a method of punishment. However, these punishments appear infrequent. In total, 74.0 per cent of girls reported that they had never seen their teacher use

[300] This question concerns the use of corporal punishment on *any* student, not the student specifically surveyed.

[301] Slightly different from social desirability, this is one example of the Hawthorne Effect, in which research participants react to being observed by altering their behaviour.

corporal punishment in the past week and only 12.7 per cent of girls said they had personally been punished physically by their teacher in the past week.

TABLE 76: TYPES OF PUNISHMENT EMPLOYED BY TEACHERS, ACROSS PROJECT LOCATIONS

Types of Punishment Employed by Teachers	Overall	Somaliland	Puntland	Galmudug	Banadir
Outcomes Based on Surveys with Female Students					
Teacher uses corporal punishment	46.3%	43.4%	38.3%	83.8%	47.8%
Teacher uses detention	36.0%	36.3%	29.8%	60.0%	32.1%
Teacher used corporal punishment in last week	26.0%	19.7%	24.6%	56.0%	27.1%
Outcomes Based on Surveys with Male Students					
Corporal Punishment	47.3%	40.6%	46.7%	63.3%	47.3%
Detention	38.5%	33.7%	33.3%	70.8%	40.0%
Teacher used corporal punishment in last week	26.8%	18.8%	28.4%	35.6%	34.7%

The rate of corporal punishment was slightly higher in public schools, with 50.2 per cent of girls in public schools saying their teacher used corporal punishment and 28.3 per cent of girls in public schools saying their teacher had used corporal punishment in the past week. Only 40.0 per cent of girls in private schools reported their teachers using corporal punishments, and the frequency of corporal punishment observed over the previous week was also lower among girls attending private schools.

Corporal punishment was far more common at the primary level than the secondary level, where only 29 per cent of girls reported that their teacher used corporal punishment on students in their class, compared to 49 per cent of primary students ($p<0.05$). It was also more common at IDP schools where 53 per cent of girls reported corporal punishment used by teachers.

Opinions regarding the frequency with which corporal punishment is used in Somali schools appear to be mixed, based on qualitative interviews with teachers and head teachers. In many cases, punishment did not seem to be something that was allowed at schools. In a discussion with a female teacher in Puntland, she said that the school had banned physical punishment.[302] Similarly, a female teacher in Galmudug stated that physical punishment was harmful to the student's learning. However, this same teacher also said some of the punishments included making the child stand face down while holding their ears.[303]

Other teachers claimed that corporal punishment is not used in their schools, but the quantitative data make clear that corporal punishment is moderately common across schools. A teacher in Somaliland said a more effective punishment instead of physical beating was calling the child's parents.[304] But still another teacher described the "murga" punishment – a common, informal disciplinary technique in South Asia – in which students bend at the waist and hold their heads

[302] KII with Female Teacher, Garowe, Puntland.

[303] KII with Female Teacher, South Galkayo, Galmudug.

[304] KII with Male Teacher, Hargeisa, Somaliland.

between their knees, though the teacher noted that this punishment should be reserved for male students.[305] According to qualitative interviewees, corporal punishment is commonly used to discipline students for getting into fights, not finishing their homework, or using social media but the murga sanction described above was typically used for students who had disrupted the classroom.[306] The qualitative data reaffirms the fact that the use of corporal punishment is relatively common, but likely varies widely from school to school, with idiosyncratic disciplinary techniques being used in some cases.

Interpretation and Discussion

These indicators – absenteeism, gender equity, punishment, preparation, and teaching approach – allowed for a strong, multidimensional measure of teaching quality in the EGEP-T baseline. While there are many ways to measure teacher quality, these were indicators that could be measured in a variety of ways – through classroom observations, student surveys, head teacher surveys, and surveys of teachers themselves. Thus, it allowed us to quasi-validate the measures for midline and endline.

At the baseline, absenteeism reported by head teachers was 1.35 full days missed by the typical teacher over the previous two weeks. It is important to note that EGEP-T is unlikely to impact teacher absenteeism rates for two reasons. First, teacher absenteeism is not a specific focus of EGEP-T programming, although better school management and teacher training programmes should reduce absenteeism, theoretically. Second, to the extent that teacher absenteeism is driven partially by late or unpaid salaries, this is not an outcome that EGEP-T can directly impact. Nonetheless, the importance of teachers' consistent attendance motivated its inclusion in this section, as it is likely to shape other downstream outcomes that EGEP-T more directly targets. We propose that EGEP-T should target absenteeism among teachers who are receiving incentives or continuous professional development through EGEP-T, though we do not set specific targets for the project to achieve in this respect.[307]

Teacher preparation was actually quite high at the baseline – classes began on time in 92.3 per cent of classrooms observed – and so the target for the endline evaluation should be to maintain on-time rates at over 91 per cent.

The teaching approach index developed by the evaluation team revealed that most teachers lecture during their classes and few engage the students with activities. A target should be set of 50 per cent of teachers using activities in their classrooms by the end of the project (up from 30 per cent at baseline). The target may be slightly less in primary schools as they used activities less often. The gender equity indicator found that only 88.2 per cent of girls had the same access to materials as boys and only 63.0 per cent of girls believe that they are treated equally in the classroom.

---

[305] KII with Male Teacher, Awdal, Somaliland.

[306] Focus Group Discussion with Male Students. Hargeisa, Somaliland; Key Informant Interview with Female Teacher. Galkacyo, Puntland; Key Informant Interview with Male Teacher. North Mudug, Puntland; Key Informant Interview with Female Teacher. Garowe, Puntland.

[307] At the time of the baseline, RI had not identified the teachers who would eventually participate in incentives and teacher training portions of their programming. The baseline absenteeism rate referenced in this discussion is absenteeism over a two-week period for a random sample of teachers from EGEP-T schools, and will only partially overlap the set of teachers participating in incentives and training programmes. In order to establish a baseline attendance rate among these subsets of teachers, we make suggestions for both the midline and pre-midline data collection in the recommendations section of this report.

Lastly, almost half of girls reported physical punishment being used in their classrooms against both girls and boys. However, corporal punishment is not used on a frequent basis in *most* classrooms, with 26.0 per cent of female respondents reporting that their teacher had used physical punishment against a student at some point in the previous week.

Teaching quality is strong in some areas but weak in others. Far too many girls feel they are not being treated equally in classrooms. Only a minority (37.4 per cent) of classrooms observed had mixed seating – in which male and female students were not segregated in the classroom – which is one indicator that can reinforce notions of girls' unequal treatment.[308] Only a few classroom observations reported mixed gender seating. In addition, many students report that their teacher is frequently absent, which clearly impedes learning. Teachers too often use physical punishment in their classrooms, which can set up an environment of fear for the students instead of an environment of learning. While teachers do often start their classes on time, over one-third are without a prepared lesson. There are many areas of improvement for teachers in Somalia across project locations and types of school. EGEP-T should focus on teacher quality as a major part of their project in order to make sure that girls have a reason to come to school and actually are in an environment that fosters learning and success.

5.4 Improved School Management and Institutional Governance

Indicator: Increase in number of Community Education Committees contributing to effective school management

Indicator Development

Another key aspect of the EGEP-T Theory of Change is the role of school management and governance. School management is seen as an intervening factor promoting student learning and sustainability. EGEP-T planning indicates that efforts to promote monitoring and support from relevant government ministries to CECs and schools will improve the management of these institutions. Moreover, if community attitudes toward education become increasingly positive – and communities take a more active role in managing local schools – the outcome should be improved school management.

To measure the quality of CEC management and governance, we developed an index of several complementary indicators. This method is analogous to a scorecard approach, in which multiple related indicators are assembled into a single score. In our case, multiple measures were aggregated to form an index that ranges from 0 to 100. These indicators capture multiple aspects of school management and institutional governance. Importantly, they are not fully comprehensive; indeed, there are many aspects of institutional governance that are intangible or difficult to conceptualize clearly. Others are idiosyncratic: specific to a particular circumstance, but not generalizable to other locations or situations. The indicators described below have the advantage of being, more or less, universally positive and achievable outcomes across all targeted schools. Further, they are measurable in a systematic way, ensuring that valid comparisons can be made over time and across space.

[308] Importantly, we do not suggest that EGEP-T target mixed-gender seating in the classroom. As RI's Monitoring & Evaluation team pointed out, mixed-gender seating is reflective of culture and the underlying religious conservatism in each area, and is not a specific target of their project. Efforts to change segregated seating led by an external actor are unlikely to be successful and may cause resentment or hostility to other aspects of EGEP-T programming.

The indicators selected can be classified into two themes. The themes and indicators are described below. Most indicators are measured through an interview with the school head teacher, and direct observation of school records. Exceptions are measured via interviews with teachers other than the head teacher, denoted in the list below with an asterisk (*):

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, 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)

The first theme concerns overall management of schools by their CECs. In order to be effective, CECs must meet on a regular basis and manage their duties effectively. Note that management quality is assessed by teachers, who may have vested interests in the action of the CECs. Despite the potential for bias in their responses, teachers are informed observers of how CECs operate, which gives them unique insight into the quality of CEC management. Moreover, randomly-sampled community members may be insufficiently familiar with the CEC to provide a valid assessment of its performance.

The second theme focuses on the extent to which the CEC actively supports the school they oversee. Support can take either financial or non-financial forms, and this is reflected in the selected sub-indicators. For instance, CECs may support girls enrolled in school through the provision of bursaries; they may also contribute financially to school projects, paying teacher salaries, or other school needs. At the same time, they may donate time or building materials to infrastructural improvements, and otherwise make in-kind contributions to the functioning of the school.

The third theme focuses on the maintenance of student records and the promulgation of critical policies at the school level. Importantly, having records is insufficient, as is claiming to have established a policy. In many cases, schools keep records but they are extremely incomplete or so severely disorganized that they are not usable. Similarly, schools may have developed a particular policy or policy document, but if it is not readily available for distribution, it is unlikely to be faithfully

implemented. For these reasons, enumerators are asked to physically review the records and policies in question and – in the case of student records – assess their completeness.

TABLE 77: SCHOOL MANAGEMENT INDICATORS, BY PROJECT LOCATION

Indicator	Somaliland	Puntland	Galmudug	Banadir	Overall
CEC Quality and Financial Management					
CEC meets monthly or more	55.3%	68.3%	90.9%	61.9%	65.0%
CEC management is very good	55.2%	74.5%	86.0%	69.8%	66.6%
CEC Support for School					
CEC provides bursary support	14.9%	25.0%	18.2%	71.4%	27.9%
CEC financial contribution	27.7%	25.0%	36.4%	38.1%	28.6%
CEC non-financial contribution	51.1%	13.3%	72.7%	33.3%	33.6%
Formal Policies and Record-Keeping					
Records of grades and enrolment (0-4)[309]	2.64	2.63	3.64	2.24	2.64
Critical policies in place and available for review (0-4)[310]	0.87	2.17	2.82	2.00	1.76
Aggregate Index Score					
Scorecard Score (0-100)	41.6	45.9	65.6	54.4	47.2

Findings

Overall, the findings on these seven indicators suggest that schools and the institutions that govern them are performing at a low-to-moderate level, with significant room for improvement. For instance, 91.4 per cent of head teachers report that there is a functioning CEC overseeing their school; however, only 77.2 per cent of those CECs have met within the last month. A substantial share of schools – 17.4 per cent – either do not have a functioning CEC or its last meeting took place over three months prior.

Results across all indicators are reported in Table 77, which provides disaggregated scores by indicator and project location. Overall scores for each indicator are presented in the right-most column, averaging across all schools or teachers in the sample.[311] As noted, the structure and data

[309] This metric captures whether schools keep records of student grades and enrolment, and the quality of those records. Please see the discussion above regarding the construction of a 4-point scale to measure the quality of school record-keeping.

[310] As noted in the discussion surrounding the development of this indicator, this metric concerns a school's mission statement, code of conduct, child protection policy, and school development plan. In the case of the mission statement, head teachers were asked whether the school had a mission statement. In the latter three cases, head teachers were asked whether their school had a given policy or document, and enumerators asked to see the document to verify its existence and the ability of head teachers to provide it.

[311] The sample for this analysis includes interviews with head teachers at 140 schools, and interviews with 522 other teachers. The mean number of teacher interviews – not including head teachers, who completed a different survey instrument – at schools was 4.2, providing a large sample of teacher opinions regarding school management, community attitudes, and other outcomes.

sources employed vary across indicators, but their construction ensures that they are comparable across project locations and other relevant sample subgroups.[312]

Disaggregating the results by project location highlights specific shortcomings in different settings. For instance, policy promulgation and documentation is weakest in Somaliland, where schools score 0.87 on a 5-point scale, compared to scores of 2.0 and above for all other locations. The gap in performance on this indicator is interesting because Somaliland does not underperform uniformly across all indicators. Separating this policy score by its individual components, Somaliland schools lag other schools in the sample on the four types of policies studied. Somaliland schools are approximately half as likely as others to have an established child protection policy, and less than half as likely to have an official mission statement.

Disaggregating the results by project location also reveals general trends across locations. The most compelling of these trends concerns the overall performance of schools in Galmudug. Across seven indicators, Galmudug achieves the highest on five, outpacing other locations with respect to CECs' meeting frequency, the quality of CEC management, the share of CECs that make non-financial contributions to the schools they oversee, record-keeping for both teacher attendance and student outcomes, and the development and documentation of critical policies.

TABLE 78: PERFORMANCE OF SCHOOL MANAGEMENT, BY CHARACTERISTICS OF SCHOOL

Indicator	Rural	Urban	IDP	Non-IDP	Conflict-Affected	Non-Conflict
CEC Quality and Financial Management						
CEC meets monthly or more	55.6%	68.3%	63.2%	100%	47.6%	68.1%
CEC management is very good	58.2%	71.1%	87.7%	64.0%	75.4%	65.4%
CEC Support for School						
CEC provides bursary support	16.7%	31.7%	14.3%	28.6%	19.1%	29.4%
CEC financial contribution	22.2%	30.8%	57.1%	27.1%	23.8%	29.4%
CEC non-financial contribution	44.4%	29.6%	57.1%	32.3%	23.8%	35.3%
Formal Policies and Record-Keeping						
Records of grades and enrolment (0-4)	2.64	2.64	3.14	2.62	2.81	2.61
Critical policies in place and available for review (0-4)[313]	1.47	1.86	2.57	1.71	2.10	1.70
Aggregate Index Score						
Scorecard Score (0-100)	42.7	48.7	64.9	46.2	44.0	47.7

[312] In the disaggregated analysis, Hirshabelle is excluded, as only one school was sampled from this area. As a result, disaggregated findings are not reliable for Hirshabelle. All 140 schools in the sample, including Hirshabelle, were included in the calculation of indicator totals in the right-most column of Table 66.

[313] As noted in the discussion surrounding the development of this indicator, this metric concerns a school's mission statement, code of conduct, child protection policy, and school development plan. In the case of the mission statement, head teachers were asked whether the school had a mission statement. In the latter three cases, head teachers were asked whether their school had a given policy or document, and enumerators asked to see the document to verify its existence and the ability of head teachers to provide it.

Beyond geographic differences, we also expect the performance of schools to vary systematically according to their institutional structure. That is, schools that are owned and managed by communities are likely to differ from those that are operated publicly, and both of these types might differ from schools oriented toward IDPs.

To assess this possibility, we analysed variation in school management indicators across categories that RI uses to classify their target schools. The first set of results, focusing on overall school management and CEC performance, are shown in Figure 22.

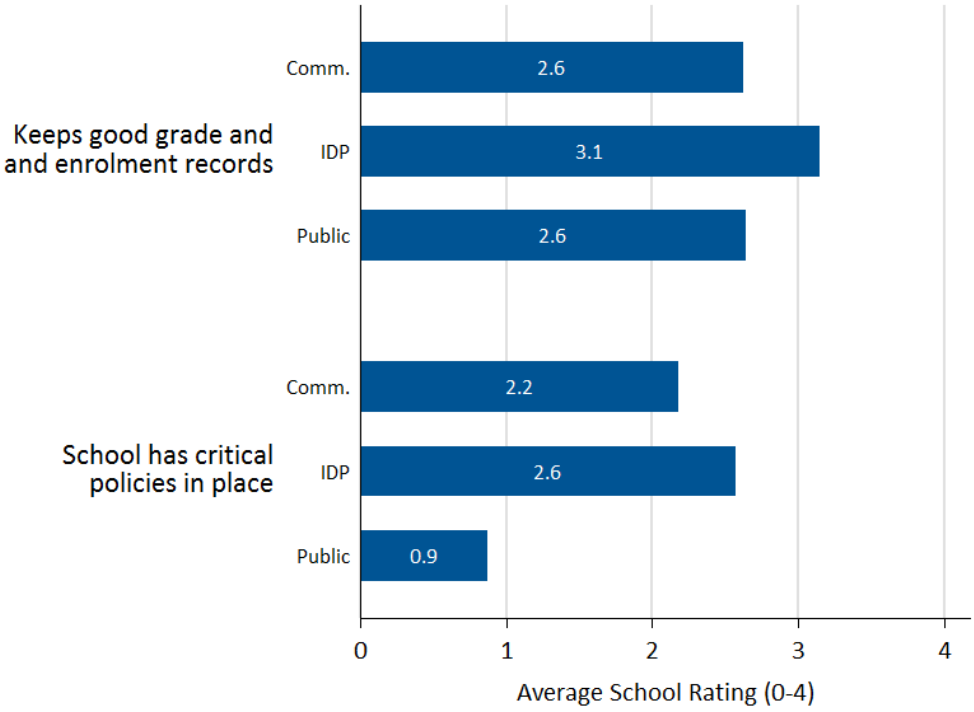
Somewhat surprisingly, IDP schools are more likely to have a properly-functioning CEC, and teachers in IDP schools are more likely to rate their respective CEC's management as "very good." It is important to note that only seven schools in the sample are included in this category, as IDP schools, so our conclusions may be driven partially by the specific sample. Nonetheless, the differences are sufficiently large that they suggest a structural difference between IDP and non-IDP schools.

Less surprisingly, CECs managing community schools outperform those managing public schools. To illustrate, community schools were more likely to have a CEC than public schools (95.1 per cent versus 85.1 per cent) and – among those schools with a CEC – the CECs were more likely to have met within the last month (70.9 per cent versus 65.0 per cent). Although the institutional composition of community schools can vary widely, we would expect community-run and/or community-funded schools to have a stronger CEC, all else equal – a hypothesis that appears to be borne out by the data.

Finally, we consider the impact of school type on record-keeping and policy promulgation and documentation. As Figure 22 shows, performance is moderate with respect to record-keeping, with outcomes measured on a 5-point scale. The biggest deficiencies in record-keeping appear to stem from the availability of grade records and the comprehensiveness of grade records. While 89.1 per cent of all schools maintain enrolment records for students, only 73.9 per cent have records available for student grades in mathematics, Somali and English. Taken further, only 47.8 per cent of schools have mostly or extremely complete grade records, though the differences across school type are not substantial.

One finding that is an outlier in Figure 22 is the performance of public schools with regard to policy promulgation and documentation, where they average a score of 0.9 on a 5-point scale. Only one-third of public schools have an official mission statement, according to their head teachers, compared to 86.8 per cent of community and IDP schools; public schools underperform systematically across all four policy types studied. Importantly, this difference does not appear to be driven only by the existence of policies or by the ability of head teachers to show documentation of the policies – rather, the deficiency of public schools is found in both cases.

FIGURE 22: QUALITY OF RECORD-KEEPING AND POLICY PROMULGATION, BY SCHOOL TYPE



Discussion and Targets

The indicators developed and reported above capture diverse aspects of school management and institutional governance. The analysis includes indicators that should be directly affected by RI programming, such as the promulgation of school-level policies, such as a Code of Conduct and Child Protection Policy. At the same time, it includes indicators that are only indirectly affected by RI programming, but which represent key aspects of school management, such as adequate record-keeping.

Critically, the indicators capture aspects of school governance related to material support provided by CECs, as well as policy support they provide to develop and implement higher-quality school processes and procedures. Well-run CECs are both valued by respondents and strong potential drivers of student outcomes. CECs take on a number of roles within schools and communities, ranging from dispute resolution to financial support. CECs in some schools have stepped in to pay teachers' salaries when the ministry has failed to do so, keeping schools open and maintaining teacher morale.[314] CEC members also fulfil a monitoring role for both students and teachers. According to one CEC member, they "check out the absence of the teachers, students, and the teaching procedures".[315] A member of a different CEC notes that the CEC itself recruited and selected the school's mathematics teachers – a common situation, especially when CECs are responsible for teacher salaries – and that they now actively monitor their performance.[316] CECs also report monitoring teacher performance in terms of the materials being used in the classroom and

[314] FGD with CEC members, Sahil, Somaliland.

[315] FGD with CEC members, Sahil, Somaliland. Members of other CECs, and teachers themselves, often emphasised the CEC's role in monitoring the attendance of teachers, specifically. See: FGD with CEC members, Bari, Puntland; FGD with CEC members, Dhusamareeb; KII with male teacher, Garbadadar, Somaliland.

[316] FGD with CEC members, Dhusamareeb.

the pedagogical methods employed.[317] Beyond teacher performance, CECs keep track of student attendance rates and exam performance, and inform students' parents when problems arise.[318] By providing an additional layer of accountability for students and teachers, CECs fill a role that is often neglected by government ministries and communities themselves.

The indicators selected have one final advantage that is worth noting: rather than focus on individual aspects of *process*, they focus on *outcomes* that can arise from varied processes, but which are critical in effectively functioning schools. An alternative approach would focus on procedural aspects of school management – for example, whether school staff have met with ministry officials to develop a Child Protection Policy. This approach assumes a uniform process for developing a policy will obtain in every school. Instead, the indicators selected focus on the outcome, however it occurred – a well-managed school should have a policy in place, whether it arose through intense coordination with ministry officials, through community discussion and action, or through the strong will of a single member of the school staff.

Looking forward to the midline and endline evaluations, targets for these sub-indicators should be ambitious, while accounting for their varied nature. That is, some indicators are "stickier" than others, because they may require coordination with other stakeholders or because there are bureaucratic hurdles to clear. On one extreme, we consider record-keeping non-sticky, because schools should already be doing this, and there are few obvious barriers to doing so, beyond a lack of time or interest. At the other extreme, we consider CEC provision of financial support to schools stickier, because CECs themselves face severe resource limitations that may limit the ability of the project to directly influence financial support for a school. Taking account of these differences, we propose targets based on standardized increases, with suggested targets reported in Table 79.[319]

TABLE 79: MIDLINE AND ENDLINE TARGETS FOR SCHOOL MANAGEMENT INDICATORS

Indicator	Baseline Level	Midline Target	Endline Target
CEC meets monthly	65.0%	70.7%	76.5%
Quality of CEC management	66.5%	72.2%	77.9%
CEC provides bursary support	27.9%	33.3%	38.7%
CEC financial contribution	28.6%	34.0%	39.5%
CEC non-financial contribution	33.6%	43.0%	52.5%
Records of grades and enrolment (0-4)	2.64	2.89	3.13
Critical policies in place and available for review (0-4)	1.76	1.97	2.19

[317] FGD with CEC members, Mogadishu.

[318] KII with female teacher, Maroodi Jeeh, Somaliland; KII with MoE official, Bari, Puntland.

[319] By standardized, we mean utilizing data on the distribution of outcomes at the baseline to set realistic goals for the midline and endline. For relatively sticky outcomes, we suggest targeting improvements of 0.08 standard deviations above the baseline in each subsequent evaluation period. For less sticky outcomes, we suggest targeting improvements of 0.2 standard deviations.

5.5 Positive community attitudinal change

Indicator: Increase in caregivers' aspirations for level of schooling they hope their girls will reach

The current cultural and social context under which EGEP is operating is not conducive to girls' academic empowerment. As such, positive community attitudinal and behavioural change was chosen as an intermediate outcome because stakeholders (e.g., parents, boys, and community leaders) first need to see the value of girls' education in order for the progress made under this project to continue in the long-term. Based on qualitative interviews with teachers and CEC members, examples of barriers to girls' education under this intermediate outcome include the ingrained belief among fathers that education for girls is not an ideal investment if their daughters will be married off early in life.[320] The same interviews also indicate that the sentiment among mothers is relatively similar: there is little value in educating girls who will most likely end up engaged primarily in housework.[321]

Building on the experiences of EGEP, efforts to promote attitudinal change in EGEP-T will focus on promoting positive attitudes toward girls' completion of schooling. That is, rather than focus on pro-education attitudes generally, EGEP-T will specifically seek to change attitudes regarding girls continuing in school through completion of secondary school.

The primary indicator selected for assessment of community attitudes toward girls' educational completion is a measure of caregiver attitudes the topic. Specifically, caregivers selected for participation in the household survey were asked to indicate the level of education to which they aspire for their daughter to complete. As we describe below, the vast majority of respondents stated that they would like their daughter to complete college or university. We focus on this level of educational completion as our metric, asking what share of respondents desire this high level of completion for their daughters.

The baseline evaluation collected a wide range of ancillary data on community attitudes, from many different populations. To gauge overall community attitudes and behavioural changes, survey questions regarding personal views and the views of other relevant stakeholders towards the importance of girls' education were directed at girls, boys, teachers, and caregivers. Randomly selected girls in target schools were asked questions about whether going to school would be important for their future careers and if it was important for children to attend school in general. To identify the barriers to education, girls were asked whether they had a choice to attend, and stay, in school or if this decision was outside of their control.

Particular attention in tool design was paid to the possibility of social desirability bias and other types of bias in the reporting of attitudes regarding girls' education. Because respondents may overstate the extent to which they or their community value girls' education when asked directly, we designed a number of survey questions that confronted respondents with hypothetical scenarios designed to elicit information about the relative value they place – or they believe their community places – on girls' education. As just one example, respondents were asked about the relative likelihood that parents would fund either boys' or girls' university education, in a hypothetical scenario.

[320] FGD with CEC members, Mogadishu, Banadir.

[321] KIIs with Female Teacher, South Galkayo.

Our measurement strategy for this indicator also relied on triangulating opinions from a number of different respondent groups, such as male and female students, caregivers of adolescent girls, teachers, and head teachers. As such, we divide the analysis in this section according to the respondent group being discussed.

Primary Indicator – Educational Aspirations

The project's primary indicator of community attitudes toward girls' educational completion focused on the attitudes and aspirations of caregivers. Specifically, the indicator asked respondents to describe the level of education they would like their daughter to achieve. In some ways, this question captures precisely the attitude that EGEP-T seeks to change: the support for girls' educational *completion*, rather than girls' education more generally. This distinction is important in the project's context, because there is relatively broad support for girls' education in general, but potentially less support for girls to delay marriage and complete secondary school. On the other hand, while the indicator's focus on aspiration is important for measuring attitudes, it may exhibit ceiling effects, as most parents support girls' educational completion in the absence of competing demands, such as marriage or household financial constraints. For this reason, we triangulated community attitudes through a number of approaches, which we describe in the sections that follow.

In general, support for girls' educational completion appears to be high in EGEP-T communities. Out of all respondents, 89.0 per cent indicate that they would like their daughter to complete college or university, and a further 2.3 per cent would like their daughter to complete upper secondary. Very few respondents report that they would like their daughter to stop schooling before completing lower secondary school, though 4.7 per cent do claim that they would like their daughter to complete primary school only. This finding is also reflected in the qualitative data: according to focus groups with girls in Sahil, "parents feel happy…when they see their daughter's education is high. Mothers encourage their daughter's education – even if she sells tomatoes, she covers her daughter's education".[322] In a focus group in Nugal, mothers said that it is important to encourage girls to complete their education, because that way they will reach university and find employment.[323]

Surprisingly, aspirations were higher, on average, among male caregivers than female, with 92.8 per cent of male caregivers supporting university education for their daughters. In contrast, only 88.6 per cent of female caregivers indicated the same – while still constituting the vast majority of female respondents, this gap is dramatic in context. While no male caregivers wished for their daughters to complete less than upper secondary school, 5.1 per cent of female caregivers wished for their daughters to stop schooling after primary school, and 0.8 per cent hoped that they do not complete any schooling at all. Fathers in one focus group in Somaliland emphasised the economic benefits of education for girls: that they will "work with agencies…they can become civil engineers, they can drive cars, they can become university teachers".[324]

There are also sizable gaps across community types and their exposure to external shocks. Respondents in rural communities express lower aspirations, on average, as do respondents in drought-affected and conflict-affected communities. In conflict-affected communities, just 77.5 per

[322] FGD with girls in Sahil, Somaliland.

[323] FGD with mothers in Nugal, Puntland.

[324] FGD with fathers in Sheikh, Somaliland.

cent of caregivers desire that their daughter complete university, compared to 90.7 per cent in non-conflict areas.

Unlike the large gaps across community types described above, there are relatively modest differences across project locations, as shown in Table 80. The table reports the baseline share of caregivers who aspire for their daughter to complete university, and sets targets for the indicator going forward. Importantly, our suggested targets reflect the potential for ceiling effects to begin influencing results at the midline, as baseline shares are relatively high across the sample. Our suggested targets consist of a 2.0 percentage point increase in the share of caregivers hoping their daughter completes university from the baseline to the midline, and a further 1.0 point increase from the midline to the endline.

TABLE 80: BASELINE COMMUNITY ATTITUDES, AND TARGETS FOR THE MIDLINE AND ENDLINE

Indicator	Baseline Level	Midline Target	Endline Target
Overall	89.0%	91.0%	92.0%
Somaliland	89.7%	91.7%	92.7%
Puntland	87.7%	89.7%	90.7%
Galmudug	85.1%	87.1%	88.1%
Banadir	91.5%	93.5%	94.5%

Unlike the large gaps across community types described above, there are relatively modest differences across project locations, as shown in Table 80. The table reports the baseline share of caregivers who aspire for their daughter to complete university, and sets targets for the indicator going forward. Importantly, our suggested targets reflect the potential for ceiling effects to begin influencing results at the midline, as baseline shares are relatively high across the sample. Our suggested targets consist of a 2.0 percentage point increase in the share of caregivers hoping their daughter completes university from the baseline to the midline, and a further 1.0 point increase from the midline to the endline.

Girls in School

Beyond the aspirations of caregivers, the evaluation sought to assess attitudes toward schooling more generally and among a wider set of targeted populations. Among randomly sampled girls from target schools, approximately 94 per cent indicated that going to school is essential for their future career, and nearly 98 per cent believe that it is important for children to attend school in general. Responses from focus group discussions with cohort girls are in agreement with these findings. In Sahil, when girls were asked whether they think the material they learn in school is good for their future, one responded that education can improve their value as mothers and caregivers: “if a girl learns something, she can help her family”.[325] Other girls in the same focus group replied that they could use the skills they acquired in school to find employment as a teacher.[326]

The extent to which girls perceive education as important is disaggregated by grade level in Table 81. Across all grades, girls overwhelmingly report that it is important that they personally attend school, and that it is important for children to go to school more generally.

[325] FGD with girls, Sahil, Somaliland.

[326] FGD with girls. Sahil, Somaliland.

TABLE 81: GIRLS' PERCEPTIONS OF THE VALUE OF EDUCATION, BY GRADE

Indicator	Grade 6	Grade 7	Grade 8	Form 1	Form 2
Going to school is important for what I want to do when I grow up	94.4%	94.6%	93.8%	93.4%	96.6%
It is important for children to go to school	99.0%	97.4%	97.8%	96.5%	98.3%

When girls were asked questions gauging barriers to their education, the results were somewhat mixed. While a total of approximately 83 per cent of sampled girls agreed to not being able to choose for themselves whether they can attend or stay in school, approximately 87.9 per cent agreed that their family provides them the support they need to stay in school and perform well. Finally, when girls were asked if they felt pressure to leave school and get married, only about 28 per cent agreed.

Table 82, below, disaggregates the results according to girls' grade level, owing to the possibility that girls' views on their own control over educational decisions and the pressure they might feel to drop out and get married is likely to vary with age. In Table 82, we report the share of girls who agree a lot with each of the first two statements, and the share of girls who *disagree* a lot with the last statement. Surprisingly, perhaps, the share of girls who report feeling pressured to drop out does not appear to rise with age.

TABLE 82: GIRLS' VIEWS ON BARRIERS TO EDUCATION

Indicator	Grade 6	Grade 7	Grade 8	Form 1	Form 2
Agree: I cannot choose to attend or stay home – I must accept what happens	69.0%	74.3%	66.7%	62.7%	63.8%
Agree: I get needed support from family to stay in school and perform well	89.9%	89.4%	87.3%	82.6%	90.0%
Disagree: I feel pressure to drop out and get married	58.7%	51.4%	61.9%	55.3%	65.0%

Responses from girls participating in focus group discussions are no more definitive. When girls were asked about their family members' stance on girls' education and marriage, some girls reported that they "make their family proud"[327] when they do well in school while other girls reported that the "community sometimes discourages girls; when they see an educated female, they ask her, why are you studying now, you will get married to someone tomorrow".[328] The findings in Table 82 are supplemented by responses from girls comparing the likelihood that boys would get university tuition funding from their families relative to girls (Table 83).

TABLE 83: GIRLS' EXPECTATIONS OF COMMUNITY SUPPORT FOR BOYS' AND GIRLS' EDUCATION

Response	The family of the boy would raise enough money to send him to university?	The family of the girl would raise enough money to send him to university?
Very Unlikely	8.1	10.7
Somewhat Unlikely	11.2	25.3
Somewhat Likely	48.7	52.0

[327] FGD with girls. Sahil, Somaliland.

[328] FGD with girls. Sahil, Somaliland.

Very Likely	32.0	12.1
Total	100	100

Shown above, approximately 81 per cent of randomly selected girls in target schools indicated that it was "somewhat likely" or "very likely" that boys would receive family funding. However, when the same question was asked about girls, this value decreases to approximately 64 per cent, a 17 per cent difference. When boys were asked the same questions, the difference was nearly identical at 16.5 per cent. The patterns observed may be symptomatic of perceived comparative advantages between girls and boys by caregivers. When fathers in Maroodi Jeeh were asked who they would send to school if they were under financial constraints, they responded that they would prioritise sending boys to school, "because girls work at home mostly and they are good at that compared to boys".[329]

Head Teachers and Teachers

When it comes to girls' education, it is important that we include teachers' insights on parents' opinions as well as the opinions of other relevant stakeholders. Teachers can provide valuable perspectives, since parents may have incentives to misrepresent their opinions about education. Admitting that one does not support educating their daughter may hurt an individual's image; they may also be concerned about judgment from field researchers working on a project specifically dedicated to girls' education. As such, teachers can provide a valuable perspective, as individuals who are embedded within communities, but who likely value education more than the average community member.

To assess community attitudes, head teachers were asked to indicate the relative levels of support offered to boys' and girls' education, respectively, from four different groups within their communities. For ease of interpretation, we combined categories from a 5-point Likert scale into a 3-point measure: unsupportive, indifferent/neutral, and supportive. Table 84 reports the results gathered from head teachers.

TABLE 84: HEAD TEACHERS' PERCEPTIONS OF SUPPORT FOR BOYS' AND GIRLS' EDUCATION AMONG COMMUNITY

Group being Assessed	Unsupportive	Indifferent	Supportive
Support for Boys' Education			
Fathers	27.9	6.4	65.7
Mothers	13.6	2.1	84.3
Religious Leaders	19.3	12.9	67.9
Clan Leaders	25.7	8.6	65.7
Support for Girls' Education			
Fathers	27.9	3.6	68.6
Mothers	16.4	0.0	84.3
Religious Leaders	17.9	15.0	67.1
Clan Leaders	25.7	18.6	55.7

When head teachers were asked to rate fathers' support for boys' education, approximately 66 per cent rated them as "supportive" This value is not much different at 69 per cent when head teachers were asked to rate fathers' support for girls' education. This trend is similar for mothers and local religious leaders, as well. While approximately 84.3 per cent of head teachers rated mothers as

[329] FGD with fathers in Maroodi Jeeh, Somaliland.

"supportive" of boys' education, 83.6 per cent provided the same rating when asked to gauge mothers' support for girls' education. Similarly, 68.9 per cent of sampled head teachers rated local religious leaders as "supportive" of boys' education, 67.1 per cent rated them as "supportive" of girls' education. Head teachers rated 65.7 per cent of clan leaders as "supportive" of boys' education but said only 55.7 per cent of clan leaders support girls' education—a 10 per cent difference.

Though the percentages of "supportive" ratings for girls' and boys' education are similar within each community member type, with the exception of clan leaders, we see notable differences across types. Over 80 per cent of sampled head teachers gave a "supportive" rating to mothers, regardless of whether her support was for boys' or girls' education. However, fathers, local religious leaders, and clan leaders lag behind, with all three community member types receiving this rating from less than 70 per cent of sampled head teachers. Interview responses from teachers and CEC members support the trends observed. When teachers in South Galkayo and CEC members in Mogadishu were asked whether community leaders and religious leaders support or engage in activity that supports girls' education, the two groups agreed that community and religious leaders do not provide noticeable support".[330]

We also asked classroom teachers about their views on community attitudes towards girls' education. Table 85 shows the views of regular teachers regarding the pressure put on girls to drop out by their community, the community's support of girls' education after marriage, its support for girls' university education, and finally, whether the community treats girls and boys equally.[331]

TABLE 85: TEACHERS' PERCEPTIONS OF ATTITUDINAL BARRIERS TO GIRLS' EDUCATION

Questions	Agree
Girls in Community Feel Pressure to Drop Out and Get Married	61.2%
Community Supports Girls to Continue Schooling After Marriage	68.9%
Community Supports Girls to Continue Schooling, even University	77.3%
Girls are Treated Equally to Boys in this Community	70.0%

Similar to cohort girls' responses on their views regarding barriers to girls' education, teacher responses are relatively mixed. While 61.2 per cent of teachers agree that girls feel pressure from the community to drop out of school to get married, 68.9 per cent of teachers believe that the community supports girls continuing their education after marriage. The majority of teachers, 77.3 per cent also indicate that this support continues even up to the university level. Furthermore, 70.0 per cent of sampled teachers believe the community treats girls and boys equally. When the same group of teachers was presented with a hypothetical scenario in which a child had been admitted to university and needed community support to attend, their responses suggested that the community would favour supporting a male student in this situation over a female student. Specifically, 75.5 per cent of teachers indicated that the community would likely raise money for the male student to attend university, compared to just 62.7 per cent who thought the community would raise money for a female student.

[330] KII with teachers in South Galkayo; FGD with CEC in Mogadishu.

[331] As with our analysis of head teachers' responses above, we collapse a 5-point likert scale into three categories to make the underlying patterns more clear.

Community Members

Beyond the opinions of students and teachers, the baseline also assessed community attitudes themselves. Using a random sample of households in EGEP-T communities, the baseline collected data on attitudes toward education among caregivers of girls aged 11-18, whether they were enrolled in EGEP-T schools or not.

Respondents were asked a number of questions regarding the value they place on girls' education. Overall, they show strong support, in theory, for girls' education, by indicating that girls are just as likely to use their education as boys, and by indicating that they would like their female child to complete schooling up to university level. Responses to questions such as these were overwhelmingly positive: 77.8 per cent of caregivers strongly agreed that a girl was just as likely to use their education as a boy, with another 19.4 per cent agreeing more weakly. And 89.0 per cent indicated that, if they had the choice, their female child would stay in school through the completion of university.

Table 86 disaggregates these findings regarding support for girls' education across project locations, urbanity, and the gender of the caregiver responding to the survey. Across locations, there are relatively small swings with regard to the share of caregivers who wish for their daughter to complete school through university. There are larger shifts across locations with respect to the second question: whether girls are just as likely to use their education as boys. On this metric, Somaliland outperforms other locations, at 85.9 per cent, compared to a low in Banadir of 70.2 per cent. There does not appear to be a systematic difference between female and male caregivers' attitudes in this respect; however, it is important to note that only 10 per cent of the available sample of caregivers are male, meaning that conclusions regarding differences across gender are based on a small sample size of just 56 men.

TABLE 86: CAREGIVER ATTIUDES TOWARD GIRLS' EDUCATIONAL COMPLETION

Subgroup	Would like daughter to complete schooling through university	Strongly agrees that girls are just as likely to use education as boys
Overall	89.0%	77.8%
Somaliland	89.7%	85.9%
Puntland	87.7%	71.6%
Galmudug	85.1%	78.4%
Banadir	91.5%	70.2%
Rural	83.5%	88.2%
Urban	92.0%	73.5%
Female Caregiver	88.6%	77.8%
Male Caregiver	92.8%	76.5%

The natural conclusion from the community-level findings, thus far, is that community members have broad and deep support for not just girls' education, but also girls' completion of higher-level education. In contrast to that apparent enthusiasm, however, respondents were less overwhelming in their support for girls' education when presented with difficult financial choices. In one hypothetical scenario, we asked caregivers to imagine a situation in which their sister was sick, and needed assistance with their medical bills. Assuming a lack of money to help their sister, we asked caregivers to make the difficult choice of how to raise funds: sell some of their household goods or an animal, or withdraw their daughter from school to avoid paying school fees. Given those two choices, 21.9 per cent of respondents indicated that they would withdraw their daughter from school; a further 23.7 per cent were not sure what action they would take.

Respondents were also confronted with a hypothetical question regarding a marriage proposal for their daughter. Given a proposal for their daughter, we asked respondents whether they would support their 15-year old daughter – who is currently enrolled in school – in getting married, continuing their education, or doing both. The majority of respondents (68.2 per cent) would encourage their daughter to continue their schooling; but 12.6 per cent would flatly encourage their child to accept the proposal and *not* continue their schooling after marriage. The idea that individuals have aspirational preferences for their children, but often face difficult choices, is consistent with how qualitative interviewees – whose responses allow more space for context and clarification – described their preferences for girls' school completion. As mothers participating in an FGD in Nugal described it, they planned to continue funding their daughter's education *if they could afford it*, and they wanted their daughters to continue through the university level.[332]

The finding that educational attainment is constrained by difficult realities is not altogether surprising. It is heartening that a majority of respondents (54.4 per cent) would prefer to sell some of their assets while still investing in their daughter's education. However, the large minority who would withdraw their daughter from school represents a significant share of the population, and the households most needing awareness-raising by projects such as EGEP-T.

Interpretation and Discussion

While there is both enthusiasm to learn and the recognition that education is important for their future careers, girls are aware that whether they are allowed to attend or stay in school is not up to them. The results in this section indicate that even though many girls are currently receiving the parental support they need to perform well in school, they must accept whatever happens should that support be rescinded as they get older. When asked about the likelihood that parents would raise enough money for boys' university tuition, more girls responded "somewhat likely" or "very likely" than when they were asked the same question but for girls' tuition. This set of questions was asked again to randomly selected boys and classroom teachers, and both groups indicated that families are more likely to raise funds to send boys to university than they are to send girls – results consistent with sampled girls' responses. The findings across groups suggest that, after secondary level education, parents' support for girls' education begins to dissipate. This information concurs with cultural norms, given that, at this point, girls have reached marriageable ages.

In general, responses from caregivers are extremely pro-girls' education, with the community supporting girls' post-secondary education. However, the picture is complicated, and tempered, by reports from girls, boys, and teachers. While the vast majority of girls – 87.9 per cent – state that they receive the support they need from their families to complete their schooling, the outlook is less optimistic when respondents are either faced with a hard choice, or are describing the expected level of tangible support they can receive from their community or families.

The notion of a mixed picture is supported by responses from focus group discussions with girls which indicate that support for girls' education further diminishes after marriage.[333] There is an entrenched belief among parents and male partners that marriage and girls' education are mutually exclusive and, at best, up to the husband. When girls were asked whether they could continue their education after marriage, most responded that their parents have reminded them that their main role is to be a wife who "stay[s] at home and manage[s] the tasks".[334] And, "they advise [girls] to stay

[332] FGD with mothers in Nugal, Puntland.

[333] FGD with girls, Sahil, Somaliland.

[334] FGD with girls, Sahil, Somaliland.

at home and not learn, that in the future they will end up being [a] wife and cooking for the spouse." As for attending school after marriage, one respondent said, "if the men want [them] to continue studies then he'll allow a tutor, but not school".[335] Others say "either you marry someone or continue your study, but to marry and study at the same time is not possible".[336]

5.6 Intermediate Outcomes as Predictors of Learning & Transition

In the previous section, we analysed EGEP-T's proposed intermediate outcomes, with a primary goal of establishing baseline levels against which future evaluations will be assessed. In addition, our analysis attempted to shed light on variation in these intermediate outcomes across relevant subgroups of the targeted population by, for instance, assessing differences in attendance rates across project locations.

This section of the report again focuses in EGEP-T's five intermediate outcomes: attendance, teacher quality, school management and institutional governance, girls' self-esteem and empowerment, and community attitudes and behaviour. However, the section centres on a different question: whether there is a verifiable relationship between the targeted intermediate outcomes and the learning and transition outcomes constituting the core focus of GEC programming. Put differently, the section investigates the veracity of EGEP-T's Theory of Change by studying whether improved intermediate outcomes are associated with increased learning and higher rates of transition.

Two important limitations to the analysis should be acknowledged. First, many of the indicators selected for measuring the impact of EGEP-T over time are poorly suited to an analysis focusing on relationships to learning or transition outcomes. For instance, to measure the quality of school management, we developed eight individual metrics, including measures of CEC activity, school record-keeping, and administrative monitoring of teachers' performance. The metrics are *fit for purpose*, in that they capture key aspects of school management, are verifiable, and are uniformly measurable across locations. That is, they are fit for the purpose of assessing change over time.

However, they are less suited to studying their effect on, or association with, learning outcomes. Assessing the effect of eight separate metrics is difficult, as they have varied and contradictory effects, and the metrics themselves may be correlated with one another. And different metrics may be related to learning via different channels of varying strength, making any results more complex to interpret. To address the problem, we focus only on those metrics which we expect to have the most direct effects on learning, and we do not limit ourselves to the metrics we employ for setting baseline levels. To illustrate, consider school management: in the analysis that follows, we focus on two measures of school management – the activity level of CECs, and length of instructional school-day (distinguishing between those that have abnormally few hours of teaching time per day). The former is one of the metrics employed in our assessment of baseline school management; the latter is not. Nonetheless, the latter factor, length of instructional school-day, is an indicator of management quality, and we expect it to have a more direct effect on learning outcomes.

The second limitation concerns the interpretation of the findings below. We employ linear regression models to study the association between a wide range of variables and learning and transition outcomes in a large cross-sectional sample of students. While we control for many factors – such as region, and urbanicity – that would otherwise produce bias in the results, we are not able

[335] FGD with girls, Sahil, Somaliland.

[336] FGD with girls, Sahil, Somaliland.

to control for unobserved factors that influence both intermediate outcomes and learning. In other words, while the results below demonstrate relationships between variables, they do not imply a causal relationship.

Learning Models and Results

We study aggregate learning outcomes by combining numeracy and literacy scores into a single dependent variable, adjusted to range from 0 to 100.[337] The mean value of the dependent variable is 65.0 per cent, and it is distributed normally. The baseline model of learning outcomes takes the form

$$y_i = B_0 + \gamma_i + \mu_i$$

where y_i is the learning score for student *I*, and γ_i is a vector of independent variables measured at the level of individual students or their respective schools. The model includes a constant, denoted by B_0, and the standard error term, μ_i. To maximize the statistical power of the models, we use data from surveys and learning assessments with cohort girls, cohort boys, and bursary girls, where possible. In some models, we use variables that were not captured for bursary girls or cohort boys, and the analysis is restricted to cohort girls only. In other cases, we utilize variables that are theoretically relevant only to girls – such as community support for girls' education – and we limit the sample to female students. We note both exceptions when they occur. As the analysis includes respondents from three related but distinct samples, we do not employ sampling weights. Our goal in this analysis is not to draw conclusions that are representative of any broader population, but to draw conclusions about relationships *within* the sample. Finally, because learning outcomes and their predictors are likely to be correlated within schools, we cluster standard errors by school.

The results of the baseline model are presented in Figure 23.[338] The model includes a standard set of demographic and geographic control variables that are likely to be correlated with learning outcomes. For instance, differences between regions are captured by including a binary variable each for Somaliland, Puntland, and combined Hirshabelle-Galmudug.[339] The interpretation of the Somaliland variable is relative to the omitted reference category, which consists of students in Banadir. Therefore, a strong negative effect of the Somaliland variable means that students in Somaliland score approximately 20 points lower than those in Banadir on average, controlling for the other factors in the model.

The results in Figure 23 are consistent with most prior expectations, and with findings elsewhere in this report. Male students score 7.9 points higher than their female peers, and student performance increases dramatically from lower to higher grades. Two other effects are notable: students from female-headed households outperform other students by 2.0 percentage points, and students in drought-affected schools underperform against their peers by 4.1 percentage points. The findings, as well as those related to grade level, student gender (female in the graph), and the binary variable representing Somaliland, are statistically significant at the 5 per cent level.

---

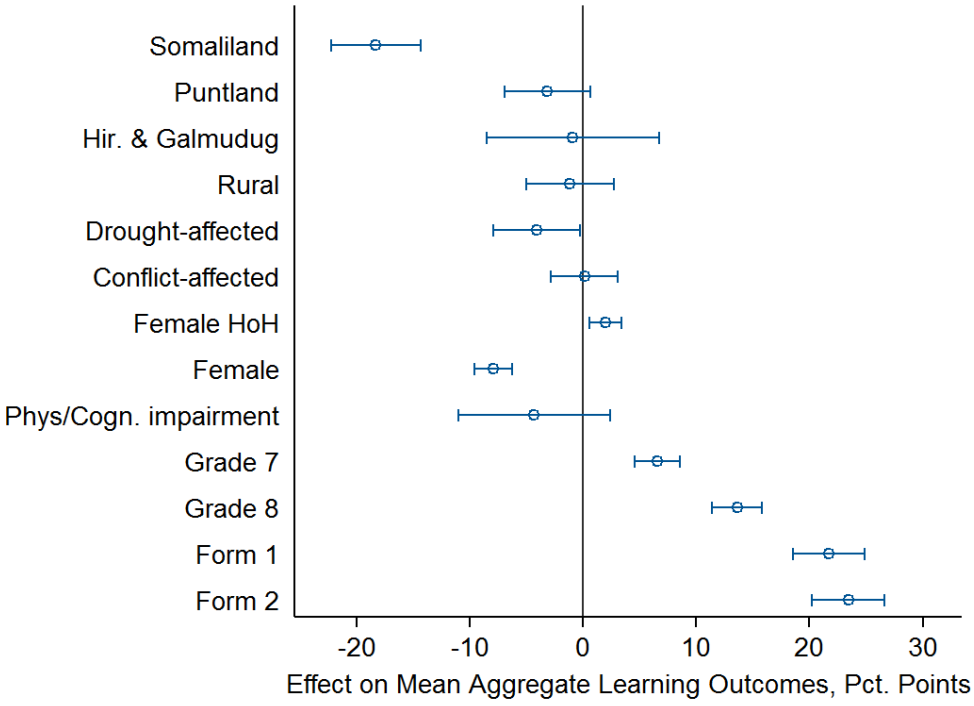
[337] More precisely stated, the dependent variable in this analysis is the mean of literacy and numeracy outcomes. For each student, their aggregate score is an unweighted mean of their two examination scores.

[338] Figure 23 plots the beta coefficients from a linear (OLS) regression model as points; the bars around each point represent the 95 per cent confidence interval for each estimated coefficient. A variable that is statistically significant in the model – i.e. a variable whose association with learning outcomes is systematic in a statistical sense – will have a 95 per cent confidence interval that does not cross the vertical line bisecting the graph at zero.

[339] As only one school was sampled in Hirshabelle, including the area as its own category results in unstable and extremely imprecise parameter estimates.

Several factors, such as urbanicity and conflict do not appear to influence assessment scores. For instance, students in rural schools score similarly to those in urban schools, once we control for the other demographic characteristics in the model. Students in conflict-affected schools do not appear to score worse than those in more peaceful locations. However, there is a strong correlation between both urbanicity and conflict, on one hand, and project location, on the other.

FIGURE 23: REGRESSION RESULTS - GEOGRAPHIC AND DEMOGRAPHIC PREDICTORS OF AGGREGATE LEARNING OUTCOMES



Building on the baseline model, we estimate additional regressions that include measures of EGEP-T's intermediate outcomes. In each case, we incorporate intermediate outcomes one at a time.[340] In other words, we estimate a new model – which includes the same set of control variables reported in Figure 23 – but add a measure of attendance. Next, we estimate a model that includes the same controls, adds measures of school management, but removes the variable measuring attendance. In Figure 24, we report coefficient estimates and 95 per cent confidence intervals for specific variables of interest, estimated across multiple models.[341]

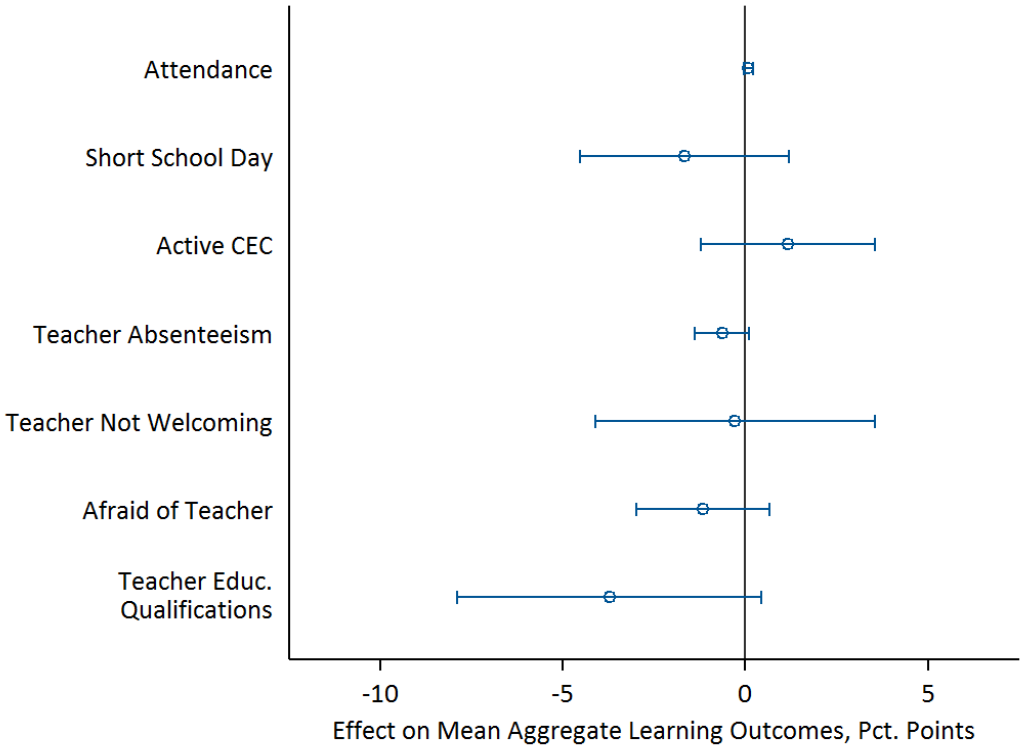
Attendance is the first intermediate outcome studied. We measure attendance at the school level, using data from classroom headcounts conducted by Forcier researchers. As described previously, Forcier researchers randomly selected up to eight classrooms (one per grade) and physically counted the students present at the time, comparing the total to the class's stated enrolment. The attendance rate was calculated for each class; the mean attendance rate at a given school is our measure of attendance.

[340] While the sample size is sufficiently large to include many more independent variables, many of the variables of interest are correlated with one another in complex ways that alter the results across models.

[341] The full results of these models, including all control variables, are provided in Annex 14.

The result shows that schools with higher rates of attendance have higher learning outcomes. However, the effect is substantively small and is not statistically significant (p = .25).[342]

FIGURE 24: REGRESSION RESULTS – EFFECT OF ATTENDANCE, SCHOOL MANAGEMENT AND TEACHER QUALITY INDICATORS ON LEARNING OUTCOMES



The next set of results focuses on school management and institutional governance. As discussed previously, we employ two measures of school management: a short school day, and CEC activity. Schools are defined as having a short school day if their head teachers report that the school-day consists of fewer than five hours of teaching time. While not included as a core metric of school management, it is nonetheless a reasonable proxy for school management, and is expected to have a more direct effect on learning outcomes than many alternative variables.[343] CEC activity is a binary variable, which takes the value "1" if a school has an active CEC that has met within the past month. Neither effect is statistically significant at conventional levels. However, the two effects are in the expected direction: the point estimate for *short school day* is negative, as we would expect less instructional time to be correlated with worse learning outcomes; the point estimate for *Active CEC* is positive, as we would expect active CECs to have a positive effect on a number of aspects of a school's functioning, including teaching quality, attendance, and financial support.

[342] A null finding for attendance is unsurprising in this context. A stronger research design would study the correlation between a specific student's attendance and their learning outcomes. Instead, our analysis aggregates attendance rates across many classes at a student's school, many of which have no direct impact on the student's performance (because they are classes from a different grade, for example).

[343] Across 140 schools, 20.7 per cent reported a school-day comprising four or fewer hours of teaching. Unfortunately, the data failed to indicate why some schools have abbreviated days, or if differences between schools are simply a function of head teachers' varying interpretations of the survey question.

Finally, our third set of results investigates the relationship between teacher quality and learning. We use four variables that capture different aspects of teacher quality:[344]

- *Teacher Absenteeism* – the number of absences, over the previous two weeks, reported by students about their teachers
- *Teacher Not Welcoming* – binary variable for students who disagree with the statement "my teachers make me feel welcome in the classroom"
- *Afraid of Teacher* – binary variable for students who agree with the statement "I am afraid of my teacher"
- *Teacher Educ. Qualifications* – the share of teachers at the school whose highest educational qualification is secondary school completion or lower

Once again, the effects of these variables, shown in the bottom half of Figure 24, are in the expected direction. Higher numbers of teacher absences reported by students is associated with poorer learning outcomes, though the results are only marginally significant ($p = .098$). Similarly, as the share of teachers with only a secondary school diploma or less increases, average learning declines ($p = .08$).[345]

The fourth intermediate outcome we study assesses the role of community attitudes in learning outcomes. We measure community attitudes in three ways:[346]

- *Community Supports Girls' Education* – the mean of perceived support for girls' education by four groups of community members (fathers, mothers, local religious leaders, and clan leaders), as reported by teachers[347]
- *Girl Perceives Gender Gap* – binary variable indicating that a girl believes their community is more likely to financially support a boys' education than a girls' education, in a hypothetical scenario
- *Pressured to Marry* – binary variable for girls who agree with the statement "I feel pressure to drop out of school and get married"

The association between community attitudes – or perceptions thereof – and girls' learning outcomes is reported in Figure 25. Teacher perceptions of support for girls' education in their communities is weakly related to improved learning outcomes, though the correlation is not statistically significant. Similarly, girls who indicate that their community would favour a boy over a girl when providing financial support for a child's education perform slightly worse on learning assessments, but the effect is not statistically distinguishable from a null or zero effect. Notably, girls who report that they feel pressure to get married actually score 2.4 points higher than other girls ($p = .03$).

[344] Because this model includes variables derived from surveys administered to cohort boys and cohort girls, bursary girls are excluded from the analysis.

[345] To clarify this finding: as the share of teachers with higher qualifications (post-secondary training or university) increases, learning outcomes improve.

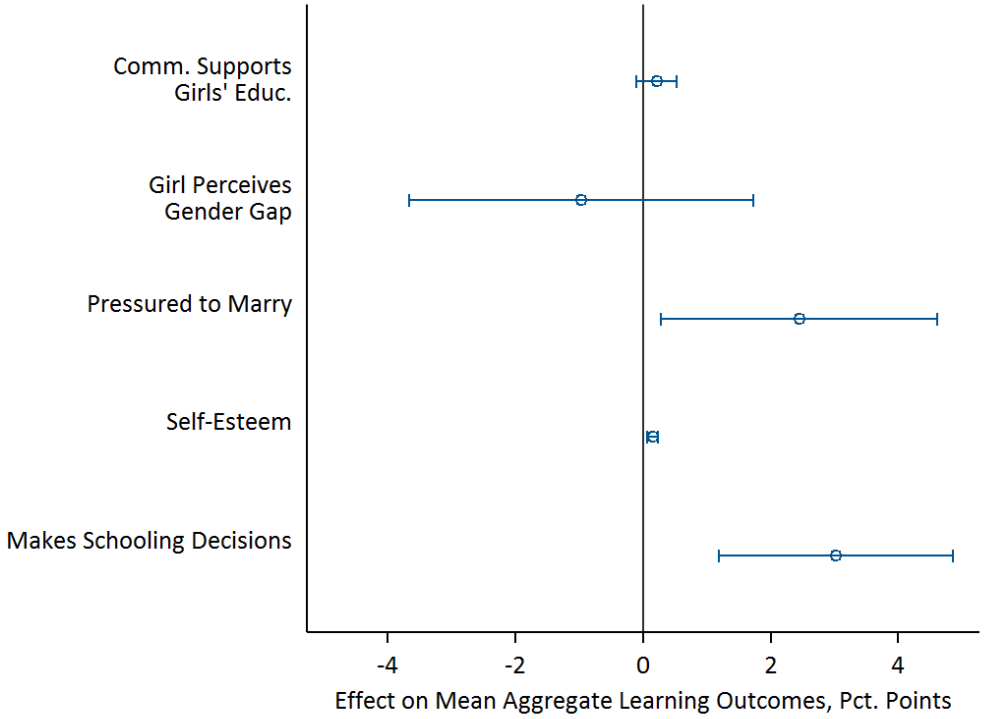
[346] The analysis focuses on community attitudes toward girls' education; as a result, we limit the sample to female students, for whom community attitudes regarding girls' education are most relevant.

[347] Teachers were asked to rate the extent to which each of the four groups in their community supported girls' education, on a 5-point Likert scale. Responses across all four groups were aggregated (forming a scale from 0 to 16) and averaged across all surveyed teachers in a school.

The final intermediate outcome we assess is girls' self-esteem and empowerment. Self-esteem was captured through a series of over 20 questions. We use a self-esteem index, created by aggregating results from all related questions.[348] In practice the scale ranges from 0 to 56, with higher values representing greater self-esteem. Empowerment was captured with a single survey question, posed to female students: "who makes decisions about whether you will go to school"? Girls who indicate that they decide solely or that they decide jointly with their family are coded as exercising (partial or full) agency over schooling decisions.

The estimated effect of self-esteem and empowerment are, arguably, the most compelling of those reported for any intermediate outcome studied in this section. Girls who report exercising some degree of agency over schooling decisions score 3.3 points higher than other girls (p = .001), and girls with higher self-esteem also perform significantly better on the baseline evaluation's learning assessment.[349]

FIGURE 25: REGRESSION RESULTS – EFFECT OF COMMUNITY ATTITUDES AND GIRLS' SELF-ESTEEM AND EMPOWERMENT ON LEARNING OUTCOMES



Transition Models and Results

Building on the brief analysis transition rates among various subgroups reported in Section 4.4, this section follows a similar approach to that described for learning outcomes previously. The section estimates a series of regression models predicting transition rates among randomly-selected households in EGEP-T communities. The analysis focuses on transition rates among both boys and girls, in an effort to determine the impact of EGEP-T intermediate outcomes – and other factors – on transition rates. By studying both boys and girls between the ages of 11 and 18, we increase the

[348] See Section 5.2 for a more detailed description of how the index was created.

[349] Both effects are robust to the inclusion of boys in the sample, using data on boys' own self-esteem and empowerment.

available sample size. The sample is drawn from households in EGEP-T communities, and not from girls and boys enrolled in EGEP-T schools. The factors influencing transition rates may differ from those that shape transition decisions among EGEP-T beneficiaries.

As in our analysis of learning outcomes, our basic regression model controls for a variety of demographic factors, such as the geographic region of the respondent, urbanicity, gender, and gender of the head of household. Our outcome variable is a binary indicator for successful transition, which is defined as progression from one grade to the next, or re-enrolling in school for children who were not enrolled the previous year. Unsuccessful cases include those who drop out, who are held back in the same grade year-on-year, and those who remain out of school in both time periods. We estimate linear models to enable a more straightforward interpretation: the coefficients that we report can be interpreted approximately as the change in probability of successful transition associated with a particular predictor.[350] Finally, to account for correlation in transition rates between children in the same household we cluster standard errors at the household level.

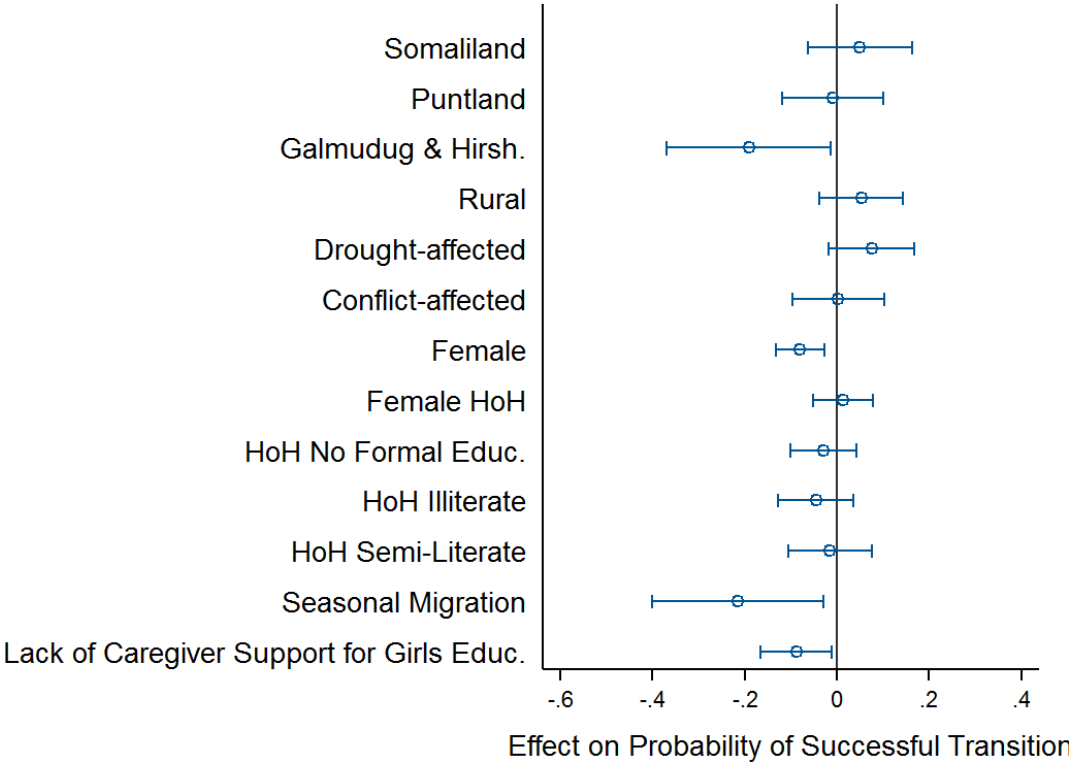
Figure 26 provides results of our preferred model, which incorporates a variety of demographic factors. As the graph illustrates, there are a number of predictors of transition rates that accord with our theoretical expectations. In families that report occasional seasonal migration, children are less likely to progress from one grade to another; female students are also about 8 percentage points less likely to transition, all else equal. Geographically, there are relatively small differences in transition rates between Somaliland, Puntland, and Banadir, which serves as the omitted reference category for the model. However, children in Galmudug and Hirshabelle (combined due to the limited sample available in Hirshabelle) are about 19 percentage points less likely to transition than those in Banadir ($p = .04$).

As with our findings regarding the predictors of learning outcomes, some of the results in Figure 26 run counter to our theoretical expectations. Our regression model allows us to assess the role of drought in transition rates. Based on the results in Figure 26, drought is not significantly associated with lower transition rates. In fact, the point estimate reported for a binary drought indicator is positive, running counter to expectations about the impact of drought. This is broadly consistent with the subgroup results reported in Section 4.4, which show that girls in drought-affected communities actually had marginally higher transition rates than the sample average.

Similarly, conflict does not exert a discernible effect on transition rates in the regression model. As we noted with regard to learning outcomes, however, it is difficult to distinguish between the effect of conflict and other factors with which it is correlated, because conflict is correlated so closely with project location. While the results do not support the notion that conflict reduces transition rates, our ability to test this proposition empirically is sufficiently limited that it should not be used to guide project decision-making.

[350] We also estimated a series of logistic regression models, with substantively similar results.

FIGURE 26: REGRESSION RESULTS – GEOGRAPHY, DROUGHT, AND DEMOGRAPHIC CHARACTERISTICS AS PREDICTORS OF SUCCESSFUL TRANSITION



Finally, we find that caregiver support for girls' education is a strong predictor of transition. Our measure of caregiver support comes from responses to a hypothetical scenario, in which caregivers were asked to imagine that their sister was sick and needed assistance with her medical bills. Given a choice between selling household assets and withdrawing their hypothetical daughter from school to save money on school fees, we ask respondents to indicate how they would handle the situation. We find that, in households where the caregiver indicates that they would withdraw their daughter from school, both girls and boys are *less* likely to successfully transition, as shown at the bottom of Figure 26.

Conclusions

The analysis in this section probed the relationship between EGEP-T's proposed intermediate outcomes and student learning. In doing so, it provided an initial, tentative assessment of RI's Theory of Change. Again, it is important to recognize the limitations of this analysis, particularly the cross-sectional nature of the data. Without an experimental or quasi-experimental design, it is not possible to attribute causation to the factors studied here. While we find a strong relationship between, for instance, girls' self-esteem and learning outcomes, we cannot eliminate the possibility that this relationship is a spurious correlation driven by one or more omitted variables, or that girls who perform well on learning assessments have higher self-esteem as a result of their performance (i.e. reverse causation).

Despite this caveat, the findings are generally consistent with RI's Theory of Change. In almost every case, proxy variables for EGEP-T intermediate outcomes had effects in the direction consistent with theoretical expectations, even if many of the effects were not sufficiently large or estimated precisely enough to attain statistical significance. In cases where the data and model were of higher

quality for testing hypotheses regarding learning – for instance, variables measured at the level of individual girls, such as their teacher's absenteeism or their own self-esteem – the results provided more conclusive evidence in support of the Theory of Change.

6. Contextual Factors – Drought in Somalia

Somalia's Drought

Somalia's geographic and environmental position have long made it susceptible to periods of poor rainfall, and its economic environment – especially the reliance on agricultural and pastoralism as a primary source of livelihood – makes drought particularly disruptive to households and communities. The region's climate is almost entirely hot and dry, with average annual rainfall over the last quarter-century averaging just 23 cm.[351] Rainfall is seasonal, with major rains (Gu) occurring between April and July, and a period of lighter rainfall (Deyr) occurring between October and early December.[352]

The geographic factors are only part of the story, however, and Somalia's susceptibility to drought – and the conversion of drought into full-fledged humanitarian crises – is exacerbated by a long history of instability, conflict, and limited investment in public infrastructure. The country has extremely poor water and sanitation infrastructure.[353] Ongoing conflict has defined Somali politics and life since the 1991 overthrow of the Siad Barre regime and the associated collapse of the central government – the conflict has weakened the government's ability to respond to early warning signs of drought, and has increased the population's vulnerability to droughts and issues of food insecurity.[354]

Somalia has experienced two severe droughts in recent history. In 2011, consecutive seasons of very low rainfall produced one of the worst droughts in recent history in the Horn of Africa, which led to a famine killing over 250,000 people in Somalia.[355] During the drought, almost half of Somalia's population faced a humanitarian crisis.[356] Last year, Somalia was again facing a severe drought. In February 2017, the president of Somalia declared that the prolonged drought in the country constituted a national disaster. The declaration was issued after the World Health Organisation (WHO) issued a warning of risk of famine. More than half of Somalia's population needed urgent humanitarian aid at the time of the announcement.[357] Moreover, almost 3 million people failed to meet their daily food needs and another 3.3 million people needed livelihood support to avoid crisis in the beginning of 2017.[358]

At the time of preparing this report, Somalia is no longer considered to be in an active drought. In November, 2017, the Food Security and Nutrition Analysis Unit (FSNAU) of the Food and Agriculture

---

[351] World Bank Climate Change Knowledge Portal, 1991-2015.

[352] Food Security and Nutrition Analysis Unit – Somalia. (http://www.fsnau.org/analytical-approach/methodologies/climate).

[353] Water Infrastructure Development Program for Resilience in Somaliland. African Development Bank Group. 2016.

[354] Social Science Research Council (SSRC). *'Crisis in the Horn of Africa'*, n.d.

[355] "Somalia famine "killed 260,000 people"". BBC News. 2013.

[356] Brookings Institution. '*Somalia: Drought + Conflict =Famine?. Chapter 3*'. 31 January 2012.

[357] Al Jazeera. *'Somalia declares 'national disaster' over drought'*, 28 February 2017.

[358] WFP. *'Somalia'*, 2017.

Organisation (FAO) reported that November rainfall totals had been better than expected.[359] At the same time, a number of areas in Somalia continue to experience drought conditions, and the need for emergency food rations is widespread. The FSNAU also suggests that drought conditions may return during 2018, as the Gu rains slated to begin in April are expected to be below normal, based on climate forecasting models.[360] And, despite the abatement of drought, the FSNAU reported in January 2018 that 2.7 million people still face nutrition-related crises in Somalia, and drought-related displacement, despite declining from early 2017, is still significant.

Impact on Learning

It is uncertain how significant is the drought and its impact in EGEP-T project areas will be. According to RI's own assessments, 23.0 per cent of the school's in the evaluation are severely impacted by the drought to date, and a further 5.6 per cent are impacted more moderately.[361] In terms of impact on EGEP-T schools, the most severe effects appear to be in Galmudug, Somaliland, and Puntland, in that order. The analysis is based on schools selected into the baseline sample; however, and does not represent the full geographic spread of EGEP-T schools across areas.

Overall, the evidence suggests that the drought has already impacted EGEP-T schools negatively. As we showed in the previous section, schools impacted by the drought have lower learning scores than unaffected schools, even when controlling for a wide range of demographic and geographic factors. Regression results suggest that a student in a drought-affected school scores 4.1 percentage points lower in terms of aggregate numeracy and literacy. Our models control for the project location, its urbanicity, grade level of the student, and various demographic characteristics of the student and their household. The drought's effect in the model is robust to the inclusion of other predictors of learning outcomes as well, such as school attendance rates, girls' self-esteem, and indicators of teacher quality – across all models of learning outcomes that we estimated, we find that exposure to drought is associated with a drop in learning outcomes of between 3.6 and 4.3 percentage points. The impact is substantively large: drought is associated with a drop in aggregate learning outcomes from 65.0 to 60.9 per cent.

Impact on Transition

This section builds briefly on the quantitative analysis presented in Section 5.6. There, we reported a series of regression models predicting transition outcomes based on demographic factors, EGEP-T intermediate outcomes – such as community attitudes – and other factors that are expected to influence transition. As part of the analysis, we considered the impact of drought on transition rates among the benchmark transition sample. Our findings do not support the idea that drought reduces transition rates; however, it is important to note that drought is sufficiently correlated with other characteristics – including living in a rural community. The implication is that it is difficult to parse the effect of drought from the effect of other factors that predict transition rates. In these circumstances, qualitative evidence can provide additional illumination where quantitative models are less suited.

The quantitative evidence regarding the drought's impact on transition is inconsistent with the, unfortunately limited, data collected from qualitative interviews regarding drought. Some interviewees report that entire schools will be closed during drought seasons:

---

[359] Climate Update. Food Security and Nutrition Analysis Unit-Somalia". FSNAU.2017.

[360] Climate Update. Food Security and Nutrition Analysis Unit-Somalia". FSNAU.2017.

[361] It is important to note that RI's assessment appears to be conservative, understating the level of impact on some schools that we expect will be affected.

> *"For example during drought seasons, no one remains at the class – all the students left from the school and the school is closed. Also, teachers left from the school and they don't wait while the students come back to the school. So, the main challenge is the drought: there's no school to monitor the students and there's no teachers to wait for them. You will see them only when they themselves come back to the school." -MoE official in Galkayo, Puntland*

Even for those who do not migrate, and those living in urban areas – where drought has a less direct impact on their household livelihoods, generally – drought can influence transition rates and enrolment. For families engaged in agro-pastoralism, especially, drought has significant financial consequences, even in the absence of migration, and these effects can prevent families from being able to afford school fees.[362] But drought can influence those in the urban areas as well. As one interviewee described it, the recent drought "destroyed everything" and "affected everybody in urban and rural areas".[363] Other urban interviewees reported that their household financial situation was also impacted, as they had to support family members living in rural areas where the impact was more direct.[364]

Discussions with other interviewees suggest that there may be dramatically different effects of the drought depending on the local context. While some rural areas are likely to experience out-migration – resulting in girls being uprooted from their schools in these areas – others, especially urban schools, are likely to see an influx of students. One interviewee indicated that the drought would bring more students to their school: "When people move here they do enrol their children and these people are usually those families that have been affected by the drought".[365] Other interviewees echoed the point.[366]

The findings illustrate an important point for the midline and endline evaluations. Drought is likely to have divergent effects on net enrolment in schools: those in areas hit by drought may see a decline in enrolment, while urban schools may see an increase. However, drought will have an unambiguously negative impact on transition rates, as measured in the evaluation. If girls who are part of the cohort migrate away from their schools or simply drop out, it will "count against" EGEP-T's performance indicators. However, if non-cohort girls migrate into an area and enrol at an EGEP-T school, this has no positive effect on the evaluation's transition measure. Therefore, while the net effect of the drought on *enrolment* in any particular school may be unclear *a priori*, the effect of the drought on transition is almost certain to be negative among project schools.

To shed further light on the role of the drought in transition at this stage of the project, we reviewed additional data collected from head teachers regarding dropout rates. At the time of fieldwork, it was too early in the school-year to definitively classify students as having dropped out, even if they had stopped attending class. However, head teachers were asked to assess the primary reasons that girls have dropped out or stopped attending school this year and compare those reasons to last year.

---

[362] FGD with fathers, Bari, Puntland; FGD with fathers, Galgadud, Galmudug; FGD with mothers, Bari, Puntland; FGD with fathers, Sahil, Somaliland.

[363] FGD with mothers, Bari, Puntland.

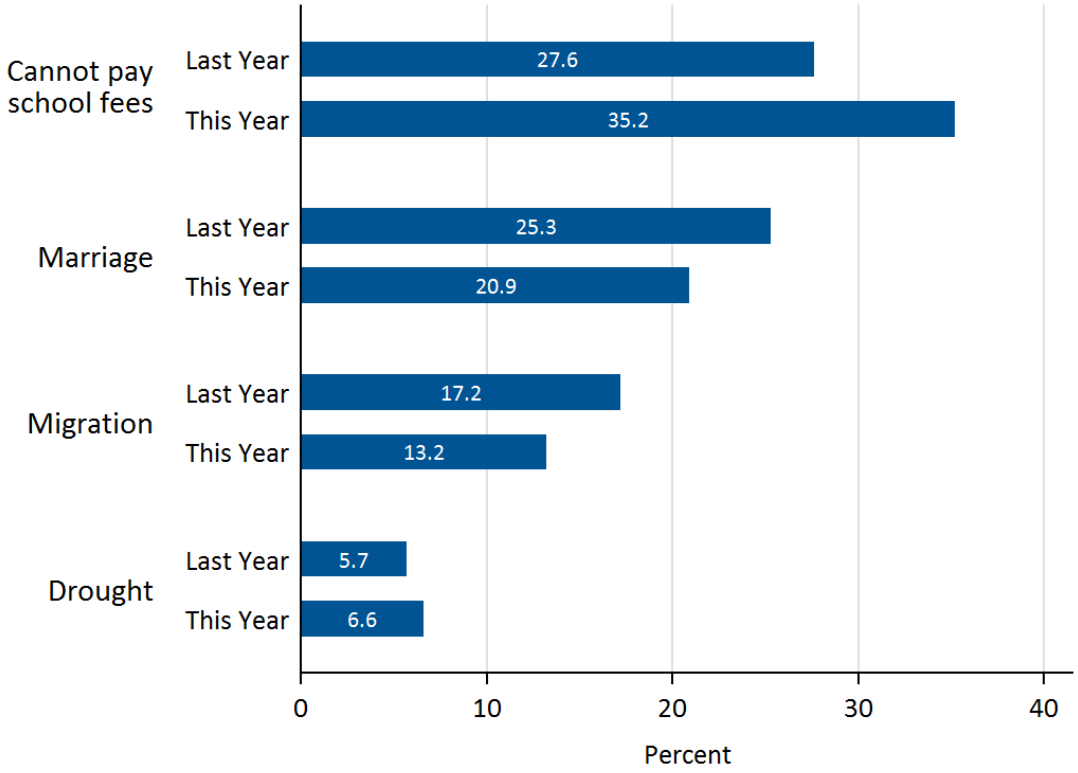
[364] FGD with fathers, Bari, Puntland.

[365] KII with male teacher, Garbadadar, Somaliland.

[366] KII with female teacher, Lafa-ruug, Somaliland.

Respondents were asked to select the single most-important reason girls dropped out in both years. We plot their responses in Figure 27, which shows that head teachers put slightly more emphasis on families' inability to pay school fees this year (35.2 per cent) compared to last year (27.6 per cent), and they also put slightly more emphasis on the impact of the drought (6.6 versus 5.7 per cent).

FIGURE 27: MOST COMMON REASONS FOR GIRLS TO DROP OUT OF SCHOOL



The results likely understate the potential importance of the drought in shaping transition rates, for three reasons. First, unsurprisingly, the drought is more commonly cited by head teachers in drought-affected schools: among schools affected by the drought, 10.5 per cent of head teachers indicate that it is the primary reason for dropouts this year. Second, as noted, many dropouts had not yet occurred at the time of the fieldwork – students may not have been classified as dropouts yet, if they have only recently stopped attending. Third, and most importantly, drought may be an underlying, but not a proximate cause, for many dropouts. The most common reasons for girls to drop out in a *typical* year, according to head teachers, are an inability to pay school fees and getting married. Lacking the money to pay school fees is the proximate cause, but drought and other economic shocks to household incomes may be the root cause of many dropouts.

Baseline data from classroom headcounts provides an additional method for investigating the potential impact of the drought on transition. In each classroom selected for our sample, enumerators collected enrolment figures, where they were available. While official school records have shortcomings – as discussed at length in Section 5.4 – the figures provide a useful measure of average enrolment rates by school, as classrooms were randomly selected within schools.

We focus on total enrolment, which is a count of students enrolled in each class.[367] We look at schools in Somaliland and Puntland only, as these are the only two project locations that a significant number of both drought-affected and unaffected schools. Our sample includes 634 classes, spread across 107 schools. When we regress total enrolment on an indicator of drought exposure, we find that classes in drought-affected schools have 6.2 fewer students, on average. When we control for urbanicity – because drought exposure is most prevalent in rural schools – we find that the correlation is weaker, but that classrooms in drought-affected schools still have 1.7 fewer students.

Our analysis of enrolment levels is merely suggestive. EGEP-T schools vary dramatically in size – exhibiting an order of magnitude differences in enrolment totals, in some cases – and thus the finding that drought-affected schools have lower enrolment may simply be spurious, reflecting the larger size of urban schools that are not drought-affected. As noted, we control for urbanity in the analysis and limit the sample to schools in Somaliland and Puntland, where there is variation *within* locations in terms of drought severity.[368] Moreover, the gaps in school size are less likely to affect the analysis because we are investigating the effect of drought on *class* enrolment, rather than total school enrolment. Although large schools may well have larger class sizes, the bulk of the difference between small and large schools is in the number of classes, rather than their size. Finally, even when we limit the sample by removing class-level outliers in terms of enrolment numbers, we still find a persistent effect of drought on enrolment numbers.

Impact on Attendance

Moving beyond primary project outcomes, it is possible that drought will impact have upstream effects as well, particularly on intermediate outcomes critical to the project's overall success. One such intermediate outcome is attendance: if drought constitutes an economic shock to households, they may respond by asking their children to work outside the home; alternatively, they may need their children to engage in additional at-home domestic work, if adult family members migrate in response to the drought.

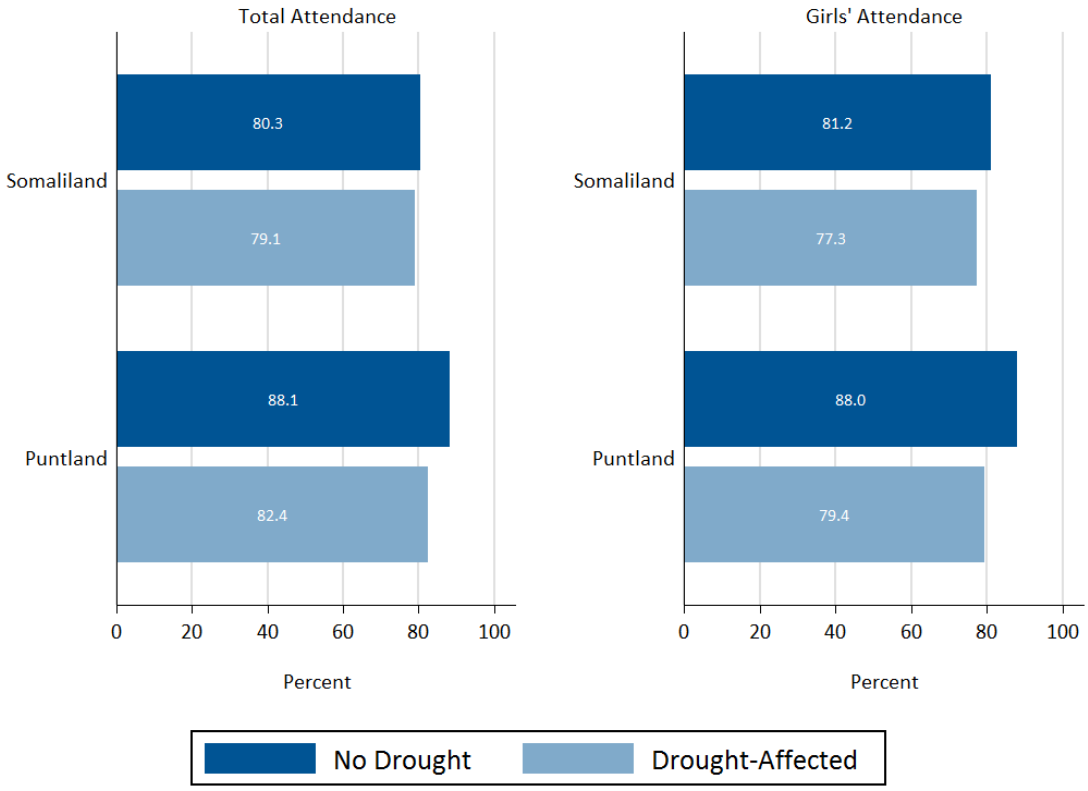
These possibilities are speculative, but they are consistent with many characteristics of Somali society and the Somali economy, in which girls are engaged in extensive housework, and migration for economic opportunities is extremely common. Moreover, there is tentative evidence that drought is already associated with lower attendance rates in drought-affected schools. Using the same classroom headcounts discussed above, we calculated attendance rates on the day of fieldwork at a given school. Focusing on Somaliland and Puntland only – as they are the only areas with both drought and non-drought schools – we find that attendance rates are somewhat lower in drought-affected schools. As shown in Figure 28, the gap between the two types of school is very small in Somaliland, but is more pronounced in Puntland, where drought-affected schools have 5.7 per cent lower attendance rates.

It is important to note that this analysis is based on attendance on a single day per classroom. However, the sample of classrooms itself is large, and sampled randomly from project schools. Moreover, the correlation between drought and lower attendance is outsized in the case of girls' attendance – as the right panel of Figure 28 illustrates.

[367] Note that enrolment totals do not account for the fact schools reside in communities of vastly different sizes, and with very different pools of potential students.

[368] Puntland and Somaliland contrast with Galmudug, for instance, where all sampled schools are drought-affected.

FIGURE 28: RELATIONSHIP BETWEEN DROUGHT AND ATTENDANCE (LEFT PANEL) AND GIRLS' ATTENDANCE (RIGHT PANEL)



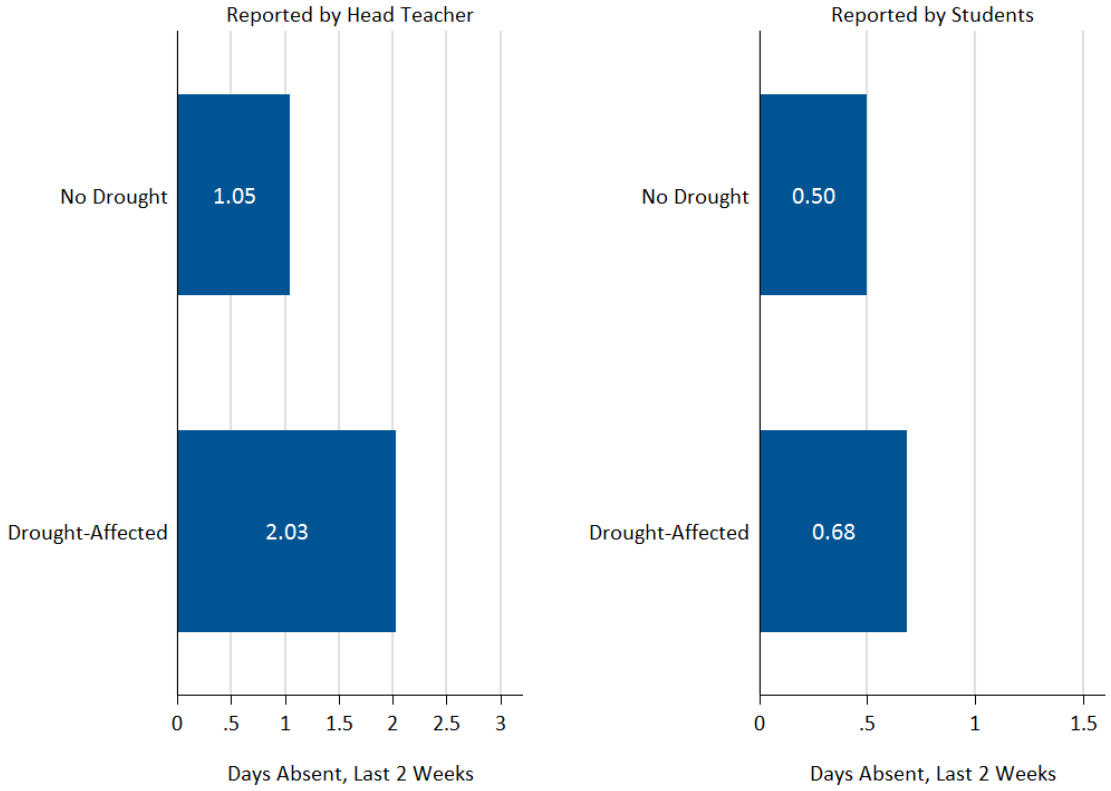
Impact on the Quality of Education

Finally, reviews of the qualitative data collected during the baseline suggested other, less obvious, potential effects of drought on educational outcomes. Specifically, one MoE official, quoted above, suggested that teacher themselves may leave an area during a drought. And, just as students' households face increased economic burden during droughts, teachers must deal with difficult economic circumstances at the same time. One possible consequence of this burden is an increase in absenteeism among teachers, which is already a well-known problem in Somalia's schools, and in EGEP-T project schools.

The baseline evaluation collected data on teacher absenteeism from head teachers and from students. In both cases, absenteeism was measured over a two-week time period; that is, head teachers were asked to indicate the number of full and partial days a teacher had missed in the previous two weeks. Focusing on full-day absences, head teachers report an average of 1.35 full-day absences in the previous two weeks, among a random sample of 514 teachers. In schools with five-day school-weeks, this is equivalent to a teacher missing over 10 per cent of all days.

Is drought correlated with higher teacher absenteeism? Focusing, again, on Somaliland and Puntland, Figure 29 shows that teachers in drought-affected schools are more likely to be absent than their counterparts in non-drought schools. Based on reports from head teachers, teachers in drought-affected schools have nearly double the absences in a two-week period. Using data collected from students about their own teachers – also over a two-week period – we find less absenteeism overall; however, we still find a correlation, albeit much weaker, between drought and teacher absenteeism.

FIGURE 29: TEACHER ABSENTEEISM AS A FUNCTION OF DROUGHT EXPOSURE



Conclusions

The analysis in this section focused on an idiosyncratic but important obstacle to project success: the ongoing and worsening drought in Somalia. The goal of our analysis was to illustrate the diverse effects that drought could have on project outcomes: far from a single, one-directional impact, we expect that the drought will have dramatically different effects across different contexts. We also expect that the drought will have unexpected consequences, some of which will not be captured in the quantitative data collected during annual evaluations. Carefully accounting for drought's effects in future project evaluations will be important. We recommend that RI and its partners continue to collect data on drought severity and classify schools according to drought severity, on a continuous basis. In addition, we also recommend that future evaluation waves target girls who have dropped out of EGEP-T project schools – even if they are not part of the learning cohort – for in-depth interviews, to ascertain the role of drought in their decision.

7. Conclusions & Recommendations

7.1 Conclusions

Profile of Project Beneficiaries

The household survey data derived from a random sample of households in the community reveal that motherhood and orphanhood may influence school enrolment and that many in the community face substantial economic challenges. In-school girls were less likely than out-of-school girls to be mothers (5.3 per cent versus 8.8 per cent) and in-school girls were less likely to be orphans than out-of-school girls (0.4 per cent versus 2.5 per cent). The survey respondents of EGEP-T communities indicated that they faced severe economic barriers. The caregivers for over quarter of survey respondents (26.3 per cent of ISGs and 29.0 per cent of OOSGs) reported that they are unable to meet basic needs without charity.

Interviews with learning cohort girls and boys support findings from the household survey regarding poverty: nearly a third of cohort girls (29.5 per cent) attend drought-affected schools. In addition to these challenges at the household level, a substantial portion of girls also faced barriers at school that put them at risk of educational marginalisation. Despite the dominance of Somali as the mother tongue of most Somalis and, indeed, of cohort girls (99.8 per cent), over a third (39 per cent) report that the language of instruction at school is different and nearly half (47.8 per cent) are afraid of their teacher.

Baseline Learning Levels

At the baseline, literacy and numeracy levels are uniformly low, but they increase with grade level, affirming the assessment's validity. In the aggregate, the mean numeracy score of cohort girls is 68.9 per cent, the mean Somali literacy score is 76.4 per cent, and the mean English literacy score is 38.6 per cent. Bursary girls score at similar levels: 67.3 per cent in numeracy, 74.6 per cent in Somali literacy, and 36.5 per cent in English literacy.

The cohort girls' learning scores consistently lag behind those of cohort boys in the same grade level and across grades on numeracy, Somali literacy, and English literacy assessments. That is, at every grade level, the average score of boys is higher than that of girls. Qualitative interviews suggest that some of the top-performing students in schools are female students, but overall, as shown by the mean test scores above, a significant gap in academic performance exists between girls and boys.

Foundational skills, particularly basic addition and subtraction skills, have been mastered by the majority of cohort girls and boys: 89.5 per cent of cohort girls and boys reached the level of a proficient learner in the first addition task, and 88.3 per cent reached the level of proficient reader on the first subtraction task. Moreover, as subtasks became more difficult, learning scores increasingly became bimodal with most cohort girls and boys demonstrating the knowledge of either a proficient learner or the complete lack of knowledge of a non-learner. For these more difficult subtasks, far fewer girls and boys scored into the emergent learners or established learners categories.

Three types of barriers emerged as correlates of lower learning outcomes: (1) deficits in school infrastructure, (2) teacher quality, and (3) other barriers. Infrastructural deficits leading girls to struggle to move around the school and to not use the toilet at school predict lower learning outcomes. Lower learning outcomes are also observed among girls exposed to poor teaching practices, high teacher absenteeism, and teachers who girls say are unwelcoming or who make them feel afraid. Other predictors of worse learning outcomes include girls feeling unsafe at school, girls feeling unsafe on the way to school, and girls travelling a long distance to arrive at school.

Baseline Transition Rates

The evaluation's sampling approach did not allow for the straightforward measurement of baseline transition rates among the primary learning and transition cohorts that will be tracked at midline

and endline. Rather, the outcome was benchmarked by assessing transition rates among a random sample of households in EGEP-T communities. The baseline transition rate of benchmark girls in this sample was 72.7 per cent. The evaluation defined a successful transition as one in which the girl surveyed graduated to the next grade or one in which she re-enrolled into an appropriate grade level in the past year. While transition rates are, arguably, higher than should be expected in the Somali context, this may have arisen due to the prior impact of EGEP and first-year interventions implemented as part of EGEP-T. Moreover, by defining the transition sample as children aged 11-18 years still present at home, the baseline may be overestimating transition rates – children who have migrated due to marriage, for schooling, or to seek employment, all of whom are less likely to be enrolled in school than children still at home, are excluded from the sample, which may produce a higher-than-expected baseline transition rate.

Transitions rates vary by age and gender. Among benchmark girls, successful transition rates are steady from age 11 to 15 at between 72.8 per cent and 79.2 per cent but begin to decline after 15 and fall to a low of 62.3 per cent at age 18. Benchmark boys have an overall higher rate than benchmark girls of successful transition, 76.8 per cent. Benchmark boys have a steady transition rate between 78.4 per cent to 86.3 per cent until the age of 18 in which transition rates drop to 61.8 per cent.

Baseline Sustainability Levels

Aggregating the scores of nine distinct sustainability indicators, EGEP-T communities taken together rate 1.45 on a scale ranging from 0 to 4. Attitudes within communities and CEC financial support of their schools are promising and will support the sustainability of the project's activities and results. Nevertheless, substantial challenges at the community and system level are likely to hinder sustainability. At the community level, while attitudes regarding girls' education are positive, community actions or behaviours are virtually non-existent, and at the system level, the promotion of child protection mechanisms and gender development strategies is largely absent.

Baseline Levels of Intermediate Outcomes

ATTENDANCE

Based on attendance rates gathered from in-person headcounts, attendance is 84.2 per cent among all students. Differences between cohort girls and boys are minor: the attendance rate is 83.3 per cent among learning cohort girls and 84.4 per cent among learning cohort boys. Across project locations, attendance rates are lowest in Somaliland and Galmudug which have mean attendance rates of 79.8 per cent and 82.0 percent, respectively, whereas in Puntland the mean attendance rate is 87.1 per cent.

Higher attendance rates are hypothesized to lead to higher learning outcomes in the Theory of Change, however this is not borne out convincingly in the data. Schools with higher rates of attendance have higher aggregate learning outcomes, but the effect is substantively small and not statistically significant.

SCHOOL MANAGEMENT AND GOVERNANCE

The majority of teachers report that their community education committees are active with 65 per cent of teachers indicating that CECs meet monthly or more frequently, and 66.6 per cent of teachers perceive CECs as good managers of the school. Schools maintain inadequate records of student grades and attendance. In many cases, schools keep records that are incomplete or disorganized such that they are not usable. Other aspects of CEC management are also found to be wanting. For instance, CECs have not set a child protection agenda for schools by developing school-level policies for child protection or a staff code of conduct.

According to the Theory of Change, better school management will have a positive effect on several aspects of a school's functioning including teacher quality, attendance, and financial support, which in turn will promote better student learning outcomes. While schools in which there is a CEC that has met within the last month did not score significantly higher than schools which did not have a

CEC that did so, the relationship between active CECs and aggregate learning outcomes is positive. The lack of a relationship suggests that CECs may not yet have become sufficiently effective to significantly influence learning outcomes, and we would expect the relationship to become stronger as the intervention proceeds.

TEACHING QUALITY

Teacher quality at the baseline is uneven. Teacher absenteeism is the most notable teacher quality issue with teachers missing an average of 1.35 days every 2 weeks of the school. The high level of absence is exacerbated by the fact that in many schools, instruction only lasts half a day, so a short amount of time spent in class is made even shorter because of teacher absenteeism. Teachers are often found not to be prepared for class and in a high proportion of classes observed during the baseline, students merely copied down what the teacher wrote on the board for much of the observation period. Despite other issues in teacher quality, students of the schools tended to perceive gender equity in the classroom.

Higher number of teacher absences and greater shares of teachers with only a secondary school diploma or less in a school are associated with worse aggregate learning outcomes for that school in accord with the TOC's link between teaching quality and learning outcomes. However, as with the relationships with other intermediate outcomes above and aggregate learning scores, this relationship is not statistically significant.

COMMUNITY-BASED ATTITUDES

Community attitudes toward girls' education are largely positive; however, when competing priorities are presented, positive support for girls' education and girls' educational completion is shown to be somewhat more tenuous. Nearly all community members (89.0 per cent) say that they wish for their daughter to complete university. However, 22.2 per cent of caregivers say that they are more likely to withdraw their daughter from school than sell their household assets if faced with financial need. There is also a common perception among girls that members of their community would be more likely to provide financial support to a boy who was enrolling at university than to a girl in the same situation.

The perception by teachers in the community of positive support for girls' education is only weakly related to improved learning outcomes, but this correlation is not statistically significant. Similarly, girls who indicate that their community would favour a boy over a girl when providing financial support for education perform worse on learning assessments, however the relationship is not statistically distinguishable from a null effect.

LIFE SKILLS AND SELF-ESTEEM

Girls' self-esteem was measured through an index of approximately two dozen questions asked of learning cohort girls. The index finds that 24.7 per cent exhibit moderately high self-esteem and just 0.4 per cent show high self-esteem. The index questions measuring the girls' sense of agency reveals girls feel the least amount of decision-making power with regard to schooling decisions in which only 25.6 per cent of girls feel they are solely responsible for deciding whether they will continue with school while 51.5 per cent feel that they make schooling decisions jointly with their family.

The effect of agency and self-esteem on aggregate learning outcomes are among the most compelling for any intermediate outcome analysed in this manner. Learning cohort girls who report that they exercise some degree of agency over schooling decisions score significantly higher (3.3 points) than girls who do not feel they have that agency. Likewise, girls with higher self-esteem index scores perform significantly better on the learning assessments.

Project approach to gender and social inequalities

The project meets and surpasses the minimum Gender Equality and Social Inclusion standards established by the FM. RI's gender analysis and learning from the first phase of EGEP has allowed the project to select interventions that are more likely to work, and which target specific barriers to girls'

educational completion. Examples of interventions that are well-targeted to specific barriers include the provision of bursaries and cash grants based on the recognition that limited financial resources to households is one of the largest barriers to education, especially for girls; similarly, the provision of solar lamps, which enable girls to study at night after completing their household chores, represents an intervention that takes into account context-specific barriers to girls' educational completion and seeks to put girls and boys on a more even footing. The project goes beyond accommodating gender inequities and, in many cases, seeks to transform them, as shown by the project's emphasis on the need for female teacher-mentors to act as role models and a source of psychological support for girls.

While the project is broadly gender-transformative, in areas other than gender – such as the barriers faced by children with disabilities – the project is less transformative. In many cases, the project is unable to address inequities specifically through targeted interventions. The best illustration of this shortcoming concerns children with disabilities, who face stigma within Somali society and a wide range of barriers to educational completion. The project targets girls with disabilities with bursary support, as part of a broader group of severely marginalised students. In other ways, however, the project is unable to address disability-specific barriers by, for instance, addressing stigma or promoting the importance of educational completion for children with disabilities in particular. In these cases, although the project is not transformative, it does accommodate inequalities, so as to avoid exacerbating them.

7.2 Evaluation Recommendations

Reduce survey length, increase sample size: Data collection for this baseline evaluation focused on collecting a broad range of data from many different populations. In most cases, this was because the indicators targeted by EGEP-T – especially its intermediate and sustainability outcomes – were relatively new, and little guidance was provided by the FM regarding measurement. The evaluation team took a conservative approach as a result, and collected data on a wider range of outcomes than was strictly necessary. At the midline and endline, evaluators and RI staff should review the baseline data and results again closely to determine which indicators they wish to maintain. We would encourage caution in cutting indicators, but there are a number of obvious choices – e.g., test grades collected from school records – which did not yield useful information. To make up for a narrower focus, the midline and endlines should increase sample sizes collected for some outcomes, where EGEP-T wants to draw more firm conclusions.

Maintain the sampling methodology: The approaches utilized at the baseline – stratification of respondents, whether they were students, or classrooms – worked well, as did completing a standard number of each instrument at each school. These approaches should be replicated at the midline and endline, to allow for valid comparisons between samples without significant post-sampling adjustment required.

Create unified indices of intermediate outcomes: The intermediate outcomes assessed at the baseline were typically triangulated across multiple types of data, multiple sub-indicators, and multiple populations, in addition to focusing on the specific individual indicators developed by RI and its implementing partners. While the analysis of a broader set of indicators was hopefully useful for learning about project design, too many indicators and too many sources of data occasionally made it difficult to understand exactly where project schools and communities stood – a problem that is likely to be exacerbated at the midline, when the evaluation team may find multiple, contradictory trends in different indicators within a single IO. To clarify progress in terms of intermediate outcomes from the baseline to midline and endline, the evaluation team should consider computing

standardized indices for each intermediate outcome, which will weight individual sub-indicators equally and produce a single numeric score. Condensing intermediate outcomes to a single summary score will not prevent future reports from providing additional detail on sub-indicators, but will increase the readability of the report. Combining multiple measures into a single index also has benefits for statistical power, making it more likely that the evaluation will be able to detect any meaningful changes that occur in intermediate outcomes from one evaluation period to the next.

Consider carefully the measurement of community attitudes: At the baseline, measures of community attitudes came from the household survey, but also a number of supplemental sources. At the midline, EGEP-T plans to conduct household surveys with members of the learning cohort. However, the parents of cohort girls are not comparable to the household survey collected at the baseline, which was randomly drawn from communities around EGEP-T schools. Therefore, the midline will need to either incorporate a supplemental sample of households that mirrors the sampling approach at the baseline, or otherwise account for the systematic differences in sampling approach.

Pilot midline learning assessment extensively: The baseline assessments were broadly well-designed. However, owing to the pre-post design that lacks a control group, it is essential that the midline assessments be piloted extensively to ensure that they are of equivalent difficulty to the baseline. Projects employing a control group have the luxury of "differencing out" any systematic difference between baseline and midline assessment difficulty; a pre-post design does not, and requires extra care in assessment design as a result.

Design midline learning assessment with ceiling effects in mind: At baseline, mild ceiling effects were observed in the assessment of numeracy. Concerns regarding ceiling effects will only become more pronounced over time – as students progress – if learning assessments are designed to be equally difficult across evaluation waves. The midline assessment should be designed as two distinct components. The first portion of the assessment should seek to match the baseline precisely in terms of difficulty, a goal which should be verified with extensive pilot testing. The second portion should incorporate two or more subtasks that are verifiably more difficult. At the midline, comparisons to the baseline can be made using only the first portion of the learning assessment; at the endline, the evaluation team can utilize the full assessment to compare endline to midline, and the truncated version to compare endline to baseline. While this approach will not completely obviate concern regarding ceiling effects, it preserves the ability to make valid, like-for-like comparisons at each stage of the evaluation.

Maintain "hypotheticals" as survey questions: The baseline survey tools asked respondents to consider a number of hypothetical scenarios, designed to assess community attitudes toward education, parental engagement in their child's education, and other socially-desirable, and often intangible, outcomes. These measures were fairly successful, exposing gender gaps in support for education that were not visible in more traditional survey questions. Where they target specific EGEP-T outcomes, these hypotheticals should be maintained, and refined.

Collect identifying teacher information: One weakness of the baseline data collection design was that the rich data collected on teachers could not be linked back to specific students. For instance, data from classroom headcounts and classroom observations could be linked to students at the level of schools, but could not be linked to learning assessments of students in that specific teacher's class. In some cases, doing so may present ethical or privacy issues; however, RI and its evaluators should consider this approach at the midline. Studying the relationship between learning outcomes

and individual-level, specific exposure to high- or low-quality teachers is a more powerful approach than is possible when teacher quality is aggregated to the school level.

Target teacher trainees with purposive sampling: Teachers surveyed at the baseline comprise a stratified random sample of teachers at EGEP-T schools. However, EGEP-T teacher training programmes are focused on a select group of teachers, who were *not* targeted. The midline evaluation should repeat the sampling strategy of the baseline, but supplement it with a purposive sample of teachers engaged in EGEP-T teacher training and other teacher-focused interventions. Once EGEP-T has selected teachers for training and the provision of teacher incentives, the evaluation team and RI's Monitoring & Evaluation staff should consider collecting data on these teachers' attendance – and potentially other indicators – immediately, to provide a valid baseline for tracking progress among this cohort. Data collection could involve telephone interviews or collecting data from head teachers, as needed.

7.3 Programming Recommendations

Improving Community Awareness: EGEP-T should focus on creating girls' education awareness events in coordination with religious leaders and other leaders not already associated with community education. Religious leaders are perceived by the majority of people, with the exception of people in Banadir, to be supportive of girls' education. They are also some of the most influential people in their communities and would most likely be some of the most convincing as well. More efforts will be in the future to improve awareness in the community, especially among families of girls who are out of school.

Girls' Empowerment: The results of the girls' empowerments surveys were generally positive and few girls suffered from low self-esteem. However, most girls did say that they become nervous when speaking in front of people and consider themselves lucky when doing well on a test. Going forward, any self-esteem or empowerment programs should take into account these two aspects of self-esteem as they seem to be the most prominent areas of lower self-esteem.

Likewise, girls' agency was seen to be relatively high but the area where girls experienced the least amount of agency was the decision to continue education. Improving agency in this area is of utmost importance for EGEP-T so that girls experience both an increase in overall agency but also an increase in the area that is most directly tied to the EGEP outcomes of learning and transition.

Gender Equity in Classrooms: While the utilization of various teaching methods seemed to be high, only 60 per cent of girls felt their teachers treated them equitably as compared to boys. The majority of classrooms did not have mixed seating among boys and girls but other measures of gender equity in the classroom were fairly high. More gender-sensitivity training should be given to teachers so that girls feel like they are treated equally to boys in the classroom.

Child Protection Mechanisms: Child protection mechanisms were not commonly found in schools and only 62 per cent of girls reported knowing an adult at the school to whom they could report abuse. While female teachers were linked to higher rates of would-be reporting, it is not feasible to recruit large numbers of female teachers during the project period, and this process falls outside the remit of EGEP-T. Gender-sensitive child protection training should be provided to the teachers already in EGEP-T schools and students, specifically girls, should be educated on what to do if they are experiencing problems in the school.

Give girls the time and space to study: It is not sufficient to focus on changing community attitudes toward enrolment of girls in school. Girls enrolled in upper-primary and lower-secondary

systematically lag behind boys in learning outcomes, and a significant reason is the burden of housework on girls. This burden impacts their attendance and their ability to study. Awareness-raising campaigns should focus on enrolment at the start of the year, but campaigns should continue during the year to encourage households to shift some of the housework burden from their daughters who are enrolled in school.

Teacher Quality: Many students did not feel like their teachers were prepared every day for class and the head teachers reported an average of one day missed by a teacher every two weeks. In addition, classroom observations showed that teachers do not always have a lesson prepared when they do come to class. Going forward, teachers should be trained in the importance of lesson planning and preparation and EGEP-T should work with schools to implement programs to reduce teacher absenteeism.

Focus on teacher quality and barriers to attendance: Many factors traditionally thought to influence student learning – such as sharing textbooks, or the lack of clean water at schools – is not associated with worse learning outcomes in EGEP-T schools. But teacher quality and physical barriers to attendance – a long or unsafe journey to school – have unambiguously negative impacts, and project efforts should be directed to these areas.

Target teacher absenteeism: Teacher absenteeism is high in project schools, and it is strongly correlated with poor learning outcomes among students. Incentives provided to teachers should target their consistent attendance at school, to encourage better learning and a more consistent learning environment.

Target interventions geographically: Different project areas appear to have fundamentally different problems. Galmudug had the worst transition rate at the baseline, but stronger learning outcomes; in contrast, Somaliland was an outlier in terms of its poor learning scores, but had the highest average transition rate in the baseline sample. In addition, Somaliland had the lowest mean score on an index of teacher quality. This suggests that interventions targeting transition and enrolment, such as bursary support and back-to-school campaigns would be most effective in Galmudug, while teacher training may have an outsized impact in Somaliland.

Men's Involvement: While men and boys showed enormously high rates of approval for girls' education, follow through in terms of actions was often quite limited. Men were not involved in their child's education and it was fully left up to the mother to keep the child in school and deal with all school-related matters. To this end, girls may not have felt their father's support in continuing their education. EGEP-T should focus on getting men more involved with their daughters' education in order to make sure girls feel fully supported in attending and staying in school.

Work with, rather than against, cultural barriers: While early marriage is one of the most common reasons for girls to drop out of school, and a significant barrier to girls' educational attainment, many community members express support for continued education after marriage. A subset of the population would balk at encouraging girls to wait until they are 18 or even older to marry, but may be sympathetic to suggestions to continue their education after marriage at a younger age. Efforts to influence perceptions regarding education after marriage should target the two main gatekeepers to the continuation of girls' education: parents and husbands. By designing messaging campaigns that address education after marriage specifically, as opposed to the strictly the importance of education in a general sense, the project may be able to make greater progress on this issue, which affects a large number of Somali girls.

Annex 4: Beneficiary tables

Table 85: 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.	30100	[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 section 1.3 for the methodology behind the calculations which is based on actual 2016/17 enrolment figures.

Table 86: 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.	0	We do not count any boys as part of the learning cohort
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.	63626	This is the number of boys in the target schools that are benefiting from the project in some way but not counted as part of the learning cohort
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.	25,586	This is the number of girls below G6 in the target schools ie. that are benefiting from the project in some way but not counted as part of the learning cohort
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.	684	This is calculated as 3 teachers per project school. One teacher per school will be trained as a Teacher Mentor and two teachers per school will be supported through the Continuous Professional Development Programme.
Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.	115,645	The project is not tracking this number. The endline for phase 1 estimated this to be the number of adults reached through the project. Phase II is working with the same communities so it is valid to estimate the same number will be reached.

Table 87: 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			
Upper primary	Y	19,178	1,321
Lower secondary	Y	5,622	316
Upper secondary	Y	5,300	-
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 88: Target groups - by age

Age Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Aged 6-8 (% aged 6-8)			
Aged 9-11 (% aged 9-11)	Y	To be completed with support from EE and FM	61
Aged 12-13 (% aged 12-13)	Y	To be completed with support from EE and FM	424
Aged 14-15 (% aged 14-15)	Y	To be completed with support from EE and FM	673
Aged 16-17 (%aged 16-17)	Y	To be completed with support from EE and FM	412
Aged 18-19 (%aged 18-19)	Y	To be completed with support from EE and FM	39
Aged 20+ (% aged 20 and over)	Y	To be completed with support from EE and FM	Not sampled
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 89: Target groups - by sub group

The project requests EE and FM guidance in completing this table as would be relying on the baseline data for populating the values.

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Disabled girls (please disaggregate by disability type)			**74**
Orphaned girls			

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Pastoralist girls			**29**
Child labourers			**24**
Poor girls			**We would consider the whole cohort to be 'poor'. Any further nuance would depend on the definition provided.**
Other (please describe)			
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 90: Target groups - by school status

The project requests EE and FM guidance in completing this table as would be relying on the baseline data for populating the values.

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			
Out-of-school girls: have attended school, but dropped out			
Girls in-school			
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Annex 5: MEL Framework

The latest version of the project's MEL Framework is provided separately.

Annex 7: Data collection tools used for Baseline

The data collection tools used during the baseline, in Open Data Kit script format, are provided separately. The quantitative tools included are:

- Household Survey
- Girls School Survey, completed with cohort girls at school
- Boys School Survey, completed with cohort boys at school
- Bursary Girls School, completed with bursary girls at school
- Teacher Survey
- Head Teacher and School Survey, completed by head teachers at school
- Classroom Observations Tool
- Headcount Tool
- Learning Assessments

The qualitative tools included are:

- Boys Focus Group Discussion Guide
- Girls Focus Group Discussion Guide
- CEC Members Focus Group Discussion Guide
- Fathers Focus Group Discussion Guide
- Mothers Focus Group Discussion Guide
- Head Teacher Key Informant Interview Guide
- Teachers Key Informant Interview Guide
- MOE Child Protection Key Informant Interview Guide
- MOE Quality Assurance Key Informant Interview Guide
- MOE Gender Unit Key Informant Interview Guide

Annex 8: Datasets, codebooks and programs

The baseline evaluation's cleaned and labelled datasets are provided separately. The datasets include the following:

- an_HeadTeacherSurvey.dta - data from the Head Teacher Survey
- an_SchoolSurveysLA.dta - combined data from the Girls School Survey, Bursary Girls Survey, and Boys School Survey, together with learning assessments conducted with each respondent. To identify different types of respondents, use the variable survey_type. Note that this dataset also contains attendance data for each cohort girl, derived from official school records. This data was captured as part of the Head Teacher Survey, but was merged with data on individual girls to facilitate analysis.
- an_HHSurveyLA.dta - data from the household survey, as well as all learning assessments conducted in conjunction with the HH survey. Note that this dataset does not provide transition information, which is provided separately (see below) due to the structure of the dataset and the looped structure of the ODK survey. In this dataset, each row is one household.
- an_TransitionHH.dta - transition data from the household survey. In this dataset, each row is an eligible child from the kish grid (girls and boys 11-18 years old). As a result, households are repeated over multiple rows. This dataset should only be used for transition benchmarking.
- an_Headcount.dta - Data from the Classroom Headcount tool.
- an_ClassObs.dta - Data from the Classroom Observation tool.
- an_TeacherSurvey.dta - Data from the Teacher Survey (paper-based)
- an_BenchmarkLA.dta - Data from learning assessments conducted with benchmark girls *and* Form 3 participants in the learning assessment pilots. To identify different types of respondents, use the variable survey_type.

Matching Individuals and Schools

The data are anonymized in line with standard ethical research protocols. To uniquely identify schools, and merge data across datasets on the basis of school, the variable school_code identifies all schools using the same numeric code across datasets. To uniquely identify individual respondents, a series of ID variables are provided:

- unique_hhID identifies households
- unique_girlID identifies members of the main girls cohort
- unique_bursaryID identifies members of the bursary girls cohort
- unique_boyID identifies members of the boys cohort

In some contexts, a variable unique_id uniquely identifies cohort boys, cohort girls, and bursary girls, when all three are included in a single dataset.

Replication Scripts

Two Stata .do files are provided to ease replication of the primary findings regarding learning and transition outcomes. The .do files are called an_learning.do and an_transition.do, respectively.

Annex 9: Learning test pilot and calibration

Piloting Learning Assessments

The learning assessments used in the EGEP-T baseline evaluation were designed jointly by Relief International (RI) and CARE, which is also operating a GEC-T project in Somalia. Joint development of the assessments allowed for greater expertise to be brought to bear on the design, including monitoring and evaluation staff from both organisations. Moreover, joint piloting of the assessments allowed for a larger sample size to be gathered for the pilot, with a total of 310 students participating in the pilot test.[278]

Assessment design took into account lessons learned from the endline studies of EGEP and SOMGEP, respectively. In particular, a key aspect of design that differs from many other GEC-T approaches to learning assessment is the dual-language requirement for assessing literacy in Somalia, given that both English and Somali are both commonly used as the language of instruction in Somali schools. Beyond the dual-language nature of literacy assessment, other linguistic challenges were addressed explicitly during the design phase. For instance, Somali words and phrases were chosen and reviewed with care to ensure that they could be used across diverse populations, who speak varying dialects of Somali. Similarly, English vocabulary was selected to ensure that the equivalent phonemes were recognisable by native speakers of Somali.

The pilot test took place in September 2017, approximately two months prior to the start of fieldwork for EGEP-T. The pilot tests took place in four districts in Galmudug, Puntland, and Somaliland:

- Galkayo District, Mudug region, Galmudug
- Garowe District, Nugal region, Puntland
- Burao District, Togdheer region, Somaliland
- Aynabo District, Sool region, Somaliland

Students included in the pilot ranged from age 8 to age 20, though the bulk of students (86.8 per cent) were aged 12-18. All students fell into one of five grade levels, selected to represent a diverse set of learning levels, as shown in Table 85.

TABLE 91: COMPOSITION OF PILOT LEARNING ASSESSMENT SAMPLE

Grade	Number of Respondents	Share of Pilot Sample
Grade 5	60	19.4%
Grade 7	63	20.3%
Grade 8	67	21.6%
Form 1	64	20.7%
Form 3	56	18.1%
Total	310	100.0%

The results of the pilot were analysed by RI and CARE, including analysis of student performance by subtask. Subtask-specific numeracy, Somali literacy and English literacy scores are presented in Tables 86, 87, and 88, below. Importantly, the total scores presented below differ from those calculated by

[278] Compared to the FM's stated guidance of 75-150 students, per the Monitoring, Evaluation and Learning Guidance, Part 2.

RI and CARE in two ways. First, their original analysis was conducted prior to specific guidance on assessment scoring from the FM, and aggregate scores in their original analysis were calculated without using equal weighting of all subtasks. Our analysis follows the approach described in Sections 2 and 4 of the report – and recommended by the FM – that each subtask in a given assessment take equal overall weight in the final score. Second, in the case of the literacy assessment, we separate Somali and English literacy explicitly when calculating aggregate scores, mirroring the method used in the primary report.

The analysis of the pilot data revealed that scores on the first two subtasks of the numeracy assessment were fairly high, at 97 and 96 per cent, respectively. With such high scores, the subtasks do not distinguish very effectively between students, and were removed from the final version of the numeracy assessment as a result. The two subtasks removed focused on number identification and quantity discrimination. Table 86 reports the mean score for the full pilot (13 subtasks), as well as the mean score for the reduced set of subtasks (11) included in the final assessment design.

TABLE 92: PILOT NUMERACY SCORES, BY SUBTASK

Subtask	Mean Score
1	97.2
2	95.5
3	58.4
4	91.8
5	87.3
6	68.6
7	55.5
8	71.9
9	68.5
10	30.7
11	50.1
12	21.9
13	56.5
Aggregate	65.7%
Aggregate, Reduced Subtasks	60.1%

TABLE 93: PILOT SOMALI LITERACY SCORES, BY SUBTASK

Subtask	Mean Score
1	63.2
2	91.1
3	81.9
4	106
5	51.5
6	58.3
7	73.9
8	59.9
Aggregate	73.2%

TABLE 94: PILOT ENGLISH LITERACY SCORES, BY SUBTASK

Subtask	Mean Score
1	79
2	62
3	47
4	95
5	18
6	19
7	17
8	14
9	6
Aggregate	39.7%

Beyond these changes, no notable adaptations were made to the assessments after piloting. The pilot assessments suggested mild, not severe ceiling effects would occur in the numeracy assessment, though this problem typically worsens over the life of an evaluation, as students' numeracy improves with time. Figure 28 plots the scores of the pilot numeracy assessment, where scoring was done according to the method used in Section 4 of the report, and the first two subtasks of the pilot are removed, to mirror the assessment used in the primary evaluation. In practice, the plot appears to overstate the extent of ceiling effects, as just two students out of 310 achieved perfect scores. At the same time, 20 students, or 6.5 per cent of the sample, achieved scores of 95 per cent or higher. Given that this assessment was used at the baseline, this suggests significant potential for ceiling effects two years later. Particularly high numeracy scores at the pilot were concentrated especially among Form 1 – and, to a lesser extent, Form 3 – respondents.

Similar ceiling effects were observed in the pilot scores of Somali literacy, as shown in Figure 29. In the pilot sample, approximately 8.4 per cent of respondents achieved particularly high scores, above 95 per cent. Importantly, it is possible that the ceiling effects observed in terms of Somali literacy were obscured by the aggregation of Somali and English subtasks into a single literacy score, as English literacy scores were relatively low and did not exhibit any ceiling effects.

At the time of the pilot and the baseline evaluation, only one version of the learning assessments was designed and available for use. As a result, no calibration between assessment versions has yet occurred. As we note in the discussion below, calibration will be a critical task for the midline and endline evaluations.

FIGURE 28: DISTRIBUTION OF PILOT NUMERACY SORES

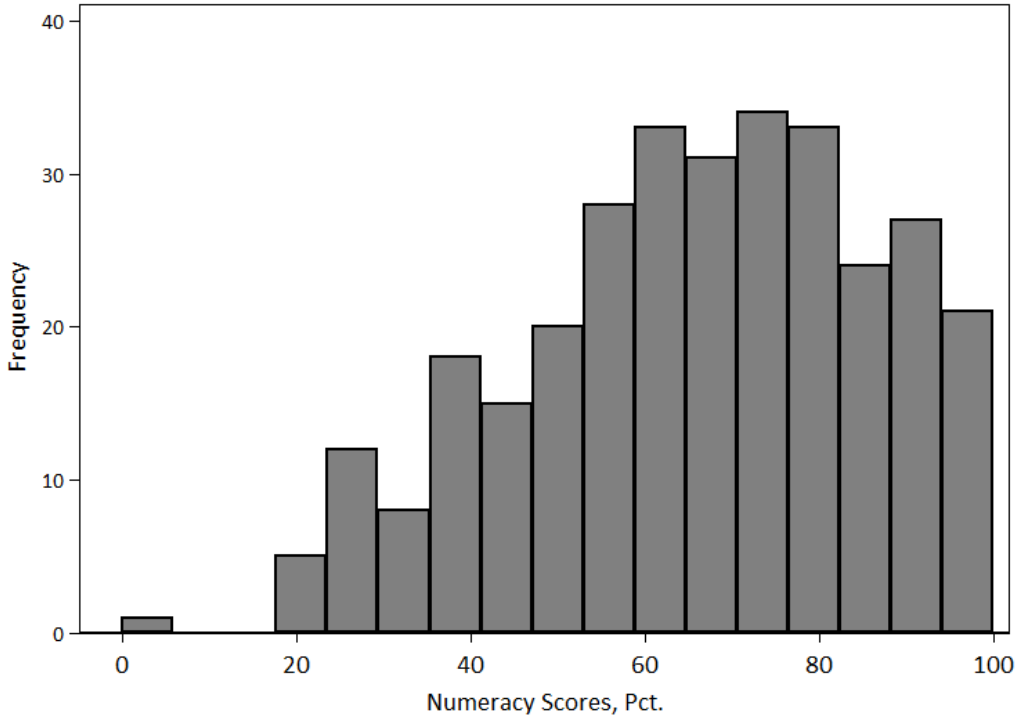
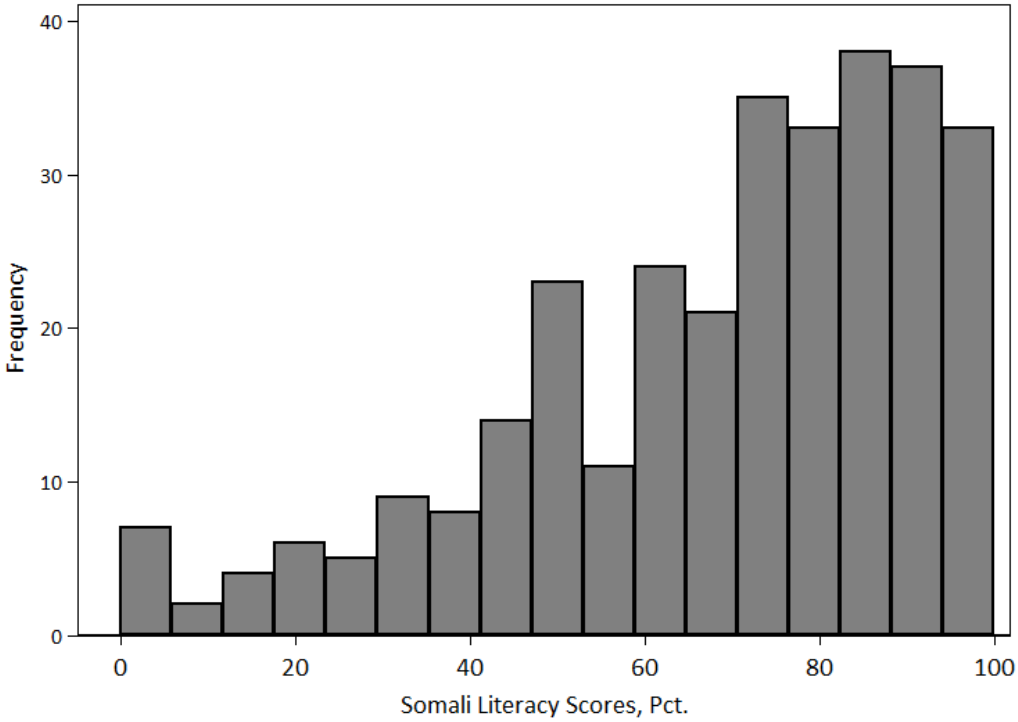


FIGURE 29: DISTRIBUTION OF PILOT SOMALI LITERACY SCORES



Assessment Scoring

In the baseline evaluation, scoring of learning outcomes followed the FM's guidance, derived from the GEC-T Baseline Report Template. To fix terminology, we call individual test questions "items"; we call a set of such items that are unified by a single theme or prompt – such as a series of addition problems, or a series of questions about a single written vignette – a subtask. Multiple subtasks comprise the full assessment.

The learning scores reported in the main baseline report were calculated such that each subtask receives equal weight in the final score. For instance, consider a subtask with four items: each student is assigned a score for the subtask, which is the percentage of those four items that they answered correctly. The total assessment score is the mean of subtask scores. In practice, this means that individual items can be weighted radically differently in the final test score, because some subtasks contain a single item and some contain as many as ten items. However, ensuring equal weight is given to each subtask is essential – this ensures that a student lacking a single skill (such as division) is not unduly punished by the fact that the division subtask may have many individual items. Moreover, weighting subtasks equally is sensible, because items vary dramatically in the time they require, while subtasks tend to require similar levels of effort and time from students.

It is important to note the special grading considerations applied to subtasks that asked students to read a prompt and assessed the words they read per minute. These subtasks, unlike the others included in the assessments, do not conform neatly to a percentage score. To score these tasks, the evaluation team applied the following steps: first, students who were unable to read the prompt at all were assigned a raw score of zero. Second, students who did not finish the complete prompt before time expired were assigned a raw score equal to the number of words they read. Third, for students who finished the prompt before the expiration of time, we calculated the number of words they *would have* read in a full minute, if they maintained the same speed over the course of the full minute allotted. That is, if a girl were allotted 60 seconds to read a 100-word prompt, but finished in 30 seconds, her score of 100 words would be multiplied by 2, to indicate the number of words read per full minute. Finally, to convert raw scores into final scores, words read per minute were capped at 100, such that scores ranged from 0 to 100, inclusive. These steps are in line with standard GEC-T guidance.

During reading speed tasks, enumerators recorded both the total number of words read, and the number of words read incorrectly. In theory, this approach would allow the evaluation team to accurately capture the number of words read *correctly*, rather than the total number of words read. However, after reviewing the data extensively, the evaluation team elected not to adjust the raw number of words read to remove words read incorrectly. In other words, the scores reported for words-per-minute tasks throughout this report are the total number of words read per minute. The evaluation elected not to adjust scores on the basis of incorrectly read words because the counts provided by enumerators did not appear to be consistent across enumerators. In some cases, enumerators appeared to misunderstand the instructions, recording nearly as many incorrectly-read words as total words. In other cases, student read widely different numbers of words incorrectly from one task to another, an outcome that is plausible in rare cases, but unlikely to occur in large numbers, given that reading tasks did not contain exclusively simple or difficult words, but contained a mix of words.

Finally, from a practical perspective, the evaluation team considered the difficulty that enumerators faced on these tasks: they were asked to keep extremely accurate time over a short (60 second time period), where a girl finishing a prompt in 40 versus 42 seconds would result in significantly different

end scores.[279] At the same time, they were tasked with keeping track of where the girl finished at the end of the timing period. Finally, they were asked to keep an accurate count of words read correctly or incorrectly. The difficulty of this final task would be exacerbated when assessing students who read quickly, or who speak with an accent unfamiliar to the enumerator. It is also important to note that enumerators were provided minimal guidance in determining which words were read incorrectly: how incorrect of pronunciation qualified a read word as "incorrect"? These factors compounded the difficulty of an already difficult task, and motivated the evaluation team to focus on raw words read, rather than adjusting scores downward to account for words read incorrectly.

[279] In a 100-word prompt, a difference in finishing speed of 40 versus 42 seconds would result in 150 versus 143 words per minute.

Annex 12: External Evaluator declaration

Name of Project: 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:

Brenton D. Peterson, Evaluation Lead

Juuso Miettunen, Technical Advisor

Anna Russell, Analyst

Sarute Vihoontien, Analyst

_____ (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: BDP)
- All data analysis was conducted independently and provides a fair and consistent representation of progress (Initials: BDP)
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed (Initials: BDP)
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by _Forcier Consulting (Company) (Initials: BDP)
- All child protection protocols and guidance have been followed (Initials: BDP)
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols (Initials: BDP)

Brenton D. Peterson

(Name)

Forcier Consulting

(Company)

09 April 2018

(Date)

Annex 13: Project Management Response

Relief International response to EGEP-T Baseline Report

Relief International, ADRA, and CISP, have read the baseline report with interest and are grateful to Forcier Consulting for the high quality product produced. The EGEP-T consortium are committed, throughout the project, to using research and learning to inform the project design. At this point in the process, Relief International and partners have read the report and are now considering what should be adjusted as a result. Such decisions need to be made through a consultative, participatory process, and then approved by the Fund Manager. Therefore, the consortium plans to convene for a workshop in May 2018. The workshop will consider, both findings from the baseline, and from internal learning gained through the first year of the project. The way forward regarding changes will be agreed and presented to FM in the upcoming RAM meeting. Hence, the following response represents key areas of consideration for the project, whilst recognising the firm decisions regarding adjustments will be made at a future date.

Comments regarding the Evaluation Approach and Recommendations on the Evaluation Approach

In general the approach and tools have successfully derived the data and conclusions helpful to the project. The majority would therefore remain similar at midline. However, the following are areas that will require further attention:

- Most importantly and unfortunately, the numeracy results included significant ceiling effects. This is very disappointing as will present challenges at midline and endline in terms of demonstrating impact. The tool was piloted (by RI and CARE) prior to use, and the results at piloting stage did not show the same pattern. The consortium is not sure what to do about this and would be grateful to discuss this matter with the Fund Manager.
- The baseline report provided substantial analysis related to impact of the drought. This is very useful to the project, and will be to the wider sector. We would continue with this level of analysis at midline in order to determine whether the effects have deepened or lessoned.
- Regarding the IO of girls' confidence and self-esteem, the indicator related to potential for leadership. Unfortunately this somehow was not measured. We will need to look at a way of measuring it at midline and possibly finding an approach to retrospectively establish a baseline. The other related data and findings shared under this IO are also useful.
- Tools measured the extent to which teachers treated boys and girls the same in the classroom. We would not necessarily advocate for them being treated in the same way. We encourage gender sensitive teaching, rather than gender blind. Unfortunately the data does not provide us with the insight we need in this area and we would therefore change the approach for midline in this regard.
- The use of hypothetical scenarios to gauge respondent's attitudes was an approach the project had not used in a previous evaluation. The results were very insightful and we would continue, and possibly expand this approach at midline.

Responses re External Evaluator Programmatic Recommendations

Improving Community Awareness: This recommendation is inline with the planned approach. The strong potential of religious leaders as attitudinal influencers was emphasised in the report. Religious leaders are already engaged during BCC activities; the project will consider if this level of engagement could be increased further;

Girls' Empowerment: This is inline with the planned approach. Increasing girls' agency is about the people surrounding the girl, as well as the girl herself. The project is delivering life skills and using girls' clubs as fora to conduct activities to help develop girls' self esteem and confidence.

Gender Equity in Classrooms: The project considers the approach to measurement in this area to have been off track. Upon discussion, the External Evaluator agrees with the points raised by project management. The evaluation has concentrated on measuring whether a gender blind approach is being taken by teachers. This recommendation implies that the project should be aiming for teachers to treat girls and boys exactly the same. The Project disagrees with this suggestion. Our own monitoring also confirms that girls and boys often report being treated differently in the classroom. However, when questioned more deeply, it is often then explained as the teacher needing to encourage the girls to raise their hands more often etc as they often lack confidence, more so than the boys. We encourage a gender sensitive approach to teaching as opposed to gender blind. The project plans to incorporate gender sensitive teaching guidance in the continuous professional development programme.

Give girls the time and space to study: Project management agrees with the suggestion and will place additional emphasis on this point in behaviour change and awareness raising messages. Qualitative data suggested the expectation for girls to complete chores and housework is negatively impacting on girls' learning. Communities and parents will be encouraged to distribute these tasks more evenly amongst boys and girls;

Teacher Quality: These findings are helpful to the project as the teachers' continuous professional development programme is just starting and these point can be incorporated.

Focus on teacher quality and barriers to attendance: It was an extremely small number of girls that reported considering the journey to school unsafe. These were also often the ones with the longest journeys. Unfortunately the project is not able to provide transport to school. However, the project will follow up on the few cases where lack of safety was raised as an issue.

The project aims to improve teaching quality through the continuous professional development programme.

Target teacher absenteeism: With regards to teacher absenteeism, project management would like to flag that this is heavily connected to the issue of teacher pay. Intuitively, if teachers do not receive their salaries then they are far less likely to attend regularly. The payment of teacher salaries is the responsibility of the respective MoEs and therefore beyond the remit of the project. The World Bank is working with the respective ministries to improve and streamline systems around teacher salary payments. The project is engaging in this initiative wherever possible.

Target interventions geographically: Project management strongly agree with this observation. Somalia/Somaliland is a complicated context and the project's learning cohort is far from homogenous, facing different challenges according to their location and marginalisation status. The respective MoEs are at very different levels of capacity, there is a range of private and public schools, locations facing differing levels of security challenges and drought impact, rural schools and urban schools, and some schools in IDP camps. It is therefore necessary to nuance the approach accordingly.

Men's Involvement: The baseline found that boys and men showed a high level of engagement and positivity towards girls' education. However, 'Fathers, despite overwhelmingly supporting girls' education, do not involve themselves with going to actual meetings or helping their daughters learn in a substantial way. Boys may support girls' education but their actions are not always supportive.' The consortium recognises that the emphasis in men and boys' workshops needs to particularly be action-focused as opposed to awareness raising alone.

Work with, rather than against, cultural barriers: Transition rates are low amongst girls whose parents suggest they would withdraw their daughter from school in order to marry. We will be continuing with BCC messaging that encourages both the completion of education before marriage, and also the continuing of education after marriage when that situation arises.

Additional comments regarding report findings

Forcier acknowledges the limitations of the analysis, particularly due to the lack of control group, however, "Despite this caveat, the findings are generally consistent with RI's Theory of Change. In almost every case, proxy variables for EGEP-T intermediate outcomes had effects in the direction consistent with theoretical expectations…". On the basis of the findings, EGEP is therefore not looking to radically change its Theory of Change. However, there are several areas that can be adjusted to further strengthen it.

Comments related to Drought

As expected, findings suggest that the drought is detrimentally impacting on educational outcomes. The report explains that girls who are severely or moderately affected by the drought showed significantly lower scores in both literacy and numeracy. Students in drought-affected schools were found to underperform by 4.1 percentage points. Attendance was also lower in drought affected schools, particularly amongst girls. The report goes on to describe the likely impact of drought on school drop-outs. The evaluator found that the most common reason for drop-out according to the Head Teachers, are an inability to pay school fees, and getting married; drought and related economic shocks to household incomes may well be the underlying causes. Furthermore, results suggest that the drought has led to increased teacher absence and teacher immigration.

The report highlights how the effects of drought may be massively different depending on the particular local context. For instance, while some rural schools and communities are experiencing significant out-migration, some urban schools are experiencing a notable influx of students. This aligns with our understanding of the differing challenges facing the schools.

The baseline found that CECs in drought affected areas are less likely to be providing bursary support to girls. This is understandable as the communities themselves are facing significant challenges financially. It is also particularly concerning as the girls are likely to be most in need of additional support to attend school.

The project has partly funded the drought response activities through match funding from Education Cannot Wait. It was not possible to predict the length of time through which the effects of drought would continue to be felt. In the existing project design the drought response activities will end in May 2018. However, the impact of the drought is evidently going to be felt beyond that point.

- ➔ Consortium to consider whether there is a critical need to continue the drought response activities in the most affected schools and to incorporate into Year 2 GEC workplan where possible
- ➔ Consortium to consider whether needs are being met in the urban schools receiving significant influx of students due to drought, eg. to confirm that they have enough desks and school resources
- ➔ Consortium to continue to explore possible avenues for further match-funding to aid in the continued drought response activities in GEC schools
- ➔ CEC training to have particular emphasis on building resilience
- ➔ Consortium to consider weighting CEC school development grants more heavily towards drought affected communities
- ➔ Consortium to consider targeting any new bursary support to girls in drought affected schools.

Comments relating to Learning

The baseline found that girls' learning scores lag behind those of boys across all grades and in numeracy and literacy.

There are several points of learning from the results. These will be used in developing the content and focus of remedial classes, information sharing with the respective ministries and schools, and in the EGEP approach to teacher training/content for digital learning platform.

- In numeracy, there appeared to be a general weakness in 'missing numbers'. The consortium will need to explore the extent to which this sub-topic is currently covered in the school curriculum/a
- In numeracy, some foundational skills – basic addition and subtraction, especially – are well-covered currently and learning time could be dedicated to more advanced content
- As expected, scores were significantly lower for English literacy than in both numeracy and Somali literacy. All students, including primary school students were tested in English. Those at primary level will not have been exposed to English to the same extent. However, even though English is the formal language of instruction at secondary level, lessons are rarely taught in English and secondary level students have therefore also had limited exposure to the language. The development of English language skills, particularly at secondary level, evidently need significant focus.
- The distribution of scores for both English and the harder numeracy tasks are interesting, with a high density of scores and the lower and higher ends of the spectrum. It may suggest that the test itself has a sudden step change in difficulty. It may also suggest that slower learners are being left behind at that stage and more effort should be made – including through remedial classes – to identify these struggling students and aid them in catching up.

Sub-groups that are lagging behind their peers include those in rural areas; children with disabilities; girls in families where the Head of Household has no education; girls who are old for their grade. This information can potentially be used to help identify children to participate in remedial classes, and also suggests that we may need to consider additional focus on remedial classes in the rural areas.

As was the case through GEC-1, Somaliland was identified as having consistently lower learning levels than other areas. There are a number of factors that could help explain this: a higher proportion of the EGEP-T schools in Somaliland are rural; a higher proportion are drought affected. It is also the case that schools are sometimes less well-resourced which may be because they don't receive school fees from pupils. Additionally, classrooms are more likely to be over-filled, again, as school fees may be

less of a barrier to enrolment. 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.

Regarding the targets established for midline and endline, EGEP aims to positively impact learning scores as much as possible and expects to see significant gains to that effect. However, it should also be noted that EGEP do not have control schools for use as comparison. Somalia is a fragile, post-conflict/conflict affected state with weak infrastructure. The project, and most significantly, the female beneficiaries, face significant challenges. Of particular note is the continuing effects of drought, as outlined above and in the main report itself. Arguably, it would be appropriate for these factors to be considered when establishing the target. EGEP would argue that where there is no control and such significant challenges are faced, the targets should be nuanced for the particular context, as opposed to being generic for all countries. We would appreciate discussing this with FM.

Comments relating to Transition

An interesting finding was that the study did not find a significant drop in transition rates from Grade 8 (upper primary) to Form 1 (lower secondary), rather, transition rates steadily declined as grades progress. The project will therefore continue with BCC messaging encouraging general continuance and completion of schooling, as opposed to concentrating messaging on any specific key transition points, grade-wise.

Regression analysis finds Somaliland to be a positive predictor of successful transition. This is not necessarily surprising as there are officially no school fees in Somaliland and an inability to pay school fees tends to be the most commonly cited reason for dropping out of school. Transition rates were found to be particularly low in Galmudug. The project takes note of this and will consider what additional focus can be made in that location to improve that situation. Unsurprisingly, transition rates are also low amongst pastoralists, families who migrated, and amongst families that would encourage marriage rather than education. We will be continuing with BCC messaging that encourages both the completion of education before marriage, and also the continuing of education after marriage when that situation arises.

Bearing in mind the study found that inability to pay school fees remains the most common reason for drop-outs, it is important that the project continues its support in the form of bursaries. However, whilst it may get quick short-term results, the consortium is reluctant to significantly increase the number of bursaries awarded. The reason for this reluctance is the eye towards sustainability. The consortium is working with CECs in an aim for them to increasingly take over the payment of bursaries. However, their capacity to do so is limited. It is not necessarily appropriate for schools to become/remain too heavily dependent on the payment of bursary support through aid agencies.

The proportion of girls that are out of school is relatively high at around 20% and, as expected due to declining transition rates, increases at higher ages. This suggests the importance of continuing activities such as back to school campaigns and bursary support to get out of school girls into school.

Regarding the targets for transition rates, we recognise the External Evaluator's point: the baseline transition rates are already moderately high; dramatic improvements are less possible than they would be with a lower baseline. It is also important to note that the drought is expected to have "an unambiguously negative impact on transition rates, as measured in the evaluation. If girls who are part of the cohort migrate away from their schools or simply drop-out, it will "count against" EGEP-T's performance indicators. However, if non-cohort girls migrate into an area and enrol at an EGEP-T school, this has no possible effect on the evaluation's transition measure." This combined with the

fact that the project is not using control comparison, leads Relief International to suggest that the transition targets should be conservative in ambition.

Comments relating to Attendance

Overall, 71.3% of schools had full attendance records available for the past five days; for Galmudug this was almost 100%. Somaliland schools were found to be weakest with regards to attendance record-keeping. The project will be looking at what the reasons for this could be and what other locations, particularly Somaliland, could learn from practices in Galmudug.

71% of schools had accurate records on the day when compared to headcounts. There is therefore still significant room for improvement and the project will consider how to support schools in improvement in this regard.

Attendance tended to be lower in rural schools. EGEP is already focusing particular interventions such as sanitary kits on girls in rural schools and IDP schools, but will consider what further additional challenges are faced in the rural areas in order to respond.

Attendance rates were found to increase as grades progress. This is an interesting finding as it could be assumed that attendance would generally drop as girls get older as they face increased challenges that may interfere with their schooling. However, it may also be the case that, unfortunately, those that face the most challenges gradually drop-out as grades increase, leaving a group of girls that may face challenges but less so. Additionally, girls' agency tends to increase as she gets older, which is likely to mean that she would have greater control over whether she attends school on a daily basis.

Regarding the baseline figures and targets for attendance, it is important to note that the learning cohort were those present on the day, due to the measurement approach utilised. This would mean that the group is biased towards students who attend more regularly. It would make sense for the targets to be connected to the general headcount, but not to the specific learning cohort for this reason. We are pleased to see that attendance rates are relatively high already. It is possible that this is partly due to this being phase II of a two phased project, ie. the high baseline may suggest that the previous interventions were successful in impacting on attendance. Arguably it is unrealistic to set targets that are much higher. The level the target of the learning cohort is similar to the UK average which does not seem feasible. The project would argue for more realistic targets. Alternatively, we would be grateful to discuss with the FM a possible alternative approach such as identifying the lowest performing groups in terms of attendance and attaching the targets specifically to those groups who are likely to be the most marginalised.

Community awareness and attitudes

The project aims to raise awareness and change attitudes and behaviours in ways that will lead to girls' retention and school completion, and to support the achievement of their potential in learning. Several pieces of learning have emerged from the baseline that will be incorporated into the messaging that the related activities focus on, including the following:

- As BCC activities had already started at baseline stage there was the possibility of measuring the extent of initial successes. It is encouraging to see that events are being led by community leaders.
- Transition rates are low amongst girls whose parents suggest they would withdraw their daughter from school in order to marry. The message encouraging education before marriage will be reinforced;

- The back-to-school campaigns were successful in reaching almost half of community members. The approach is therefore effective in terms of reach and should be continued in a similar way.
- The report found less prevalence of awareness raising events taking place in Benadir and Galmudug. This is as expected as, in many areas of South Central, it is not advisable to encourage public gatherings of people due to security concerns. However, the baseline found transition rates to be particularly low in Galmudug so it is important for RI and CISP to find alternative ways of reaching communities with the BCC messages.
- The External Evaluator highlights that awareness raising is often conducted through school events. It should be recognised that, to some extent, this approach may be preaching to the converted, ie. it is the parents that are keen to actively support their children's education, that will have made the effort to contend. The project will ensure that there is significant emphasis placed on activities conducted outside the school in order to reach parents whose attitudes may be more challenging on the issue.

Points emerging for advocacy with wider stakeholders

Results show that there is still a significant gap between the genders across results. As described in the report, girls' learning scores lag behind those of boys in the same grade, across all grade levels and in both numeracy and literacy. Similarly, aggregate transition rates are higher amongst boys than girls, as are attendance rates. EGEP-T records of enrolment figures for all schools also show that enrolment levels for girls are consistently lower than boys across the grades. This is not necessarily in line with assumptions at community or ministry level. It is often said that girls have now caught up with boys and there is therefore no further need for girls to receive special attention. Having the evidence is important to enable us to demonstrate to the respective ministries, to schools, communities, and to the wider education sector, that more needs to be done to ensure the gap is closed.

Evidence shows that girls' attendance rates are lower in schools with no female teacher. We will use these results to demonstrate to schools and the respective Ministries, how important it is to increase the number of female teachers in schools.

The External Evaluator notes the importance of EMIS and the need for its strengthening. RI is engaging in this process, demonstrating the benefits of electronic data collection, and of maintaining database/s for data management. The process is being led by the respective ministries, on different timelines, and with the support of World Bank and UN. We aim to contribute wherever possible, though are not able to incorporate that as an indicator in the project as is ultimately outside the project's control.

Comments relating to Child Protection

Ensuring children are protected of course, is important for it's own sake. It is also important as findings suggest impact on learning results; though the numbers are very small, if a girl feels unsafe on her way to school and feels unsafe at school, her learning results are lower than the average. There are several insightful findings in the report relating to child protection. The results will influence the content of our training workshops with Head Teachers and Teacher Mentors. Of particular note are the following:

- In 79.35 of schools, Head Teachers report that child protection cases should be reported to a man. As noted in the report, this reduces the likelihood of female students, especially, reporting instances of abuse. Additionally, 36.9% of girls do not consider there to be an adult in the school they could talk to in the event of abuse by a teacher.

  → Schools will be further encouraged to have a female focal point for whom children can report issues to.

- The report recognises a high need for strengthening of child protection systems in school. It also finds that a higher proportion of schools have a teacher Code of Conduct in place than we expected, although only half of them had reference to child protection. The report also notes a general lack of child protection policies in place and in use at the school level.

  → The project will continue with the planned approach for strengthening of school capacity in relation to child protection. Where teacher Codes of Conduct exist they will reviewed to ensure the inclusion of child protection. Where teacher's Codes of Conduct don't yet exist, their development will be supported by the project. Training workshops will also help ensure child protection policies are in place and implemented.

44.3% of teachers were found to be implementing physical punishment. This was more common in IDP schools.

→ This is as expected and challenging to tackle as so ingrained in regular practice of teachers. However, it will certainly be raised as a key issue in the training workshops with associated follow-up to track progress in reduction in use.

Comments relating to Community Management of Schools

The baseline found CECs to be at varying levels of capacity. 12 CECs were reported as 'not functioning'. The project will follow up with these schools urgently, whilst also looking at the reasons that the CECs have ceased functioning. This does align with the consortium's recent experience which has seen significant migration due to drought leading to the loss of many roles and institutional knowledge schools and communities. The project is needing to do significant refresher training, whilst utilising the knowledge of those who have remained in the position.

It is interesting to note that cash and non-cash support from CECs to schools was found to be highest in Galmudug. In general, CECs were found to be performing most strongly in Galmudug. The project will be exploring what practices and lessons can be drawn from Galmudug to share with CECs in other locations.

Somaliland performed least well in terms of CEC performance. In some ways this could be expected; children officially do not pay school fees so the CECs have less resources to manage. Expectations are arguably higher on the MoE in Somaliland, which may lead to lower expectations on the CECs.

Comments relating to Marginalisation

1.8% of the in-school learning cohort were identified as having a disability of some kind. The report notes that 70% of all documented physical and cognitive impairments were visual impairments. Efforts should therefore be made in the work with teachers to highlight how best to support children with disabilities, particularly those with visual impairments.

IDP girls – the consortium will discuss with FM, extending the definition of direct beneficiaries to lower grades in IDP schools. The report notes only 5% of learning cohort to be IDPs but that doesn't reflect our work. The project will consider adjust beneficiary numbers and cohort definition. Most IDP schools only have the lower grades and therefore there are girls in lower grades receiving the direct support of the project. RI regrets not considering this as part of the MELF and will consider revising accordingly.

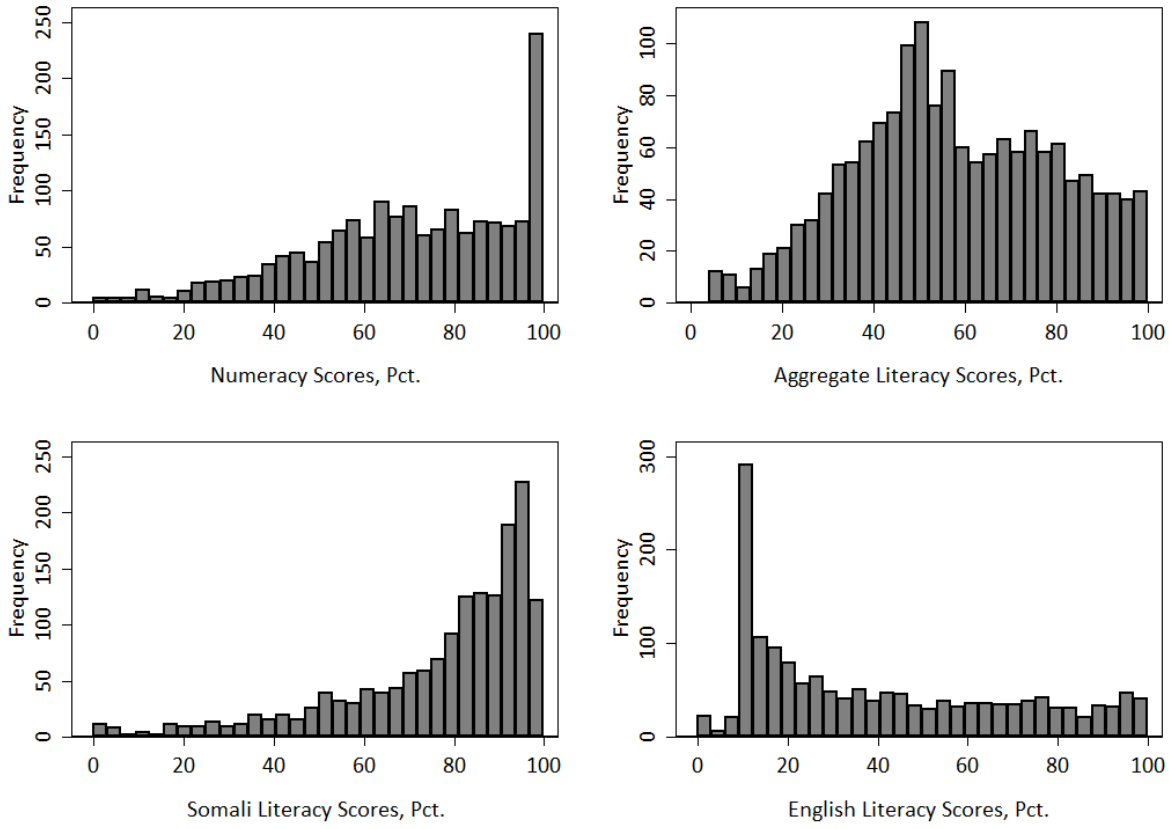
Annex 14: Assessment Validity and Design

Distribution of Scores

In the primary results provided in Section 4.1, we report the distribution of learning assessment scores among the main girls learning cohort and the boys learning cohort. Figure XXX replicates the distribution of scores, limiting the analysis to the main girls learning cohort alone. As with the combined results, we observe significant ceiling effects in the case of numeracy: 10.8 per cent of girls achieved a perfect numeracy score, while a total of 17.7 per cent scored achieved very high scores of above 95 per cent.

The Somali literacy assessments also exhibited ceiling effects, though these effects were much less dramatic. As shown in the bottom-right panel of Figure 30, Somali literacy scores were skewed left; 2.9 per cent of respondents achieved perfect scores in Somali literacy. While a much lower share than in numeracy, a distressingly large number of students achieved sufficiently high scores in Somali that they are likely to be impacted by ceiling effects at the endline. In total, 14.6 per cent of students scored above 95 per cent on Somali literacy.

FIGURE 30: DISTRIBUTION OF LEARNING ASSESSMENT SCORES, MAIN GIRLS COHORT



Validation Checks

In Section 4.1 of the core report, we discussed our attempts to validate the quality of the learning assessment implemented as part of this baseline evaluation. We provided several pieces of evidence that suggested the assessments were both relatively well-designed and well-implemented by field researchers. In this section of the annex, we elaborate on those findings, particularly those related to test item discrimination and "enumerator effects."

TEST ITEM DISCRIMINATION

One important aspect of test design is the extent to which the assessment is able to distinguish between low- and high-performing students, or learners. As we noted in Section 4.1 of the primary report, individual subtasks of the learning assessments had high values on a discrimination index, suggesting that they were able to distinguish between low- and high-performing students. In this annex, we provide methodological background on the discrimination index and report more detailed results, by both subtask and by individual test items.

The goal of assessment is to classify or rank those completing the assessment, and – in the case of repeated testing over time – to gauge improvement among the same respondents over time. An effective test item (i.e. a single question, part of a subtask, using GEC terminology) should distinguish low- and high-performing learners by allowing correct responses from high-performers and forcing incorrect responses from low-performers, on average. That is, an effective test item is one that many low-performing students answer incorrectly, and one that many high-performing students answer correctly.

This idea has a long history in the literature of educational and psychological measurement, in which researchers developed the discrimination index, *d*, to quantify the ability of individual test items to distinguish or *discriminate* between types of learners.[280] The most robust common metric for *d* is the point-biserial correlation coefficient, which is numerically identical to the Pearson correlation in the context of a single dichotomous variable and a single continuous variable. The foundational idea is that correct responses on an effective test item should be correlated with higher scores on the assessment as a whole. That is, a correct answer on a particular test item should predict higher scores on the overall assessment. A test item for which correct answers do not predict higher overall scores is not effectively discriminating between high- and low-performing students, because high-performing students (measured via the full assessment) are no more likely to correctly respond to the item than low-performing students.

For our purposes, we utilize two instantiations of this correlation. When assessing the discriminatory power of a single test item – which is binary, correct or incorrect, by construction – we utilize the point-biserial correlation, to assess the correlation between the test item and overall test scores.[281] When assessing the discriminatory power of an entire subtask – all of which include multiple test items, with scoring on a 0 to 100 scale – we use the standard Pearson correlation coefficient between the subtask and overall test scores.

---

[280] See, e.g., Brennan, Robert L. 1972. "A Generalized Upper-Lower Item Discrimination Index." *Educational and Psychological Measurement* 32(2): 289-303.

[281] The exceptions to this approach with regard to test items are those that assess reading speeds of students in words-per-minute (WPM). Due to the continuous nature of such responses, we utilize the standard Pearson correlation in these unusual cases.

There are a variety of rules of thumb for interpreting discrimination indices. A common metric is that a discrimination index of 0.3 represents a highly-discriminating test item. However, the size of *d* is related to the length of the test; at the extreme, in an assessment with a single item, *d* = 1. Relative to many assessment environments, the learning assessments employed in this evaluation are short – therefore, we may require a higher discrimination index than the 0.3 standard.

The results of the discrimination analysis by subtasks and individual test items are presented in Tables 89 and 90, respectively. Each table performs discrimination analysis for all three assessments. As the results in Table 89 show, nearly every subtask used in the learning assessments are highly-discriminating; the single exception is subtask 1 in the English literacy assessment. In many cases, *d* is greater than 0.7, representing extremely well-designed test items, in terms of discriminating power.

Table 90 extends this analysis to the level of individual test items. Naturally, *d* is often lower in this case, because individual test items are less informative than subtasks that aggregate multiple items. Nonetheless, the assessments continue to perform well at this level. In the numeracy assessment, only 8 out of 66 items fell below 0.3. In the English literacy assessment, only 1 out of 23 items fell below 0.3; in the Somali literacy assessment, the smallest value of *d* observed was 0.42, out of 23 test items. These scores indicate a test that is extremely effective at distinguish between low- and high-performing students.

TABLE 95: DISCRIMINATION INDEX OF LEARNING ASSESSMENTS, BY SUBTASK

Numeracy Subtask	D Index	Somali Literacy Subtask	D Index	English Literacy Subtask	D Index
1	0.57	1	0.60	1	0.10
2	0.44	2	0.62	2	0.63
3	0.52	3	0.68	3	0.83
4	0.69	4	0.64	4	0.81
5	0.77	5	0.75	5	0.89
6	0.59	6	0.76	6	0.87
7	0.74	7	0.78	7	0.84
8	0.79	8	0.77	8	0.82
9	0.79			9	0.79
10	0.80				
11	0.79				

TABLE 96: DISCRIMINATION INDEX OF LEARNING ASSESSMENTS, BY TEST ITEM

Numeracy Item	D Index	Somali Literacy Item	D Index	English Literacy Item	D Index
Sub-Task 1 - Q1	0.17	Sub-Task 1 - Q1	0.60	Sub-Task 1 - Q1	0.10
Sub-Task 1 - Q2	0.15	Sub-Task 2 - Q1	0.43	Sub-Task 2 - Q1	0.63
Sub-Task 1 - Q3	0.32	Sub-Task 2 - Q2	0.42	Sub-Task 3 - Q1	0.68
Sub-Task 1 - Q4	0.41	Sub-Task 2 - Q3	0.45	Sub-Task 3 - Q2	0.73
Sub-Task 1 - Q5	0.43	Sub-Task 2 - Q4	0.48	Sub-Task 3 - Q3	0.75
Sub-Task 1 - Q6	0.31	Sub-Task 3 - Q1	0.47	Sub-Task 3 - Q4	0.76
Sub-Task 1 - Q7	0.48	Sub-Task 3 - Q2	0.49	Sub-Task 4 - Q1	0.81
Sub-Task 1 - Q8	0.49	Sub-Task 3 - Q3	0.51	Sub-Task 5 - Q1	0.80
Sub-Task 1 - Q9	0.47	Sub-Task 3 - Q4	0.52	Sub-Task 5 - Q2	0.78
Sub-Task 1 - Q10	0.44	Sub-Task 4 - Q1	0.64	Sub-Task 5 - Q3	0.78

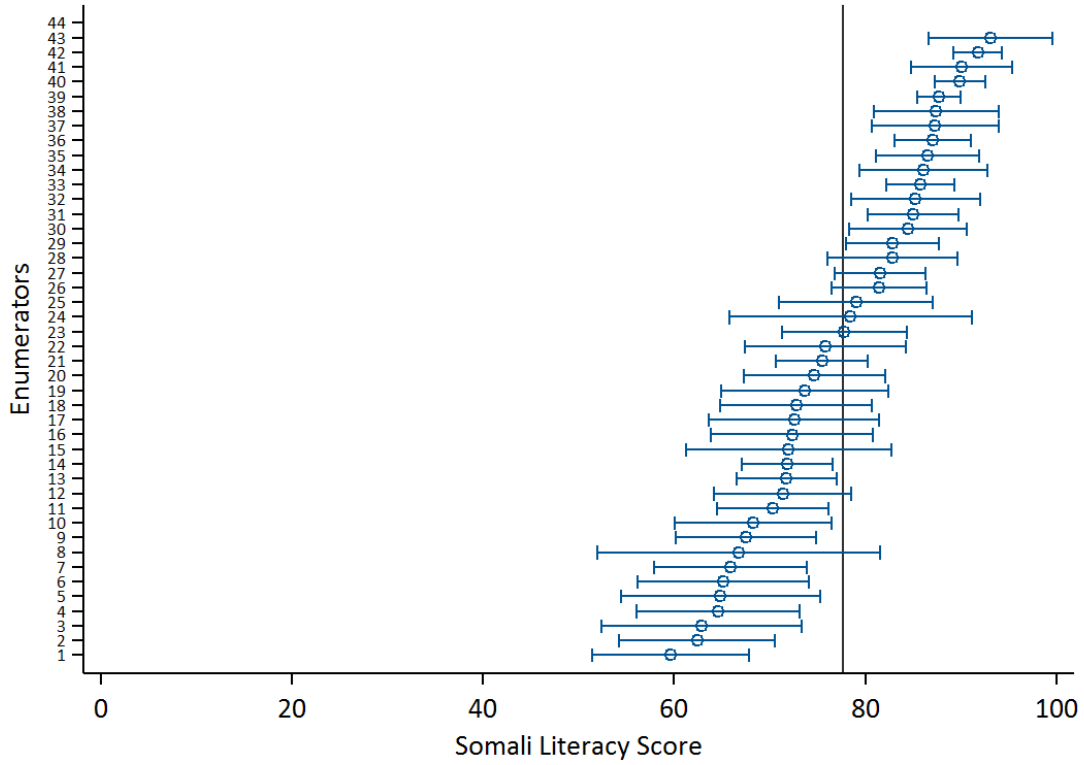
Sub-Task 2 - Q1	0.27	Sub-Task 5 - Q1	0.55	Sub-Task 5 - Q4	0.71
Sub-Task 2 - Q2	0.27	Sub-Task 5 - Q2	0.55	Sub-Task 6 - Q1	0.78
Sub-Task 2 - Q3	0.27	Sub-Task 5 - Q3	0.60	Sub-Task 6 - Q2	0.78
Sub-Task 2 - Q4	0.22	Sub-Task 5 - Q4	0.57	Sub-Task 6 - Q3	0.76
Sub-Task 2 - Q5	0.28	Sub-Task 6 - Qa	0.63	Sub-Task 6 - Q4	0.77
Sub-Task 2 - Q6	0.28	Sub-Task 6 - Qb	0.62	Sub-Task 7 – Q. A	0.76
Sub-Task 2 - Q7	0.30	Sub-Task 6 - Qc	0.63	Sub-Task 7 – Q. B	0.74
Sub-Task 2 - Q8	0.32	Sub-Task 7 – Q. A	0.69	Sub-Task 7 – Q. C	0.75
Sub-Task 2 - Q9	0.34	Sub-Task 7 – Q. B	0.71	Sub-Task 7 – Q. D	0.73
Sub-Task 2 - Q10	0.34	Sub-Task 7 – Q. C	0.71	Sub-Task 8 – Q. A	0.77
Sub-Task 3 - Q1	0.35	Sub-Task 8 – Q. A	0.71	Sub-Task 8 – Q. B	0.74
Sub-Task 3 - Q2	0.39	Sub-Task 8 – Q. B	0.68	Sub-Task 9 – Q. A	0.75
Sub-Task 3 - Q3	0.40	Sub-Task 8 – Q. C	0.69	Sub-Task 9 – Q. B	0.72
Sub-Task 3 - Q4	0.38				
Sub-Task 3 - Q5	0.39				
Sub-Task 3 - Q6	0.40				
Sub-Task 3 - Q7	0.39				
Sub-Task 3 - Q8	0.41				
Sub-Task 3 - Q9	0.44				
Sub-Task 3 - Q10	0.41				
Sub-Task 4 - Q1	0.43				
Sub-Task 4 - Q2	0.48				
Sub-Task 4 - Q3	0.56				
Sub-Task 4 - Q4	0.55				
Sub-Task 4 - Q5	0.57				
Sub-Task 5 - Q1	0.53				
Sub-Task 5 - Q2	0.57				
Sub-Task 5 - Q3	0.64				
Sub-Task 5 - Q4	0.58				
Sub-Task 5 - Q5	0.65				
Sub-Task 6 - Q1	0.41				
Sub-Task 6 - Q2	0.42				
Sub-Task 6 - Q3	0.43				
Sub-Task 6 - Q4	0.48				
Sub-Task 7 - Q1	0.49				
Sub-Task 7 - Q2	0.58				
Sub-Task 7 - Q3	0.61				
Sub-Task 7 - Q4	0.54				
Sub-Task 7 - Q5	0.66				
Sub-Task 8 - Q1	0.66				
Sub-Task 8 - Q2	0.66				
Sub-Task 8 - Q3	0.66				
Sub-Task 8 - Q4	0.68				
Sub-Task 8 - Q5	0.69				
Sub-Task 9 - Q1	0.60				
Sub-Task 9 - Q2	0.64				
Sub-Task 9 - Q3	0.70				
Sub-Task 9 - Q4	0.71				
Sub-Task 9 - Q5	0.70				
Sub-Task 10 - Q1	0.73				
Sub-Task 10 - Q2	0.72				

Sub-Task 10 - Q3	0.69
Sub-Task 10 - Q4	0.68
Sub-Task 10 - Q5	0.67
Sub-Task 11 - Q1	0.53
Sub-Task 11 - Q2	0.49

ENUMERATOR EFFECTS

As noted in Section 4.1 of the report, we performed a number of checks regarding the quality of data collection procedures, looking for anomalous assessment results of all kinds. One such analysis was an investigation of "enumerator effects", or the influence enumerators might have on learning outcomes through any of a number of mechanisms, including incorrect test administration. In the report, we presented results showing a lack of enumerator effects in numeracy scores. That is, no enumerators appear to have systematically anomalous numeracy scores, once we account for the geographic clustering inherent in each enumerator's unique sample of students.[282]

FIGURE 31: ENUMERATOR EFFECTS – MEAN SOMALI LITERACY SCORES, BY ENUMERATOR

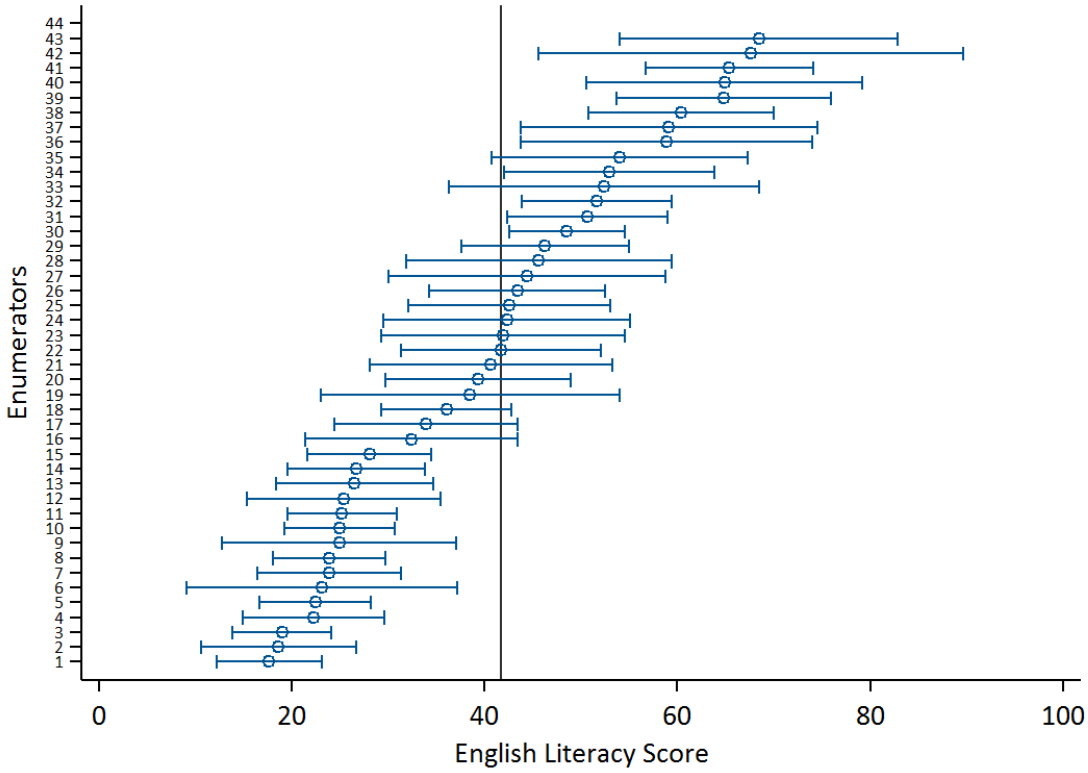


The primary report documented the lack of enumerator effects in the case of numeracy scores. For the sake of completeness, we report the results of a similar analysis, focused on Somali literacy and English literacy, respectively, in Figures 31 and 32. In each graph, we report the mean assessment

[282] If student learning outcomes are correlated within districts and schools, then we would expect some enumerators to report systematically different outcomes than the median enumerator. However, in this case, we would expect enumerators in the same district and on the same team to show a similar "shift" in test scores. Therefore, our analysis looked for enumerators whose sampled students performed significantly better or worse than all other enumerators, especially those on their same team.

score for each enumerator, with the 95 per cent confidence interval for their scores reported as a band around each mean. The graphs are sorted by highest to lowest mean score to ease comparisons across enumerators. These graphs can be interpreted by comparing the 95 per cent confidence intervals across enumerators – if one or more enumerators are clear outliers in terms of their respondents' performance, their 95 per cent confidence interval should not overlap significantly with those of other enumerators. The graphs show that none of the enumerators appear distinctly different in terms of the distribution of scores among students they assessed.

FIGURE 32: ENUMERATOR EFFECTS – MEAN ENGLISH LITERACY SCORES, BY ENUMERATOR



Methodology of the Midline and Endline Assessments

The results of the baseline evaluation suggest several steps that need to be taken to ensure successful midline and endline evaluations, and to allow for rigorous assessment of learning outcomes over time. The first, and most important, is extensive piloting and calibration of the midline and endline assessments against the baseline. To assess changes in learning outcomes over time, the project will compare assessment scores from baseline to midline to endline among the learning cohort. However, if assessment difficulties differ across the evaluation waves, the comparisons being made will be systematically biased.

The FM has recommended a calibration exercise, in which students complete both the baseline and midline assessment and scores are compared between the two tests. Our primary concern with this approach is that the calibration exercise will occur with a relatively small set of respondents – likely fewer than 300 – and will not have sufficient power to detect small but important difficulty differences in the assessment. Importantly, this issue is less problematic in the case of evaluations using a control group approach, as the same changes in difficulty will affect both the treatment and comparison groups equally. As such, the evaluations will still be able to identify the causal effect of the project itself – under standard difference-in-differences assumptions regarding parallel trends and non-

interference – because any change in gaps between treatment and comparison groups can be attributed to the project itself, rather than changes in assessment difficulty.

This is not similarly true for projects relying on pre-post comparisons without a control group. If we consider a hypothetical situation in which no true change in learning outcomes occurred, but the midline assessment was 2 per cent easier than the baseline, the midline evaluation would conclude that the project impacted learning by improving learning scores 2 percentage points. There is no credible method for distinguishing between changes in assessment difficulty and changes in learning outcomes. During meetings with the FM and project and evaluation team members from all the GEC-T projects, the EGEP-T evaluation team raised concerns of this type, and suggested that all three planned assessments be designed in advance of the baseline. Then, students could be randomly assigned to take one of the three versions at the baseline, and different versions in later waves; using this method, the comparison of baseline to midline to endline would be unbiased with respect to changes in assessment difficulty. However, given the logistical challenges that this would entail (fielding multiple different tests) and the short timeline available prior to the start of baseline data collection, this approach was not adopted. For this reason, the calibration exercise outlined by the FM is of particular importance to drawing conclusions regarding the impact of EGEP-T.

The second revision to the learning assessments, going forward, should focus on avoiding and counteracting the impact of ceiling effects in numeracy and, to a lesser extent, Somali literacy. While there is no way to correct the ceiling effects observed at the baseline, it is possible to ameliorate their effects over time. Ceiling effects tend to get worse over the life of an evaluation in which students are tracked and re-contacted over time, since assessments need to be of similar difficulty but student scores experience natural growth as they get older and progress through the grade levels.

We recommend that the midline evaluation be split into two constituent parts. The first, and main, portion of the midline should be of similar, calibrated difficulty, as the baseline, to ensure valid comparisons over time. However, a second portion of the midline should be designed at a higher difficulty level. At the midline, comparisons to the baseline can be made using only the first portion of the learning assessment; at the endline, the evaluation team can utilize the full assessment to compare endline to midline, and the truncated version to compare endline to baseline. While this approach will not completely obviate concern regarding ceiling effects, it preserves the ability to make valid, like-for-like comparisons at each stage of the evaluation.

Annex 15: Grade-Level Learning Competencies

In this annex, we provide a fuller mapping of the competencies students are expected to achieve in English and mathematics at each primary school grade level. The mapping is based on two sets of curricula, drawn from Puntland and Somaliland. As noted in the main report, the curricula available for analysis are substantially incomplete in two ways. First, the Federal Government of Somalia has not assembled a unified curriculum for students under its jurisdiction. A curriculum development process is on-going at the time of this writing. As a result, students in different regions or even districts may be held to substantively different standards in terms of grade-level competencies, a fact complicated further by the differences that exist across project locations. Second, the curricula reviewed by the evaluation team include only competency in English and mathematics; neither Puntland nor Somaliland define clear standards for grade-level competency in Somali literacy. As a result, the assessment of grade-level competency in the main report was limited to English and mathematics performance only.

Table 91 lists the specific sets of skills each educational jurisdiction defines as constituting a grade-level competency in Englist. For instance, for students to have achieved Grade 3-level competency in English in Puntland, they are expected to have mastered simple greeting, naming classroom objects, and have developed an active vocabulary of 200 words, in addition to other skills. In Somaliland, a student in the same grade level are expected to be able to draw and label simple sentences, write cursive letters, understand antonyms and pronouns, among other skills. Note that, by reviewing the tables provided in the primary report (e.g., at the end of Section 4.1), one can see how specific skills tested in the baseline literacy and numeracy assessments map to the skills described in this annex. Table 92 provides a similar mapping of grade-level competencies in mathematics.

TABLE 97: GOAL GRADE-LEVEL COMPETENCIES IN ENGLISH

Grade Level	Puntland	Somaliland
1	Not Classified	Not Classified
2	Not Classified	Reading, Writing & Grammar: • Lower- and upper-case letters • Simple vocabulary (animals and animal sounds) • Parts of speech – nouns and verbs
3	Listening & Speaking: • Simple greetings • Name classroom objects • Active vocabulary of 200 words Reading, Writing & Grammar: • Letter formation • Word recognition • Phonetic awareness • Reading words and simple sentence • Singular and plural forms • Personal pronouns • Question words	Reading, Writing & Grammar: • Draw and label simple sentences • Write cursive letters • Verbs ending in –ing • Antonyms • Pronouns • Time-related adverbs • Indefinite articles (a, an)
4	Listening & Speaking:	Listening & Speaking:

	• Asking questions • Objects in home and parts of body • Active vocabulary of 400 words Reading, Writing & Grammar: • Copying and labelling simple sentences • Reading common words and simple sentences • Present and present continuous tenses • Descriptive adjectives	• Practice speech using simple dialogue Reading, Writing & Grammar: • Writing simple sentences • Adjectives • Present and present continuous tenses • Adverb use
5	Listening & Speaking: • Dialogue • Listening comprehension of short stories • Active vocabulary of 650 words Reading, Writing & Grammar: • Writing short descriptive paragraphs (2-3 sentences) • Context-driven reading strategies • Reading short stories • Past tenses • Forming grammatical questions • Comparisons	Listening & Speaking: • Answering questions with simple sentences Reading, Writing & Grammar: • Past tenses • Reading and understanding simple stories
6	Listening & Speaking: • Active vocabulary of 900 words Reading, Writing & Grammar: • Writing descriptive paragraphs • Description and narration about self • Punctuation • Reading for context and implied meanings • Future tense • Possessive forms • Adverbs • Quantities	Reading, Writing & Grammar: • Writing simple and medium-length paragraphs • Writing short introduction letters • Reading comprehension for medium-length stories • Synonyms • Vocabulary for shopping • Gender nouns • Future tense • Present participle tense
7	Listening & Speaking: • Active vocabulary of 1200 words Reading, Writing & Grammar: • Full compositions • Write answers to questions	Reading, Writing & Grammar: • Writing full letters • Present perfect tense • Relative pronouns • Antonyms and synonyms • Passive voice • Reflexive words

	• More detailed description and narrative • First conditional sentence structure • Pronouns • Adjectives • Modals (will, should, ought to) • Riddles and tongue-twisters	• Auxiliary words • Conditionals • Coordinate conjunctions • Comparison words
8	Listening & Speaking: • Active vocabulary of 1500 words Reading, Writing & Grammar: • Full compositions • Write answers to questions • More detailed description and narrative • Subject-specific reading comprehension • Simple passive and past passive tenses • Prepositions and conjunctions • Present perfect tense • Riddles and proverbs	Listening & Speaking: • Practice with long dialogue Reading, Writing & Grammar: • Writing formal letters • Past participle tense • Past perfect tense

TABLE 98: GOAL GRADE-LEVEL COMPETENCIES IN MATHEMATICS

Grade Level	Puntland	Somaliland
1	• Numbers 0-99 • Place values • Addition up to 99 • Addition without carrying numbers • Adding and subtracting number patterns • Length and capacity comparisons using informal units • Currency and buying/selling scenarios • Lines and complex shapes	• Numbers 0-99 • Identification of shapes • Written forms of numbers 1-9 in words • Place values of 1's and 10's • Addition and subtraction without carrying/borrowing • Length, weight, capacities using arbitrary units • Time and local currency
2	• Mental math • Numbers up to 999 • Addition and subtraction while carrying one number • Multiplying and dividing 2-digit numbers by 1-digit numbers • Length, weight, capacity comparisons	• Numbers up to 999 • Comparisons of two numbers • Written forms of numbers 1-999 in words • Addition and subtraction of 3-digit numbers with carrying and borrowing

	• Shape names and identification • Basic fractions (halves, quarters)	• Patterns in addition and subtraction • Addition and subtraction word problems • Multiplication of single-digit numbers • Simple division (no remainder) • Identify rectangles, triangles, circles, ovals • Measurements using standard units (meters, etc.) • Telling time
3	• Numbers up to 9,999 • Adding and subtracting numbers up to 9,999 • Multiplying and dividing 3-digit numbers by 1-digit numbers • Length and capacity comparisons using formal units • Addition and subtraction of lengths • Perimeters of shapes • Complex and 3-dimensional shapes • Medium-level fractions (thirds, fifths) • Drawing angles	• Numbers up to 9,999 • Addition and subtraction of 4-digit numbers with borrowing and carrying • Multiplication of 3-digit numbers • Division of 3-digit numbers • Multiplication and division word problems • Basic fractions (halves, quarters) • Addition and subtraction of decimals to 2 places • Multiplication of decimals • Identify 3-dimensional shapes • Circumference/perimeter of shapes • Addition and subtraction of basic time units
4	• Numbers up to 99,999 • Number patterns • Addition and subtraction of numbers up to 99,999 • Multiplying 3-digit numbers by 3-digit numbers • Dividing 3-digit numbers by 2-digit numbers • Weights and subtraction of weights using formal units • Lengths and length conversion • Area, volume • Time conversions • Currency addition, subtraction and multiplication • Improper and mixed fractions • Addition and subtraction of fractions and decimals • Drawing angles	• Numbers up to 99,999 • Addition and subtraction of 5-digit numbers • Multiplication and division of 4-digit numbers • Multiplication and division word problems • Improper fractions, mixed numbers, and simplification of fractions • Addition and subtraction of fractions and decimals • Identify types of angles • Conversion of standardized measurements

5	• Numbers up to 999,999 • Number patterns • Roman numerals • Adding and subtracting up to 1 million without borrowing • Multiplying 3-digit by 3-digit numbers • Dividing 4-digit by 2-digit numbers with/without carrying • Multiplication and division word problems • Commutative and other properties of operations • All operations on and conversions of lengths/weights • Perimeter formulas • All operations on decimals • Complex fractions • Ratios • Statistical tables and graphs • Algebraic expressions and equations in one variable	• Lowest common multiples, greatest common divisors, etc. • Addition and subtraction of 6-digit numbers • Multiplication and division of 5-digit numbers • Addition and subtraction of more complex fractions • Multiplication and division of decimals • Measure angles; addition of angles • Types of triangles; properties of squares and rectangles and perpendicular lines • Units of measurement squared • Areas and volumes of common polygons, cubes, etc. • Data presented in tabular and graphical forms
6	• Numbers up to the millions • Number patterns • Addition and subtraction up to the millions • Multiplication of 6-digit numbers by 4-digit numbers • Dividing millions by up to 3-digit numbers • Solving complex word problems • Estimation of lengths, etc. • Area and other geometric formulas (circ. of a circle) • Measurement of speed • Fractions, exponents, square roots • Calculation of percentages • Bisecting lines and angles; complementary and supplementary angles; properties of angles • Ratios and proportions • Plotting graphs • Algebraic expressions and equations in one variable	• Numbers up to the millions • Rounding of numbers and decimals • Squares and simple square roots • Addition and subtraction of 7-digit numbers • Multiplication and division of 6-digit numbers • Identify simple number sequences • Word problems involving lowest common multiple, etc. • Simplify algebraic expressions • Solve equations in one variable • Inequalities • Reciprocals; squares and roots of fractions involving perfect squares • Convert fractions to decimals • Draw and know properties of types of lines (parallel, etc.) • Opposite and supplementary angles • More complex measurement units • Conversion of cubic measures • Measures of speed

		• Read and interpret graphical and tabular data • Mean (average)
7	• Integers • Base 2 and Base 10 numbers • Sets • Squares and roots • Multiplying and dividing numbers of any size • Complex word problems • Word problems involving fractions, decimals and percentages • Transversals and angles; Pythagorean theorem • Direct and indirect ratios • Means (averages) • Solving equations • Algebraic inequalities • Algebraic substitution	• Squares and square roots of perfect squares • Conversion of linear scale to ratio form • Word problems with scale drawings • Ratios and proportions, including word problems • Simplify algebraic expressions, including use of substitution • Solve equations in one unknown • Inequalities in one unknown • Properties of parallel lines, common polygons • Pythagorean theorem • Perimeter formulas of common polygons • Area of a circle • Surface area of 3-d shapes • Measurement problems involving discounts, interest, etc. • Word problems involving graphs and tables • Complex graphs (pie, line, etc.) • Mean and mode
8	• Base 2, Base 5, Base 10 numbers • Indices • Logarithms • Set operations (union, intersection) • Complex measurement conversions • Units of time, speed, distance, etc. • Complex, mixed, improper fractions • Operations on fractions and decimals • Angles in polygons • Types and properties of triangles • Nets of cubes, cuboids, pyramids, etc. • Probability or chance • Linear and simultaneous equations	• Conversion between fractions, decimals, percentages • Word problems involving combined operations and number sequences • Direct and indirect proportions • Form algebraic expressions and equations in one unknown • Parallelograms and rhombuses, including word problems • Nets of pyramids and prisms • Mean, median and mode

- Solving for slope
- Quadratic equations
- Sine, Cosine, Tangents

Annex 16: Topline Tables of Primary and Intermediate Outcomes

A set of tables, which report topline findings for learning, transition, and intermediate outcomes, are provided separately. Among the intermediate outcomes, the tables focus on the key indicators identified in EGEP-T's logical framework. Each outcome is disaggregated by relevant school or community characteristics and, where appropriate, individual-level characteristics.