LASER PULSE

Long-term Assistance and Services for Research (LASER)

Partners for University-Led Solutions Engine (PULSE)

Bar ama Baro - "Teach or Learn"

Somalia's Accelerated Quality Learning Program

Baseline Evaluation

SUPPLEMENT TO AGREEMENT NO. AID-7200AA18CA00009

AOR Name: Kevin Roberts

May 16, 2022

Wilella Burgess
Ann Bessenbacher
Christiana Akande
Weiling Li
Bhagyashree Katare
Abdirisak Dalmar
Farah Bashiir
Julius Ssentongo
Anthony Ssebagereka

This publication was produced for review by the United States Agency International Development (USAID). It was produced for the LASER PULSE Project, managed by Purdue University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGEMENTS

This project is supported by USAID/Somalia through USAID/DDI/ITR/R's Long-Term Assistance and Services for Research (LASER) Project, with Purdue University as the prime. The buy-in activities will build the capacity of Higher Education Institution (HEI) partners to conduct research that advances USAID and its user groups' ability to provide evidence-based Accelerated Basic Education Programs (ABEs) by understanding the effectiveness of the BAB ABE model (or components of the model), learning outcomes of diverse Somali learners across different education options, and factors that influence scalability, effectiveness and cost-effectiveness of ABE at scale. This activity is undertaken by a multi-disciplinary research team through the LASER PULSE consortium members, with Purdue University leading the effort, in collaboration with Makerere University and Somali Research and Development Institute (SORDI) in Somalia, a member of the Makerere University's Resilient Africa Network (RAN). The authors also acknowledge the invaluable contributions of the implementing partner to this work, Creative Associates, International.

About LASER PULSE

LASER (Long-term Assistance and SErvices for Research) PULSE (Partners for University-Led Solutions Engine) is a five-year, $70M program funded through USAID's Innovation, Technology, and Research Hub, that delivers research-driven solutions to field-sourced development challenges in USAID partner countries.

A consortium led by Purdue University, with core partners Catholic Relief Services, Indiana University, Makerere University, and the University of Notre Dame, implements the LASER PULSE program through a growing network of 2,500+ researchers and development practitioners in 61 countries.

LASER PULSE collaborates with USAID missions, bureaus, and independent offices and other local stakeholders to identify research needs for critical development challenges, and funds and strengthens capacity of researcher-practitioner teams to co-design solutions that translate into policy and practice.

TABLE OF CONTENTS

Acknowledgements 2
Abbreviations and Acronyms 6
Table of Figures 8
Table of Tables 10
Executive Summary 12
Project Background 12
Evaluation Approach 12
Methods 13
Key Findings 13
Sample selection considerations 14
Conclusions 15
1 Introduction 17
1.1 Project Context 17
1.2 Project Beneficiaries 17
1.3 Implementation Description and Theory of Change 17
2 Evaluation Goals and Design 18
2.1 Key Evaluation Questions 19
2.2 Overall Evaluation Design 20
2.3 Baseline Evaluation Design 21
3 Methodology 21
3.1 Data Collection Approach 21
Longitudinal Cohort Sample 21
Cross-sectional Cohort Sample 24
3.2 Data Collection Tools 24
3.3 Instruments and Instrument Development 24
3.4 Enumerator Training, Field Testing, and Quality Assurance 27
Training 27
Field Testing 28
Data Collection and Field Notes 28
Quality Assurance 31
3.5 Quantitative Analysis 35

3.6 Cost Analysis 36
4 Baseline Findings -- Learners 36
4.1 Learner Sample Characteristics 36
4.2 Sample Distribution 37
4.3 Learner Demographic Characteristics 39
Gender Disaggregation 39
Age Disaggregation 40
Prior Educational Experiences 42
Summary of Learner Demographic Characteristics 44
4.4 Educational Marginalization (SES indicators) 44
Maternal Literacy 45
Dialects Spoken 45
Family Resources 46
4.5 Learner Equity Perceptions 48
4.6 Safety 49
4.7 Learning Indicators 52
Learner Emotional Engagement in School 52
Learner Quality of Life 54
Numeracy and Literacy Levels 56
Baseline Oral Reading Fluency vs. Reading Comprehension 65
4.8 Intersection between learner Characteristics, School Types, Locations, and academic preparation 67
5 Intermediate Outcomes as Predictors of Learning 71
6 Baseline Results -- Quality of Instruction 73
6.1 Teacher Sample Characteristics 73
6.2 Class Size 76
Teacher Perceptions of Safety 78
6.3 Teacher's Feelings of Self-efficacy 79
Teacher Instructional Self-Efficacy Scale 80
6.4 Teacher Disciplinary Self-Efficacy Scale 81
Teacher Knowledge and Support 82
Teacher External Support Dimension Subscale 84
Head Teacher Sample Characteristics 86

Head Teacher education and experience 86
School Attributes and Policies 87
Support for Education 87
Access to Education 91
Head Teacher Mindsets and Perceptions of Equity Attitudes and Beliefs 92
7 Impact of Sample Selection Methodology (Cross Sectional and Longitudinal Samples) 96
Baseline EGRA Overall Density Plot Results Longitudinal and Cross-Sectional Learner Data . 102
Baseline EGRA Overall Density Plot Results Longitudinal and Cross-Sectional Learner Data . 102
Baseline Longitudinal and Cross Sectional Oral Reading Fluency vs. Reading Comprehension on EGRA 103
7.1 Cost Analysis 106
Cost Analysis Questions 106
Cost Expenditures 108
Cost Drivers 109
Preliminary Cost Analysis 111
Challenges and Limitations 113
8 Conclusions and Recommendations 114
8.1 Populations served by BAB and non-BAB Schools 114
8.2 Student Perceptions of equity, safety, and psychosocial indicators 114
8.3 Geographic Location 114
8.4 Baseline Academic Measures 115
8.5 Quality of Instruction 115
8.6 Sampling Effects 115
9 Appendix 1: Evaluation Questions, Indicators, and Data Collection and Analysis Tools and Methods 117
Abbreviations: 117
Appendix 2: Longitudinal Cohort Sampling Scheme 125
BAB Longitudinal Cohort 125
Non-BAB Longitudinal Cohort 127
Appendix 3: Item Analysis, Survey Reliability and Validity 130
Instrument Validity and Reliability 130
Factor Analysis 130
Reliability 131

Item Analysis (Difficulty and Discrimination) 132
Appendix 4: Enumerator Training 169
Appendix 5: Survey instruments 184
Longitudinal Learner Survey 184
Longitudinal Teacher Survey 184
Longitudinal Head Teacher Survey 184
Appendix 6: Field Notes 185
Data Annex 186
Data Annex 186
References 187

ABBREVIATIONS AND ACRONYMS

ABEs	Accelerated Basic Education Programs
AGES	Adolescent Girls Education in Somalia
AMELP	Activity Monitoring Evaluation Learning Plan
ANOVA	Analysis of Variance
AY	Advancing Youth
BAB	Bar ama Baro
CA	Community Assessment
CDCS	Country Development Cooperation Strategy
CEC	Community Education Committee
CMA	Center Management Assessment
CPE	Community and Parent Engagement
cwpm	Correct Words Per Minute
DEO	District Education Officers
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
ESSP	Education Sector Strategic Plan

FGD	Focus Group Discussions
FGS	Federal Government of Somalia
IDP	Internally Displaced Persons
INGOs	International Non-Governmental Organizations
IRs	Intermediate Results
IRT	Item Response Theory
ISELA	International Social and Emotional Learning Assessment
L	Level
LFI	Learning Facilitator Interviews
LFS	Learning Facilitator Survey
LKFAP	Learning Facilitator Knowledge-Attitudes and Practices
LP	Learner Profiles and Reading, Math, and SEL Assessments
MEL	Monitoring, Evaluation, and Learning
MOE	Ministry of Education
MoECHE	Ministry of Education, Culture, and Higher Education
NDP	Somali National Development Plan
NGOs	Non-Governmental Organizations
OOSCY	Out-of-school children and youth
ORF	Oral Reading Fluency
PGCI	Parent/Caregiver Interviews
PGCS	Parent/Caregiver Survey
RLA	Rapid Learning Assessments
RTI	Research Triangle Institute
SEL	Social Emotional Learning
SES	Social Economic Status
SLEC	Student Learning in Emergency Checklist
SORDI	Somali Research and Development Institute
ToC	Theory of Change

UNFPA	United Nations Populations Fund
USAID	United States Agency for International Development

TABLE OF FIGURES

Figure 1: BAB Program Theory of Change 19
Figure 2: Longitudinal Cohort Baseline Sample Distribution 23
Figure 3: Simplified Learner Model Showing Relationships Among Key Constructs and Intermediate Outcomes 25
Figure 4: Longitudinal Cohort Baseline EGRA Mean 32
Figure 5: Longitudinal Cohort Baseline EGRA Mean Variation by Enumerator 33
Figure 6: Longitudinal Cohort Baseline EGMA Mean 33
Figure 7: Longitudinal Cohort Baseline EGMA Mean Variation by Enumerator 34
Figure 8: Box Plot Analysis of EGRA Baseline Enumerator Scores 34
Figure 9: Box Plot Analysis of EGMA Baseline Enumerator Scores 35
Figure 10: Baseline Longitudinal Sample Distribution by Gender 40
Figure 11: Baseline Longitudinal BAB Cohort Distribution by Age 41
Figure 12: Baseline Longitudinal non-BAB Grade 1 Cohort Distribution by Age 41
Figure 13: Baseline Longitudinal non-BAB Grade 2 Cohort Distribution by Age 41
Figure 14: Baseline Longitudinal Sample Age Distribution by Location and Program Funding Type 42
Figure 15: Baseline Longitudinal Learner Prior Experience with Qur'anic School by Program, Location, and Funding Type 43
Figure 16: Baseline Longitudinal Learner Still Attending Qur'anic School by Program, Location, and Funding Type 43
Figure 17: Baseline Longitudinal Learner Prior Other School Attendance by Program, Location, and Funding Type 44
Figure 18: Baseline Longitudinal Sample Maternal Literacy by Program, Location, and Funding Type 45
Figure 19: Baseline Longitudinal Sample Primary Dialect by Program, Location, and Funding Type 46
Figure 20: Baseline Longitudinal Sample SES by Program Type 47
Figure 21: EGRA Learner Level at Baseline by Program and Grade 61
Figure 22: EGMA Learner Level at Baseline by Program and Grade 61
Figure 23: EGRA and EGMA Density Plots of Learner Baseline Scores by Program Type and Grade 62
Figure 24: EGRA and EGMA Radar Graphs of Learner Baseline Subscale Scores by Program Type and Grade 62
Figure 25: Baseline Longitudinal Learner EGRA and EGMA Scores by Gender 63
Figure 26: Baseline Longitudinal Learner EGRA and EGMA Scores by Location Type 64
Figure 27: Baseline Longitudinal Learner EGRA and EGMA Scores by School Type 64
Figure 28: Baseline Longitudinal Learner EGRA and EGMA Scores by State 65

Figure 29: Baseline Oral Reading Fluency for Longitudinal Cohort Learners by Program and Grade 66
Figure 30: Longitudinal Cohort Oral Fluency vs Comprehension (Literacy Benchmarking) 66
Figure 31: Learner SES, Safety and Psychosocial Perceptions by School Type 68
Figure 32: Numeracy and Literacy Baseline Scores by School Type 69
Figure 33: Learner SES, Safety, and Psychosocial Perspectives by School Location 70
Figure 34: Numeracy and Literacy Baseline Scores by School Location Type 70
Figure 35: Sample Teacher Distribution by Program Type across States 74
Figure 36: Teacher Gender by Program Type, Location, or School Funding 75
Figure 37: Teacher State by Gender 75
Figure 38: Teacher Education Level 76
Figure 39: Teacher Experience in Years 76
Figure 40: Teacher Reported BAB Class Sizes 77
Figure 41: Teacher Reported non-BAB Class Sizes 77
Figure 42: Teacher Perceptions of Safety by Program 78
Figure 43: Teacher Perceptions of Knowledge, Preparation, and Skills by Program 83
Figure 44: Head Teacher Age and Gender 86
Figure 45: Head Teacher Education Level 87
Figure 46: CEC Activities 88
Figure 47: Head Teacher Perception of Community Support for Education 89
Figure 48: Student Less Likely to Receive Community Support for Education 89
Figure 49: Head Teacher Perception of Teacher Support for All Learners 90
Figure 50: Head Teacher Perception of School Accommodations for Menstruating Girls 91
Figure 51: Head Teacher Attitudes Towards Equitable Teaching 92
Figure 52: Head Teacher Attitudes Towards Student Ability 94
Figure 53: Head Teacher Attitudes Towards Student Intelligence 95
Figure 54: Head Teacher Attitudes toward Capcity to Learn 96
Figure 55: EGRA Baseline Scores by Cohort Type 100
Figure 56: EGMA Baseline Scores by Cohort Type 101
Figure 57: EGRA and EGMA Density Plots for Longitudinal, Cross-Sectional, and Non-BAB 1st and 2nd Grade Cohorts 102
Figure 58: Effect of Sampling Cohort on Literacy Proficiency Scores 103
Figure 59: Student Age and Frequency for Longitudinal vs Cross-section Cohorts 104
Figure 60: Longitudinal Baseline EGRA and EGMA Scores by Age 104
Figure 61: Cross-sectional Baseline EGRA and EGMA Scores by Age 105
Figure 62: Longitudinal EGRA and EGMA Score Comparison by Age 105
Figure 63: Cross-Sectional EGRA and EGMA Score Comparison by Age 106
Figure 64: Expenditures by Category 110

TABLE OF TABLES

Table 1: Longitudinal Cohort Classroom Distribution 23
Table 2: Longitudinal Baseline Learner Variables and Measurement Item Sources 25
Table 3: Longitudinal Cohort Sample by State, District, School and Program 29
Table 4: Data Collection Challenges, Remedies, and Recommendations 30
Table 5: Longitudinal Baseline BAB Learners Sample Vs. Total BAB Population 37
Table 6: Longitudinal Study Learner Counts by School Funding Type 37
Table 7: Longitudinal Study Learner Counts by School Location Type 38
Table 8: Longitudinal School and Learner Counts by State, Location and Program Type 38
Table 9: : Non-BAB Program Learner Counts by Location and School Funding Type 39
Table 10: Non-BAB Program Learner Counts by Location and Grade Level 39
Table 11: Baseline Longitudinal Sample SES by Location Type, Gender, Age and State 47
Table 12: Baseline Longitudinal Learner Perceptions of Equity by Location, School Funding Type, Gender, Age, and State 49
Table 13: Baseline Longitudinal Learner Sample Perceptions of Safety by Location, Gender, Age, and State 50
Table 14: Baseline Longitudinal Learner Sample Perceptions of Safety by Program and Grade Level 51
Table 15: Baseline Longitudinal Learner Sample Perceptions of Emotional Engagement by Program and Grade Level 53
Table 16: Baseline Longitudinal Learner Sample Emotional Engagement by Location, Gender, Age, School Funding Type, and State 53
Table 17: Baseline Longitudinal Learner Sample Composite Quality of Life Scores by Program and Grade Level 55
Table 18: Baseline Longitudinal Learner Sample Composite Quality of Life Scores by Location Type, Gender, Age, School Funding Type, and State 55
Table 19: EGRA and EGMA Subtasks 57
Table 20: EGRA (Early Grade Reading Assessment) Overall Longitudinal Sample Test Results 58
Table 21: EGMA (Early Grade Math Assessment) Overall Longitudinal Sample Test Results 59
Table 22: Learner Characteristics by School Type 67
Table 23: Learner Characteristics by Location Type 69
Table 24: Multiple Regression and Correlation Results for EGRA 72
Table 25: Multiple Regression and Correlaton Results for EGMA 72
Table 26: Spearman's Correlations Between EGRA/EGMA and Key Constructs 73
Table 27: Number of Teachers by Type of Class Taught 74
Table 28: Teacher Perceptions of Safety by Location, School Funding Type, Gender, Age, and State 79
Table 29: Teacher Instructional Self-Efficacy 80
Table 30: Teacher Disciplinary Self-Efficacy 82
Table 31: Teacher Perceptions of Knowledge, Preparation and Skills by Program, Location Type, School Funding, Gender, Age, and State 83

Table 32: Teacher Perceptions of Materials and Support by Program, Location, School Funding, Gender, Age, and State 85
Table 33: Longitudinal and Cross-Sectional Sample Composition by State 97
Table 34: Longitudinal and Cross-Sectional Sample Composition by District 97
Table 35: Longitudinal and Cross-Sectional Sample Composition by Location Type 98
Table 36: Sample Comparison of Longitudinal vs. Cross-sectional 98
Table 37: Cost Analysis Questions 106
Table 38: BAB expenditure data using cost categories reported from April 2020 to September 2021 108
Table 39: Top Ten Cost Drivers 111

EXECUTIVE SUMMARY

PROJECT BACKGROUND

The Bar ama Baro - "Teach or Learn" program (BAB), supported by USAID/Somalia, aims to increase access to quality accelerated basic education for out of school children and youth ages 9-16 in targeted areas of Somalia. An international consortium, led by Creative Associates, International, began program implementation in August 2021 in 197 schools with 808 classrooms across 11 target districts in Somalia. They recruited 808 teachers (24% female) and enrolled 39,930 learners (48% female). Over the course of this 5-year project, BAB plans to establish 500-700 Accelerated Basic Education (ABE) centers/schools using the BAB model across 11-15 districts in Somalia and reach 100,000 unique learners.

EVALUATION APPROACH

USAID contracted with a multidisciplinary evaluation team led by the Evaluation and Learning Research Center (ELRC) at Purdue University, in collaboration with Makerere University (Uganda) and the Somali Research and Development Institute (SORDI) in Somalia, a member of Makerere University's Resilient Africa Network (RAN), to conduct an external evaluation of the BAB program. The evaluation has three primary objectives, namely to: (1) understand the effectiveness of the Bar ama Baro (BAB) ABE program in urban, rural, and IDP contexts; (2) document learning outcomes of diverse learners and examine the impact of learner, community, and school characteristics on learning outcomes; and (3) supply feedback to improve program operations and inform evidence-based decision making. **This baseline evaluation lays the groundwork for program evaluation by: (1) establishing starting values for project indicators and serving as a baseline for calculating growth in learning outcomes after one and two years of BAB and formal school instruction; (2) identifying/verifying contextual factors that may affect outcomes; and (3) informing sample selection and program implementation strategies.**

In addition to investigating the effectiveness of the BAB program, this evaluation is examining learning outcomes for children and youth attending public, community, and private education programs that are co-located with or near BAB sites. Data from this investigation will provide information to the Ministry of Education, Culture, and Higher Education (MoECHE) and other education stakeholders to inform education policy, benchmarking for reading and math, and understanding of the impacts of different educational models on diverse learner populations in Somalia. Because a variety of internal and external factors can influence learning outcomes, this evaluation takes a holistic approach that examines the interplay among learner, family, teacher and school, community, and environmental factors.

The external evaluation uses a multiple measure longitudinal cohort design to examine learner growth over time. This design will allow us to document changes in outcomes for individual learners through two years of programming. By following individual learners over time, the external evaluation will be able to examine the impacts of educational interventions on each learner -- both BAB and non-BAB -- (including learning outcomes, retention, and dropout rates) based on gender, age, SES (Social Economic Status), location, and other relevant demographic and contextual factors.

METHODS

The external baseline evaluation focused on describing learner, teacher, and head teacher characteristics for both BAB and non-BAB components of a longitudinal cohort; testing the quality and performance of data collection tools; and examining expected correlational associations between variables of interest. Data for this baseline evaluation came from two sources, primary longitudinal cohort data collected by the evaluation team (teacher, head teacher, and learner surveys – including EGRA and EGMA) and secondary data collected by the ABE implementing partner, Creative Associates (including community and household surveys and cross-sectional cohort EGRA and EGMA data). Data collection for both the longitudinal and cross-sectional samples took place in September-October 2021 using random sampling methodologies.

The longitudinal evaluation includes children and youth beginning as level one (BAB) or grade 1 or 2 (non-BAB) learners in late August 2021. The longitudinal cohort used a random sampling approach that purposefully oversampled rural and IDP locations to mitigate expected challenges in these areas due to population mobility, access, and other considerations. The sampling framework also considered state and district distribution, as well as funding type for non-BAB schools (community, public, or private). The evaluation team also randomly selected formal school classrooms from each site hosting a BAB longitudinal sample classroom to create the non-BAB longitudinal cohort. In cases where a BAB longitudinal site did not host a non-BAB grade 1 or grade 2 class, we identified a non-BAB sample from a nearby school in the same community.

The longitudinal baseline sample included 2912 learners (1714 BAB learners and 1198 non-BAB learners), 54 teachers, and 42 head teachers from three states (Jubaland, Southwest, and Hirshabelle) and the Benadir region and 11 districts. Additional contextual information and baseline evaluation measures (Household survey, Community Survey, cross-sectional learner measures) derived from data collected by BAB as part of their AMELP, program records, or other secondary sources.

KEY FINDINGS

The longitudinal evaluation cohort showed **little variation between BAB and Non-BAB learner samples at baseline on learner characteristics including age distribution, gender distribution, and prior school attendance.** Both the accelerated program learners and learners attending non-accelerated schools varied in age from below 5 to above 19 years of age with median ages for grade 1, BAB ABE learners, and grade 2 at 10, 11, and 12 years, respectively. While overall, both the BAB and non-BAB cohorts showed gender balance, disaggregation by school funding type revealed higher enrollment of male learners in community and private schools, while public schools enrolled more females than males and BAB schools were gender balanced.

Examining the baseline data for differences across school type (community, public, private, and BAB) revealed distinct patterns suggesting that **the various school types serve different populations of Somalia children and youth.** Cohort learners attending private schools scored highest on all baseline indicators. Private school learners are relatively affluent with more boys than girls, speak Maxaa tiri as their primary language and are more likely to have both prior personal educational experiences and more educated mothers. Not surprisingly, private school learners also scored highest on baseline assessments of literacy and numeracy.

Community schools, at the other end of the spectrum, also serve more boys than girls, but they are significantly poorer, speak predominately Maay, and their mothers are less likely to be literate. Learners attending community schools also scored lowest on all psychosocial scales, except safety, and had the lowest baseline literacy and numeracy scores.

Public schools in the longitudinal cohort sample enrolled more girls than boys and served children that are only slightly more affluent than their community-school peers, although they report maternal literacy rates that fall midway between those of private and community school children. Public school learner scores on psychosocial indicators and baseline numeracy and literacy scores are also intermediate to private and community schools.

BAB schools served boys and girls in about equal proportions. Children attending BAB schools report among the lowest levels of maternal education and SES scores. Despite these indicators of economic deprivation, BAB learner scores on psychosocial measures and baseline literacy and numeracy scores are very similar to those measured for public school learners and between community and private school learners.

Where learners live also influences socioeconomic indicators, psychosocial measures, and baseline skills and competencies. While we saw little difference in gender balance, median learner age, or prior educational experience based on location type, learners from IDP and rural areas were far more likely to report low maternal literacy rates and fewer family resources than their urban counterparts. Rural learners reported the lowest perceptions of equity, engagement, and quality of life indicators of all groups and all scored significantly lower on baseline literacy and numeracy tests than learners from either IDP or urban areas. Contrary to expectations, learners from IDP areas, while resource poor, outperformed both rural and urban learners on baseline literacy and numeracy tests – perhaps indicative of other support resources available in these areas.

SAMPLE SELECTION CONSIDERATIONS

The external evaluation compared literacy and numeracy scores at baseline for the longitudinal sample with a cross-sectional sample collected by Creative Associates. Sample selection for both cohorts used a random sampling approach, however there were key differences. The longitudinal cohort purposefully oversampled rural and IDP schools and included all learners in selected classrooms. The cross-sectional cohort drew sample classrooms proportionally and randomly selected 10-12 learners in each. The differences in sample selection methodologies reflected the different purposes for the data and resulted in quite different samples. The longitudinal sample is significantly different from the cross-sectional sample in every demographic aspect except gender. Most notably, the longitudinal sample includes a larger proportion of lower age learners, rural learners, and IDP learners.

The external evaluation compared results from BAB's cross-sectional EGRA and EGMA baseline assessments with results from the longitudinal cohort assessments to examine the effect of sampling on baseline measures of learner skills and competencies. The longitudinal sample has many more

learners categorized as non-learner or emergent learners on both the EGRA (64%) and EGMA 61%) at baseline than the cross-sectional sample (43% on both)[1].

The baseline evaluation verified expected correlations between a variety of factors and learner performance on baseline measures of skills and competencies (literacy and numeracy). Baseline learner skills and competencies in our analysis correlate most strongly with student age, but they also correlate significantly with learner SES, maternal literacy, and quality of life indicators. Urban learners outperformed rural and IDP learners on all measures. As the longitudinal cohort over-represents rural learners, who score lower on these constructs in our analysis, it is not surprising that longitudinal baseline scores at all age levels are lower than the cross-sectional scores at the same age.

Sample characteristics are important when considering education targets. For example, benchmark scores for literacy are commonly set using the first quartile of the Oral Reading Fluency score (in correct words per minute) plotted against reading comprehension scores at the 80% correct level. These measurements yield varying results based on cohort characteristics. The score derived by this method for the longitudinal sample is 49 correct words per minute, compared to 32 correct words per minute for the cross-sectional sample.

CONCLUSIONS

This baseline evaluation 1) established starting values and identified key differences among learners based on location, school type, demographics and family characteristics, and psychosocial measures; 2) verified correlations among these variables and measures of learner skills and competencies, and 3) identified sample selection and program implementation considerations. Key conclusions and recommendations include:

- Both formal (community, public, and private) and ABE classes in the baseline sample included learners across a broad spectrum of ages from 5-19 in entry level classes. This indicates a strong need for ABE to help older learners gain basic education skills rapidly, while making room in formal classrooms for larger numbers of younger learners.
- Baseline data reveals resource inequities across location types, with rural learners under-resourced compared to both urban and IDP learners. As skills and competencies at baseline are strongly correlated with socioeconomic and psychosocial indicators, directing resource allocations to these high need communities should yield positive outcomes.
- Natural disasters, violence, or unrest in the community resulted in depressed learner scores on psychosocial indicators. These results highlight the need to equip teachers through ongoing training and mentoring to meet the educational needs of all learners while supporting positive socioemotional development and health, especially during times of uncertainty, stress, or trauma.

[1] As Somalia does not currently have defined national literacy and numeracy performance levels, we adopted the performance levels described in the Leave No Girl Behind, AGES project (page 97) (Machova, Miettunen and Peterson 2020)

- Moreover, these findings underscore the need to collect data that captures the role of external factors on student outcomes.
- Although learner gender balance varied by school type, males dominated the teacher workforce by a ratio of nearly 3:1. Developing and implementing school policies and procedures that prioritize gender-balanced teacher recruitment will enhance education and role models, especially for girls.
- Finally, variations in baseline measures resulting from different sampling strategies illustrates the need to use holistic data collection strategies that provide adequate representation across location and populations and to understand the strengths and limitations of data sources when developing educational standards, benchmarking, or making other policy decisions.

1 INTRODUCTION

1.1 PROJECT CONTEXT

Somalia has one of the lowest rates of school enrollment in the world with estimates of out-of-school children and youth (OOSCY) ranging from 60-70% of the total school-aged population (Somalia Education Cluster, 2017; USAID, 2016). The combination of widespread violence, food insecurity, and recurring droughts and floods over more than two decades has resulted in many internally displaced persons (IDPs) and disrupted Somalia's educational infrastructure. Currently, the gross school enrollment rate in Somalia is only 21% for primary learners and 18% for secondary learners. Even those children in school often do not learn (Wafula 2020). Furthermore, there are sharp regional disparities in educational opportunities for children, with South Central Somalia being the most affected by the collapse of Somalia's formal education systems. Earlier studies found barriers to primary school education that include finances (prohibitive costs of tuition, textbooks, uniforms, and transportation), child labor (especially within poor households, pastoralist communities and the female gender), and lack of parental support for education (Machova, Miettunen and Peterson 2020).

1.2 PROJECT BENEFICIARIES

The Somali National Development Plan (NDP) for 2020-2024 (The Ministry of Planning, Investment and Economic Development 2020) calls for improvements in the education sector that support the development of an adequate and well-educated workforce that can help move the country forward both economically and socially. In line with the 2018-2020 Federal Government of Somalia (FGS) Education Sector Strategic Plan (ESSP) (Federal Government of Somalia Ministry of Education, Culture and Higher Education 2018) and the country's vision and mission for education priorities, USAID aims to support increased access to quality education for OOSCY ages 9-16 years in targeted regions of Somalia by implementing effective Accelerated Basic Education Programs (ABEs).

In Somalia, international non-governmental organizations (NGPs), non-governmental organizations (NGOs), and donors are the primary implementers and funders of accelerated basic education programming -- which is part of the non-formal education system. Accelerated basic education programming aims to reintegrate and fast-track over-aged OOSCY by giving them the opportunity to complete two years of the basic education curriculum in a single year -- typically done by compressing the curriculum, thus reducing the amount of classroom time required. Accelerated basic education programming is consistent with the vision and mission stated in Somalia's national education priorities.

1.3 IMPLEMENTATION DESCRIPTION AND THEORY OF CHANGE

To increase access to quality education for OOSCY, USAID/Somalia is supporting Bar ama Baro (BAB) – "Teach or Learn." BAB is a 5-year USAID-funded program implemented by an

international consortium led by Creative Associates, Intl. that aims to increase access to quality ABE for OOSCY ages 9-16 in targeted regions and districts of Somalia.

BAB estimates it will establish 500-700 ABE centers/schools using the Bar ama Baro model across 11-15 districts[2] in Somalia, although numbers of centers overall, and by district, remain under discussion. ABE sites will represent a range of community settings (urban, rural, IDP). Over the course of the project, Bar ama Baro hopes to reach about 100,000 "unique" learners. All ABE centers will use the same basic BAB curriculum and approach; however, BAB will tailor program details to meet community needs.

Working closely with relevant government authorities and other partners, BAB aims to provide relevant, flexible, safe, and quality basic education opportunities for OOSCY in Somalia through ABE centers with the **goal of increasing access to quality education for Somali OOSCY ages 9-16.** Four intermediate results (IRs) and one crosscutting result support this goal.

- IR1: Enrollment in ABE is increased
- IR2: Safety of ABE learning environments is improved
- IR3: Student learning outcomes in ABE are improved
- IR4 Government capacity to regulate ABE is enhanced
- Crosscutting IR: Youth civic engagement is strengthened.

BAB's program theory of change (ToC) builds on the development hypothesis that if enrollment in ABE is increased; if the safety of ABE learning environments is improved; if youth's learning outcomes are improved; and if federal, state, and local government capacity to regulate ABEs using an evidence-based framework as a guide is enhanced; then greater numbers of OOSCY will access quality education. Figure 1 captures the BAB theory of change.

In late summer/fall 2021, BAB initiated ABE in 197 schools with 808 classrooms across 11 target districts in Somalia. They identified 808 teachers (24% female) and enrolled 39,930 learners (48% female) (BAB FY21 Annual Report).

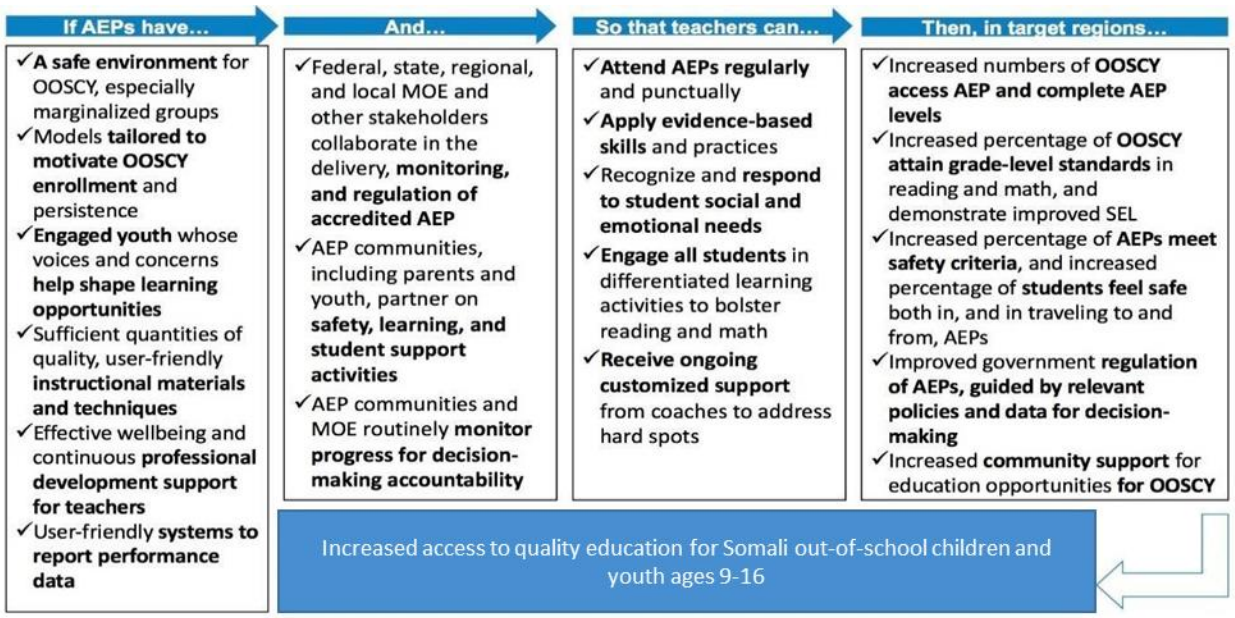
2 EVALUATION GOALS AND DESIGN

The LASER PULSE external evaluation, led by researchers in Purdue University's Evaluation and Learning Research Center (ELRC), leverages the knowledge and expertise of partners from Purdue University, Somali Research and Development Institute (SORDI) and the Resilient Africa Network (RAN). The evaluation has three primary objectives, namely to: (1) understand the effectiveness of the Bar ama Baro (BAB) ABE program across different implementation contexts;

[2] Baseline project intervention areas include: Daynile, Kahda, Shibis, and Hamarweyne in Benadir region; Baidoa and Dinsoor in the Bay region; Balcad and Jowhar districts in Middle Shabelle region; and Baraawe and Wanlaweyn districts in the Lower Shabelle region, and Kismayo in Lower Juba region. In discussion with USAID and the Ministry, the project will expand to additional districts of implementation in Year 3-4.

(2) document learning outcomes of diverse learners across community, public, private, and ABE options and examine the impact of contextual and demographic indicators on learning outcomes; and (3) supply prompt feedback to USAID and Creative (BAB) about effectiveness and cost effectiveness to inform decisions around continuous improvement, replication and scale up.

Figure 1: BAB Program Theory of Change



The external evaluation of BAB is examining the relationships between the BAB program and key outcomes to provide evidence for improving program operations and informing evidence-based decision making. The evaluation aids accountability and planning, as well as continuous improvement and learning. This evaluation also examines learning outcomes for learners attending public, community, and private education programs that are co-located with or near BAB sites. This data will provide information to the Ministry of Education, Culture, and Higher Education (MoECHE) and other education stakeholders to help set standards or benchmarks for reading and math and examine the impacts of different educational models on diverse learner populations in Somalia.

Understanding that a variety of internal and external factors influence learning outcomes, the BAB external evaluation takes a holistic approach that seeks to examine the interplay among learner, family, teacher and school, community, and environmental factors.

2.1 KEY EVALUATION QUESTIONS

Specific external evaluation questions for this project include:

1. To what extent is BAB effective in improving access to quality education for Somali out-of-school children and youth ages 9-16?

Evaluation questions in this category seek to understand the access, retention, safety, and learning outcomes achieved for targeted beneficiaries of the BAB ABE program; how and why outcomes differed across BAB implementation contexts; and differential impacts of BAB based on learner characteristics.

2. How do learning outcomes of diverse learners differ across private, community, and public-school options and interventions in the Country Development Cooperation Strategy (CDCS) focal zones? What is the impact of contextual and demographic indicators on learning outcomes?

   This evaluation question will provide data for benchmarking learning outcomes across educational models co-located with or near BAB schools with the goal of documenting variations in populations served, as well as impacts of educational models based on contextual and demographic variables. Results from this evaluation question can serve as a reference within which the BAB results can be situated and can inform USAID and the Somali governments' decision-making about programming and education models for different populations.

3. What can we learn from the BAB implementation to inform USAID and MoECHE decision-making for scale-up and sustainability?

   The final set of evaluation questions seeks to better understand the interactions among BAB components and outcomes by context, demographic group, and the costs associated with outcomes. This information can inform continuous improvement and decision making regarding what aspects of the BAB program to replicate, scale, and/or integrate this accelerated basic education model into MoECHE policies. These lessons learned may also contribute to future accelerated education programming in Somalia or other crisis and conflict-affected settings. Specifically, these questions examine the costs of implementation, the outcomes across different populations and contexts, the effects of various community and programmatic variations on program effectiveness, and unintended consequences (both positive and negative).

Appendix 1 supplies a detailed list of sub-questions for each evaluation question, as well as target indicators, and data collection and analysis strategies for the full evaluation.

2.2 Overall Evaluation Design

The BAB program external evaluation uses a theory-based, program-oriented evaluation approach with a mixed-methods design that employs a range of both qualitative (e.g., interviews, focus group discussions (FGD), and case studies) and quantitative (e.g., surveys, standardized assessments, program data) data collection tools to draw conclusions about project outcomes and impacts over time. **The evaluation's multi-measure longitudinal cohort design follows two cohorts (BAB ABE and a reference cohort from co-located or nearby non-BAB programs) of learners for two years.**

Cross-sectional cohort data collected by Creative Associates from a unique random sample of BAB learners at the beginning and end of each program year will provide additional opportunities for analysis and model validation.

Evaluation of both BAB and non-BAB cohorts will examine changes in key indicators and outcomes over time with primary data collection at baseline (September-October 2021), midline (May-June 2022), and endline (May-June 2023). The evaluation will track learning (literacy and numeracy), psychosocial measures, and retention for all cohort learners. Further, recognizing that a host of factors (both programmatic and non-programmatic) affect learner outcomes, the external evaluation examines the influences of learner, teacher, school, family, and community characteristics and context on learner outcomes.

2.3 BASELINE EVALUATION DESIGN

The external evaluation created a longitudinal cohort sample by selecting a random sample of BAB level 1 and non-BAB grade 1 and grade 2 learners to provide historical points of reference or starting points for measuring key project indicators that will form the basis for later stage examinations of BAB and non-BAB educational impacts/outcomes at midline and endline. The baseline evaluation:

- establishes baseline values for project indicators and serves as a reference for calculating growth in learning outcomes after one and two years of BAB and formal school instruction;
- identifies/verifies contextual factors that may affect outcomes; and
- informs sample selection and program implementation strategies.

3 METHODOLOGY

The external baseline evaluation focused on describing learner, teacher, and head teacher characteristics for both BAB and non-BAB components of the longitudinal cohort; testing the quality and performance of data collection tools; and examining expected correlational associations between variables of interest.

3.1 DATA COLLECTION APPROACH

Data for this baseline evaluation comes from two sources, primary longitudinal cohort data collected by the external evaluation team (teacher, head teacher, and learner surveys – including EGRA and EGMA) and secondary data collected by Creative Associates (including cross-sectional cohort learner data). Community assessment and household survey findings, collected by Creative Associates as part of their site selection process, provided contextual background.

Longitudinal Cohort Sample

The external evaluation uses a multiple measure longitudinal cohort design to examine learner growth over time. This design will allow us to document changes in learning outcomes for individual learners through two years of programming. By following individual learners over time, the external evaluation will be able to examine differential impacts of the intervention -- both BAB and non-BAB -- (including learning outcomes, retention, and dropout rates) based on gender, age, SES (Social Economic Status), location, and other relevant demographic and contextual factors.

The evaluation will follow the BAB cohort from entry into the BAB program as level 1 learners through completion of the level 2 curriculum (equivalent to grade 4).

The non-BAB reference cohort consists of two sub-samples. Sub-sample A will follow learners from entry into grade 1 through the end of grade 2, while sub-sample B will follow a different group of learners from the beginning of grade 2 to the end of grade 3.

Including both first and second graders at baseline extends the value of the reference cohort by supplying benchmarking opportunities for first time learners at baseline, as well as learning growth across one and two years of both standard and accelerated instruction. As ABE aims to compress two years of the basic education curriculum into a single year, the reference cohort provides both a mechanism for examining similarities and differences between learners enrolled in standard instruction and those enrolled in ABE at baseline and an opportunity to examine differences in rate of learning growth for learners in each group. The evaluation will examine non-BAB subgroups separately to better understand potential impacts of age/experience on constructs of interest.

The external evaluation used an independent sampling plan to select a random sample of BAB classes for the evaluation longitudinal cohort study that is distinct from the BAB cross-sectional sample (See Appendix 2). The longitudinal cohort sampling protocol focuses on all children in selected classrooms. The evaluation will compare data from the independent longitudinal study with BAB's cross-sectional data for a more robust understanding of the impact of sampling on project outcomes across cohorts (see section 7 for baseline comparisons and sampling differences).

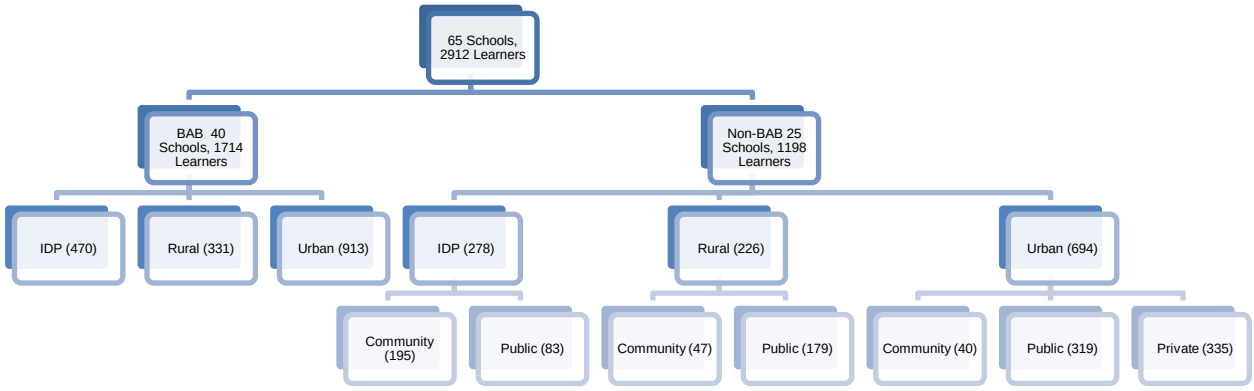
To create a longitudinal study cohort for non-BAB schools (public, private, and community), the external evaluation team randomly selected a formal school classroom from each site hosting a BAB longitudinal sample classroom. In cases where a BAB longitudinal site did not host a non-BAB grade 1 or grade 2 class, the evaluation team, in consultation with the district education officer and state/regional authority, identified a non-BAB sample from a nearby school in the same community. This sampling scheme simplified data collection logistics and minimized variations associated with community characteristics and differences in physical facilities associated with schools. However, it may introduce spillover effects resulting from BAB activities with teachers, schools, and the community. External evaluation midline and endline data collection will explore both sources of spillover and impacts.

Based on sample size calculations and oversampling considerations to account for expected attrition over the 2-year study,[3] external evaluation baseline data collection targeted an initial sample of 1200-1500 BAB learners representing all learners in 40-50 classes, and 1000-1200 non-

[3] The external evaluation used attrition rate calculations of 20% over the course of the study based on reported attribute rates from similar studies in Somalia and Somaliland (Machova, Miettunen, and Peterson, 2020, Leave No Girl Behind: Adolescent Girls' Education in Somalia (AGES) Baseline Evaluation, Consilient Research pg. 51)

BAB learners, representing all learners in 30-40 classes. Actual external evaluation baseline data collection exceeded targets, with data collected from 2,912 learners from 65 schools across 4 states and 11 districts. As illustrated in Figure 2, the baseline sample included 1,714 learners from 40 BAB schools and 1,198 learners from 25 non-BAB schools. Both BAB and non-BAB sample schools represented IDP, rural, and urban settings with non-BAB schools further encompassing community, public, and private school learners.

Figure 2: Longitudinal Cohort Baseline Sample Distribution



The external evaluation sampling scheme used a site-based sampling approach that purposefully oversampled rural and IDP classrooms recognizing that rural and IDP areas in Somalia have fewer educational resources and so more OOSCY (as a percent) than urban areas (Federal Government of Somalia Ministry of Education, Culture and Higher Education 2018). Table 1 shows target and actual longitudinal classroom distributions for the baseline sample.

Table 1: Longitudinal Cohort Classroom Distribution

Community Type	Number of classrooms		
	Target BAB	Actual BAB	Actual non-BAB
IDP	12	11	4
Rural	8	8	5
Urban	20	21	16

TOTAL	**40**	**40**	**25**

Cross-sectional Cohort Sample

The BAB MEL plan employs a random cross-sectional data collection approach that draws a **unique** set of Level 1 learners **for each data point** (beginning and end of each program year) to monitor program effectiveness at the population level. The BAB sampling protocol draws 10-12 learners per classroom to maximize the number of classrooms in the sample.

Creative Associates selected a random cross-sectional sample of 1208 learners across 92 schools from the same 4 states and 11 districts. The cross-sectional sample includes only BAB learners from urban, rural, and IDP locations. The evaluation team and BAB coordinated sample selection to prevent duplicate testing of individual learners by both the evaluation longitudinal study and BAB cross-sectional samples.

3.2 DATA COLLECTION TOOLS

The external evaluation conducted baseline longitudinal learner surveys at the sample school sites using a tablet, smartphone, or other electronic device running Tangerine™ electronic data collection software (RTI International, 2018). Longitudinal learner baseline surveys included demographic and psychosocial indicators, as well as measures of learner literacy and numeracy skills (Early Grade Reading Assessment -- EGRA, Early Grade Math Assessment -- EGMA). Trained enumerators read learner questions orally and recorded responses electronically. Enumerators also conducted teacher and head teacher surveys orally and recorded responses electronically using Tangerine™ software. In-country experts from SORDI led all external evaluation primary data collection activities, including recruiting and training both male and female in-country research associates to conduct baseline learner and teacher data collection activities.

3.3 INSTRUMENTS AND INSTRUMENT DEVELOPMENT

The external evaluation baseline longitudinal cohort primary data collection focused on survey and learning data from learners and survey data from teachers and head teachers. Learner surveys examined four latent constructs (socio-economic status, safety, engagement, and quality of life) as well as learner performance on numeracy and literacy skills.

Building on BAB's log frame, the external evaluation team identified latent constructs for learner baseline data collection from a desk review of prior educational work in a Somali context and educational theory. Prior research supports the importance of these constructs, represented in relationship to three primary BAB intermediate results in Figure 3, in supporting positive learning environments. The external evaluation team developed a longitudinal learner survey based on validated scales to examine these proximal indicators that may impact learning, including changes in learner attitudes and perceptions of education, barriers to education, perceptions of safety, and psycho-social indicators including engagement and quality of life indicators.

Figure 3: Simplified Learner Model Showing Relationships Among Key Constructs and Intermediate Outcomes

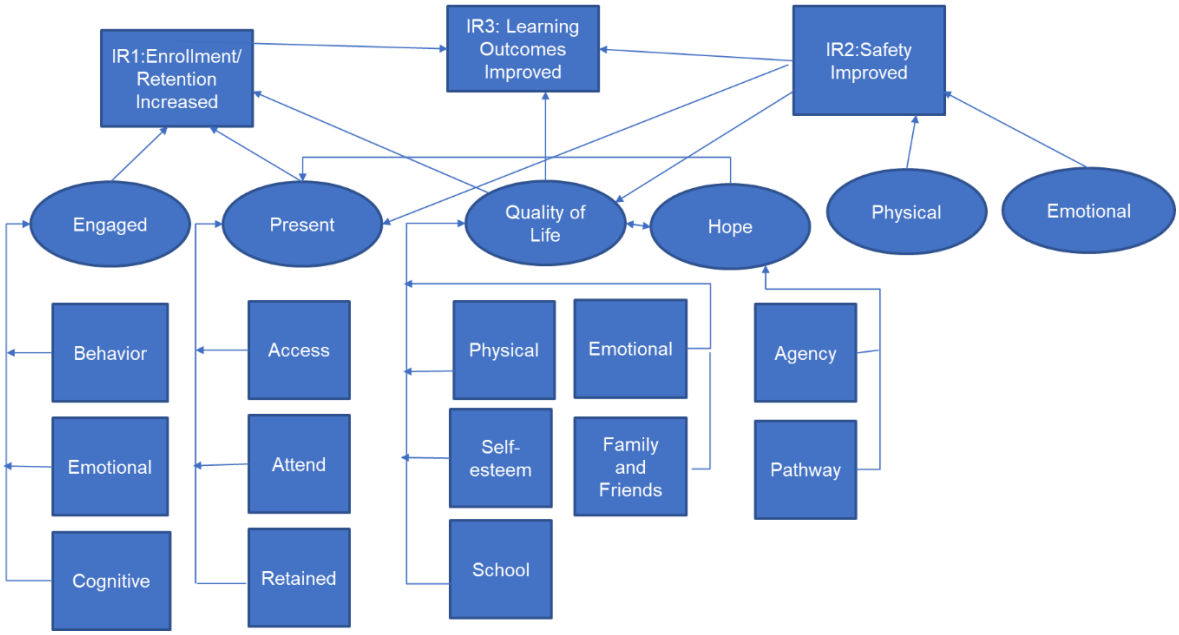


Table 2 describes specific latent variables included in the learner baseline survey and the source for questions/scale items for that variable included in the Longitudinal Learner Survey. The final column in Table 2 shows links between these variables and learning and life competencies included in BAB level 1 and level 2 curricula.

Table 2: Longitudinal Baseline Learner Variables and Measurement Item Sources

Variable	Definition	Measurement Tool	Link to BAB Life and Learning Competencies
Socioeconomic Status (SES)	Social or class standing of an individual or group. Measures often include a combination of education, income, and occupation.	Maternal literacy question (BAB Learner Survey) SES Items from ISELA and BAB Learner Survey	
Safety	Physical and emotional deriving from any combination of environmental, internal,	Perceptions of safety -- Questions from various sources	

	and external threats to well-being.		
Engagement	Behavioral --- access, attendance, and retention (persistence). Emotional -- the extent to which children feel they have value and belong in school. Cognitive – the level of intellectual effort a learner devotes to mastering tasks.	School Engagement Scale (Fredricks, et. al.; 2005) 14 items. Only Emotional Scale items are included at baseline.	Cognitive Domain Emotional Domain
Quality of Life	Perception of well-being and functioning, including self-esteem, relationships with family and friends, and attitudes towards school and the future.	KINDL Quality of Life Survey (Ravens-Sieberer & Bullinger (1998) Three subscales included at baseline (Emotional**,** **S**elf-Esteem, Friends)	Social Domain Emotional Domain Values Domain
Learning	Numeracy and Literacy	EGMA and EGRA Baseline longitudinal data collection used the EGRA and EGMA adapted by Creative Assoc. and used with their cross-sectional sample.	Learning

The evaluation team administered the learner survey to both BAB and non-BAB longitudinal learner cohorts at baseline. An expanded learner survey will be administered at midline and endline to examine changes in learner experiences and perspectives regarding academic, physical, emotional, and social need satisfaction.

The external evaluation incorporated the Level 1 Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA) instruments in the longitudinal learner survey. BAB adapted these instruments for the Somali context with the support of the MoECHE to measure learner outcomes in their cross-sectional sample. The evaluation will also use the BAB level 1 EGRA and EGMA at midline to facilitate comparisons across educational models and between the

cross-sectional and longitudinal samples. The evaluation team will develop new EGRA and EGMA tools for BAB longitudinal cohort Level 2 data collection to enable measurement of learning across 2 years of BAB ABE instruction.

The external evaluation also collected baseline information about teachers' background, training, professional attitudes and beliefs, self-efficacy, well-being, confidence, and feelings of support with a survey instrument developed using validated scales. The longitudinal teacher survey, developed using the same process as the learner survey. The evaluation will employ an expanded version of this survey at midline and endline to 1) capture changes in key teacher constructs of interest, and 2) gain teacher perceptions of learner, curricular, school, family, and community characteristics that may affect their learners' achievement.

The learner and teacher surveys draw from existing validated instruments. The evaluation team checked survey reliability by conducting item analyses (Factor Analysis and estimation of Cronbach's Alphas) for each construct of interest to verify that survey items function as intended in the Somali context. Results of the item analysis showed overall good reliability and will inform minor modifications to midline and endline surveys to ensure a strong basis for valid inferences and interpretations of our results (See Appendix 3 for a complete discussion of survey validity and reliability).

Community characteristics, including population attributes, educational infrastructure, community support, attitudes and beliefs about education, safety concerns, and other attributes that might affect learner recruitment, retention and success were touched on in the Head Teacher Survey and collected as part BAB's community assessment used for site selection purposes and summarized by SORDI researchers for the baseline report.

3.4 Enumerator Training, Field Testing, and Quality Assurance

SORDI researchers, with support from RAN (Resilient Africa Network), recruited, trained, and oversaw experienced Somali enumerators to carry out baseline evaluation data collection.

Training



Enumertors practice using data collection software during training. (SORDI, 2021)

Enumerator training used a mixed methodology that combined presentations with extensive practice sessions to develop both theoretical and practical expertise. Overall, the mode of conduct was participatory with the use of reflection sessions, question and answer sessions, and comments by both the facilitators and trainees. Enumerators took part in seven days of training followed by pilot testing. A full description of enumerator training and training schedules are included in Appendix 4.

Field Testing

The last step in enumerator training included a pilot test that allowed enumerators to practice their skills in the field with children and teachers that represented the target audience. Enumerators conducted the pilot test in non-BAB primary schools that were not among the evaluation sample schools but fit the category of examination. Pilot testing took place on the 9th and 10th of October 2021, in Mogadishu, Kismayo, and Baidao. Prior to deployment, supervisors divided enumerators into groups with team leaders. Senior supervisors and researchers were present at each pilot site to observe, answer questions, and ensure quality. Enumerators administered final versions of all tools to pilot audiences to determine the average administration time for learners, teachers, and head teachers. The teams found that, on average, a single learner took around 40 minutes, a teacher took 40 minutes, and a head teacher took 50 minutes.

Field-testing revealed the need for some minor adjustments to the surveys in the software application. The evaluation team updated all surveys accordingly and finalized them for data collection activities (Appendix 5).

Data Collection and Field Notes

Enumerators administer oral surveys to each learner cohort participant (SORDI, 2021)



Teams of SORDI enumerators collected all evaluation primary data for this baseline study, including Learner Surveys (including demographics, SES, safety, and psychosocial scales, and EGRA and EGMA assessments), Teacher Surveys, and Head Teacher Surveys. Enumerators, drawn from the sample communities, worked in mixed-gender teams whenever possible. Enumerators collected data from schools in three states (Jubaland, Hirshabelle and Southwest) and the Benadir region, and 11 districts. The original sampling protocol called for selecting non-BAB classes from the same school as the BAB sample classes. However, enumerators found that some BAB schools did not have associated non-BAB classes. When this occurred, enumerators worked with the district education officers and state/regional education authority to identify and access non-BAB schools in the same community. Table 3 lists cohort schools by district with the total number of learners sampled broken down by BAB and non-BAB classes.

Table 3: Longitudinal Cohort Sample by State, District, School and Program

State	Districts/Schools	BAB L1 Learner	% Total L1 Learner	Long. Sample	% Long. Sample	BAB	NON-BAB
Benadir	Deynile (7)	3198	11	299	10	177	12
	Kahada (4)	2374	8	244	8	149	9
	Hamarwayne (1)	647	2	77	3	45	3
	Shibis (1)	1060	4	51	2	34	1
Southwest	Baidoa (10)	3965	13	610	21	383	227
	Barawe (2)	1905	6	140	5	88	5
	Diinsor (2)	1602	5	106	4	40	6
	Walanweyn (3)	2623	9	176	6	152	24
Jubaland	Kismayo (7)	4643	16	650	22	316	334
Hirshabelle	Balcad (4)	3575	12	262	9	160	102
	Jowhar (4)	4309	14	297	10	178	119
Total		**29,901**	**100**	**2912**	**100**	**1722**	**1190**

SORDI enumerators collected longitudinal learner cohort data on a whole class basis, which often required multiple visits to each school to accommodate large classes, classes with disruptions, and absences. Enumerators noted a variety of challenges with the data collection process. Table 4 describes several, along with remediation strategies. Field notes in appendix 6 detail additional insights into data collection challenges.

Table 4: Data Collection Challenges and Remedies, and Recommendations

Challenge	**Assumption**	**Findings**	**Remedy**	**Recommendation**
Class Selection	Sample randomization assumed all BAB classes co-located with non-BAB classes	A number of schools did not include both BAB and non-BAB classes	SORDI enumerators worked with the district education officers and state/regional education authority to identify and access non-BAB schools in the same community	Better communicate with implementing partner to understand site characteristics
Learner Registration	All learners are registered, and registration lists are accurate	Frequent mismatch between rosters and actual learners in the class.	Enumerators updated registration logs and collected data from all learners currently in the class.	Document all learners in class and conduct periodic verifications on learners and registers to minimize discrepancies
Infrastructure and Safety	All classrooms are safe and learning ready for 35-50 learners.	Many schools (especially IDP) had poor infrastructure, poor or insufficient resources for the number of children (e.g., chairs, tables, blackboards), and unsafe conditions.	All learners in the class (regardless of number) were included in the sample. Data collection approaches were modified based on field conditions.	Learner hazards (e.g., open wells) documented and remediated. Target more resources for infrastructure and resources to under-resourced areas.
Environmental Challenges	Schools and classrooms are accessible for data collection	Schools in some areas (e.g., Hirshabelle) were	Data collection was delayed until the area became accessible.	Build time and redundancy into the sample frame to accommodate needed changes.

		inaccessible due to flash floods.		
Security	All schools are located in a safe location	The security situation, especially in some of the remote areas, was volatile. Urban roads were sometimes blocked for security reasons.	Project team communicated with communities prior to data collection to assess and consider safety and security concerns that might affect data collection personnel. The safety office provides frequent updates to data collection teams about current and potential threats.	Increase linkages between BAB staff and communities to foster better information exchange.

Quality Assurance

As the quality of data is a direct reflection of enumerator skill and training, the evaluation team used a multistep process for ensuring enumerator and data quality that encompassed candidate selection, continuous monitoring, and pretesting before deployment, as described below:

a) SORDI initially selected enumerator candidates based on their prior experience in the relevant field, competency, and relevance to the process of data collection.
b) Training facilitators continuously checked the trainees throughout the training to ensure understanding of both the material and the importance of collecting the data ethically and according to protocol.
c) Senior SORDI researchers tested all trainees on the final day of the training to ensure they met all requirements prior to piloting. The test examined the trainee's ability to carry out and follow protocols accurately, adhere to ethical standards, and effectively and efficiently use data collection tools. The test proficiency cut off percentage was 80%. Trainees scoring below the cut off were eliminated from the training program. One candidate in Mogadishu, who could not conduct the assessment as expected and scored less than the cutoff percentage, was eliminated. A second trainee, in Baidoa, was unable to effectively use the data collection tools and was also eliminated.
d) SORDI field supervisors reviewed data in real time to identify and address quality issues during data collection.

The evaluation team also examined all data for quality and integrity post collection. During analysis, the research team noticed large variations in EGRA and EGMA scores (all longitudinal baseline data together) between enumerators. The overall EGRA mean was 82.72 (Figure 4), while the mean of data collected by enumerators ranged from 35.64 to 272.72 (Figure 5); the overall EGMA mean was 38.48 (Figure 6), while the mean of data collected by enumerators ranged from 21.18 to 79.25 (Figure 7). These results raised questions about the potential for data bias resulting from errors or inconsistencies in the way enumerators were using the data collection tools. To test for possible enumerator bias, we used box plots to identify outliers and replaced learner IDs with the corresponding enumerator IDs. Analysis revealed four outliers for EGRA, three collected by enumerator ID 48, and one collected by enumerator ID 91 (Figure 8). All four outliers were remarkably close to the maximum value of the boxplot. Analysis detected no outliers for EGMA (Figure 9). Thus, we concluded that variation is within acceptable levels and evidence of enumerator bias is insignificant.

Figure 4: Longitudinal Cohort Baseline EGRA Mean

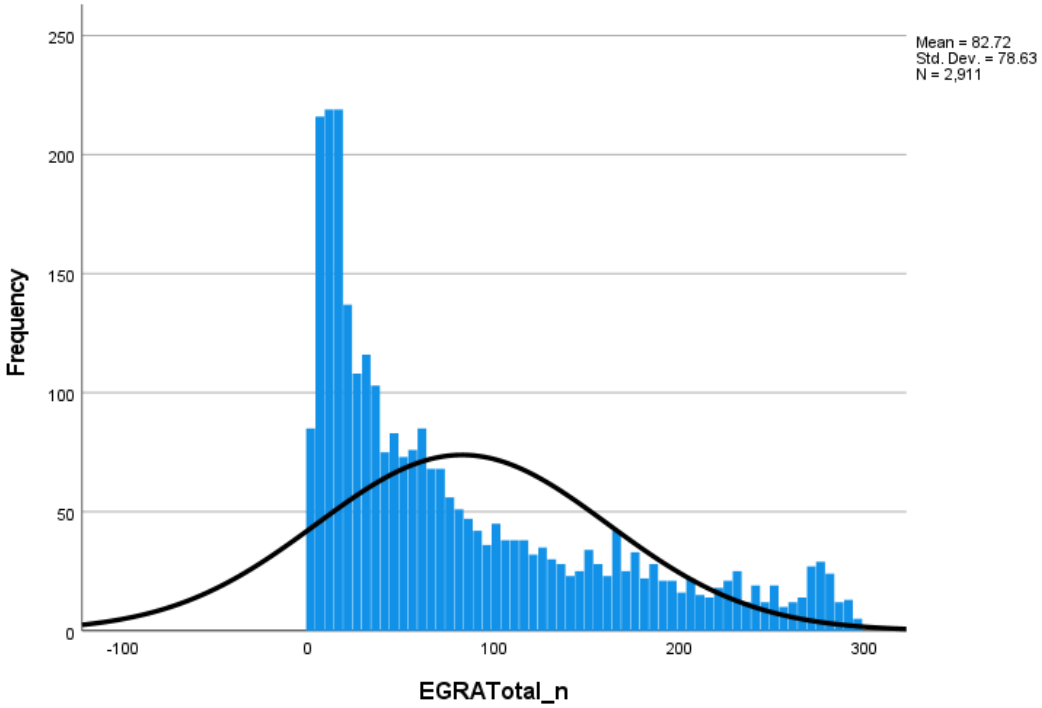


Figure 5: Longitudinal Cohort Baseline EGRA Mean Variation by Enumerator

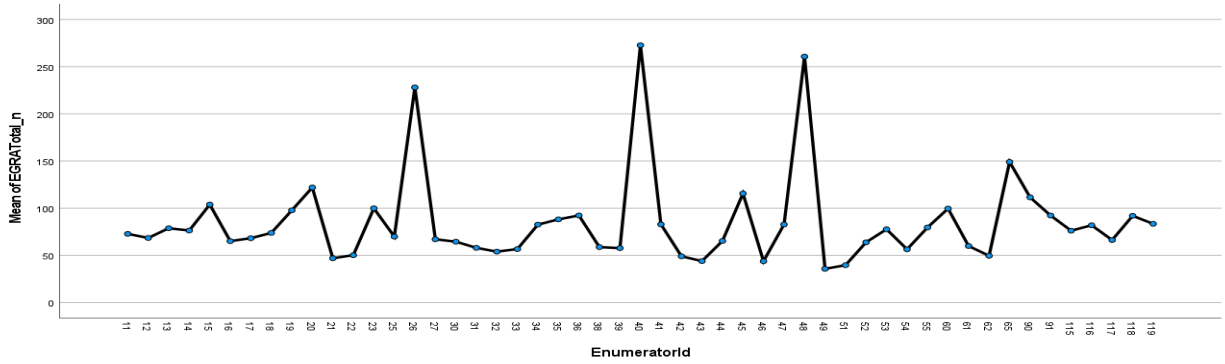


Figure 6: Longitudinal Cohort Baseline EGMA Mean

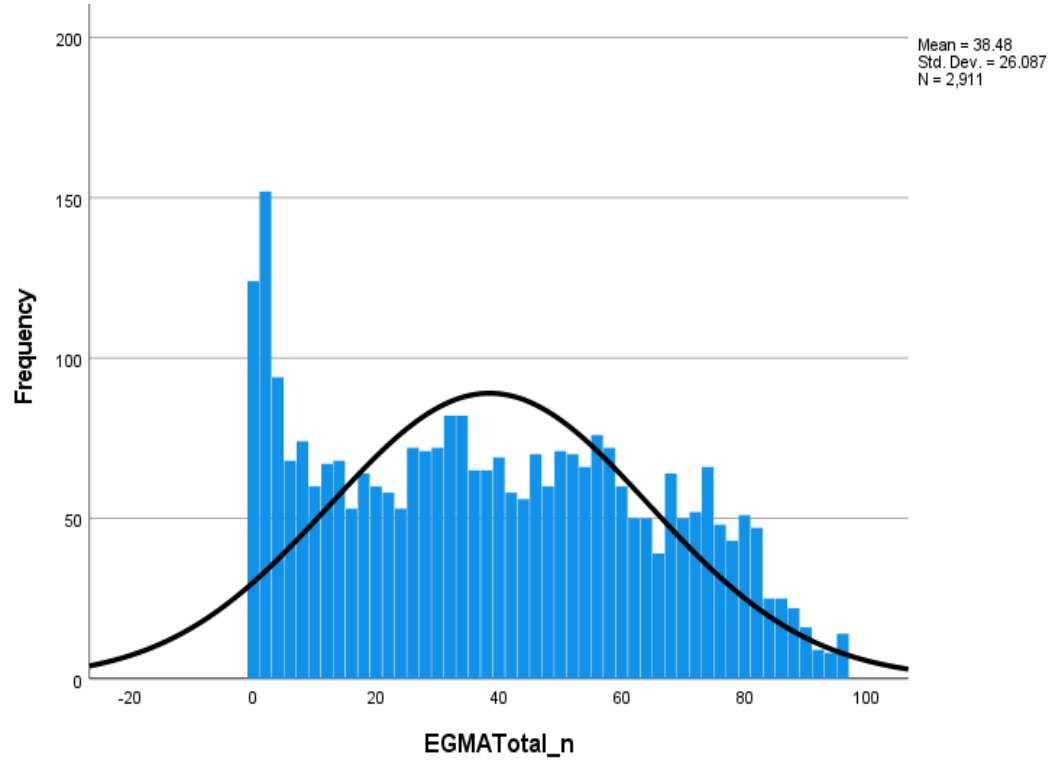


Figure 7: Longitudinal Cohort Baseline EGMA Mean Variation by Enumerator

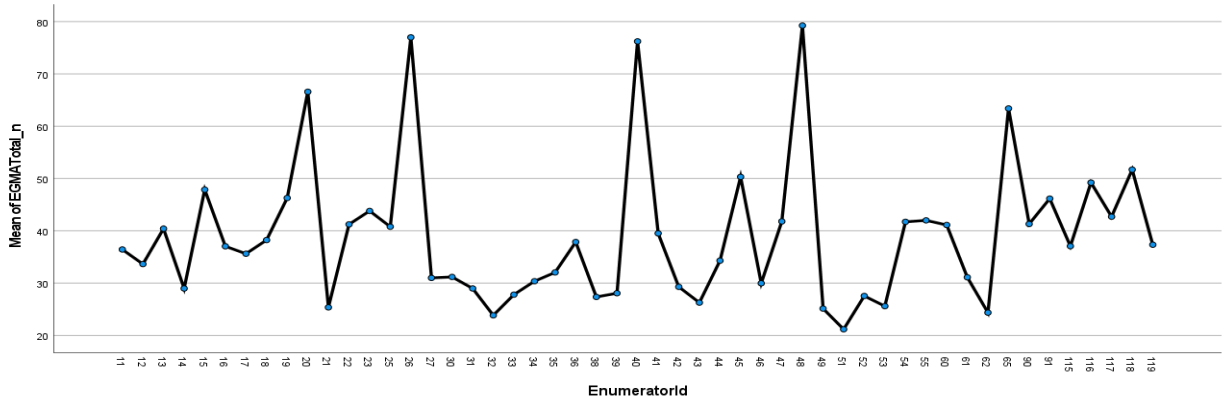


Figure 8: Box Plot Analysis of EGRA Baseline Enumerator Scores

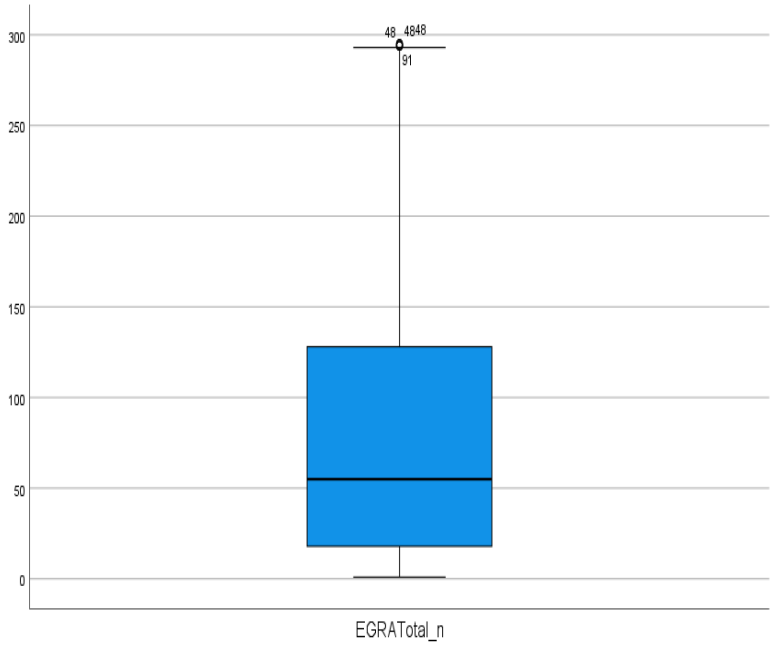
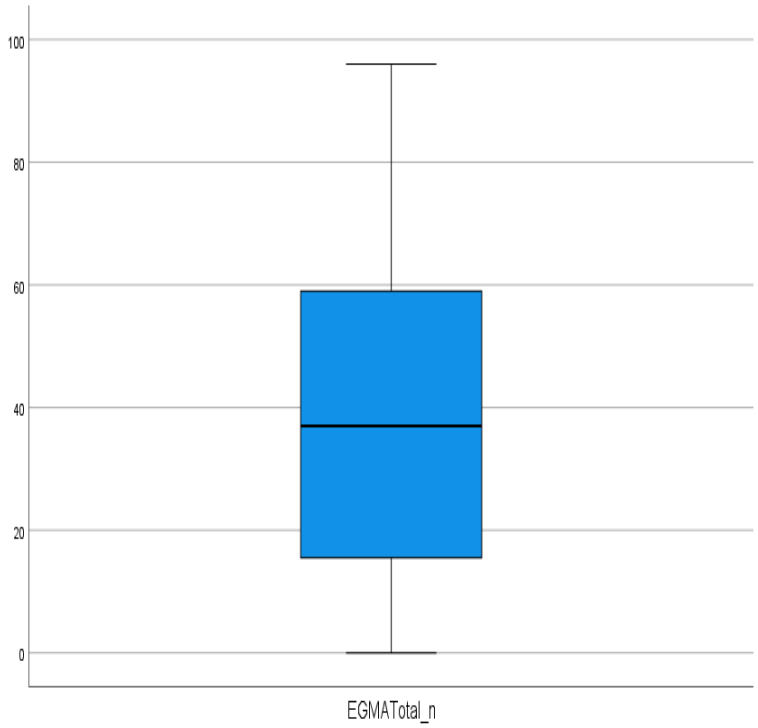


Figure 9: Box Plot Analysis of EGMA Baseline Enumerator Scores



3.5 QUANTITATIVE ANALYSIS

The baseline evaluation focused on quantitative survey data collected from learners, teachers, and head teachers. We used quantitative analysis to describe the sample/population, establishing baseline values for project indicators to serve as references for calculating growth in learning outcomes after one and two years of BAB and formal school instruction and identify the relationship between contextual factors and baseline measures.

We used frequencies, means, standard deviations, graphs, tables, and figures from descriptive analysis to describe, illustrate, or summarize variables. We used factor analysis to develop subscales for each survey and determine their reliability and validity of interpretation in the Somali context. Correlation analysis, cross tabulation/chi-square, t-tests, ANOVA, and regression analysis provide comparisons and show relationships between different subgroups (e.g., girls vs. boys). Density plots show the differential distribution of the same outcome across different settings (e.g., IDP vs. urban vs. rural).

3.6 Cost Analysis

External evaluation of the BAB ABE program includes cost-economy, cost-efficiency, and cost-effectiveness analysis that examines the costs associated with observed gains in educational outcomes of interest (e.g., math and literacy) and the variation of these costs across implementation sites, learning levels, participant characteristics and other constructs of interest. Additional prospective cost analysis will estimate costs associated with expanding the BAB model and/or replicating it in other areas. This information will help future planning and budget management. BAB will provide the evaluation team with activity-related expenditures throughout the project life cycle in line with USAID cost-capture guidance using standard reporting categories for education activities. Cost effectiveness analysis will first calculate cost per output (e.g., cost to train a learning facilitator) and the effectiveness of the intervention (e.g., improvement in math literacy outcomes per facilitator trained).

4 Baseline Findings -- Learners

4.1 Learner Sample Characteristics

The longitudinal baseline sample purposefully oversampled rural and IDP learners to ensure adequate representation of these learners at midline and endline for this repeated measure design. Thus, the baseline sample over-represents learners from IDP/Rural locations. Table 5 illustrates the difference between total population and longitudinal sample population. While increased representation of rural and IDP learners in the sample increases sensitivity of the results for these populations, it skews average results (see discussion in section 4.2). Further, the evaluation drew both the BAB and non-BAB samples only from those areas currently served by the BAB program. Therefore, results are not generalizable at the national level or to other geographical areas. The longitudinal sample included 65 school sites. Of these sites, 16 BAB sites had non-BAB classes co-located. For sites without non-BAB schools, SORDI field coordinators and supervisors reached out to district education officers (DEOs) and the state/region level education authority to provide a list of schools with similar socio-economic characteristics in the area. SORDI randomly selected non-BAB schools and classes from the list of schools and contact information provided. This approach maintained the representativeness of the sample.

LONGITUDINAL SAMPLE

Total Schools	65
BAB School	40
Non-BAB Schools	25
Total Learners	2912
BAB Learners	1714
Non-BAB Learners	1198
Non-BAB grade 1	620
Non-BAB grade 2	578

Table 5: Longitudinal Baseline BAB Learners Sample Vs. Total BAB Population

Location	Longitudinal Baseline Sample		Total BAB Population	
	# of learners	# of sites	# of learners	# of sites
IDP	470 (27.4%)	11 (27.5%)	2659 (8.9%)	52 (8.7%)
Rural	331 (19.3%)	8 (20.0%)	4101 (13.7%)	80 (13.4%)
Urban	913 (53.3%)	21 (52.5%)	23141 (77.4%)	466 (77.9%)
Total	1714 (100.0%)	40 (100.0%)	29901 (100.0%)	598 (100.0%)

4.2 SAMPLE DISTRIBUTION

As discussed previously, and described in detail in Appendix 2, external evaluation of the BAB project employs a longitudinal cohort derived using a stratified sampling approach that ensures accurate sampling and proper representation across the population (including rural, urban, and IDP locations). The external evaluation purposefully oversampled rural and IDP locations to mitigate expected challenges in these areas due to population mobility, access, and other considerations. The sampling framework also considered state and district distribution, as well as funding type for non-BAB schools (community, public, or private). The following tables show longitudinal sample breakdown for each program (BAB or non-BAB) by school funding type (Table 6), location type (Table 7) school and learner counts by state, location, and program type (Table 8), location and school funding type for non-BAB learners (Table 9) and breakdown of first and second grade learners by location type for non-BAB learners (Table 10).

Table 6: Longitudinal Study Learner Counts by School Funding Type

Funding Type	School Total	Learner Total
BAB	40	1,714
Community	6	282
Public	10	581
Private	9	335

Table 7: Longitudinal Study Learner Counts by School Location Type

Location Type	BAB Schools	NonBAB Schools	Total Schools	BAB Learners	NonBAB Learners	Total Learners
IDP	11	4	15	470	278	748
Rural	8	5	13	331	226	557
Urban	21	16	37	913	694	1,607

Table 8: Longitudinal School and Learner Counts by State, Location and Program Type

State	Location Type	Program Type	School Subtotal	School Total by Location	School Total by State	Learner Subtotal	Learner Total by Location	Learner Total by State
Benadir	IDP	BAB	4			169		
	IDP	Non-BAB	2	6		128	298	
	Urban	BAB	6			236		
	Urban	Non-BAB	5	11	17	137	373	671
Hirshabelle	Rural	BAB	4			166		
	Rural	NonBAB	2	6		50	216	
	Urban	BAB	4			168		
	Urban	NonBAB	4	8	14	175	343	559
Jubaland	Rural	BAB	2			82		
	Rural	NonBAB	2	4		152	234	
	Urban	BAB	5			234		
	Urban	NonBAB	4	9	13	182	416	650
Southwest	IDP	BAB	7			301		
	IDP	NonBAB	2	9		149	450	
	Rural	BAB	2			83		
	Rural	NonBAB	1	3		24	107	
	Urban	BAB	6			275		
	Urban	NonBAB	3	9	21	200	475	1,032

Table 9: Non-BAB Program Learner Counts by Location and School Funding Type

Location Type	Funding Type	School Subtotal	School Total	Learner Subtotal	Learner Total
IDP	Community	3		195	
	Public	1	4	83	278
Rural	Community	2		47	
	Public	3	5	179	226
Urban	Community	1		40	
	Private	9		335	
	Public	6	16	319	694

Table 10: Non-BAB Program Learner Counts by Location and Grade Level

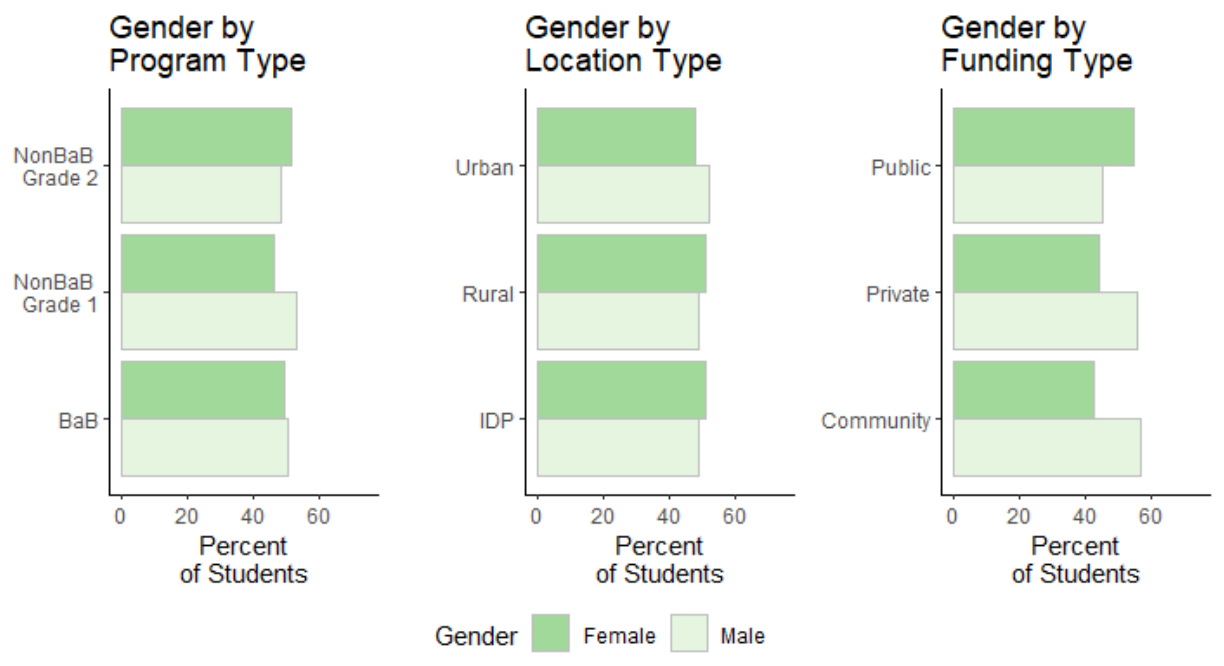
Location Type	Grade Level	School Subtotal	Learner Subtotal	Learner Total
IDP	Grade 1	4	140	
	Grade 2	4	138	278
Rural	Grade 1	5	144	
	Grade 2	3	82	226
Urban	Grade 1	13	336	
	Grade 2	15	358	694

4.3 LEARNER DEMOGRAPHIC CHARACTERISTICS

Gender Disaggregation

Gender balance varied by program, location, and school funding type (Figure 10). Overall, the BAB sample included an almost even number of girls and boys (in-line with BAB reported female enrollment of 48%), while the non-BAB classes had slightly more boys at baseline in grade 1 and more girls in grade 2. Gender differences by location are relatively small, with slightly more boys overall in urban classes and more girls in rural and IDP classes. The largest gender differences occurred when disaggregating data by school funding type. While the BAB classes were gender balanced, public schools enrolled significantly more girls than boys, while the converse was true for both private and community schools, where boys predominated.

Figure 10: Baseline Longitudinal Sample Distribution by Gender



Age Disaggregation

Sample disaggregation by age (Figures 11-13) shows similarly broad age distributions for learners enrolled in BAB level 1 and for learners enrolled in grades 1 and 2 at public, private and community schools -- ranging from below 5 years of age to over 19. The median age for BAB level 1 learners at baseline was 11 years old, falling between the medians for non-BAB learners in grade 1 (10 yrs.) and grade 2 (12 yrs.).

There was no difference in median learner age for the combined BAB and non-BAB longitudinal baseline cohort across location types, although the urban sample showed less age variation than the IDP and rural samples. The median age for learners in community schools was slightly lower than BAB, public, and private school samples (Figure 14).

Figure 11: Baseline Longitudinal BAB Cohort Distribution by Age

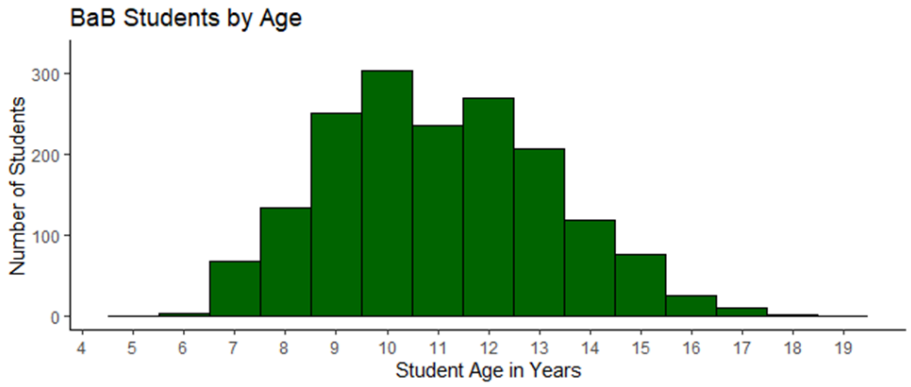


Figure 12: Baseline Longitudinal non-BAB Grade 1 Cohort Distribution by Age

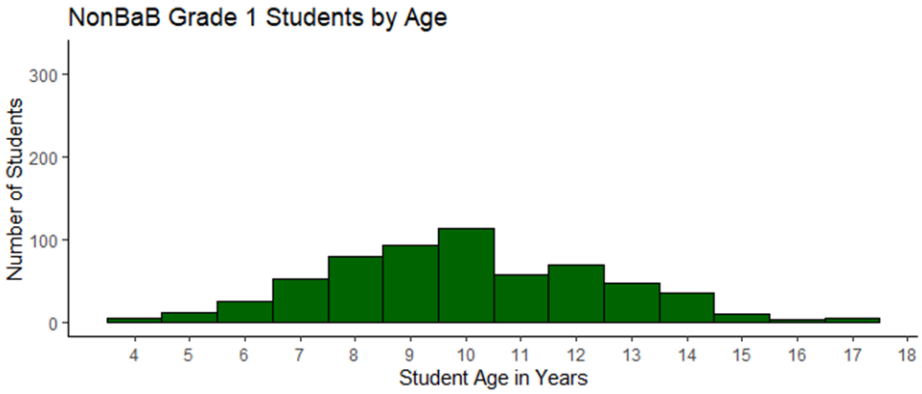


Figure 13: Baseline Longitudinal non-BAB Grade 2 Cohort Distribution by Age

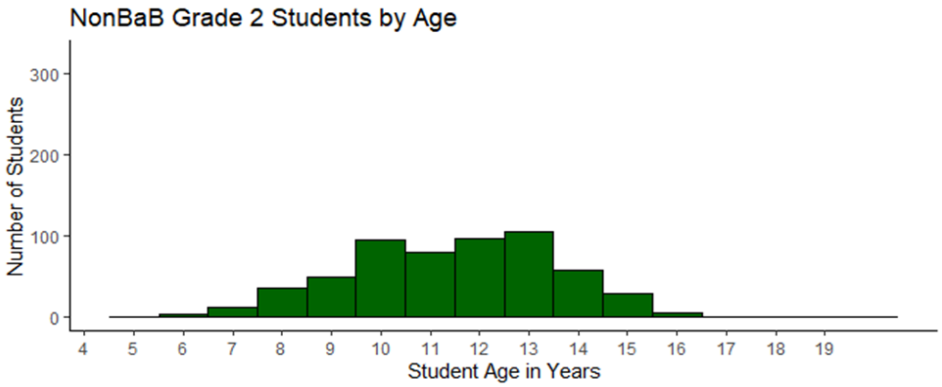
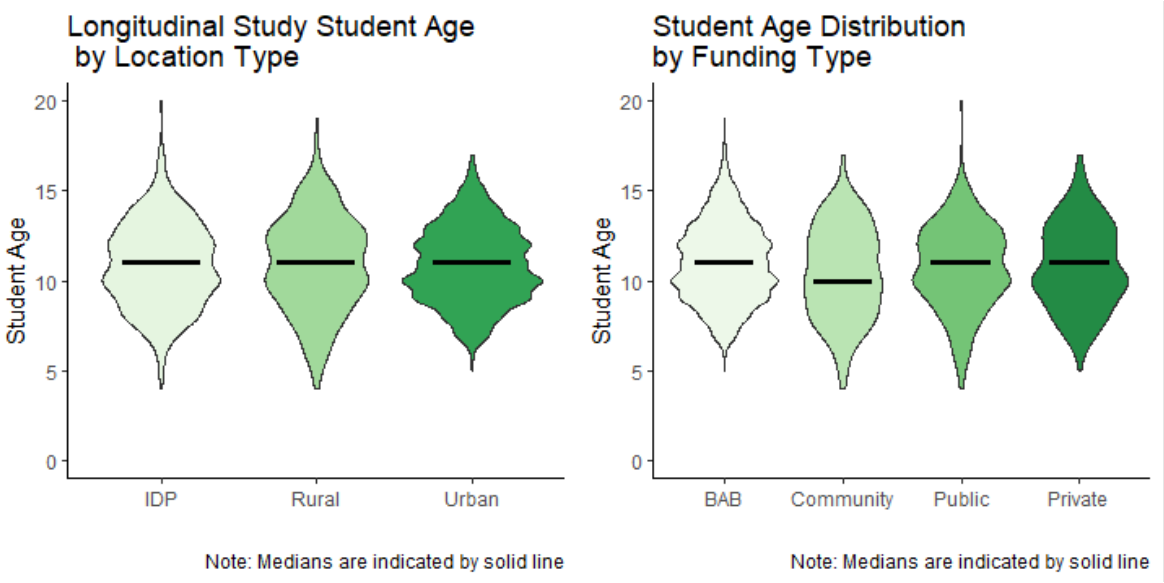


Figure 14: Baseline Longitudinal Sample Age Distribution by Location and Program Funding Type



Prior Educational Experiences

While many Somali children and youth have little or no formal academic experience, it is common for Somali families to seek religious instruction for their children through attendance at Qur'anic School. As illustrated in Figure 15, more than 80% of learners in the longitudinal cohort sample responded in the affirmative when asked about Qur'anic School attendance. Results for this question show insignificant variation across program, location, or school funding type. Similarly, the graph in Figure 16 shows that most of the learners in the sample are still attending Qur'anic school with little variation amongst the program, location, or school funding type. The learner survey also asked longitudinal cohort sample learners if they had attended any other type of school prior to their BAB enrollment. Figure 17 show that fewer than 30% of learners report school attendance other than Qur'anic prior to the current enrollment.

Figure 15: Baseline Longitudinal Learner Prior Experience with Qur'anic School by Program, Location, and Funding Type

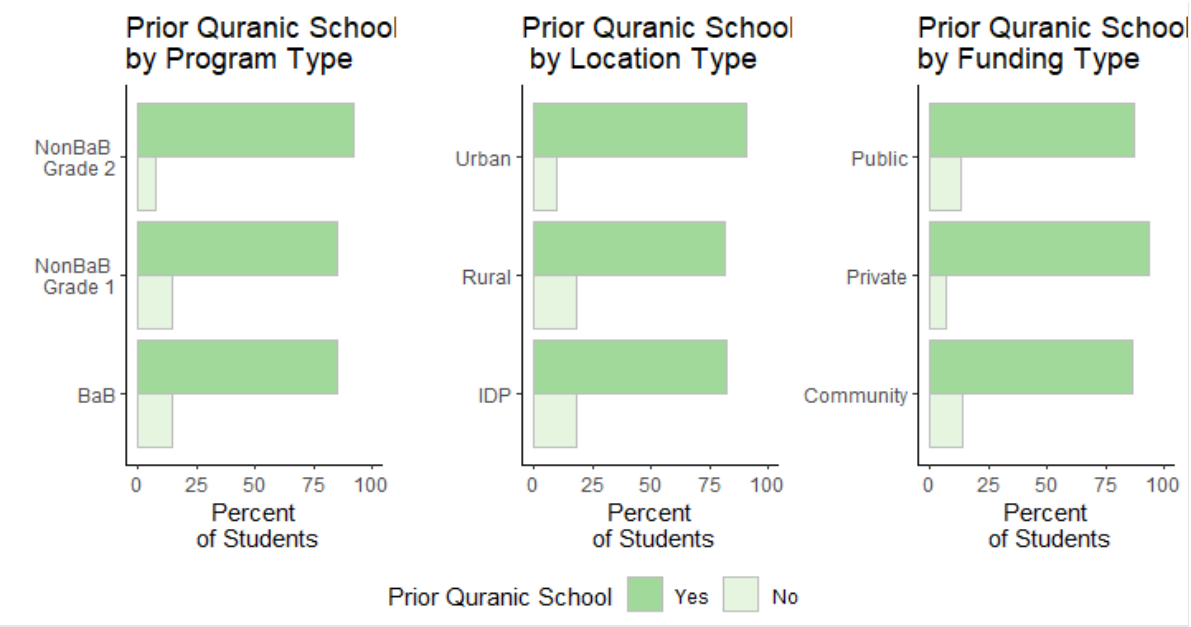


Figure 16: Baseline Longitudinal Learner Still Attending Qur'anic School by Program, Location, and Funding Type

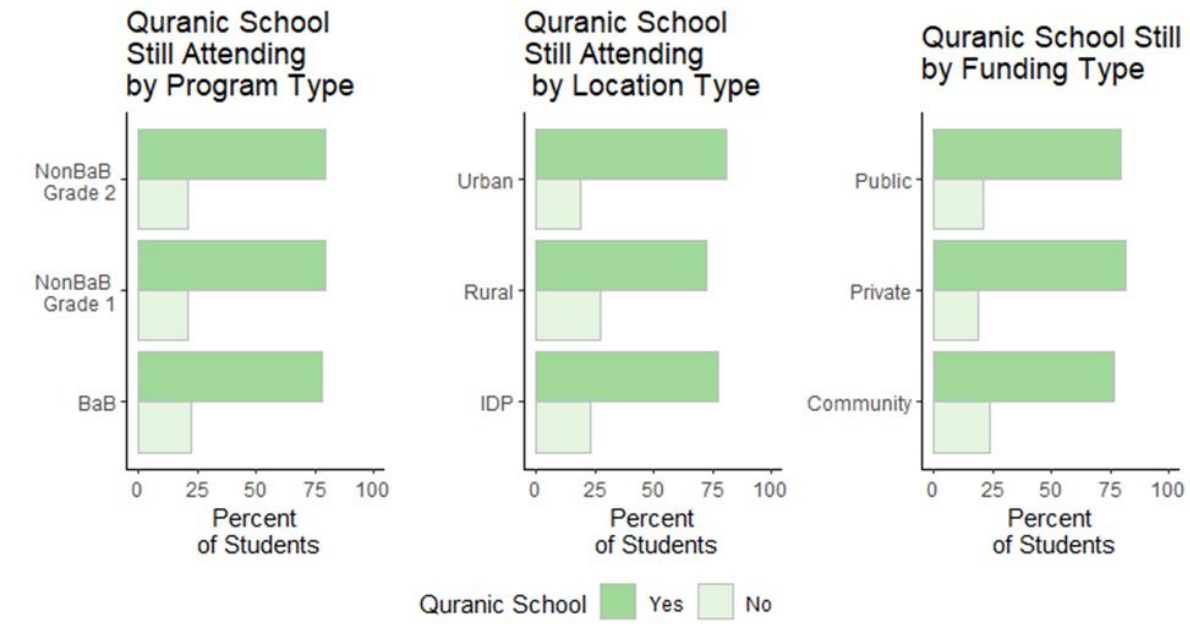
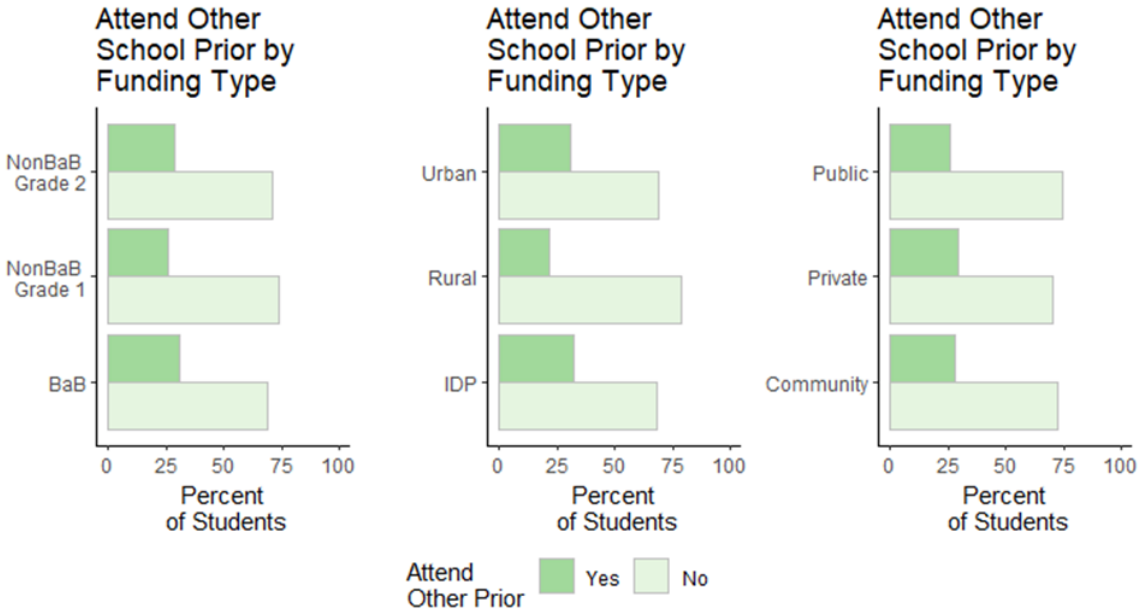


Figure 17: Baseline Longitudinal Learner Prior Other School Attendance by Program, Location, and Funding Type



Summary of Learner Demographic Characteristics

The longitudinal evaluation cohort showed little variation between BAB and Non-BAB learner samples at baseline on learner characteristics including age distribution, gender distribution, and prior school attendance. Both the accelerated program learners and learners attending non-accelerated schools varied in age from below 5 to above 19 years of age with median ages for grade 1, BAB ABE learners, and grade 2 at 10, 11, and 12 years, respectively. While overall, both the BAB and non-BAB cohorts showed gender balance, disaggregation by school funding type revealed higher enrollment of male learners in community and private schools, while public schools enrolled more females than males and BAB schools were gender balanced.

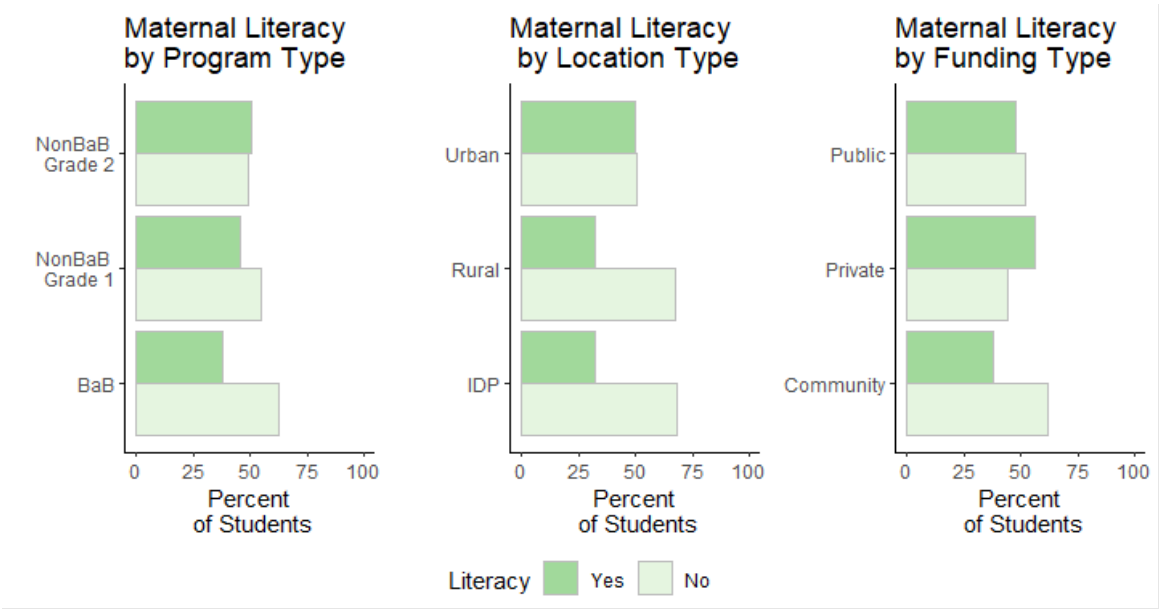
4.4 EDUCATIONAL MARGINALIZATION (SES INDICATORS)

Socio-Economic Status (SES) is widely correlated with educational outcomes. Children from lower SES backgrounds typically underperform compared to more affluent peers (Sirin 2005). Indicators of SES in the context of this study include maternal education level, dialect spoken at home (as an indicator of learner's clan affiliation), and home resources (e.g., electricity, indoor plumbing, radios, mobile phones, books). The evaluation sourced SES questions for the longitudinal survey from the BAB learner survey and the ISELA.

Maternal Literacy

Maternal education, recognized as a key element of socioeconomic status, is strongly associated with predictors of children's well-being and cognitive development (Jackson, Kiernan and McLanahan 2017), and school retention (Kamanda, Madise and Schnepf 2016). The learner survey included a question asking learners to indicate if their mother is literate (Can your mother read and write?). Responses to this item show significantly lower rates of reported maternal literacy for BAB learners compared to non-BAB learners. Nearly 75% of learners in IDP and rural schools responded that their mother could not read, as compared to about 50% of the learners from urban schools. Finally, private school learners reported the highest levels of maternal literacy (over 60%) and community schools the lowest (less than 40%) (Figure 18).

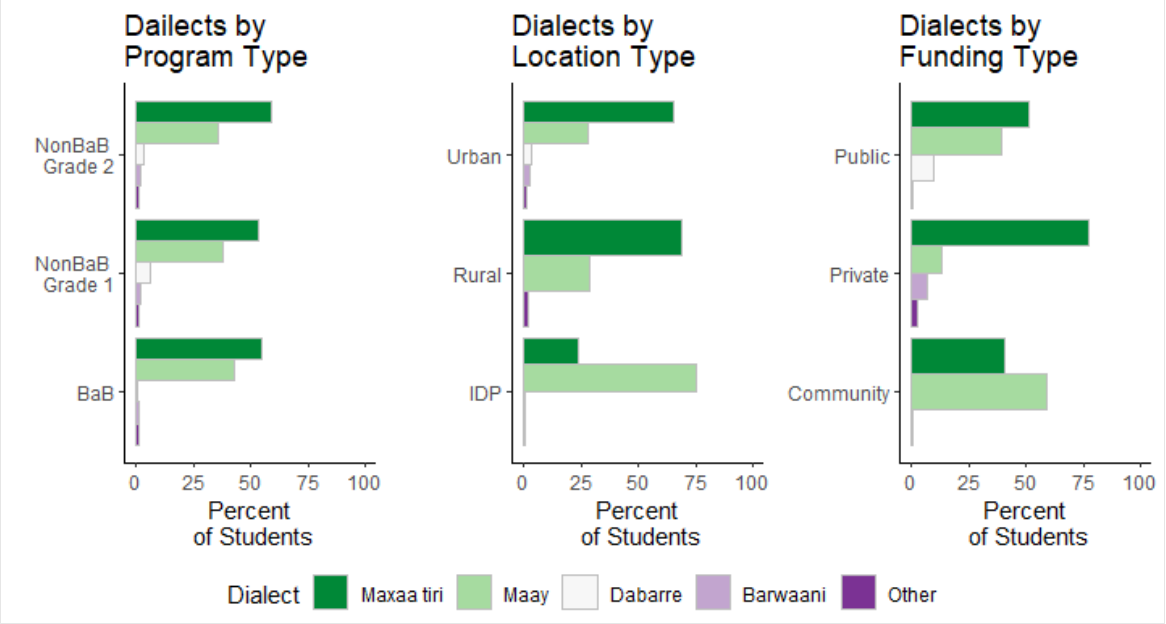
Figure 18: Baseline Longitudinal Sample Maternal Literacy by Program, Location, and Funding Type



Dialects Spoken

In Somalia, the dialect learners speak at home is an indicator of clan, social status, and economic status. Maxaa tiri is the predominant dialect and Maay the second most common dialect spoken in both the BAB and non-BAB samples. The diversity of dialects spoken is slightly lower in the BAB sample than the non-BAB sample. Prevalence and diversity of dialects spoken in a school varied by both school location and school funding type. Maxaa tiri was the dominant dialect spoken in both urban and rural schools, with Maay more common in IDP schools. Community school learners spoke Maay more often than Maxaa tiri, while over ¾ of private school learners speak Maxaa tiri (Figure 19).

Figure 19: Baseline Longitudinal Sample Primary Dialect by Program, Location, and Funding Type



Family Resources

The Longitudinal Learner Survey also includes four questions that together make up a scale intended to measure home resources as a proxy for SES. Each item was a no=0 and yes=1 question with a higher value indicating a higher perceived SES level. The scale comprises the summation of the following 4 items with a range of 0 (none of the 4 resources are present in the home) to 4 (all 4 resources are present in the home):

- Do you have electricity at home?
- Do you have a radio at home?
- Does your house have an indoor bathroom/toilet?
- Does your house have a telephone or mobile phone?

Learners attending the BAB program averaged statistically fewer (2.1) resources compared to non-BAB learners (2.36), as illustrated in Figure 20. We noted significant differences for both BAB and non-BAB learners based on location, with IDP learners significantly under-resourced (1.53 BAB/1.67 non-BAB) compared to rural learners (1.82 BAB/1.92 non-BAB) and urban learners (2.49 BAB/2.78 non-BAB). Older BAB learners (12-20 yrs.) reported significantly fewer resources than their younger (4-11 yr. old) counterparts did, but we found no differences by age for the non-BAB group. Finally, family resources varied significantly for both BAB and non-BAB learners based on state, with learners in Jubaland scoring lowest with just 1.54 (BAB) or 1.81 (non-BAB), while learners in Hirshabelle reported the highest number of resources, with 2.71 (BAB) or 3.17 (non-BAB). Figure 20 illustrates the difference in SES for BAB and non-BAB students. Table 11 summarizes SES differences by

location, gender, age, and state for both BAB and non-BAB students. The Data Annex (Section 2.2; 8.1) includes additional figures.

Figure 20: Baseline Longitudinal Sample SES by Program Type

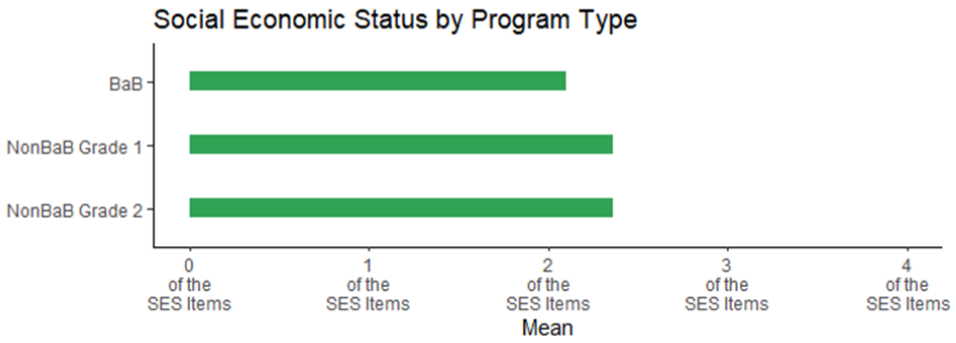


Table 11: Baseline Longitudinal Sample SES by LocationType, Gender, Age, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Students Social Economic Status			**2.10**	**1.19**	
Location Type	*	IDP	1.53	0.91	-0.57
		Rural	1.82	1.16	-0.28
		Urban	2.49	1.19	0.39
Gender		Male	2.13	1.19	0.03
		Female	2.06	1.19	-0.03
Age	*	4 to 11 year old	2.16	1.21	0.06
		12 to 20 year old	2.01	1.16	-0.08
State	*	Benadir	2.19	1.03	0.09
		Hirshabelle	2.71	1.22	0.61
		Jubaland	1.54	1.24	-0.56
		Southwest	2.00	1.10	-0.09
Group		**Subgroup**	Mean	SD	Diff from Overall Mean
All Non-BAB Students Social Economic Status			**2.36**	**1.25**	
Location Type	*	IDP	1.67	1.03	-0.69
		Rural	1.92	1.14	-0.44
		Urban	2.78	1.19	0.42
School Funding Type	*	Community	2.03	1.20	-0.34
		Public	2.22	1.26	-0.14
		Private	2.89	1.08	0.53
Gender		Male	2.38	1.24	0.02

Group		Subgroup	Mean	SD	Diff from Overall Mean
		Female	2.34	1.26	-0.02
Age		4 to 11 year old	2.36	1.26	0.00
		12 to 20 year old	2.36	1.23	0.00
State	*	Benadir	2.58	1.14	0.21
		Hirshabelle	3.17	1.02	0.81
		Jubaland	1.81	1.25	-0.55
		Southwest	2.22	1.15	-0.14

4.5 LEARNER EQUITY PERCEPTIONS

Productive learning environments provide equitable support for learners of all backgrounds and abilities. The Organization for Economic Cooperation and Development (OECD 2012) defines equity in education in the following way, "equity in education means that personal or social circumstances, such as gender, ethnic origin or family background, are not obstacles to achieving educational potential…". The Longitudinal Learner Survey included two items designed to assess learner feelings of equity in their classroom. The scale is composed of the average of the following 2 items:

- My teacher treats me fairly at school.
- Reverse of item: In my classroom, some children are treated better than others.

All items were measured on a 6-point Likert scale (0-None of the time, 1-A little of the time, 2-Some of the time, 3-A lot of the time, 4-Most of the time, 5-All the time) where learners indicated the frequency of perceived equity with each statement.

All learners at the baseline expressed positive feelings regarding equitable treatment in their classroom. No significant differences were noted in perceptions of equity based on gender, age, or program (BAB, non-BAB). However, significant differences in perception of equity did emerge for both BAB and non-BAB learners based on location, with learners in Hirshabelle reporting significantly less favorable feelings of equity than learners in Benadir, Jubaland or Southwest. At the program level, BAB learners in rural areas reported significantly lower perceptions of equity than their peers in IDP and urban areas. No differences were noted by location type for non-BAB learners, however community school children expressed significantly lower perceptions of equity than learners in private or public schools. Table 12 shows perceptions of equity by location, school funding type, gender, age, and state. Additional results are included in the Data Annex (Section 2.3; 8.2; 9.4).

Table 12: Baseline Longitudinal Learner Sample Perceptions of Equity by Location, School Funding Type, Gender, Age, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Students Student Equity Perceptions			**3.73**	**1.06**	
Location Type	*	IDP	3.86	1.08	0.12
		Rural	3.53	1.00	-0.20
		Urban	3.74	1.07	0.01
Gender		Male	3.70	1.08	-0.03
		Female	3.76	1.05	0.03
Age		4 to 11 year old	3.75	1.07	0.02
		12 to 20 year old	3.71	1.06	-0.02
State	*	Benadir	3.76	1.07	0.02
		Hirshabelle	3.30	0.87	-0.43
		Jubaland	3.69	1.02	-0.04
		Southwest	3.95	1.10	0.22
Group		**Subgroup** Abs	Mean	SD	Diff from Overall Mean
All Non-BAB Students Student Equity Perceptions			**3.78**	**0.99**	
Location Type		IDP	3.80	1.03	0.02
		Rural	3.79	0.95	0.01
		Urban	3.77	1.00	-0.01
School Funding Type	*	Community	3.66	0.98	-0.12
		Public	3.77	1.00	-0.01
		Private	3.89	1.00	0.11
Gender		Male	3.78	1.01	0.00
		Female	3.78	0.98	0.00
Age		4 to 11 year old	3.81	0.97	0.02
		12 to 20 year old	3.74	1.02	-0.04
State	*	Benadir	3.92	0.88	0.14
		Hirshabelle	3.44	0.85	-0.34
		Jubaland	3.78	1.01	0.00
		Southwest	3.89	1.09	0.11

4.6 SAFETY

USAID's Safer Learning Environments Working Group (USAID 2016) defined a safe learning environment as, "a place where structured learning is free from environmental, internal, and external threats to the learners' and educators' well-being … where infrastructure of a learning environment (and also the people within a learning environment) is deemed safe." Threats to safety can be internal

(e.g., bullying, corporal punishment, or gang violence) or external (e.g., attacks on the school or natural disasters). All threats to safety have the potential to significantly decrease the academic performance of learners. Learners that feel physically or emotionally unsafe at school or on the way to school are prone to poor attendance and increased dropout rates (Dunne et. al., 2005; Mullis, et. al.,2012b; Kibriya et. al., 2018). Research suggests that learner perceptions of safety may have a greater impact on learner success than actual measures of safety (Goldstein, S.E., Young, A., and Boyd, C., 2008). Safety items on the Learner Baseline Survey derive from several instruments and the BAB Learner Survey, as no appropriate brief quantitative measure was found.

The Safety scale items were used to assess learners' perception of safety in the school environment. The scale is composed of the average of the following 3 items:

- I feel safe at school.
- I feel safe on my way to school.
- Reverse of item: I am picked on or bullied at school.

All items were measured on a 6-point Likert scale (0-None of the time, 1-A little of the time, 2-Some of the time, 3-A lot of the time, 4-Most of the time, 5-All the time) where learners indicated their perceived safety regarding each statement.

All learners felt relatively safe both at school and on their way to school (mean of 3.67) with no significant differences based on gender or learner age. However, some significant differences appeared based on grade, school type, location type and state (Table 13; Data Annex, Section 2.5; 8.4; 9.6). Learners in non-BAB grade 2 reported significantly higher perceptions of safety than learners in either grade 2, or BAB – perhaps reflecting their prior experience in school (Table 14). Urban learners also reported significantly higher perceptions of safety than learners in rural or IDP schools and private school learners felt safer than learners attending public, community, or BAB schools. Finally, learners in Hershabelle reported significantly lower levels of safety than learners in other states – perhaps due to the diversity of context including environmental and social challenges present in that state during sampling.

Table 13: Baseline Longitudinal Learner Sample Perceptions of Safety by Location, Gender, Age, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Students Student Perceptions of Safety in the School Environment			**3.62**	**1.02**	
Location Type	*	**IDP**	**3.52**	**1.11**	**-0.11**
		Rural	**3.54**	**1.00**	**-0.08**
		Urban	**3.71**	**0.97**	**0.08**
Gender		**Male**	**3.63**	**1.04**	**0.00**
		Female	**3.62**	**1.00**	**0.00**

Group		Subgroup	Mean	SD	Diff from Overall Mean
Age		**4 to 11 year old**	**3.59**	**1.03**	**-0.03**
		12 to 20 year old	**3.67**	**1.00**	**0.05**
State	*	**Benadir**	**3.81**	**0.88**	**0.18**
		Hirshabelle	**3.16**	**0.94**	**-0.46**
		Jubaland	**3.69**	**0.94**	**0.07**
		Southwest	**3.71**	**1.10**	**0.09**

Group		Subgroup Abs	Mean	SD	Diff from Overall Mean
All Non-BAB Students Student Perceptions of Safety in the School Environment			**3.74**	**0.93**	
Location Type		IDP	3.75	0.96	0.01
		Rural	3.70	0.83	-0.04
		Urban	3.75	0.96	0.01
School Funding Type	*	Community	3.65	0.94	-0.09
		Public	3.67	0.99	-0.08
		Private	3.94	0.78	0.20
Gender		Male	3.76	0.90	0.02
		Female	3.72	0.97	-0.02
Age		4 to 11 year old	3.71	0.92	-0.03
		12 to 20 year old	3.78	0.95	0.04
State	*	Benadir	3.95	0.75	0.21
		Hirshabelle	3.34	0.93	-0.40
		Jubaland	3.74	0.87	0.00
		Southwest	3.83	1.03	0.09

Table14: Baseline Longitudinal Learner Sample Perceptions of Safety by Program and Grade Level

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Longitudinal Study Students Student Perceptions of Safety in the School Environment			**3.67**	**0.99**	
Program Inclusion	*	BAB	3.62	1.02	-0.05
		NonBAB Grade 1	3.68	0.98	0.01
		NonBAB Grade 2	3.80	0.88	0.13

4.7 LEARNING INDICATORS

In-line with the Somali National Education curriculum, the BAB curriculum includes instruction in numeracy, literacy, and socio-emotional learning (SEL). The baseline external evaluation examined indicators of psycho-social well-being that map to the SEL skills included in the BAB curriculum and are known to correlate with education outcomes, as well as baseline numeracy and literacy skills. These will also be measured at midline and endline.

Learner Emotional Engagement in School

For learners to benefit from a learning environment, they must be both present and engaged. School engagement is commonly considered a learner characteristic that can be influenced by the school environment in ways that are either harmful or protective. In a literature review on the subject, Fredricks and her coauthors identified three types of engagement, namely: behavioral, emotional, and cognitive (Fredricks, et al. 2005, Fredricks, et al. 2005). Behavioral engagement may include good behavior, following the rules, or taking part in academic, social, or extracurricular activities. Emotional engagement includes positive and negative reactions to the school and people within the school environment, feelings of belonging and being valued, and appreciation for success in school-related outcomes. Cognitive engagement encompasses the learner's investment in learning, desire to go beyond minimum requirements, willingness to accept challenges, and ability to self-regulate. High levels of behavioral and emotional engagement positively correlate with academic achievement (Ladd and Dinella 2009). Engagement items on the Learner Survey derive from the 14-item School Engagement Scale (Fredricks, et al. 2005). Only three of the five items associated with emotional engagement were included in the Baseline survey. We excluded behavioral and cognitive engagement items for baseline, as they need experience of the classroom that may not be present early in the school year. Items from all three scales will be included in the Midline and Endline surveys. The baseline emotional engagement scale is composed of the average of the following three items:

- I like being at school.
- I feel happy at school.
- I am interested in the work at school.

All items were measured on a 6-point Likert scale (0-None of the time, 1-A little of the time, 2-Some of the time, 3-A lot of the time, 4-Most of the time, 5-All the time) where learners indicated the frequency of each indicator of emotional engagement at school associated with each statement.

Emotional engagement scores at baseline showed no significant difference based on gender (3.16 for males and 3.19 for females) or between BAB (3.17) and non-BAB learners (3.16) overall. However, non-BAB grade 1 learners scored statistically lower on emotional engagement at baseline (2.99) than the grade 2 learners (3.33). We also noted significant differences for rural learners (2.69 and 2.98, respectively for BAB and non-BAB learners) who scored lower than IDP (3.27/3.26) and urban (3.30/3.18) learners. Learner emotional engagement also varied by school funding type with the lowest scores among community school learners (2.90) followed by public school (3.09) and BAB learners (3.17), with private school learners scoring highest (3.49). As with other baseline indicators, learners in Hirshabelle scored significantly lower at baseline on the emotional engagement scale than

learners in Benadir, Julaland, or Southwest for cohort learners in both BAB and non-BAB classes. Tables 15 and 16 summarize emotional engagement scores across sample subgroups. This data is presented graphically in the Data Annex (Section 2.4; 8.3; 9.5).

Table 15: Baseline Longitudinal Learner Sample Emotional Engagement by Program and Grade Level

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Longitudinal Study Students Student School Engagement Perceptions			**3.17**	**1.09**	
Program Inclusion	*	BAB	3.17	1.09	0.01
		NonBAB Grade 1	2.99	1.13	-0.17
		NonBAB Grade 2	3.33	0.99	0.17

Table 16: Baseline Longitudinal Learner Sample Emotional Engagement by Location, Gender, Age, School Funding Type, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Students Student School Engagement Perceptions			**3.17**	**1.09**	
Location Type	*	IDP	3.27	1.13	0.10
		Rural	2.69	1.10	-0.48
		Urban	3.30	1.03	0.12
Gender		Male	3.16	1.10	-0.02
		Female	3.19	1.09	0.02
Age	*	4 to 11 year old	3.12	1.11	-0.05
		12 to 20 year old	3.25	1.07	0.07
State	*	Benadir	3.57	1.01	0.40
		Hirshabelle	2.41	1.02	-0.77
		Jubaland	3.31	0.93	0.13
		Southwest	3.25	1.07	0.08
Group		Subgroup	Mean	SD	Diff from Overall Mean
All Non-BAB Students Student School Engagement Perceptions			**3.16**	**1.08**	
Location Type	*	IDP	3.26	1.04	0.10
		Rural	2.98	1.05	-0.18
		Urban	3.18	1.09	0.02
School Funding Type	*	Community	2.90	1.14	-0.25

Group		Subgroup	Mean	SD	Diff from Overall Mean
		Public	3.09	1.09	-0.07
		Private	3.49	0.91	0.33
Gender		Male	3.15	1.08	0.00
		Female	3.16	1.08	0.01
Age		4 to 11 year old	3.14	1.05	-0.01
		12 to 20 year old	3.18	1.11	0.02
State	*	Benadir	3.40	0.98	0.24
		Hirshabelle	2.53	1.06	-0.63
		Jubaland	3.23	0.99	0.07
		Southwest	3.30	1.09	0.14

Learner Quality of Life

Quality of Life is a general term that encompasses the learner's perception of well-being and functioning in physical, emotional, mental, social, and everyday areas of life. Quality of life indicators for children and youth include indicators of self-esteem, relationships with family and friends, and attitudes towards school and the future. Self-esteem, defined as an individual's subjective evaluation of their own worth (Harter 1988), includes both beliefs about self-worth and emotional states, like triumph, despair, pride, and shame. Low self-esteem can decrease desire to learn, ability to focus, and willingness to take risks – thus negatively impacting learning outcomes. Educational research suggests that increasing self-esteem is one of the best ways to improve academic achievement (Rubie, Townsend and Moore 2004). Likewise, increased feelings of school belonging are positively correlated with positive educational outcomes and negatively correlated with absence and dropout rates (Korpershoek, et al. 2020). Quality of life items on the Longitudinal Learning Survey derive from the 24-item KINDL^{R} Quality of Life Survey (Ravens-Sieberer and Bullinger 1998). This tool was developed to measure children and youth perceptions of quality of life related to physical and emotional well-being, self-esteem, family and friends, and school. It includes versions for children ages 7-17 with questions that are straightforward and easy to understand. The evaluation uses quality of life as a proxy for global self-worth and self-efficacy in this study. The baseline Learner Survey includes subscales for self-esteem, friendship, and school.

All items were measured on a 6-point Likert scale (0-None of the time, 1-A little of the time, 2-Some of the time, 3-A lot of the time, 4-Most of the time, 5-All the time) where learners scored each statement over the past week.

Learners in all groups scored relatively positively for quality of life indicators across all three subscales with no significant differences by gender or age for either BAB or non-BAB learners. In-line with findings for learner emotional engagement in school, learners in non-BAB 2nd grade classes scored significantly higher on quality of life indicators (3.43) than either non-BAB first graders (3.22) or BAB learners (3.24). Rural (3.07/BAB, 3.22/non-BAB) and IDP learners (3.19/BAB, 3.24/non-

BAB) had significantly lower Quality of Life scores than their urban counterparts (3.34/BAB, 3.39/non-BAB). As with other indicators, private school learners scored significantly higher on quality of life scales (3.56) than community school learners (3.12) with public (3.29) and BAB (3.24) learners intermediate. Composite quality of life scores for learners in Hirshabelle (3.08/BAB, 3.21/non-BAB) and Southwest (3.12/BAB, 3.23/non-BAB) were significantly lower than for learners in Benadir (3.42/BAB, 3.42/non-BAB) and Jubaland (3.45/BAB. 3.42 non-BAB).

Patterns of scores on quality-of-life subscales (friendship, self-esteem, and emotional well-being) generally show little deviation from the overall patterns of the composite scores, except for quality-of-life subscales by state. Aggregates of the three subscales show learners in Hirshabelle and Southwest scoring significantly lower than learners in Benadir and Jubaland. However, there is considerable variation across subscales. Learners in Hirshabelle scored low on friendship and self-esteem scales, but matched Benadir learners with the highest scores on measures of emotional well-being. Learners in Southwest, on the other hand, had among the highest scores for friendship, but mixed scores for self-esteem (low for BAB and average for non-BAB learner) and low scores for measures of emotional well-being.

Composite quality of life results are shown in Tables 17 and 18 below. Complete composite and subscale data is included in the Data Annex (Section 2.6; 8.5-8.8; 9.7).

Table 17: Baseline Longitudinal Learner Sample Composite Quality of Life Scores by Program and Grade Level

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Longitudinal Study Students Student Quality of Life			**3.28**	**0.64**	
Program Inclusion	*	BAB	3.24	0.64	-0.03
		NonBAB Grade 1	3.22	0.63	-0.05
		NonBAB Grade 2	3.43	0.60	0.15

Table 18: Baseline Longitudinal Learner Sample Composite Quality of Life Scores by LocationType, Gender, Age, School Funding Type, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Students Student Quality of Life			**3.24**	**0.64**	
Location Type	*	IDP	3.19	0.64	-0.06
		Rural	3.07	0.63	-0.17
		Urban	3.34	0.63	0.09
Gender		Male	3.22	0.63	-0.02
		Female	3.26	0.65	0.02

Group		Subgroup	Mean	SD	Diff from Overall Mean
Age		4 to 11 year old	3.25	0.65	0.01
		12 to 20 year old	3.23	0.63	-0.01
State	*	Benadir	3.42	0.60	0.18
		Hirshabelle	3.08	0.62	-0.16
		Jubaland	3.45	0.64	0.21
		Southwest	3.12	0.63	-0.13
Group		**Subgroup** Abs	Mean	SD	Diff from Overall Mean
All Non-BAB Students Student Quality of Life			**3.32**	**0.62**	
Location Type	*	IDP	3.24	0.63	-0.08
		Rural	3.22	0.70	-0.11
		Urban	3.39	0.59	0.07
School Funding Type	*	Community	3.12	0.58	-0.21
		Public	3.29	0.65	-0.04
		Private	3.56	0.53	0.24
Gender		Male	3.33	0.61	0.01
		Female	3.32	0.64	-0.01
Age		4 to 11 year old	3.31	0.62	-0.01
		12 to 20 year old	3.34	0.64	0.02
State	*	Benadir	3.42	0.53	0.10
		Hirshabelle	3.21	0.61	-0.11
		Jubaland	3.43	0.64	0.10
		Southwest	3.23	0.65	-0.10

Numeracy and Literacy Levels

Numeracy and literacy baselines were measured using the Early Grade Reading Assessment (EGRA) (RTI International, Early Grade Reading Assessment (EGRA) Toolkit, Second Edition 2015) and Early Grade Math Assessment (EGMA) (RTI International 2014) models. These assessments are oral exams that focus on the basic skills underpinning reading and numeracy skill development. The reading exam includes letter recognition, phonemic awareness, reading simple words, and listening comprehension. The math test includes number recognition, comparisons, and ordering sets of objects. The BAB program, in collaboration with the MoCHE, adapted EGRA and EGMA versions appropriate for Level 1 learners in Somalia during a workshop held in summer, 2021. The evaluation team incorporated these same assessments into the learner survey for longitudinal cohort data collection at baseline to allow direct comparison with BAB's cross-sectional cohort. Literacy was assessed for the Somali (Af-Maxaa tiri) dialect.

The assessments test the general skills listed in Table 19 (see full text versions of the assessments in Appendix 5). In addition to the factor analysis, reliability, and item differentiation and discrimination, we calculated ceiling and floor effects by the percentage frequency of lowest or highest possible scores achieved by respondents on both the EGRA and EGMA (Appendix 3). For EGRA, 10.9% of respondents scored in the lowest 10%, while just 2.1% scored in the upper 10%. For EGMA, 25.8% scored in the lowest 10%, while nobody scored in the upper 10%. Ceiling and floor effects are generally considered significant if they are higher than 15% (Terwee, et al. 2007). For our sample, no ceiling effects were noted for either EGMA or EGRA. Although the EGRA showed a floor effect, it is not concerning, as learners in our cohort are new learners.

This section summarizes key results from the EGMA and EGRA baseline survey of longitudinal cohort learners. Additional longitudinal cohort results are included in the Data Annex (Section 3; 9.8-9.9). Comparison of the longitudinal cohort baseline results with BAB cross-sectional cohort results are included in section 7 below and in the Data Annex (Sections 7.1-7.5).

Table 19: EGRA and EGMA Subtasks

EGRA subtask	EGMA subtask
Listening Comprehension	Number Identification
Phonemic Awareness	Number Discrimination
Letter Sound Identification	Missing Number
Invented Words	Addition Level 1
Familiar/Real Words	Addition Level 2
Oral Reading Fluency	Subtraction Level 1
Reading Comprehension	Subtraction Level 2
Dictation 1	Word Problems
Dictation 2	

Each subtask included a set of one to 100 items. Subtasks were scored as the percentage of items correct out of the total number of items (from 0-100%). The aggregate score is the mean of all subtask scores.

Enumerators conducted numeracy assessments in the primary dialect spoken by the learner. Thus, we do not expect dialect to impact learner outcome scores for numeracy. For the literacy exam, numerators give instructions in the learner's primary dialect, but tasks were conducted in the official Somalia language (Af-Maxaa tiri) – thus, learners with a primary dialect other than the language of instruction may be disadvantaged.

We summed the 9 sub-sections of the EGRA in equal weights to create a Total EGRA Score out of 100 possible points. Likewise, we summed the 8 sub-sections of the EGMA in equal weights to create a Total EGRA Score out of 100 points. The means by demographic groupings are provided in Tables 20 and 21 below. EGRA and EGMA data disaggregated by subscale is included in the Data Annex (Section 3; 9.8-9.9).

Table 20: EGRA (Early Grade Reading Assessment) Overall Longitudinal Sample Test Results

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Longitudinal Study Learners EGRA Total Percent Correct			**38.18**	**23.50**	
Program Inclusion	*	BAB	37.22	22.97	-0.96
		Non-BAB Grade 1	31.10	22.47	-7.08
		Non-BAB Grade 2	48.63	22.64	10.45

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Learners EGRA Total Percent Correct			**37.22**	**22.97**	
Location Type	*	IDP	40.99	26.40	3.77
		Rural	31.80	19.02	-5.42
		Urban	37.25	21.98	0.03
Gender	*	Male	38.76	22.49	1.54
		Female	35.65	23.35	-1.57
Age	*	4- to 11-year-old	32.12	20.97	-5.11
		12- to 20-year-old	44.37	23.73	7.15
State	*	Benadir	36.45	23.16	-0.77
		Hirshabelle	33.04	18.88	-4.18
		Jubaland	34.82	21.74	-2.40
		Southwest	40.97	24.72	3.74

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Non-BAB Learners EGRA Total Percent Correct			**39.56**	**24.19**	
Location Type	*	IDP	32.24	23.96	-7.32
		Rural	30.89	19.44	-8.67
		Urban	45.31	24.06	5.75
School Funding Type	*	Community	31.02	21.07	-8.54
		Public	39.57	24.00	0.01
		Private	46.73	24.69	7.17

Gender		Male	40.30	24.01	0.74
		Female	38.79	24.36	-0.77
Age	*	4- to 11-year-old	34.28	23.72	-5.28
		12- to 20-year-old	47.68	22.62	8.12
State	*	Benadir	38.23	23.43	-1.33
		Hirshabelle	43.07	22.30	3.51
		Jubaland	32.10	22.61	-7.46
		Southwest	45.07	25.39	5.51

Table 21: EGMA (Early Grade Math Assessment) Overall Longitudinal Sample Test Results

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Longitudinal Study Learners EGMA Total Percent Correct			**33.88**	**23.13**	
Program Inclusion	*	BAB	33.79	22.92	-0.09
		Non-BAB Grade 1	25.75	22.21	-8.13
		Non-BAB Grade 2	42.86	21.39	8.98

Group		Subgroup	Mean	SD	Diff from Overall Mean
All BAB Learners EGMA Total Percent Correct			**33.79**	**22.92**	
Location Type	*	IDP	37.54	25.31	3.75
		Rural	28.34	21.68	-5.45
		Urban	33.83	21.66	0.04
Gender	*	Male	36.68	22.39	2.89
		Female	30.84	23.09	-2.95
Age	*	4- to 11-year-old	27.79	21.28	-6.00
		12- to 20-year-old	42.19	22.51	8.40
State	*	Benadir	35.53	22.51	1.74
		Hirshabelle	27.53	19.77	-6.26
		Jubaland	27.73	19.48	-6.06
		Southwest	38.80	24.75	5.01

Group		Subgroup	Mean	SD	Diff from Overall Mean

All Non-BAB Learners EGMA Total Percent Correct			**34.00**	**23.43**	
Location Type	*	IDP	27.29	23.37	-6.71
		Rural	23.68	21.13	-10.33
		Urban	40.05	22.25	6.05
School Funding Type	*	Community	28.60	22.37	-5.40
		Public	32.21	23.28	-1.80
		Private	41.67	22.71	7.67
Gender	*	Male	36.22	23.29	2.22
		Female	31.70	23.37	-2.30
Age	*	4- to 11-year-old	27.44	22.88	-6.57
		12- to 20-year-old	44.10	20.51	10.10
State	*	Benadir	33.05	23.02	-0.95
		Hirshabelle	36.32	21.35	2.32
		Jubaland	24.64	22.14	-9.36
		Southwest	41.67	23.10	7.67

BAB learners entered level 1 with higher overall skills in both literacy and numeracy than learners entering grade 1 in the same communities, despite similar reported prior educational experiences. As seen in Figures 21 and 22, at baseline, nearly 2/3 of BAB learners scored in the non-learner or emergent learner range on EGRA and EGMA assessments (64.2 and 61%, respectively). While nearly ¾ learners beginning first grade in a non-BAB class fell into these categories (75.2 and 72.2%, respectively) [4]. As expected, beginning grade 2 learners in the longitudinal sample scored significantly higher at baseline on EGRA and EGMA assessments, with zero non-learners and only 43.4% scoring as an emergent learner on the EGRA and only 0.3% scoring as a non-learner and 43.4% as emergent on the EGMA assessment. Density plots of EGRA and EGMA total scores by subgroup (BAB, grade 1, and grade 2) provide another view of learner baseline skills and competencies and clearly position BAB learner mean scores between non-BAB learners on baseline measures (Figure 23). The pattern of scoring, with BAB learners between grade 1 and grade 2 learners, repeats for all EGRA and EGMA subtasks, as illustrated in the radar plots in Figure 24.

[4] As Somalia does not currently have defined national literacy and numeracy performance levels, we adopted the performance levels described in the Leave No Girl Behind, AGES project (page 97) (Machova, Miettunen and Peterson 2020).

Figure 21: EGRA Learner Level at Baseline by Program and Grade

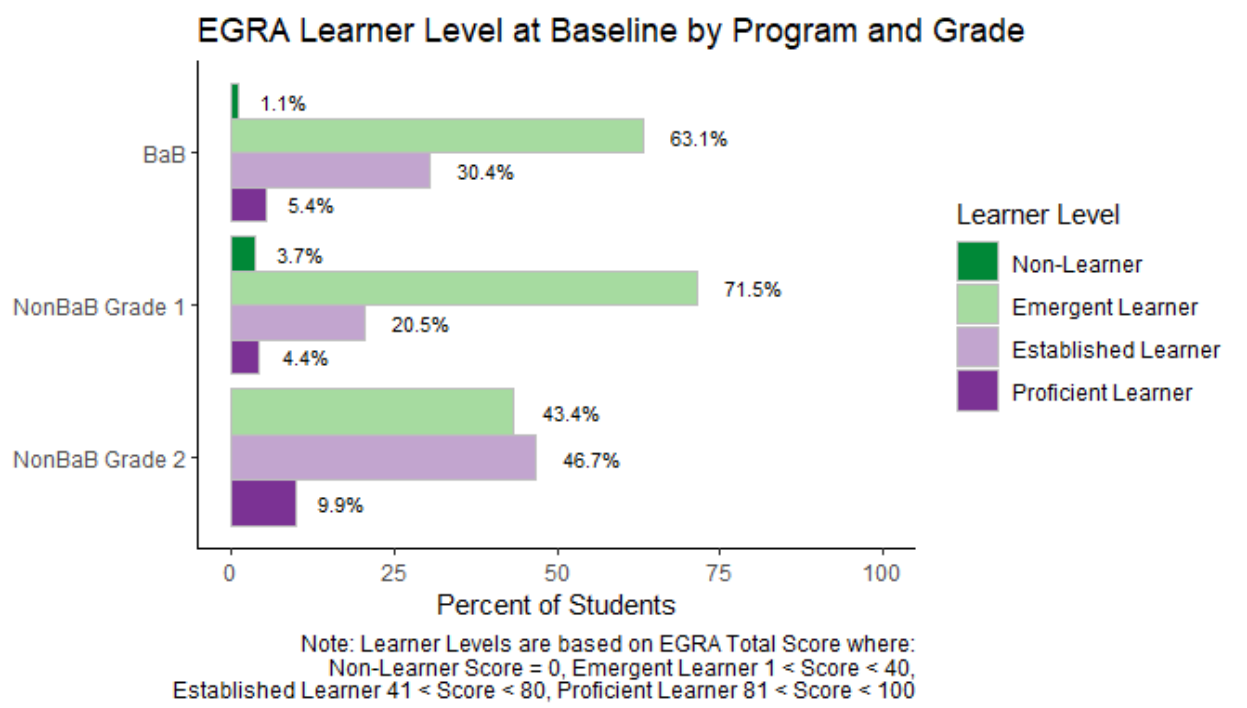


Figure 22: EGMA Learner Level at Baseline by Program and Grade

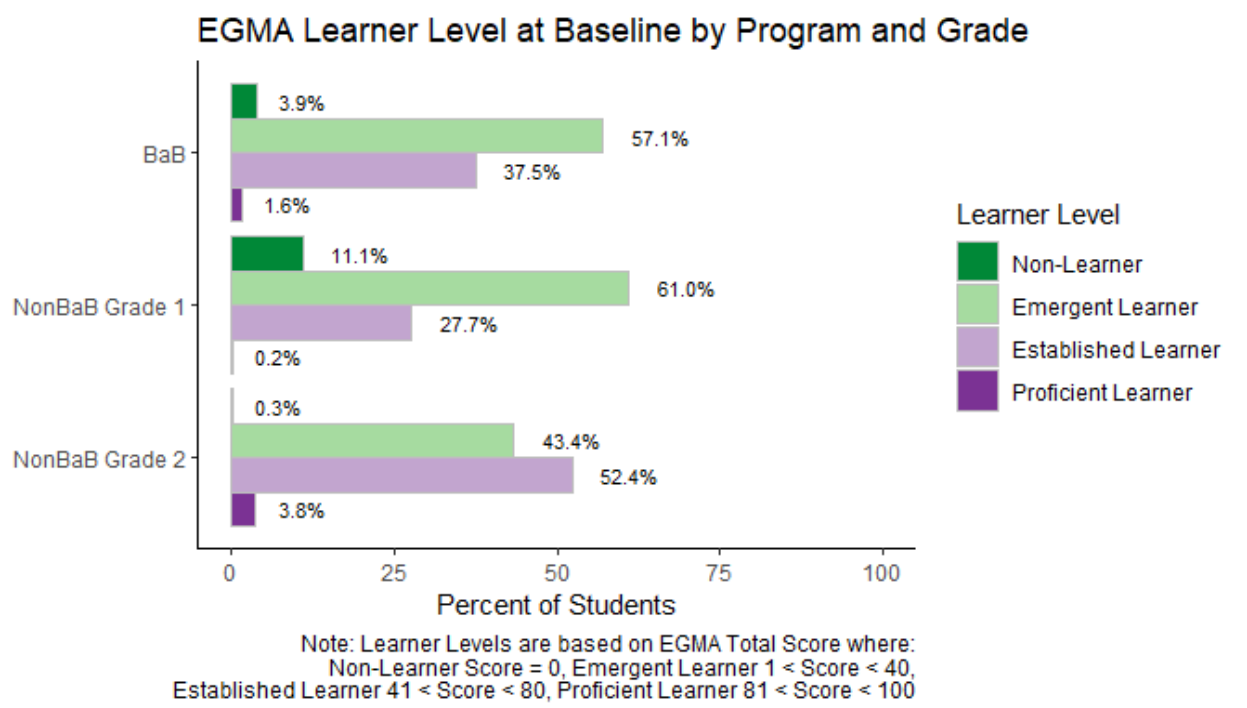


Figure 23: EGRA and EGMA Density Plots of Learner Baseline Scores by Program Type and Grade

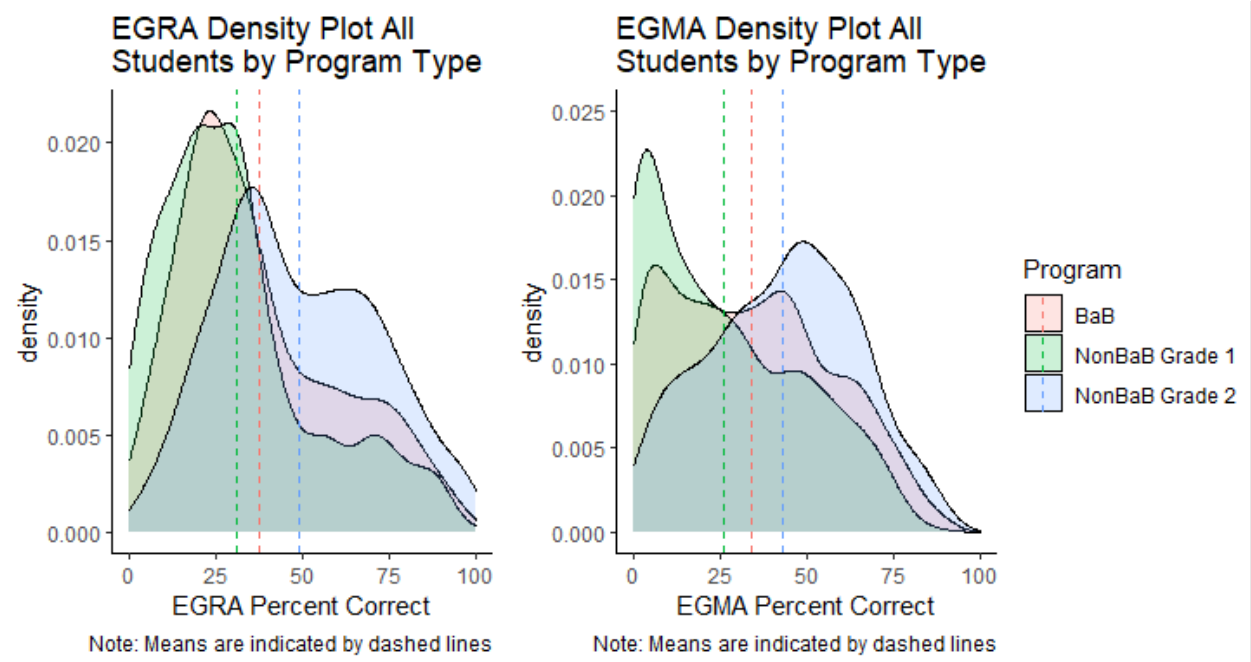
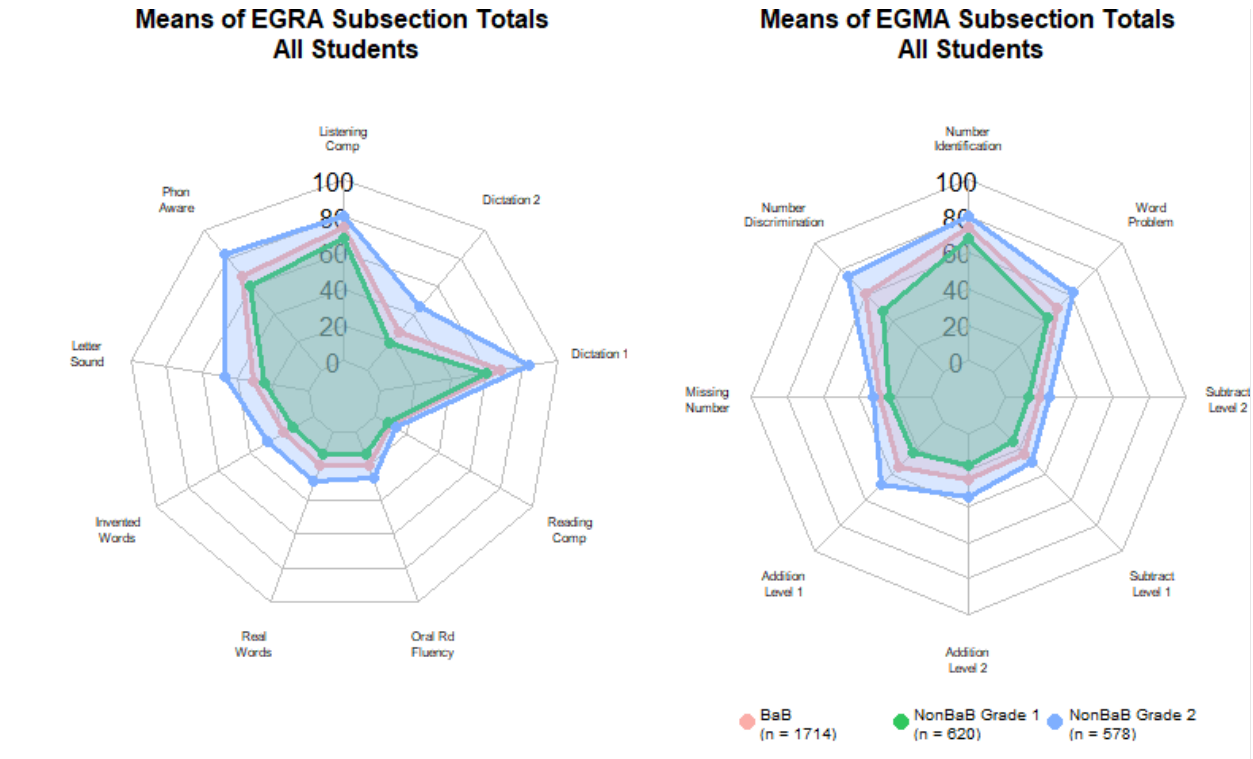


Figure 24: EGRA and EGMA Radar Graphs of Learner Baseline Subscale Scores by Program and Grade



Density plots of learner baseline performance on EGRA and EGMA exams show variations by gender (Figure 25), location type (Figure 26), school funding type (Figure 27), and state (Figure 28). At baseline, boys scored slightly higher than girls on literacy exams, but significantly higher than girls on the numeracy assessment.

Rural learners underperformed learners from urban and IDP areas on both literacy and numeracy assessments. Interestingly, learners in IDP schools outperformed their urban counterparts at baseline.

Private school learners significantly outperformed learners in all other school types at baseline, while community school learners had the lowest baseline scores. Public school and BAB learners had similar baseline scores for EGMA and EGMA – intermediate to community and private schools.

Learner baseline skill and competency levels also varied by state, with learners in Southwest significantly outperforming learners in other states while learners in Jubaland lagged.

Figure 25: Baseline Longitudinal Learner EGRA and EGMA Scores by Gender

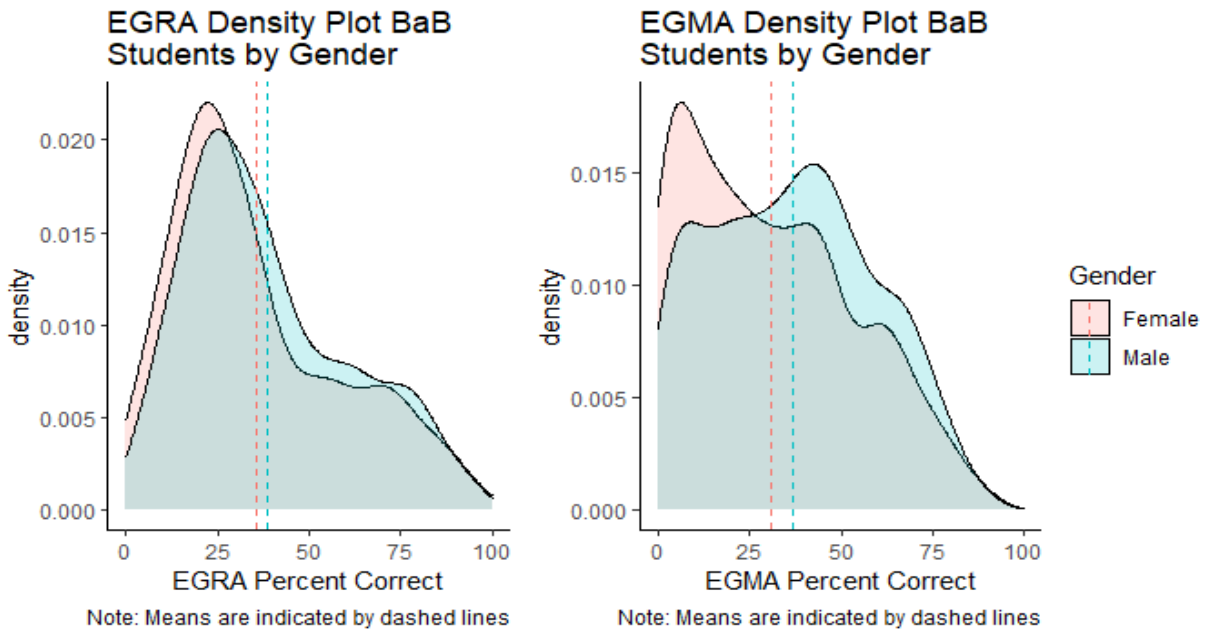


Figure 26: Baseline Longitudinal Learner EGRA and EGMA Scores by Location Type

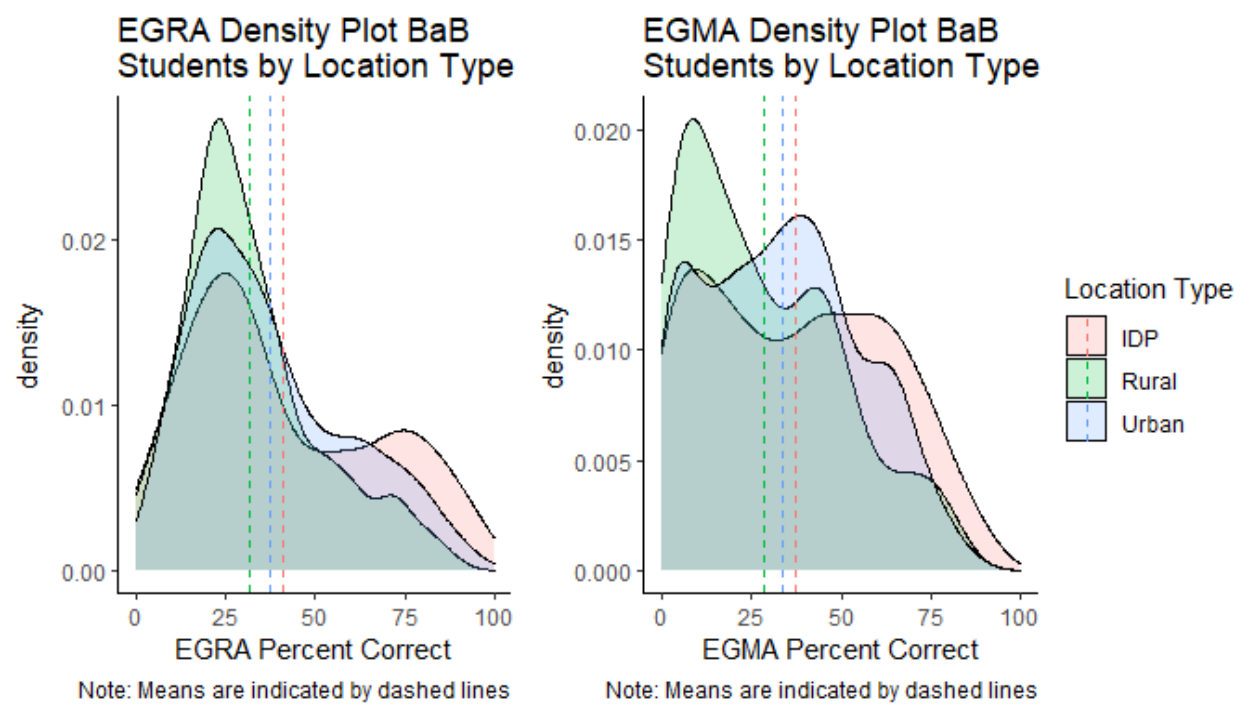


Figure 27: Baseline Longitudinal Learner EGRA and EGMA Scores by School Type

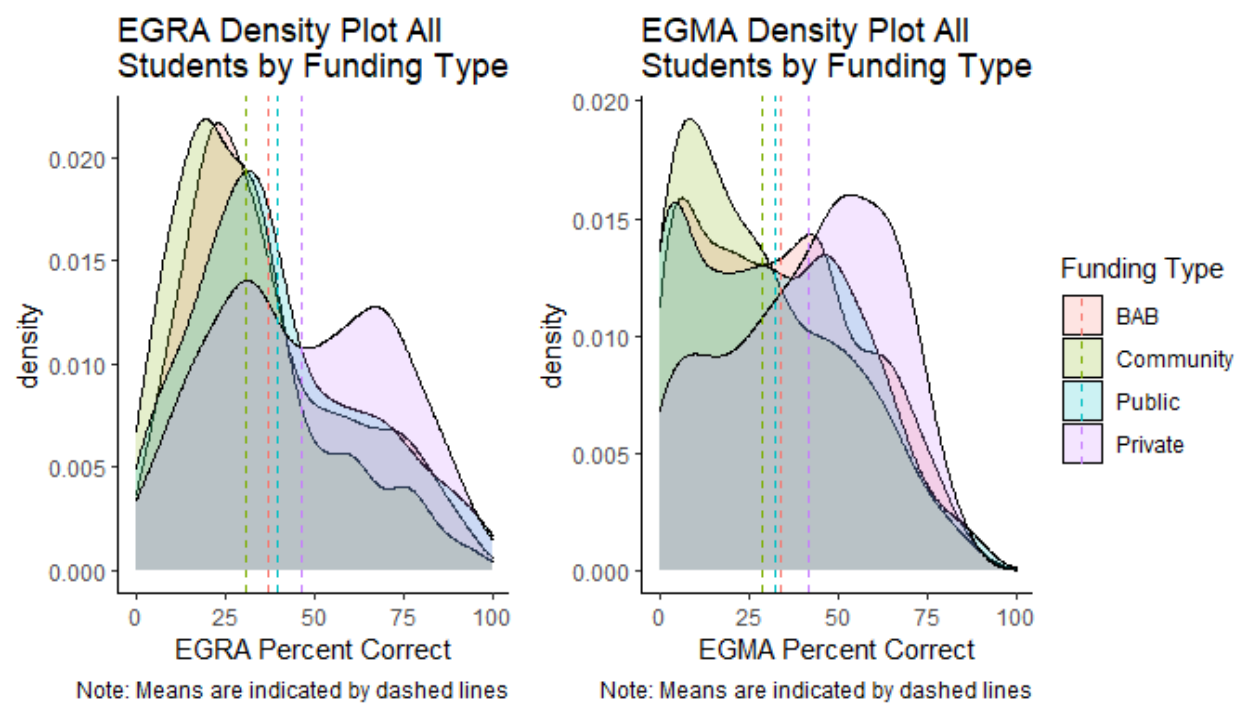
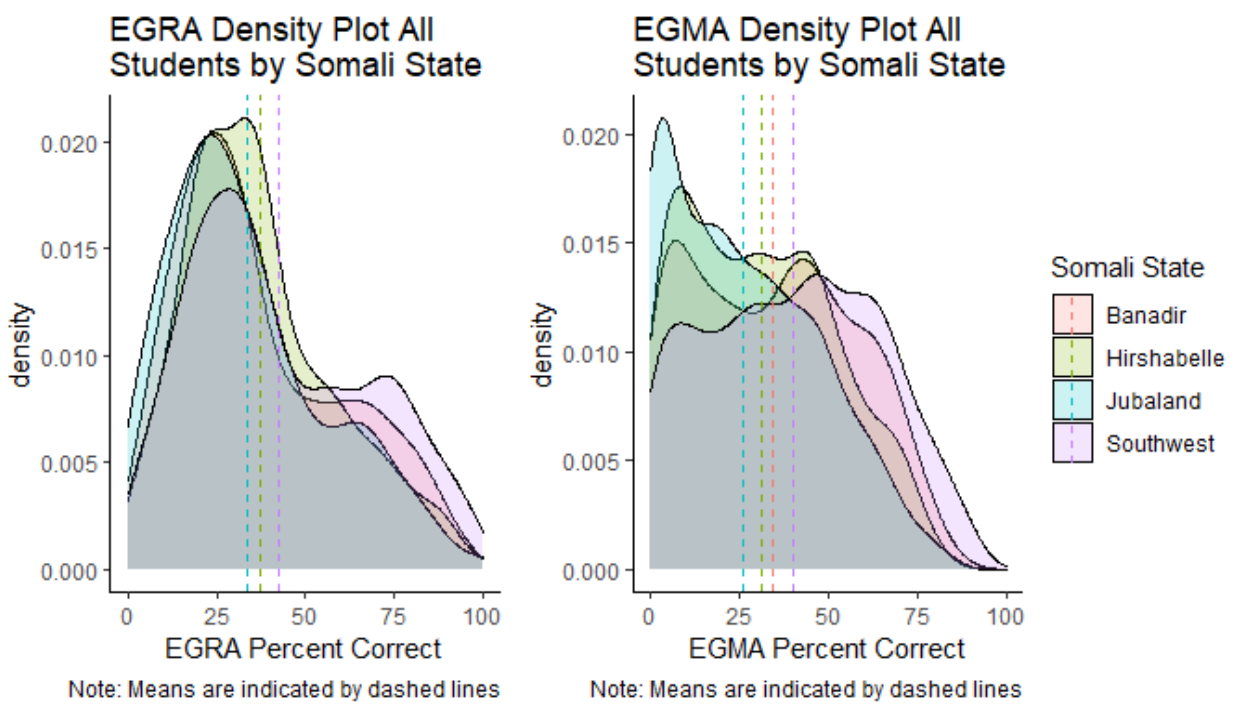


Figure 28: Baseline Longitudinal Learner EGRA and EGMA Scores by State

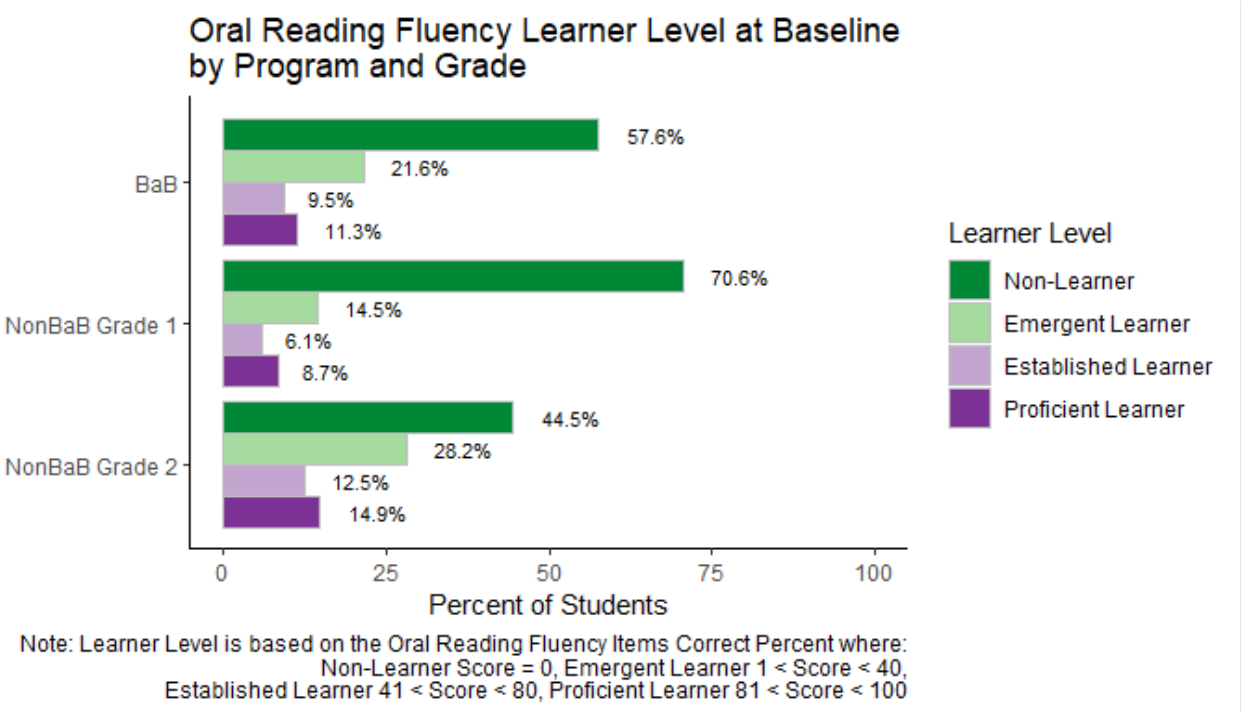


Baseline Oral Reading Fluency vs. Reading Comprehension

Minimum reading proficiency benchmarks are often set as the level of reading fluency, measured by correct words per minute (cwpm), where comprehension is at an acceptable level. Typically, 80% comprehension is taken as the standard for learners to demonstrate acceptable comprehension. While no benchmark currently exists for the country of Somalia, Ethiopia set a benchmark for the Af Somali language in either grade 2 or 3 at 50 cwpm (RTI International 2017). This corresponds to literacy in cwpm where 80% comprehension is reached.

The Oral Reading Fluency subsection of the EGRA measures how quickly and accurately a learner can read. It contains a paragraph with 66 words for the learner to read. This is a timed subsection and learners are given one minute to read the paragraph. Fluency results are depicted in Figure 29 by learner fluency level at baseline, where learners scoring zero (no correct answers) are categorized as non-learners, learners scoring between 1-40 cwpm as emergent readers, learners scoring between 41-80 cwpm as established readers, and learners scoring above 81 cwpm are categorized as proficient. As previously described, since Somalia does not currently have defined national literacy and numeracy performance levels, we adopted the performance levels described in the Leave No Girl Behind, AGES project (page 97) (Machova, Miettunen and Peterson 2020).

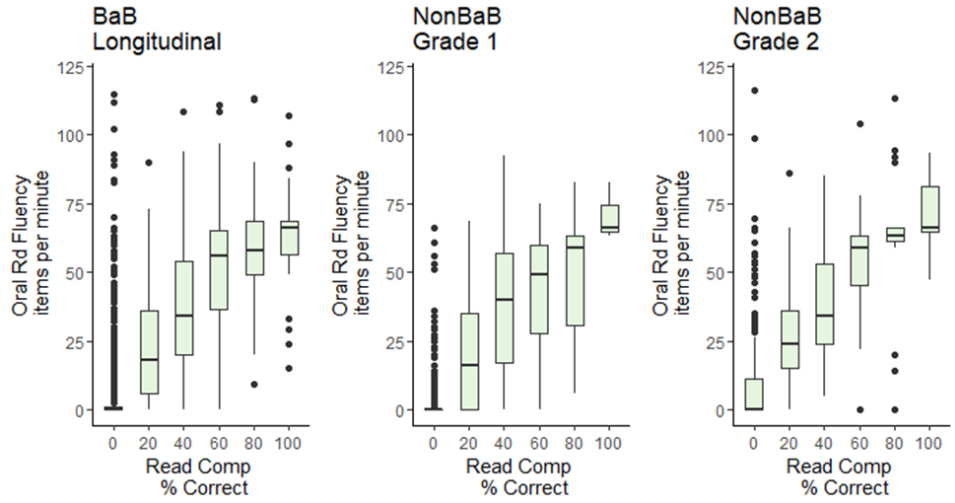
Figure 29: Baseline Oral Reading Fluency for Longitudinal Cohort Learners by Program and Grade



We plotted oral reading fluency scores against reading comprehension scores (Figure 30) to find the oral reading fluency items per minute at the first quartile where reading comprehension scores reached 80% correct, as a measure of proficiency.

The proficiency scores vary by cohort subgroup with BAB learners at 49 cwpm, Non-BAB Grade 1 learners at 31 cwpm, and Non-BAB Grade 2 learners at 62 cwpm.

Figure 30: Longitudinal Cohort Oral Fluency vs Comprehension



4.8 INTERSECTION BETWEEN LEARNER CHARACTERISTICS, SCHOOL TYPES, LOCATIONS, AND ACADEMIC PREPARATION

The External Evaluation Longitudinal sample data was disaggregated by gender, age, primary dialect, and prior educational experiences to understand the background and characteristics of learners. The summary table of learner demographic characteristics by school funding type (Table 22) and the bar graph showing SES, and learner safety and psychosocial perceptions (Figure 31) show distinct patterns and suggest that different populations of Somalia children and youth are served by the various school types.

Cohort learners attending private schools scored highest on all indicators in our analysis. They are relatively affluent with more boys than girls, speak Maxaa tiri as their primary language and are more likely to have both prior personal educational experiences and more educated mothers. Not surprisingly, private school learners also scored highest on baseline assessments of literacy and numeracy (Figure 32).

Community schools, at the other end of the spectrum, also serve more boys than girls, but they are significantly poorer, speak predominately Maay, and their mothers are less likely to be literate. Learners attending community schools also scored lowest on all psychosocial scales, except for safety, and had the lowest baseline literacy and numeracy scores.

Public schools in the longitudinal cohort sample enrolled more girls than boys and served children that are only slightly more affluent than their community-school peers, although they report maternal literacy rates that fall midway between those of private and community school children. Public school learner scores on psychosocial indicators and baseline numeracy and literacy scores are also intermediate to private and community schools.

BAB schools served boys and girls in about the same proportions. Children attending BAB schools report among the lowest levels of maternal education and SES scores. Despite these indicators of economic deprivation, BAB learner scores on psychosocial measures and baseline literacy and numeracy scores are very similar to those measured for public school learners.

Table 22: Learner Characteristics by School Type

	School Type			
Student Characteristic	**BAB**	**Community**	**Public**	**Private**
Predominant Gender	Male (50.5%)	Male (57.1%)	Female (54.9%)	Male (55.8%)
Median Age	11	10	11	11
Prior Ed*	88.7%	89.0%	89.8%	95.2%
Maternal Ed	37.7%	37.9%	48.0%	56.1%

	School Type			
Student Characteristic	**BAB**	**Community**	**Public**	**Private**
Primary Dialect	Maxaa tiri (54.8%)	Maay (59.2%)	Maxaa tiri (51.1%)	Maxaa tiri (77.6%)

*Including both religious and non-religious

Figure 31: Learner SES, Safety and Psychosocial Perceptions by School Type

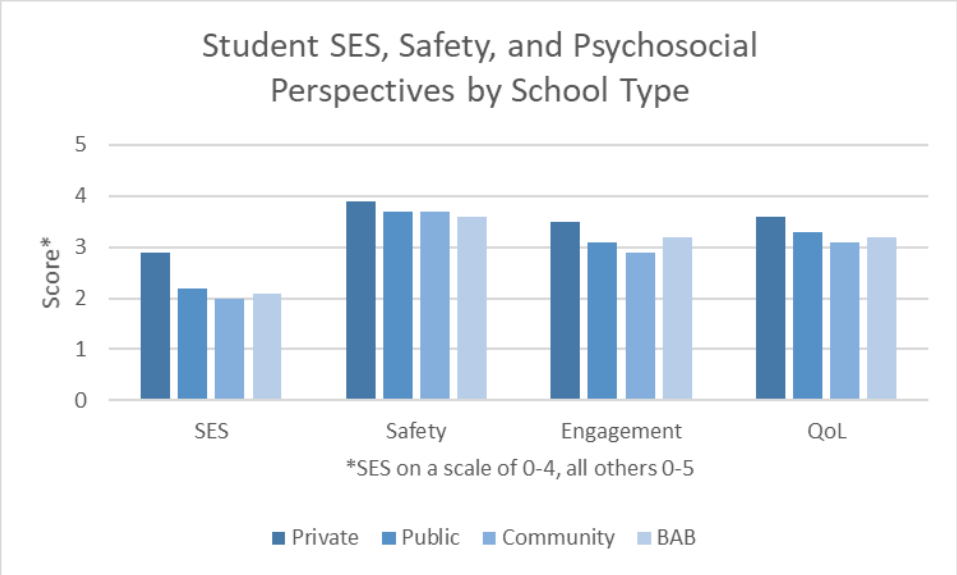
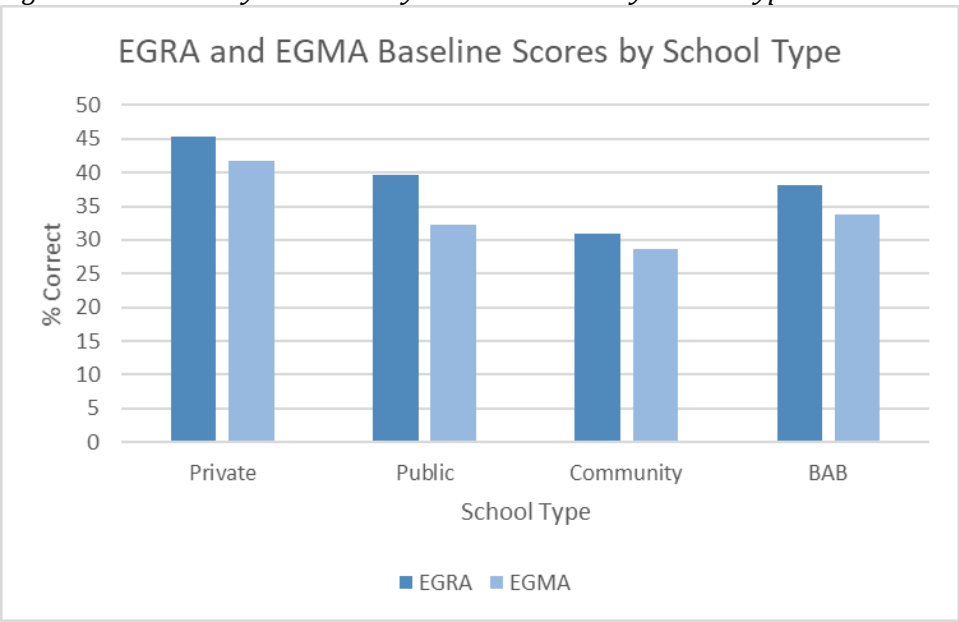


Figure 32: Numeracy and Literacy Baseline Scores by School Type



Socio-economic indicators, psychosocial measures, and baseline skills and competencies are also influenced by where learners live. While we saw slight difference in gender balance, median learner age, or prior educational experience based on location type, learners from IDP and rural areas were far more likely to report low maternal literacy rates and fewer family resources than their urban peers. (Table 23 and Figure 33). Rural learners reported the lowest perceptions of equity, engagement, and quality of life indicators of all groups and all scored significantly lower on baseline literacy and numeracy tests than learners from either IDP or rural areas. Contrary to expectations, learners from IDP areas, while resource poor, outperformed both rural and urban learners on baseline literacy and numeracy tests – perhaps indicative of other support resources available in these areas (Figure 34).

Table 23: Learner Characteristics by Location Type

	School Type		
Student Characteristic	**IDP**	**Rural**	**Urban**
Predominant Gender	Female (50.9%)	Female (51.0%)	Male (52.0%)
Median Age	11	11	11
Prior Ed	86.4%	86.0%	92.5%
Maternal Ed	32.0%	32.5%	49.8%
Primary Dialect	Maay (75.3%)	Maxaa tiri (69.1%)	Maxaa tiri (65.2%)

Figure 33: Learner SES, Safety, and Psychosocial Perspectives by School Location

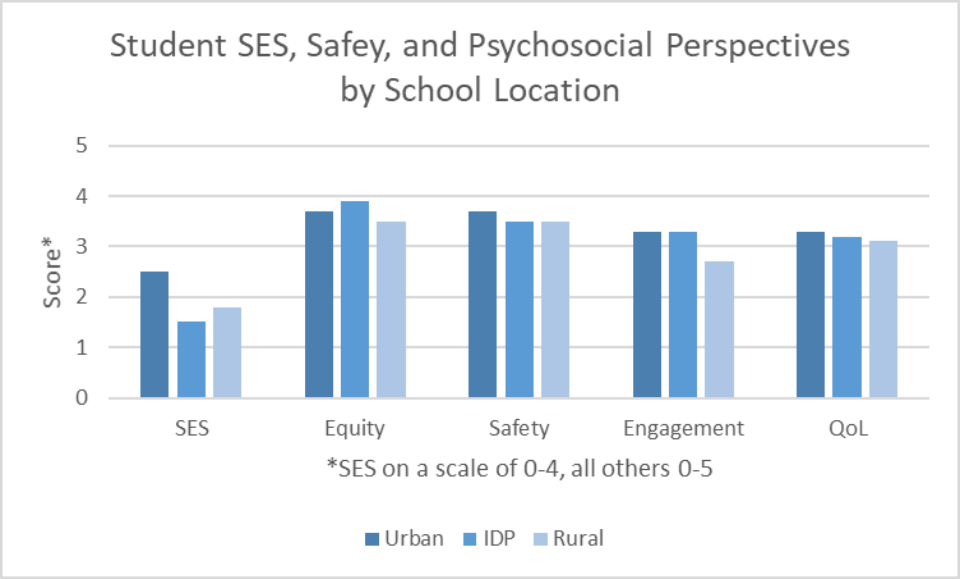
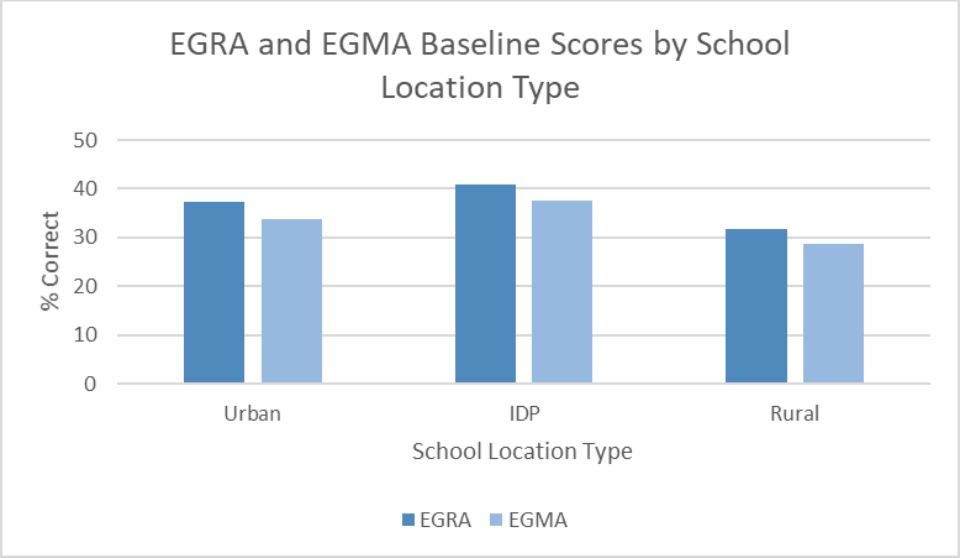


Figure 34: Numeracy and Literacy Baseline Scores by School Location Type



Learner baseline characteristics are described in more detail in the following sections and in the Data Annex (Section 1.1; 9.1-9.3).

BAB learners, on average, scored significantly lower on SES indicators, including maternal literacy and family resources, than non-BAB learners suggesting that the BAB program is serving financially marginalized learners in the target communities. SES indicators are further affected by location, school funding type, and geographic location (state). Baseline data suggests stratification of learners

by SES indicators across school types (community, BAB, public, and private). Community school learners reported the lowest levels of family resources, maternal literacy rates, and perceptions of equity in the classroom with public school learners intermediate and private school learners scoring highest on all indicators. BAB learners scored about the same as public school learners on measures of SES and perceptions of equity but reported low levels of maternal literacy that mirrored community school learners. Learners in urban areas appear more affluent than their rural and IDP counterparts, with significantly higher rates of maternal literacy and more resources. Finally, indicators of SES varied significantly by state, with learners in Hirshabelle and Benadir reporting higher SES indicators than learners in Southwest and Jubaland. Of note, while Hirshabelle learners had the highest reported SES indicators, they also had the lowest perceptions of equity.

5 INTERMEDIATE OUTCOMES AS PREDICTORS OF LEARNING

Correlation and multiple regression analyses were conducted to examine the relationship between EGRA/EGMA and potential predictors. Table 24 summarizes the descriptive statistics and analysis results for EGRA, while Table 25 summarizes the descriptive statistics and analysis results for EGMA. Results presented in Table 26 indicate that baseline EGRA scores are significantly positively correlated with all intermediate outcome variables of interest (socio-economic status [SES], learner emotional engagement, learner age, maternal literacy, and quality of life), indicating that learners with higher scores on these variables tend to have higher baseline EGRA scores.

The multiple regression model with all five predictors produced R-squared = 0.134, $F(5, 2768) = 85.649$, $p < .001$, indicating that the model is a significant predictor of baseline EGRA scores and 13.4% of the variance in these scores is accounted for by the variables in the model. A close look at these results reveals that the intermediate outcome variables have significant positive regression weights, indicating that learners who score higher on these variables/scales are expected to have higher baseline EGRA scores, after controlling for other variables in the model. It is important to note that the low R-squared could be a result of a combination of low correlations between the intermediate outcomes and the baseline EGRA scores, on one hand, and the wide variability in the data as reflected in the large standard errors.

The results presented in Table 27 reveal that the baseline EGMA scores are also significantly positively correlated with all intermediate outcome variables in the model except maternal literacy which turned out not to be statistically significant. The regression model significantly predicts baseline EGMA scores with these intermediate outcomes, R-squared = .146, $F(5, 2768) = 95.768$, $p < .001$, indicating that 14.6% of the variance in baseline EGMA scores is accounted for by the variables in the model. Just like the regression results for the baseline EGRA scores, the R-squared here is low because of small correlation coefficients associated with the intermediate outcomes and the wide variability in the EGMA scores as evident in the standard errors.

To perform multiple linear regression, some key assumptions must be met to have precise parameters estimates. However, some minor deviations from these assumptions may not affect the precision of these estimates. The assumptions of normality and linearity were minimally violated in our data, but that did not significantly affect the parameter estimates. Typically, when there are

violations in any or all the assumptions of linear regression (and Pearson's correlation, reported in a linear regression analysis), a non-parametric method like Spearman's rho is recommended for estimating the strength of association between variables and is interpreted the same way. Therefore, we performed Spearman's rho, also known as Spearman's rank order correlation analysis, to validate the estimates from the regression analysis. Results from that analysis are presented in Table 26, and it shows that the correlations estimates are not significantly different from the Pearson's correlation coefficients presented in the regression analysis. Please, note that Spearman's rho estimated here is comparable to Pearson's zero-order correlations in Tables 24 and 25.

Table 24: Multiple Regression and Correlation Results for EGRA

	Regression Coefficients			Correlations		
	B	*SE B*	β	Zero-order	Partial	Part
Constant	-3.383	2.575				
SES	10.292***	1.410	.134	.159***	.137***	.129***
Engagement	1.763***	.453	.081	.138***	.074***	.069***
Age	13.805***	.858	.287	.273***	.292***	.285***
Mother Literacy	5.310***	.875	.111	.118***	.115***	.107***
Quality of Life	2.692**	.785	.072	.139**	.065**	.061**

Notes: R-squared = .134 (p< .001), * p < .05, ** p < .01, ***p<.001

Table 25: Multiple Regression and Correlaton Results for EGMA

	Regression Coefficients			Correlations		
	B	*SE B*	β	Zero-order	Partial	Part
Constant	-2.674	2.508				
SES	9.990***	1.373	.132	.138***	.137***	.128***
Engagement	2.241***	.442	.105	.126***	.096***	.089***
Age	15.983***	.836	.338	.331***	.342***	.336***
Mother Literacy	3.048***	.852	.065	.060***	.068***	.063***

Quality of Life	.088	.765	.002	.079	.002	.002

Notes: R-squared = .147 (p< .001), * p < .05, ** p < .01, ***p<.001

Table 26: Spearman's Correlations Between EGRA/EGMA and Key Constructs

	EGRA	EGMA
	Spearman's rho	Spearman's rho
SES	0.173***	0.150***
Engagement	0.132***	0.122**
EGRA	1.000	0.791***
EGMA	0.791***	1.000
Age	0.277***	0.328***
Mother Literacy	0.109***	0.061**
Quality of Life	0.163***	0.089***

Notes: * p < .05, ** p < .01, ***p<.001

6 BASELINE RESULTS -- QUALITY OF INSTRUCTION

Teacher quality is considered a cornerstone of effective education. In 2014, the Director-General of UNESCO articulated the importance of teachers with these words, "A quality universal primary education will remain a distant dream for millions of children living in countries without enough trained teachers in classrooms" (UNESCO 2014). The Longitudinal Teacher Baseline Survey includes items that describe teacher background and preparation and measures key elements associated with learning outcomes, including teacher engagement, self-efficacy, content and pedagogical knowledge, attitudes and beliefs about learning, and perceptions of support and safety.

6.1 TEACHER SAMPLE CHARACTERISTICS

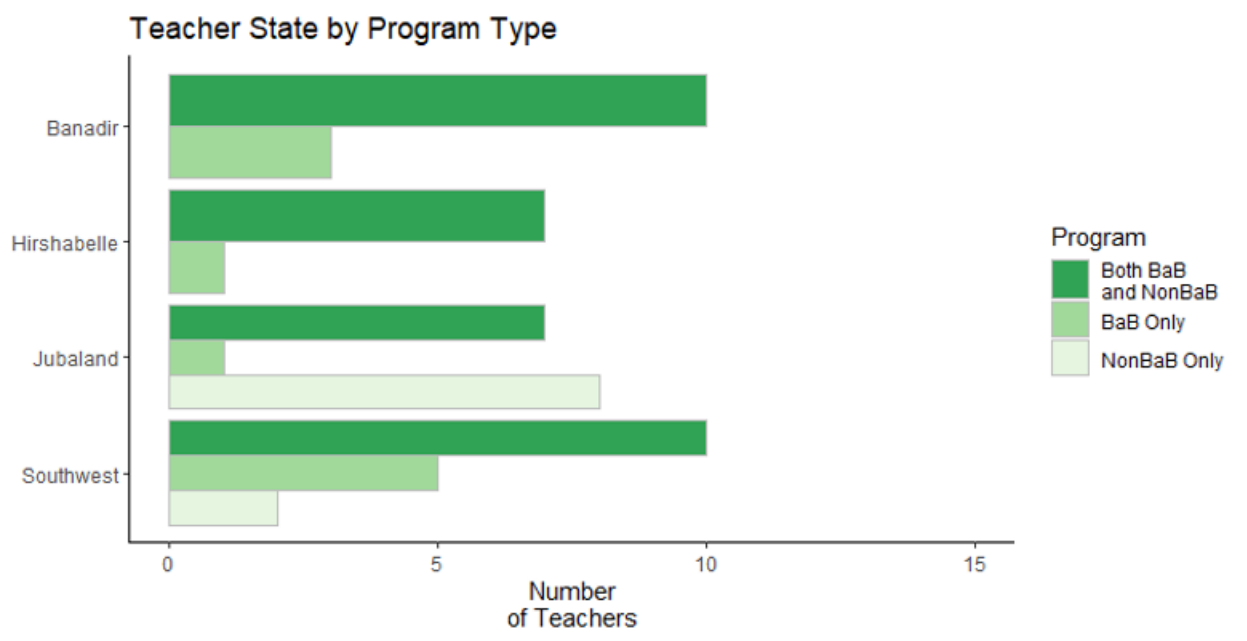
SORDI enumerators tried to collect survey data from teachers for each longitudinal cohort learner classroom. The sampling frame purposefully included teachers that teach both BAB and non-BAB classes (typically working a double shift), and teachers that teach only BAB or non-BAB classes. A total of 54 teachers from 44 schools completed the Teacher Survey, as shown in Table 27. Some of the non-BAB teachers were not willing to be interviewed for this project.

Table 27: Number of Teachers by Type of Class Taught

Type of Class Taught	# Teachers	# Schools
Total	54	44
Both BAB and non-BAB classes	34	33
Only BAB classes	10	10
Only non-BAB classes	10	8

Teacher subgroup distribution by state is illustrated in Figure 35. Note that, although the longitudinal cohort sample includes non-BAB classrooms in all four states (Benadir, Hirshabelle, Jubaland, and Southwest) sample teachers come exclusively from Jubaland and Southwest.

Figure 35: Sample Teacher Distribution by Program Type across States



In general, longitudinal cohort teachers are experienced, well-educated, and male. The following figures present baseline teacher characteristics for all longitudinal cohort teachers. Additional characteristics and data disaggregated by subgroup (both BAB and non-BAB, BAB only, or non-BAB only) are included in the Data Annex (Sections 4.1-4.3). Note that the small sample size for teacher subgroups and unequal subgroup distribution (e.g., all teachers in the non-BAB group teach in either Jubaland or Southwest states), limit interpretation and generalizability of these data.

Teachers in all classroom types are most often male and in their 20's for both BAB and non-BAB classes, regardless of program type, location, funding type (Figure 36), or state (Figure 37). Teachers overwhelmingly have at least a grade 12 education, with nearly ¾ reporting additional education

levels, as shown in Figure 38. Moreover, more than half of the sampled teachers have five or more years of teaching experience and only 3 of the 54 reported less than one year of experience (Figure 39).

Figure 36: Teacher Gender by Program Type, Location, or School Funding

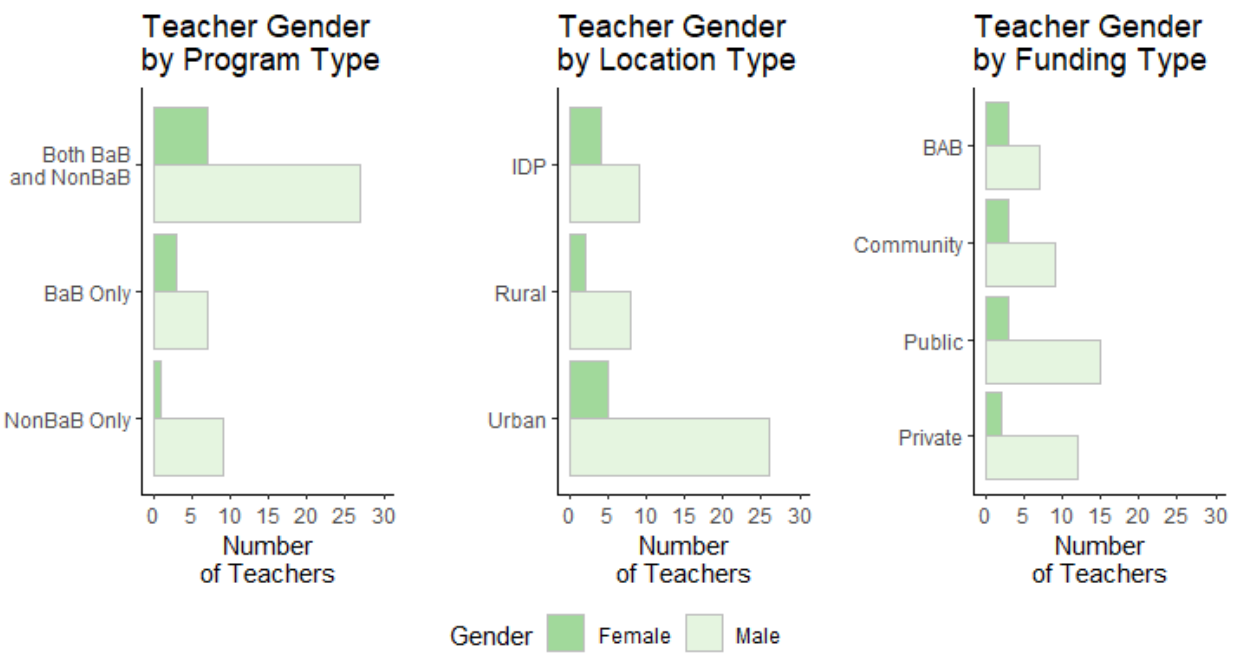


Figure 37: Teacher State by Gender

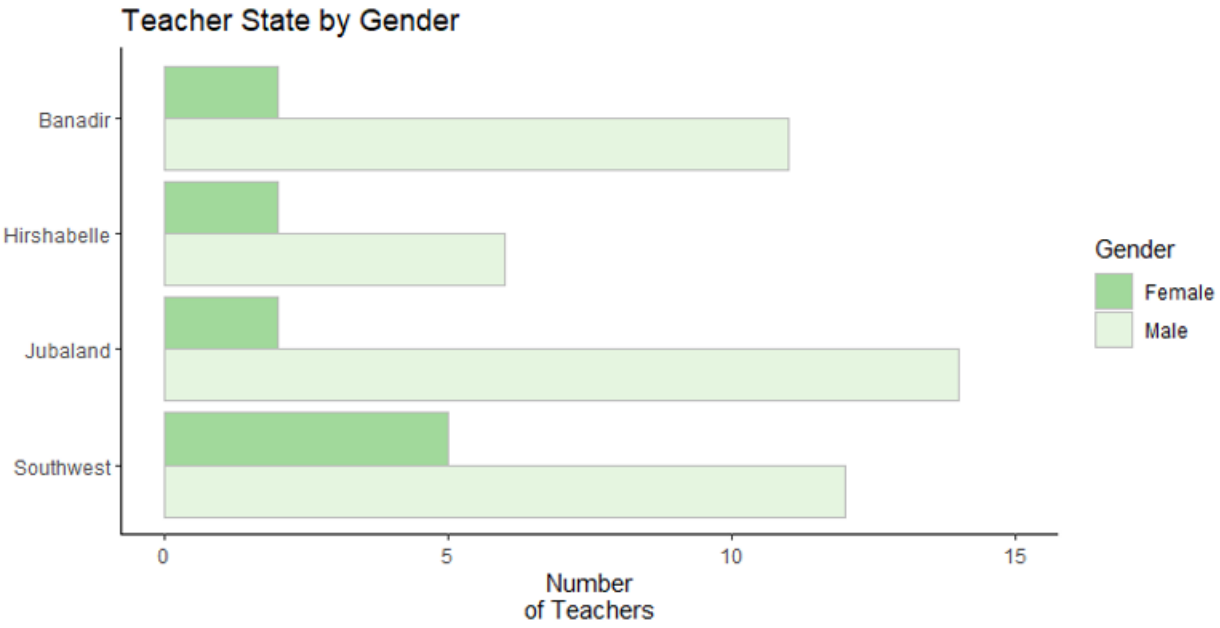


Figure 38: Teacher Education Level

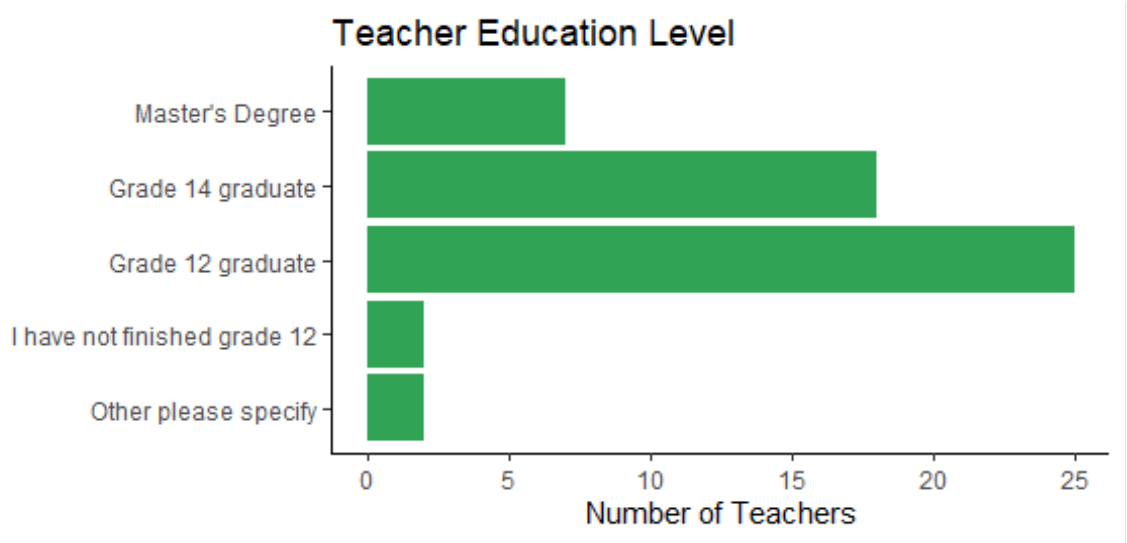
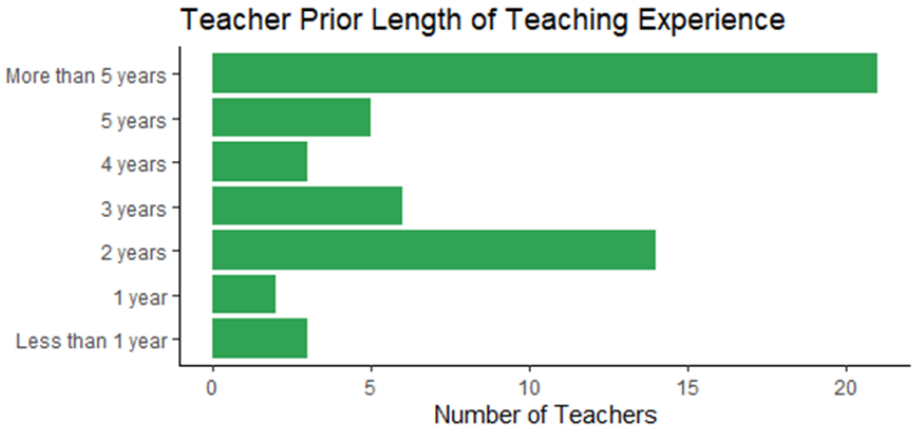


Figure 39: Teacher Experience in Years



6.2 CLASS SIZE

Class size, or learner/teacher ratio, correlates strongly with learner outcomes. Overall, research shows that learners in smaller classes are more engaged and perform better in all subjects than their peers in larger classes, particularly in early elementary school. Small classes are most beneficial for marginalized learners (Bedard and Kuhn 2006) (Dee and West 2011) (Flemming, Toutant and Raptis 2002). Longitudinal cohort teachers reported a wide range of class sizes for both BAB and non-BAB classes (Figures 40 and 41). Some non-BAB classes experienced extremely large numbers of learners

(greater than 100). Enumerator field records noted insufficient tables and chairs to accommodate all learners in some classes. In some cases, this led to learners being sent home.

Figure 40: Teacher Reported BAB Class Sizes

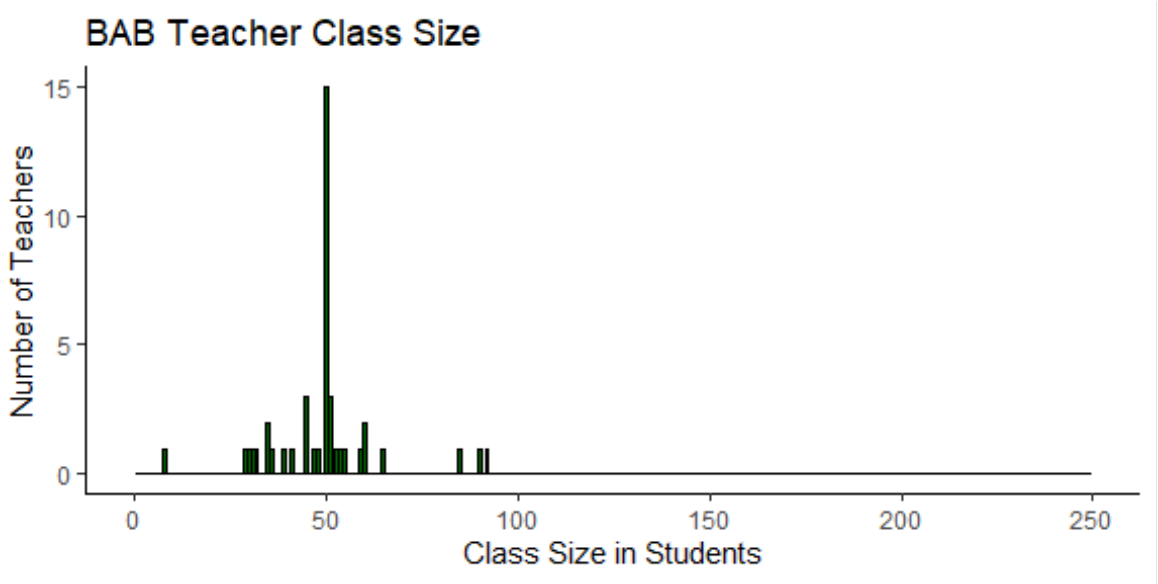
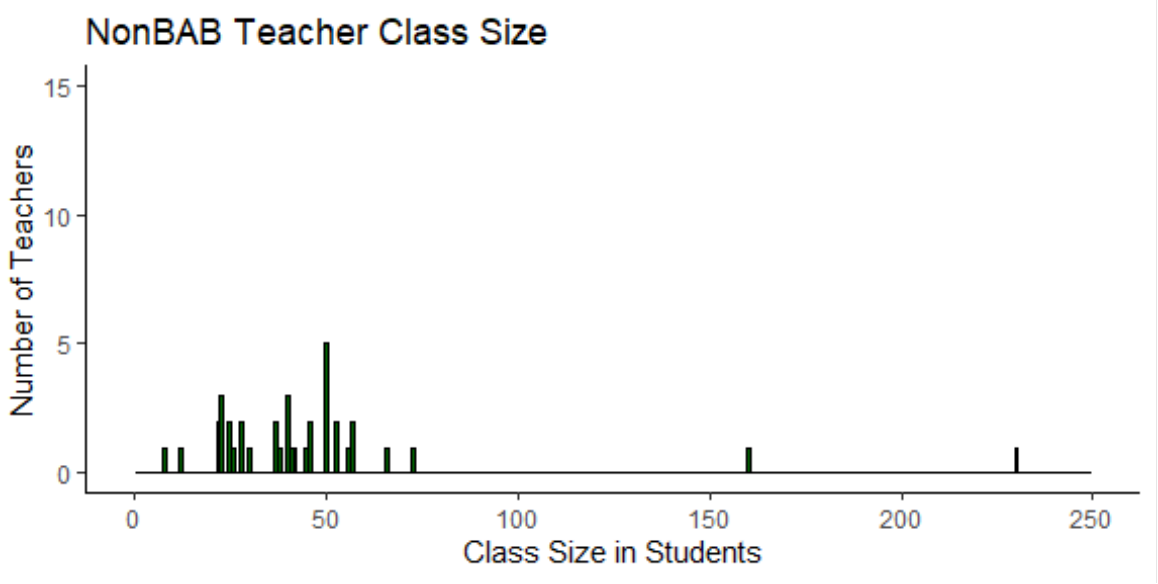


Figure 41: Teacher Reported non-BAB Class Sizes



TEACHER PERCEPTIONS OF SAFETY

As previously described, threats to safety can significantly influence learner academic, social, and emotional outcomes. The Teacher Baseline Survey included items to assess the Teacher's perception of their own and their learners' physical and emotional safety in the school environment. The scale is composed of the average of the following four items modified from the BAB Teacher Survey:

- I feel physically safe at school.
- All my learners are physically safe at school, regardless of their gender, age, family background, disability, or other characteristics.
- All my learners, regardless of gender, age, disability, family background, or other characteristics are safe on their way to school.
- All my learners, regardless of gender, age, disability, family background, or other characteristics are accepted and emotionally supported in my school.

All items were measured on a 6-point Likert scale (0-strongly disagree, 1-disagree, 2-somewhat disagree, 3-neither agree nor disagree, 4-somewhat agree, 5-agree, 6-strongly agree) where teachers indicated their perceived safety of themselves and their learners.

Like their learners, teachers reported positive feelings of perceived safety regardless of program, location, school type, gender, teacher's age, or state (Figures 42 and Table 28). None-the-less, despite generally positive feelings of safety overall, baseline results showed variations in perceptions of safety based on location and program. Teachers in rural areas were less positive than their colleagues in urban or IDP areas. Likewise, those teachers that only taught in BAB classes reported less positive perceptions of safety. Mirroring learner perceptions, teachers from Hirshabelle reported the lowest levels of perceived safety (although still positive). The Data Annex (Section 5.1; 8.9) includes complete teacher safety data.

Figure 42: Teacher Perceptions of Safety by Program

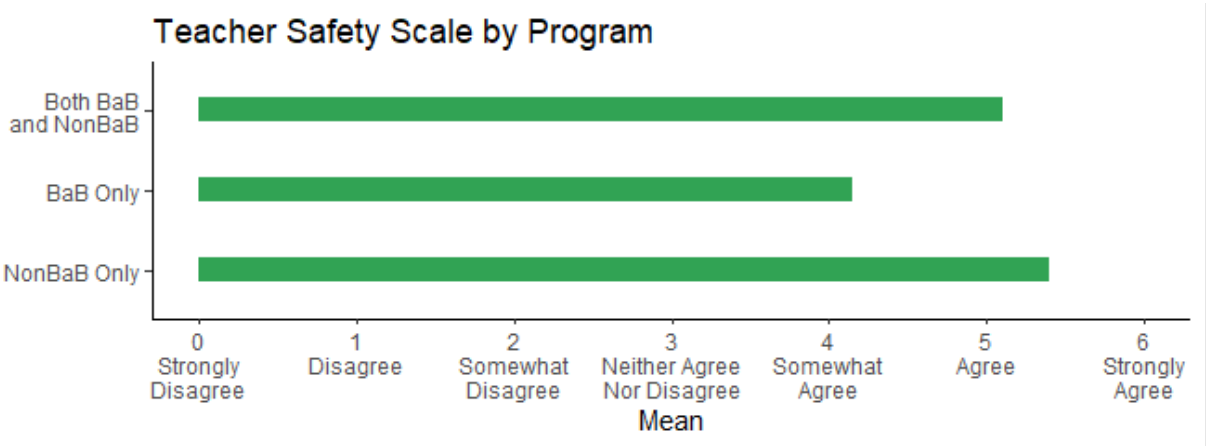


Table 28: Teacher Perceptions of Safety by Location, School Funding Type, Gender, Age, and State

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Teachers Safety Scale			**4.99**	**1.01**	
Program	*	Both BAB and NonBAB	5.11	0.80	0.12
		BAB Only	4.15	1.56	-0.84
		NonBAB Only	5.40	0.50	0.41
Location Type		IDP	4.75	0.95	-0.24
		Rural	4.50	1.87	-0.49
		Urban	5.24	0.49	0.26
Funding Type		BAB	4.15	1.56	-0.84
		Community	5.19	0.69	0.20
		Public	5.11	0.94	0.12
		Private	5.25	0.51	0.26
Gender		Male	5.01	1.07	0.02
		Female	4.91	0.80	-0.08
Age Group		Under 30 Year Olds	5.05	0.89	0.06
		30 - 40 Year Olds	5.19	0.40	0.20
		40 - 65 Year Olds	4.50	1.75	-0.49
State		Benadir	5.10	0.50	0.11
		Hirshabelle	4.69	1.84	-0.30
		Jubaland	5.17	0.93	0.19
		Southwest	4.87	0.91	-0.12

6.3 TEACHER'S FEELINGS OF SELF-EFFICACY

Teacher perception of self-efficacy refers to a teacher's beliefs in their ability to effectively handle the tasks and challenges they encounter in their work and positively impact learner learning. Teachers with a keen sense of self-efficacy exhibit behaviors and strategies in the classroom that support positive learning outcomes (Jerald 2007). Albert Bandura, a leader in the development of self-efficacy theory, links experience, or performance accomplishments, with self-efficacy and argues that self-efficacy may vary across activity types and tasks (Bandura 1977). For example, a person may feel high self-efficacy for teaching reading, but low self-efficacy for teaching math. Bandura created a 30-item teacher self-efficacy scale (unpublished) that includes 7 subscales. Self-efficacy questions on the Longitudinal Teacher Survey are modified from two of Bandura's subscales, instructional and disciplinary self-efficacy.

The Self-Efficacy scale questions were used to assess teacher's confidence in their abilities to effectively teach and manage the behavior of all children in their classroom. Aggregate measures that

include both the composite of instructional and disciplinary subscales are included in Data Annex Section 5.4. Subscale results, which show the type of task-based variation Bandura described, are presented in the following sections.

Teacher Instructional Self-Efficacy Scale

We included 7 questions derived from Bandura's Teacher Self-Efficacy Scale that assess Teacher's self-efficacy regarding instruction. The scale is composed of the items below:

- How much can you do to get through to the most difficult learners?
- How much can you do to promote learning when there is a lack of support from the home?
- How much can you do to increase learners' memory of what they have been taught in previous lessons?
- How much can you do to motivate learners who show low interest in schoolwork?
- How much can you do to get learners to work together?
- How much can you do to overcome the influence of adverse community conditions on learners' learning?
- How much can you do to get children to do their homework?

All items were measured on a 5-point Likert scale (0-nothing, 1-very little, 2-some influence, 3-quite a bit, 4-a great deal) where teachers indicated their perception of self-efficacy.

Longitudinal cohort teachers expressed moderate to high levels of instructional self-efficacy. Male teachers expressed significantly higher levels of self-efficacy compared to female teachers. Interestingly, teachers in rural areas expressed higher feelings of instructional self-efficacy than teachers in IDP or urban areas. While community and private school teachers felt more self-efficacy than public or BAB teachers. Finally, instructional self-efficacy varied by state (Table 29). Figures of results are included in the Data Annex (Section 5.4.1; 8.15).

Table 29: Teacher Instructional Self-Efficacy

Group	Subgroup	Mean	SD	Diff from Overall Mean
All Teachers Instructional Self-Efficacy Scale		**3.15**	**0.62**	
Program	Both BAB and NonBAB	3.25	0.67	0.10
	BAB Only	2.83	0.59	-0.32
	NonBAB Only	3.14	0.32	-0.01
Location Type	IDP	3.03	0.67	-0.12
	Rural	3.50	0.45	0.35
	Urban	3.09	0.62	-0.06

Group		Subgroup	Mean	SD	Diff from Overall Mean
Funding Type		BAB	2.83	0.59	-0.32
		Community	3.42	0.45	0.27
		Public	3.03	0.71	-0.12
		Private	3.31	0.54	0.16
Gender	*	Male	3.27	0.54	0.11
		Female	2.70	0.72	-0.45
Age Group		Under 30 Year Olds	3.11	0.63	-0.04
		30 - 40 Year Olds	3.41	0.68	0.26
		40 - 65 Year Olds	3.07	0.52	-0.08
State		Benadir	3.34	0.50	0.19
		Hirshabelle	3.21	0.81	0.06
		Jubaland	3.15	0.40	0.00
		Southwest	2.97	0.76	-0.18

6.4 TEACHER DISCIPLINARY SELF-EFFICACY SCALE

We included 7 questions derived from Bandura's Teacher Self-Efficacy Scale that assesses teacher feelings of self-efficacy as it relates to managing classroom behavior. The scale is composed of the three items below:

- How much can you do to get children to follow classroom rules?
- How much can you do to control disruptive behavior in the classroom?
- How much can you do to prevent problem behavior on the school grounds?

All items were measured on a 5-point Likert scale (0-nothing, 1-very little, 2-some influence, 3-quite a bit, 4-a great deal) with higher numbers indicating higher feelings of self-efficacy.

As with instructional self-efficacy, longitudinal cohort teachers expressed moderate to high levels of disciplinary self-efficacy overall. Male teachers again felt significantly more confident than their female counterparts. Unlike instructional self-efficacy, no differences were noted based on school location type, but significant differences were noted based on the state in which the teacher was located. Teachers in BAB schools expressed the lowest perceptions of disciplinary self-efficacy compared to all other school types (community, public, private). Data are displayed in Table 30 below and in the Data Annex (Section 5.4.2; 8.16).

Table 30: Teacher Disciplinary Self-Efficacy

Group		Subgroup	Mean	SD	Diff from Overall Mean
All Teachers Disciplinary Self-Efficacy Scale			**3.11**	**0.73**	
Program		Both BAB and NonBAB	3.22	0.73	0.11
		BAB Only	2.73	0.61	-0.38
		NonBAB Only	3.13	0.81	0.02
Location Type		IDP	3.10	0.74	-0.01
		Rural	3.10	0.82	-0.01
		Urban	3.12	0.73	0.01
Funding Type		BAB	2.73	0.61	-0.38
		Community	3.22	0.70	0.11
		Public	3.11	0.91	0.00
		Private	3.29	0.54	0.17
Gender	*	Male	3.23	0.69	0.12
		Female	2.64	0.72	-0.47
Age Group		Under 30 Year Olds	3.13	0.77	0.02
		30 - 40 Year Olds	3.38	0.57	0.26
		40 - 65 Year Olds	2.75	0.64	-0.36
State	*	Benadir	3.33	0.53	0.22
		Hirshabelle	3.00	0.96	-0.11
		Jubaland	3.37	0.72	0.26
	*	Southwest	2.75	0.65	-0.37

Teacher Knowledge and Support

The Longitudinal Teacher Survey included sets of questions that shed light on teachers' feelings of preparation and support. These questions include both internal measures of support (content and pedagogical knowledge) as well as external measures of support (materials, resources, people). The two separate dimensions can be found below. Composite scale results are included in the Data Annex (Section 5.2; 8.10).

6.4.1.1 Teacher Internal Support Dimension Subscale

The internal support subscale examines teachers' perceptions of the pedagogical and content knowledge they need to be an effective teacher. This scale included four statements, as follows:

- I have the content knowledge I need to effectively teach my class.
- I have a range of techniques to effectively teach all learners in my class.

- I have the knowledge and skills I need to effectively teach all children in my class, regardless of their gender, age, or family background, disability, or other characteristics.
- I have various strategies to effectively manage my classroom.

All items were measured on a 7-point Likert scale (0-very untrue, 1-untrue, 2-somewhat untrue, 3-neutral, 4-somewhat true, 5-true, 6-very true) where teachers indicated their perceived level of support.

Teachers generally had favorable beliefs about their knowledge, preparation, and skills for effective teaching. Those teachers that only have BAB classes were somewhat less positive than their peers teaching only non-BAB or a combination of classes (Figure 43). Women, rural, and teachers in Hirshabelle also reported slightly lower perceptions of internal support (Table 31). A table of all data is included in the Data Annex (Section 5.2.1).

Figure 43: Teacher Perceptions of Knowledge, Preparation, and Skills by Program

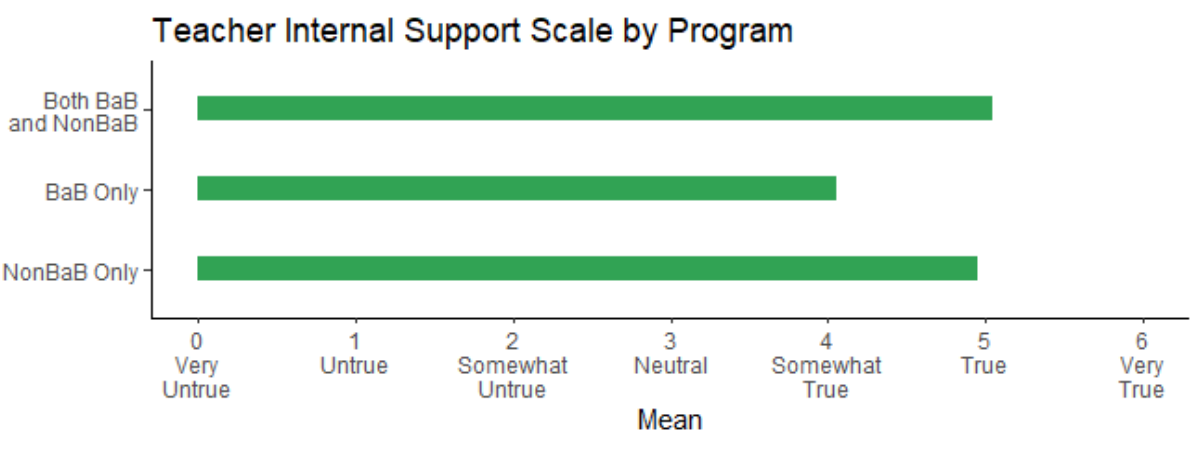


Table 31: Teacher Perceptions of Knowledge, Preparation and Skills by Program, Location Type, School Funding, Gender, Age, and State

Group	Subgroup	Mean	SD	Diff from Overall Mean
All Teachers Internal Support Scale		**4.85**	**1.05**	
Program	Both BAB and NonBAB	5.05	0.76	0.20
	BAB Only	4.05	1.69	-0.80
	NonBAB Only	4.95	0.80	0.10
Location Type	IDP	4.88	0.94	0.04

	Rural	4.68	1.86	-0.17
	Urban	4.89	0.73	0.04
Funding Type	BAB	4.05	1.69	-0.80
	Community	5.06	0.81	0.22
	Public	5.11	0.77	0.26
	Private	4.89	0.75	0.05
Gender	Male	4.90	1.05	0.05
	Female	4.66	1.07	-0.19
Age Group	Under 30 Year Olds	4.81	0.87	-0.04
	30 - 40 Year Olds	5.44	0.44	0.59
	40 - 65 Year Olds	4.44	1.87	-0.41
State	Benadir	5.25	0.27	0.40
	Hirshabelle	4.25	1.95	-0.60
	Jubaland	4.86	0.77	0.01
	Southwest	4.81	1.03	-0.04

Teacher External Support Dimension Subscale

Teacher measures of external support include those materials and resources (human and non-human) that a teacher needs to effectively teach. This can include desks and chairs, books, pens and paper, curricula, administrative and peer support, and a host of other resources. The external support scale included three questions, as follows:

- I have the support I need to effectively teach my class.
- I have the materials I need to effectively teach my class.
- I have people and resources I can draw on when I have challenges in my classroom.

All items were measured on a 7-point Likert scale (0-very untrue, 1-untrue, 2-somewhat untrue, 3-neutral, 4-somewhat true, 5-true, 6-very true) where teachers indicated their perceived level of support.

While teachers at community, public, and private schools had positive responses, teachers that taught only BAB classes and rural teachers had less favorable perceptions of resource availability and external support. Teacher perceptions of external support data are shown in Table 32 and graphically in the Data Annex (Section 5.2.2).

Table 32: Teacher Perceptions of Materials and Support by Program, Location, School Funding, Gender, Age, and State

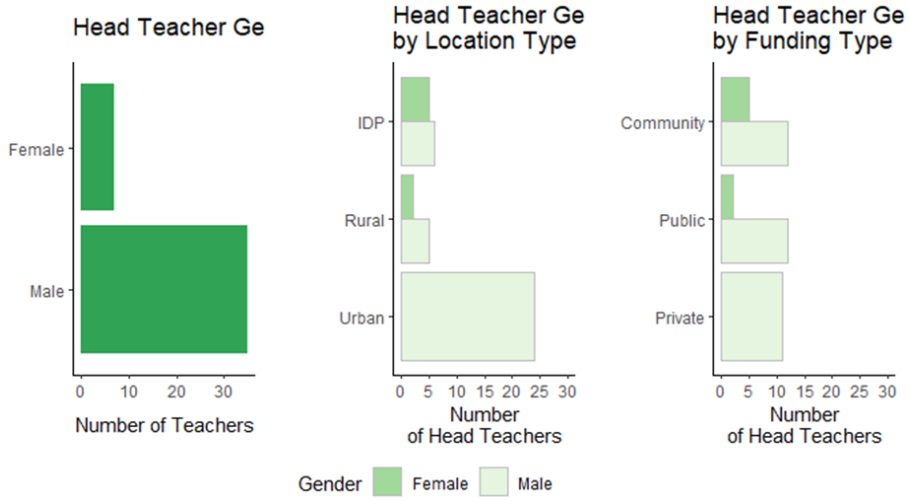
Group		Subgroup	Mean	SD	Diff from Overall Mean
All Teachers External Support Scale			**4.56**	**1.05**	
Program	*	Both BAB and NonBAB	4.83	0.86	0.27
		BAB Only	3.77	1.44	-0.80
		NonBAB Only	4.43	0.83	-0.13
Location Type	*	IDP	4.67	0.95	0.10
		Rural	3.77	1.60	-0.79
		Urban	4.77	0.74	0.21
Funding Type		BAB	3.77	1.44	-0.80
		Community	4.70	0.95	0.13
		Public	4.74	0.96	0.18
		Private	4.79	0.70	0.22
Gender		Male	4.63	1.11	0.07
		Female	4.30	0.78	-0.26
Age Group		Under 30 Year Olds	4.68	0.85	0.12
		30 - 40 Year Olds	4.58	0.89	0.02
		40 - 65 Year Olds	3.96	1.81	-0.60
State		Benadir	4.64	0.62	0.08
		Hirshabelle	4.13	1.97	-0.44
		Jubaland	4.60	0.94	0.04
		Southwest	4.67	0.86	0.10

Head Teacher Sample Characteristics

School leadership has a strong impact on nearly every aspect of teaching and learning and good school leadership has been empirically linked to improved learner achievement (Wallace Foundation. (2013). The Longitudinal Head Teacher baseline survey included questions aimed at understanding indicators of head teacher quality (e.g., preparation, training, attitudes and beliefs), questions related to school policies and procedures, and questions related to learner and community characteristics.

The head teacher baseline sample included 42 head teachers from 42 different schools across all four states. Head teachers are overwhelmingly male, with mean ages ranging from 28 (female) to 32 (male). Rural head teachers were older (35), on average, than urban head teachers (30) and head teachers in IDPs were significantly younger (25) (Figures 44).

Figure 44: Head Teacher Age and Gender

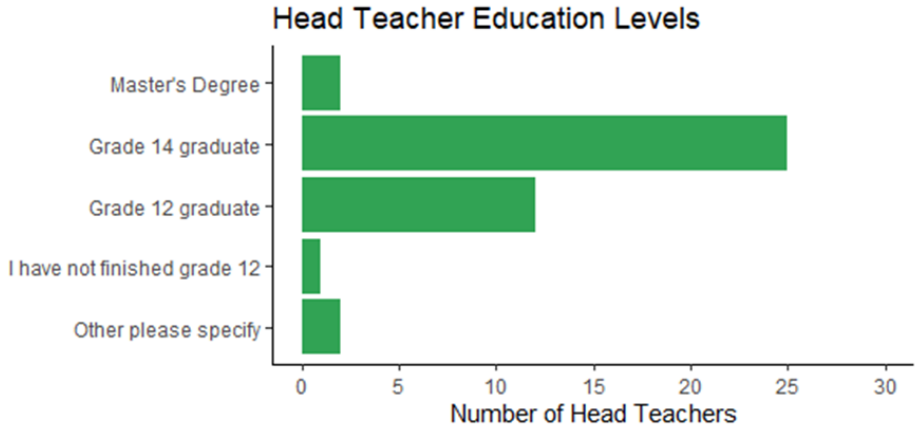


Head Teacher education and experience

All but one head teacher reported at least a 12th grade education, with nearly 2/3 graduating from grade 14 or acquiring a master's degree (Figure 45).

All but 4 head teachers report at least a year of prior classroom teaching experience, see Data Annex (Section 4.3). Nearly 2/3 of the head teachers reported receiving training on leadership and management and the same number reported receiving training on child rights.

Figure: 45 Head Teacher Education Level



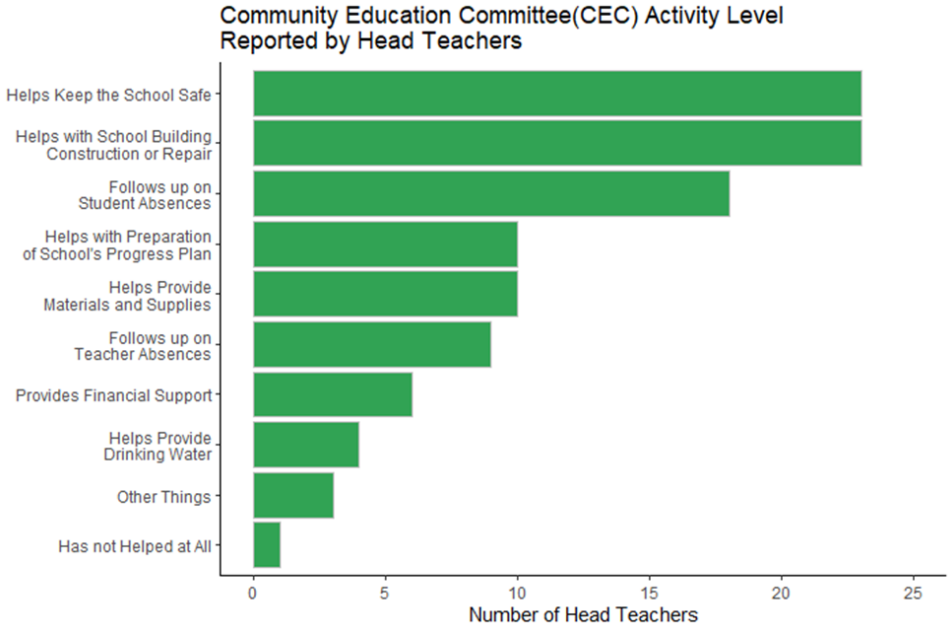
School Attributes and Policies

Sixty percent (25/42) of the sample head teachers reported that their BAB school includes learners with disabilities, while only 50% (21/42) reported that their non-BAB school includes learners with disabilities.

Support for Education

Nearly all head teachers (37/42) reported the presence of a Community Education Committee (CEC) in their school. The top two reported activities for the CECs include school safety and building construction and repair. Other common activities include following up on learner absences, preparing school progress plans, providing materials and supplies, and following up on teacher absences (Figure 46).

Figure 46: CEC Activities



Although most head teachers reported community support for the education of all learners, some teachers found this to be very untrue in their community (Figure 47). Learners identified most often as less likely to receive community support for education were girls and children with disabilities (Figure 48).

Figure 47: Head Teacher Perception of Community Support for Education

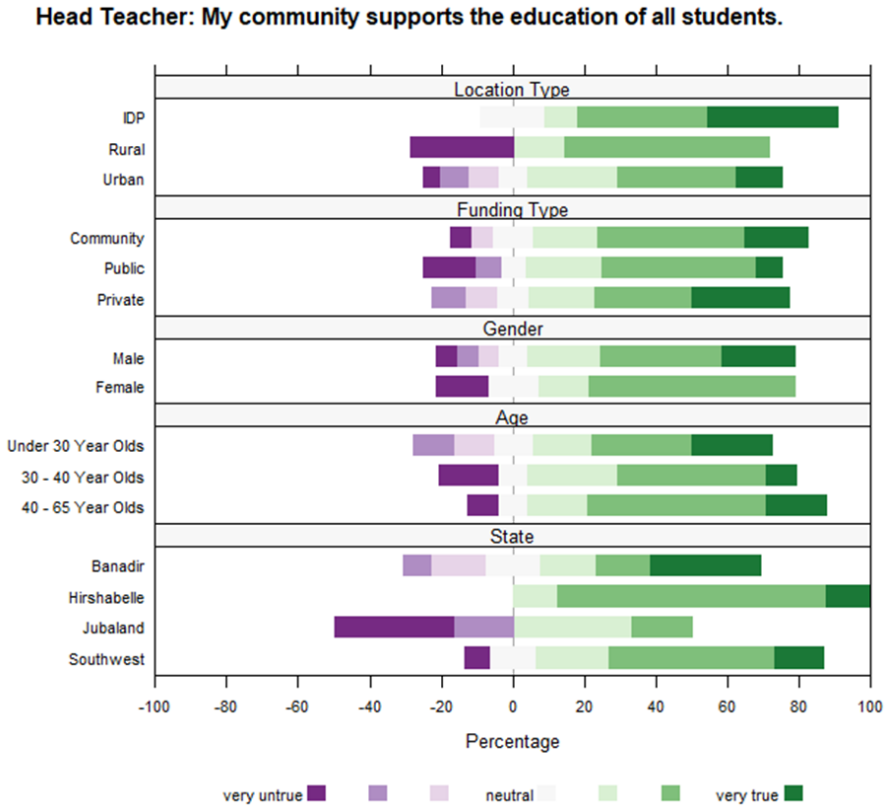
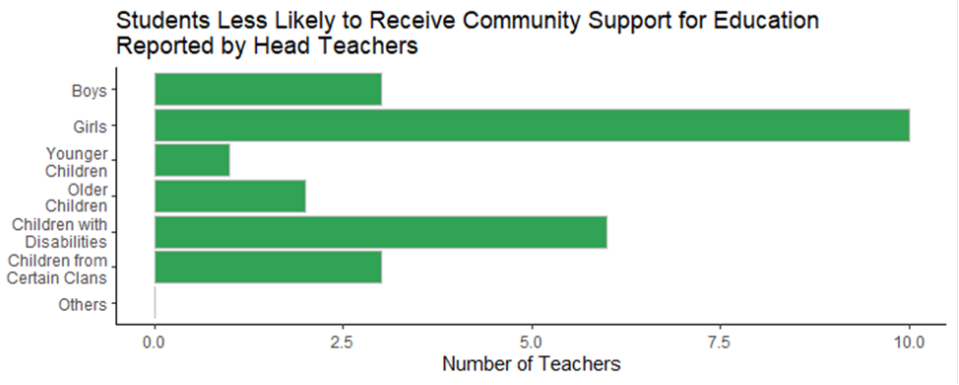


Figure 48: Student Less Likely to Receive Community Support for Education



On the contrary, head teachers felt very confident that their teachers provide the necessary support for all learners to be successful at school (Figure 49) but were split in their perception of their school's ability to accommodate the needs of girls when menstruating (Figure 50).

Figure 49: Head Teacher Perception of Teacher Support for All Learners

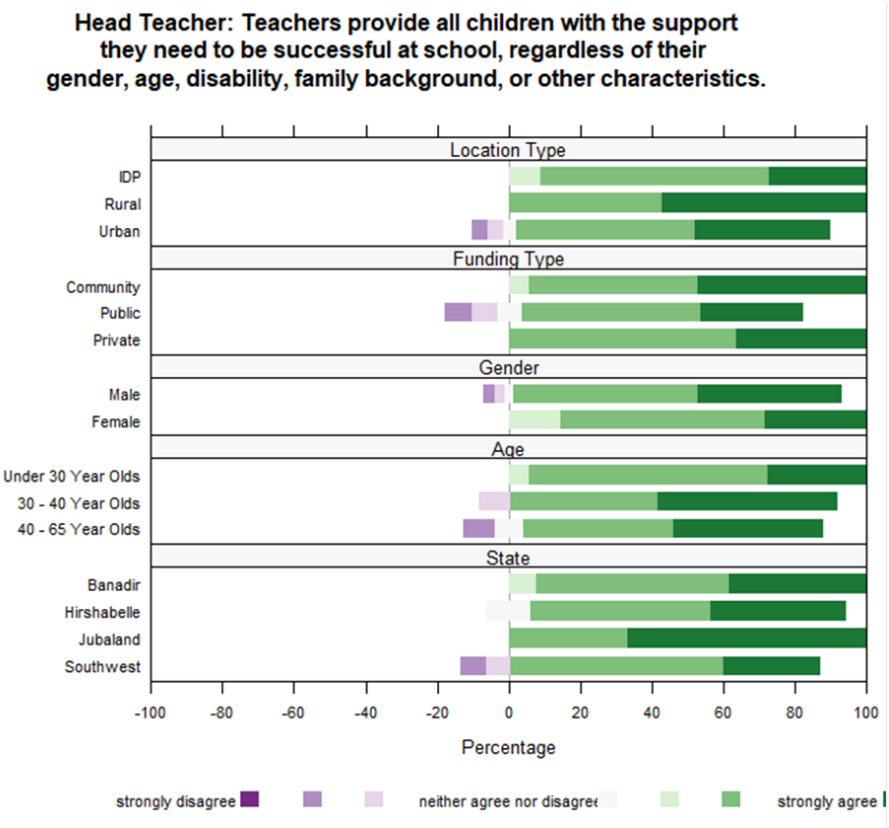
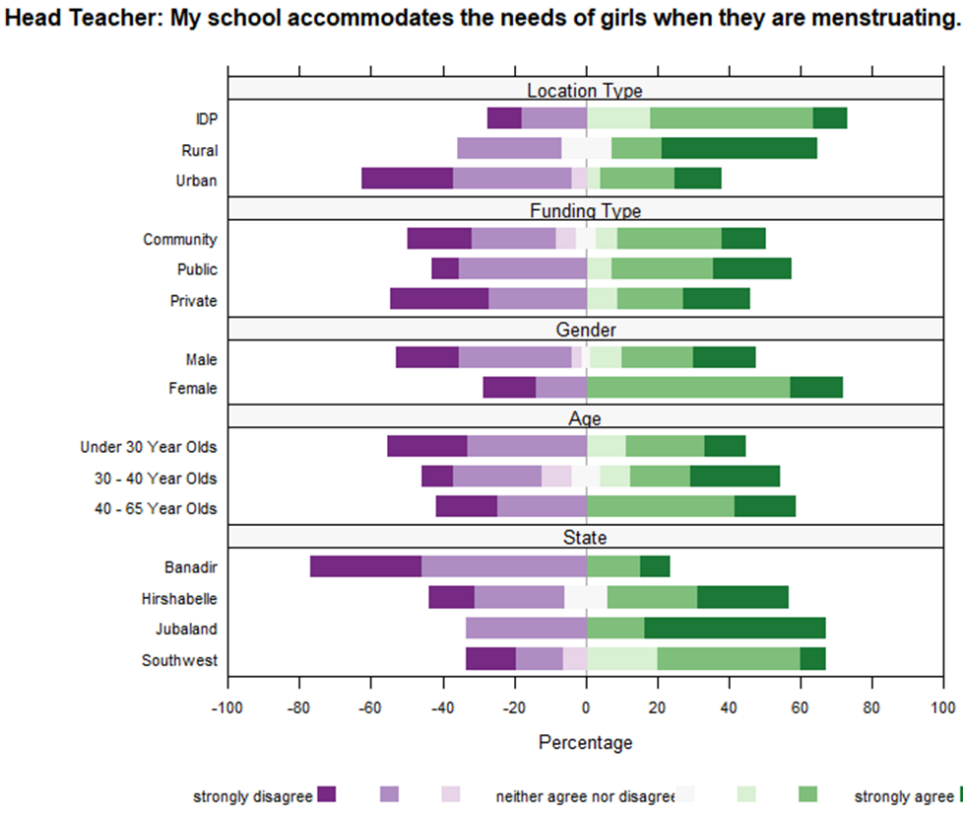


Figure: 50: Head Teacher Preception of School Accommodations for Mentstrating Girls



Access to Education

The Head Teacher survey asked for responses to a series of prompts regarding barriers to education in their community. For each potential barrier, head teachers were asked to identify the group(s) affected including boys, girls, boys and girls, children with disabilities, children from certain clans, all children, others, or not a barrier in my community. Several barriers were identified by most head teachers as primarily affecting all children, including school fees, frequent absence, family migration, malnutrition, lack of family support, lack of community support, safety concerns, transportation or distance to school, insufficient infrastructure, insufficient materials and supplies, lack of teacher support, and poor performance. Barriers identified as primarily affecting girls include culture/tradition, marriage, family chores or work, and abuse by classmates. Barriers identified as disproportionately affecting children with disabilities included culture/tradition, school fees, transportation or distance to school, and abuse by classmates. Finally, barriers identified as particularly relevant to children from certain clans included culture and tradition,

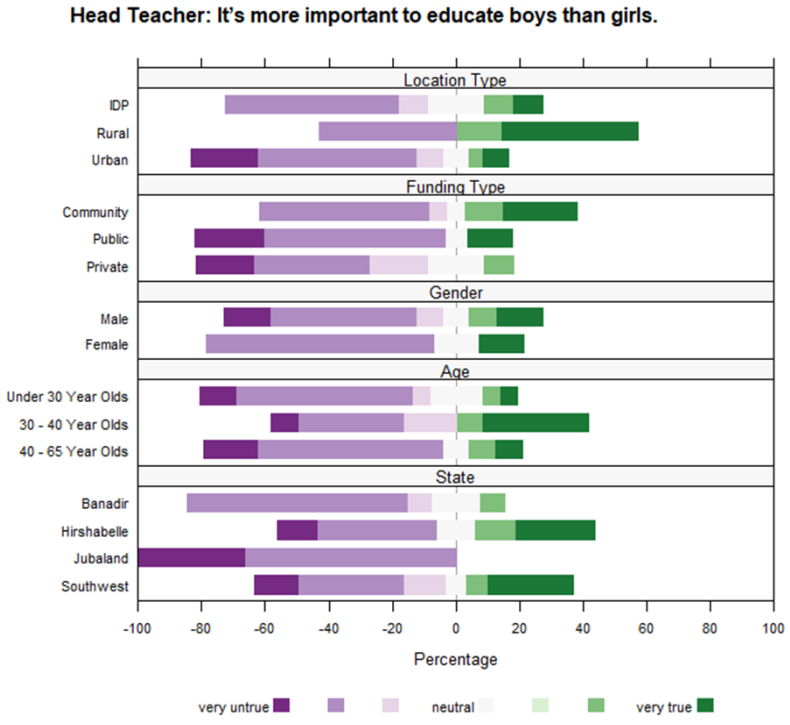
school fees, and abuse by classmates. Graphs of head teacher responses to common barriers for different groups are found in Section 6.7 of the Data Annex.

Head Teacher Mindsets and Perceptions of Equity Attitudes and Beliefs

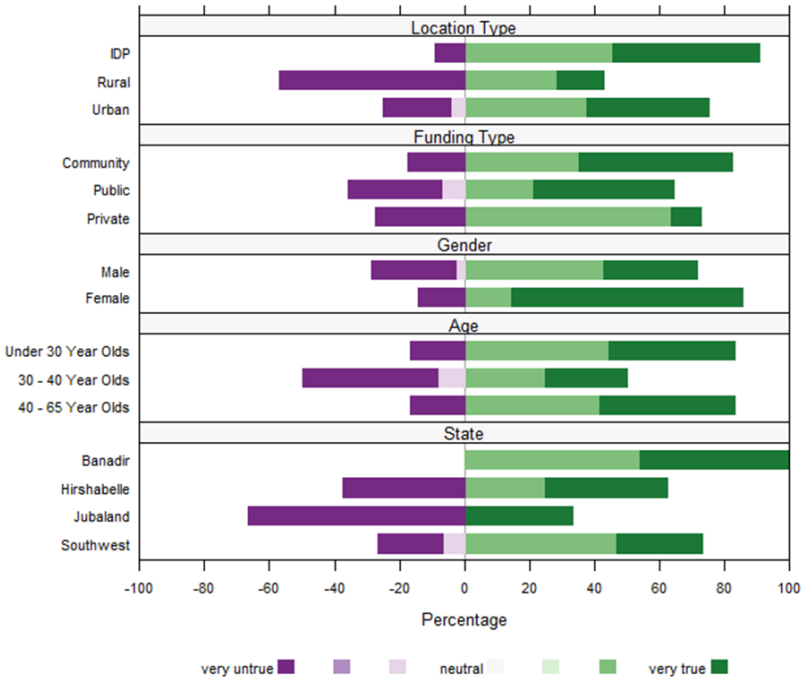
Head teachers were asked several questions that assess their attitudes towards equity and mindset relative to learning in general and for specific groups. Individuals with a static mindset believe that a person's intelligence, talent, and other qualities are innate and cannot be changed. These individuals believe that if you are not good at something, you will never be good at it. In contrast, people with a growth mindset believe that talent and intelligence can be developed with practice and effort.

Head teacher attitudes towards equity and mindsets varied considerably. Most head teachers did not agree that "boys are naturally better at school than girls" (Figure 51a) and did agree that "educating girls is important for society's development" (Figure 51b) and completing primary education is equally important for all children, regardless of gender, disability, or family characteristics" (Figure 51c). Taken together, answers to these three questions suggest commitment to equitable teaching of all learners. It is notable, however, that head teacher beliefs around equity vary by location and by state.

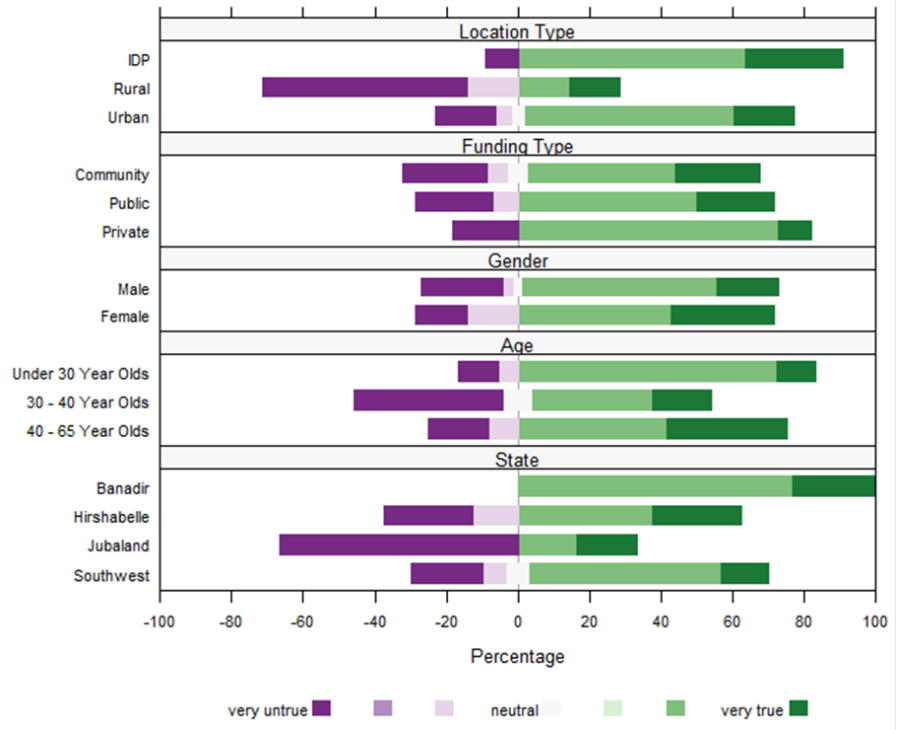
Figure 51: Head Teacher Attitudes Towards Equitable Teaching



Head Teacher: Educating girls is important for society's development.



Head Teacher: Completing primary education is equally important for all children, regardless of gender, disability, or family characteristics.



Several questions focused on head teacher's mindset. Responses to the statement, "I can predict with accuracy the ability of learners in my school to solve problems if I know their gender, family background, and disability status," varied from very untrue to very true (Figure 52). For this question, responses that disagree could be indicative of more growth mindsets, while answers affirming this statement indicate a more fixed or static mindset towards learning potential. On the other hand, responses might simply reflect a pragmatic view of the impact of resource limitation on learning. Responses were less variable for the statement, "Learners have a certain amount of intelligence and teachers can't really do much to change it," which most head teachers felt was false (Figure 53). Most head teachers agreed with a final statement, "All learners have the capacity to learn," (Figure 54). Midline and endline evaluations will explore these themes in more detail and examine links between head teacher mindsets, teacher attitudes and beliefs, and learner outcomes. Additional head teacher data is included in the Data Annex (Section 6).

Figure 52: Head Teacher Attitudes towards Student Ability

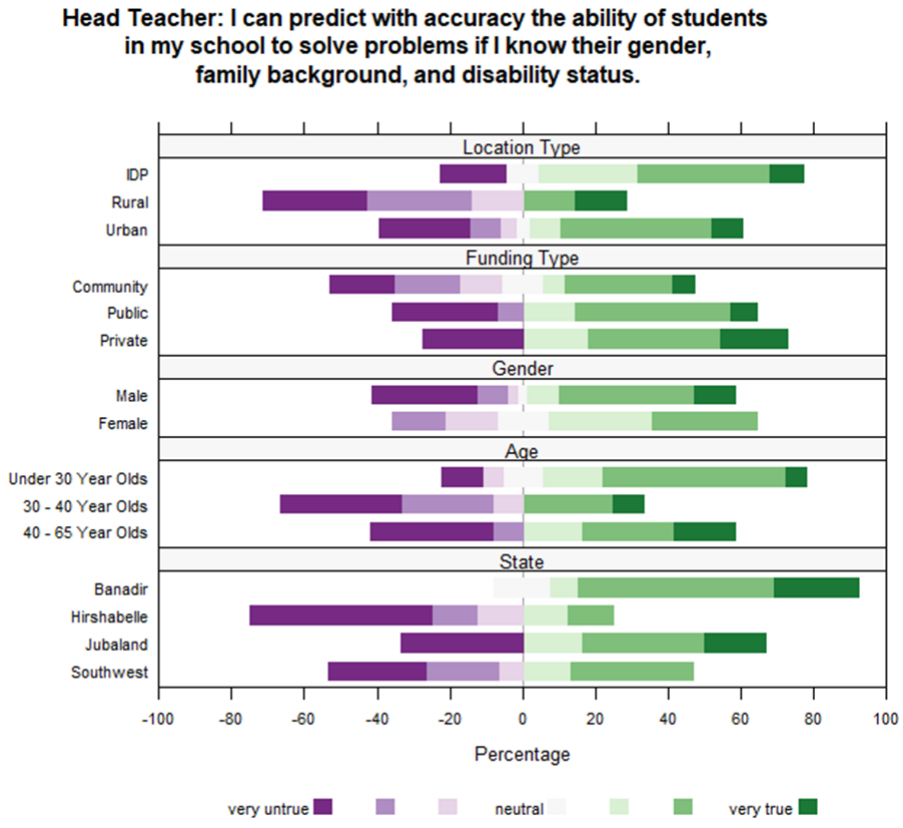


Figure 53: Head Teacher Attitudes toward Student Intelligence

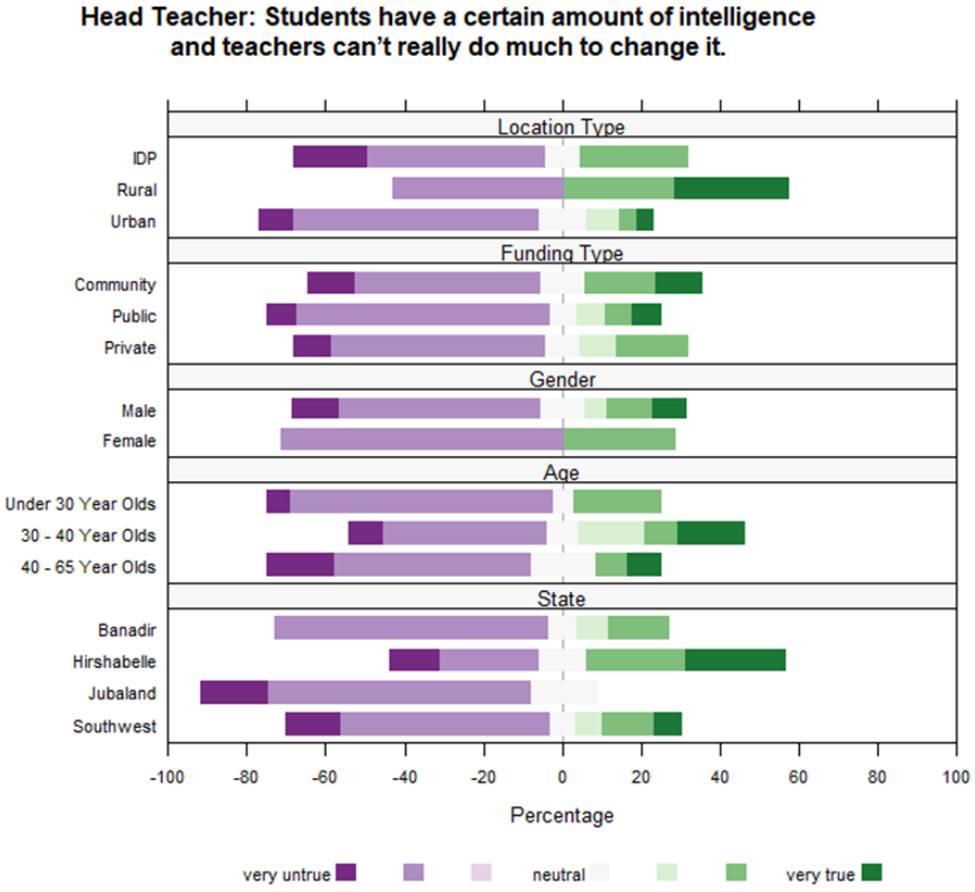
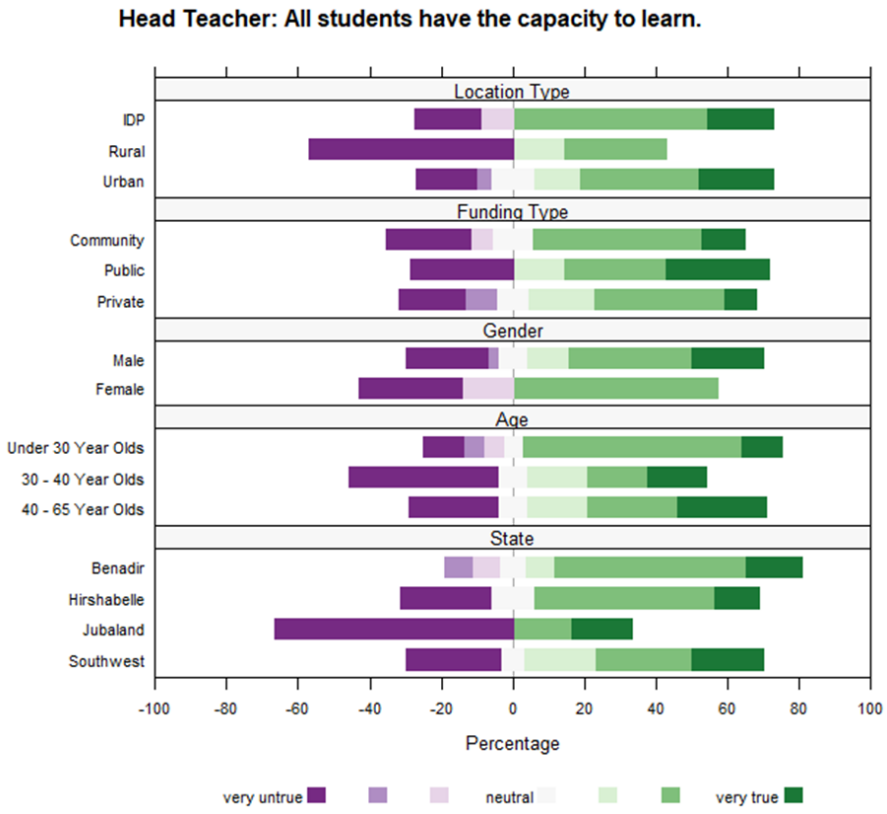


Figure 54: Head Teacher towards Student Capacity to Learn



7 IMPACT OF SAMPLE SELECTION METHODOLOGY (CROSS SECTIONAL AND LONGITUDINAL SAMPLES)

As previously described, the external evaluation longitudinal sample and BAB's cross-sectional sample both employed a random sampling approach to cohort selection. However, two key differences in sampling approach should be noted: 1) the evaluation purposefully oversampled rural and IDP schools while BAB drew sample classrooms randomly proportional to size, and 2) the longitudinal sample included all children in selected classrooms, while BAB randomly selected 10-12 learners in each of the randomly selected classrooms. The differences in sample selection methodologies reflect the different purposes for the data.

The longitudinal sample will follow cohort learners through two years of participation to measure changes in outcomes and indicators of interest over time. The evaluation employed a cluster sampling approach whereby we randomly selected classrooms as clusters and included all members from those clusters in the sample. This classroom level selection protocol provides several affordances. First, it minimizes the risk of bias that might be introduced by selecting a sample of individuals from within classroom. Second, it ensures representation of all learners in a heterogeneous classroom, Third, it decreases risks associated with perceptions that some children

are chosen while others are not. Finally, the smaller number of classrooms simplifies data collection logistics.

BAB's cross-sectional cohort used stratified sampling to select classrooms and then chose 10-12 learners from each classroom. This stratified approach ensures representation across a greater number of schools but might miss some crucial population differences within a very heterogeneous classroom. For example, learners from different ability levels might react differently under the same teacher's instruction; with just 10-12 learners in one classroom, the sample size is not large enough to examine different ability levels in the same classroom. Moreover, because of the high heterogeneity of the BAB learners in one classroom, whether the 10-12 could be a representative sample of each classroom is questionable.

The following tables compare longitudinal and cross-sectional cohorts by state (Table 33), district (Table 34) and location type (Table 35). Analysis shows that the two sampling techniques resulted in quite different samples. In Table 36, the p values from Pearson's chi-square tests showed the longitudinal sample is statistically significantly different from the cross-sectional sample in every demographic aspect except gender. Most notably, the longitudinal sample has higher components of lower age learners, rural learners, and IDP learners. The cross-sectional sample has 73.6% Urban learners, 16.6% IDP learners, 9.8% Rural learners, while longitudinal sample has 53.3% Urban, 27.4% IDP and 19.3% Rural, as expected.

Table 33: Longitudinal and Cross-Sectional Sample Composition by State

State	Longitudinal		Cross-sectional	
	# of learners	# of sites	# of learners	# of sites
Benadir	405 (24%)	10 (25%)	289 (24%)	21 (23%)
Hirshabelle	334 (20%)	8 (20%)	310 (26%)	19 (20.5%)
Jubaland	316 (18%)	7 (17.5%)	180 (15%)	16 (6.5%)
Southwest	659 (38%)	15 (37.5%)	429 (35%)	36 (39%)
Total	1714 (100%)	40 (100%)	1208 (100%)	92 (100%)

Table 34: Longitudinal and Cross-Sectional Sample Composition by District

District	Longitudinal		Cross-sectional	
	# of learners	# of sites	# of learners	# of sites
Baidoa	379 (22.1%)	9 (22.5%)	179 (14.8%)	15 (16.3%)
Balcad	155 (9.0%)	4 (10.0%)	133 (11.0%)	7 (7.6%)

Barawe	88 (5.1%)	2 (5.0%)	83 (6.9%)	7 (7.6%)
Deynile	177 (10.3%)	4 (10.0%)	129 (10.7%)	9 (9.8%)
Diinsor	40 (2.3%)	1 (2.5%)	60 (5.0%)	5 (5.4%)
Hamarwayne	45 (2.6%)	1 (2.5%)	24 (2.0%)	1 (1.1%)
Jowhar	179 (10.4%)	4 (10.0%)	177 (14.7%)	12 (13.0%)
Kahada	149 (8.7%)	4 (10.0%)	92 (7.6%)	8 (8.7%)
Kismayo	316 (18.4%)	7 (17.5%)	180 (14.9%)	16 (17.4%)
Shibis	34 (2.0%)	1 (2.5%)	44 (3.6%)	3 (3.3%)
Walanweyn	152 (8.9%)	3 (7.5%)	107 (8.9%)	9 (9.8%)
Total	1714 (100.0%)	40 (100.0%)	1208 (100.0%)	92 (100.0%

Table 35: Longitudinal and Cross-Sectional Sample Composition by Location Type

Location	Longitudinal		Cross-sectional	
	# of learners	# of sites	# of learners	# of sites
IDP	470 (27.4%)	11 (27.5%)	201 (16.6%)	17 (18.5%)
Rural	331 (19.3%)	8 (20.0%)	118 (9.8%)	10 (10.9%)
Urban	913 (53.3%)	21 (52.5%)	889 (73.6%)	65 (70.7%)
Total	1714 (100.0%)	40 (100.0%)	1208 (100.0%)	92 (100.0%

Table 36: Sample Comparison of Longitudinal vs. Cross-sectiona

Descriptions	Longitudinal Sample	Cross Sectional Sample	p value
State			<.001***
Benadir	405 (23.6%)	289 (23.9%)	–
Hirshabelle	334 (19.5%)	310 (25.7%)	–
Jubaland	316 (18.4%)	180 (14.9%)	–

Southwest	659 (38.4%)	429 (35.5%)	–
Location			<.001***
IDP	470 (27.4%)	201 (16.6%)	–
Rural	331 (19.3%)	118 (9.8%)	–
Urban	913 (53.3%)	889 (73.6%)	–
Gender			.18
Female	848 (49.5%)	579 (47.9%)	–
Male	866 (50.5%)	627 (51.1%)	–
Age			<.001***
4 to 11 years old	1000 (58.3%)	503 (41.6%)	–
12 to 20 years old	714 (41.7%)	705 (58.4%)	–

Note: * = significant at p < .05; ** = significant at p < .01; *** = significant at p <.001.

The evaluation compared results from BAB's cross-sectional EGRA and EGMA baseline assessments with results from the longitudinal cohort assessments to examine the effect of sampling on results. The higher percentage of rural and IDP learners in the longitudinal sample also led to differences in average cohort performance on EGRA and EGMA assessments. As illustrated in Figures 55 and 56, the longitudinal sample has many more learners categorized as non-learner or emergent learners on both the EGRA (64%) and EGMA 61%) at baseline than the cross-sectional sample (43% on both).

Additionally, the evaluation team performed item analyses (difficulty and discrimination) on the cross-sectional data and results reveal significant differences between the longitudinal data and cross-sectional data. Item difficulty indices on EGRA ranged from 0.07 to 0.99, indicating a wide range of difficulty levels; from very difficult to very easy. Item discrimination indices were mostly satisfactory. However, nine (9) items out of two hundred and ninety-five (295) have discrimination indices below the acceptable threshold of 0.3 or greater. Results from the item analysis on EGMA reveal a range of difficulty indices from 0.02 - 0.9, also indicating a wide range of difficulty similar to EGRA. Of the one hundred and seven (107) items in this assessment, twelve (12) have discrimination indices below the acceptable threshold, out of which six (6) were problematic and pose a threat to validity of the assessment. Others were satisfactory. Details of the item analysis

results from the cross-sectional cohort data are presented in Table A3-9 and Table A3-10 in Appendix 3.

Furthermore, the evaluation estimated floor and ceiling effects on the EGRA and EGMA data from the cross-sectional cohort and found no significant floor or ceiling effects present. On the EGRA, approximately 2.2% of learners scored 10% or less and no one scored a zero (0). Meanwhile, approximately 5.4% of learners scored 90% or greater, while only 0.08% of the learners scored 100%. On the EGMA, approximately 8.4% scored 10% or less, while 1.7% scored a zero (0). No learners scored 90% or greater. See details in Table A3-11 in Appendix 3.

Overall, the items performed well in both the longitudinal and cross-sectional cohorts, however, items with discrimination indices highlighted in red require review to understand the reason(s) they performed poorly and to determine the need for deletion or modification for future data collection. We recommend excluding problematic items while making benchmarking decisions due to validity concerns.

Figure 55: EGRA Baseline Scores by Cohort Type

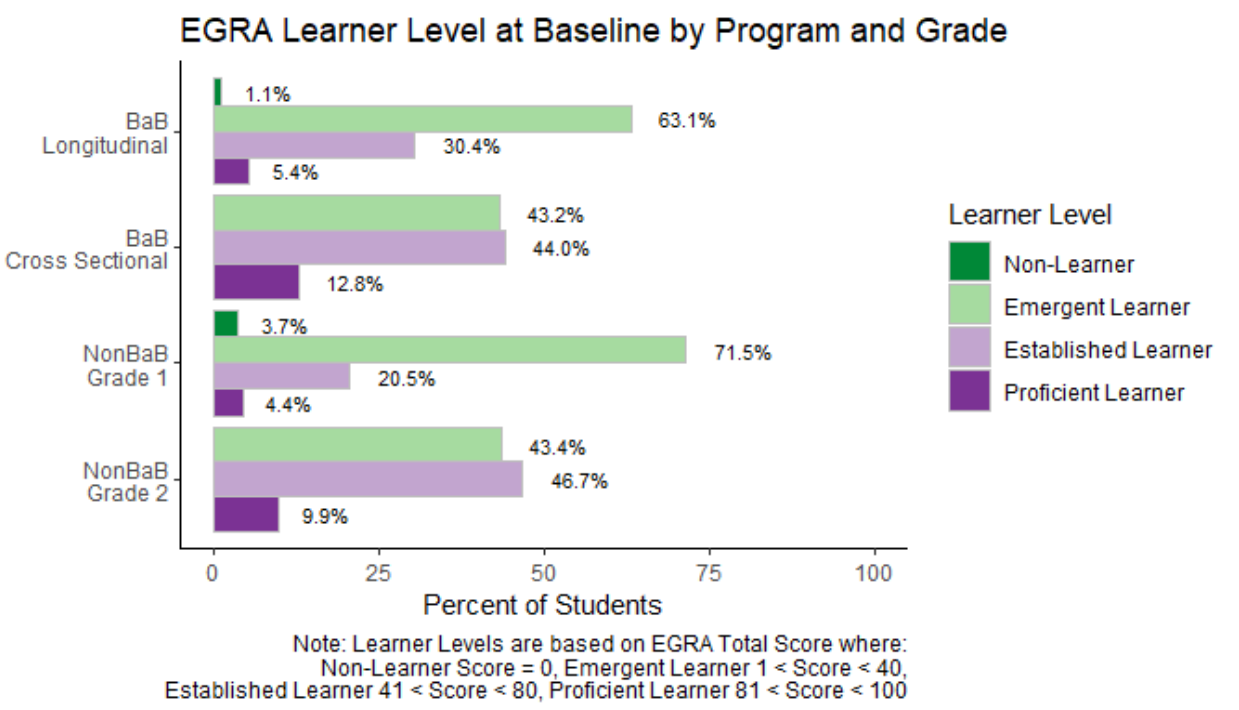
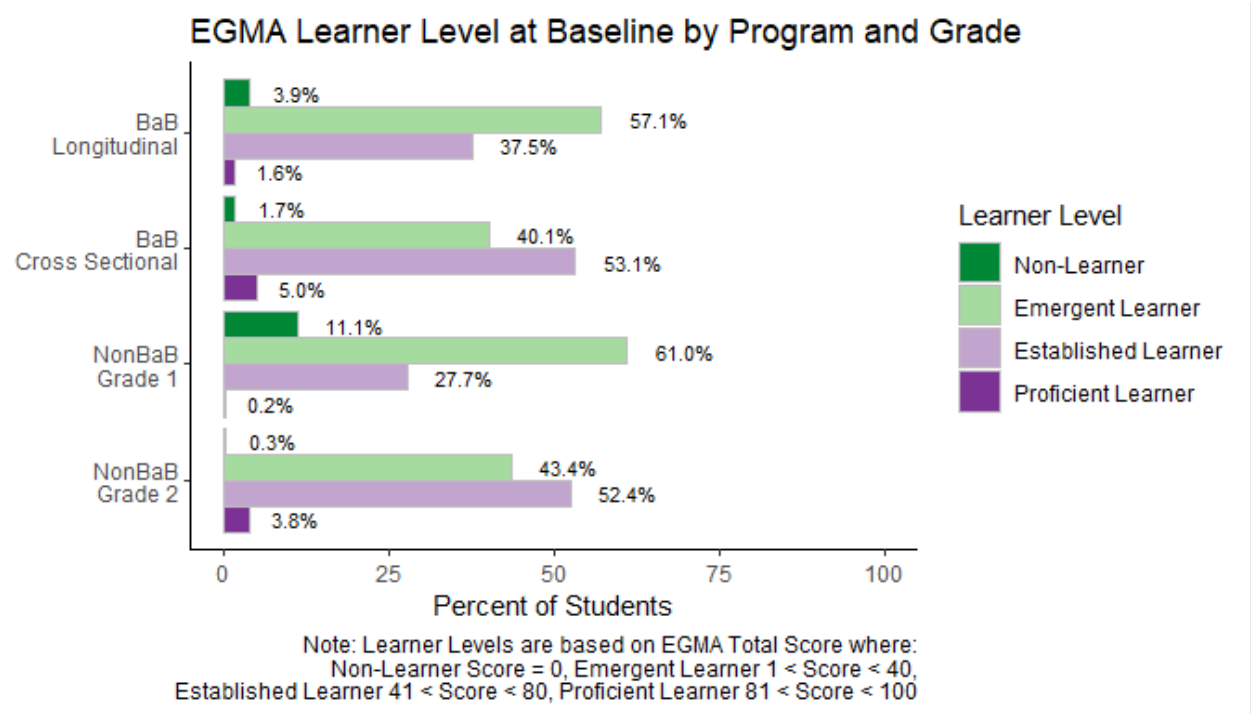


Figure 56: EGMA Baseline Scores by Cohort Type

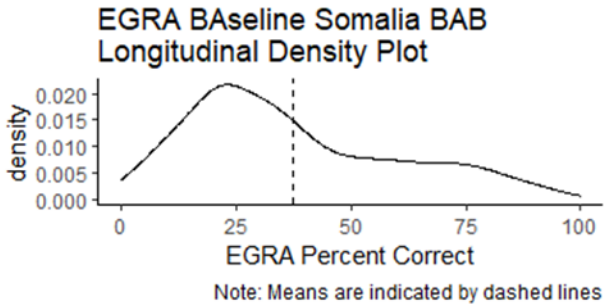


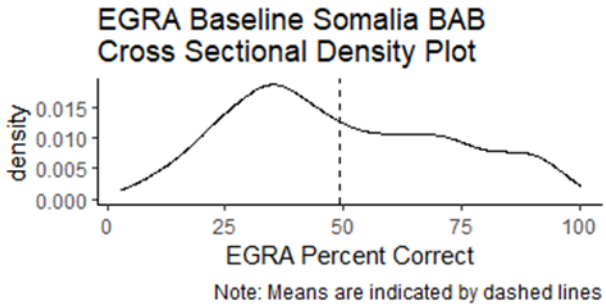
Density plots (Figure 57) show the differences in total EGRA and EGMA scores for each ABE cohort type as well as for non-BAB first and second grade samples. The mean baseline EGRA and EGMA total scores for the cross-sectional sample are the same as the mean baseline EGRA and EGMA scores for the non-BAB grade 2 sample.

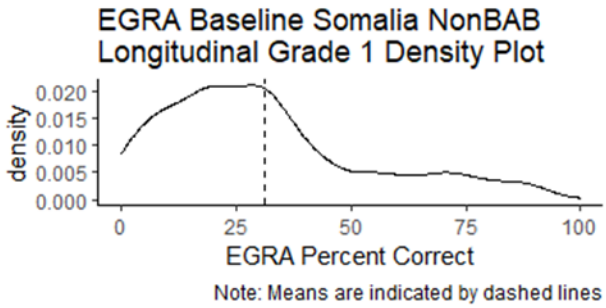
The series of oral reading fluency vs comprehension box and whisker plots in Figure 58 illustrate the impact sampling can have on criteria selection setting literacy standards and benchmarking, as the different samples provide a range of oral fluency scores at 80% comprehension.

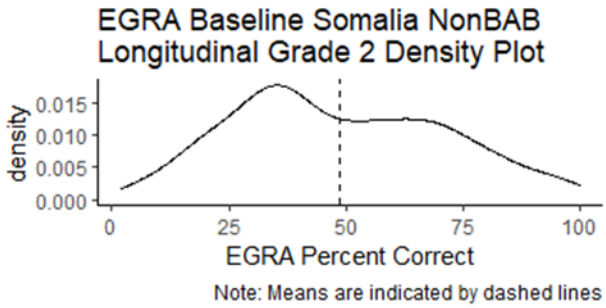
Figure 57: EGRA and EGMA Density Plots for Longitudinal, Cross-Sectional, and Non-BAB Grade 1 and 2 Cohorts

Baseline EGRA Overall Density Plot Results Longitudinal and Cross-Sectional Learner Data

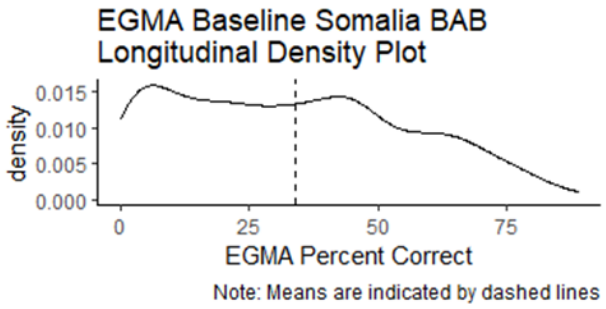


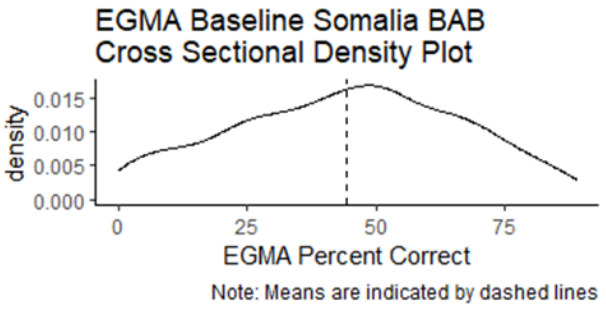


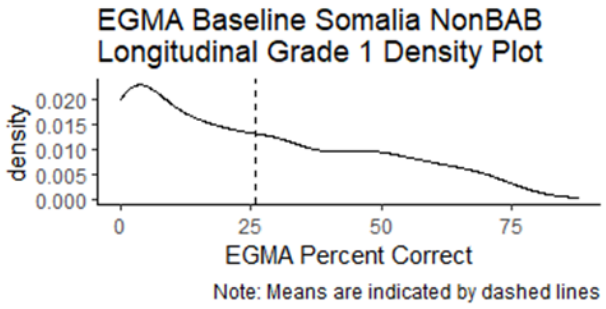




Baseline EGRA Overall Density Plot Results Longitudinal and Cross-Sectional Learner Data







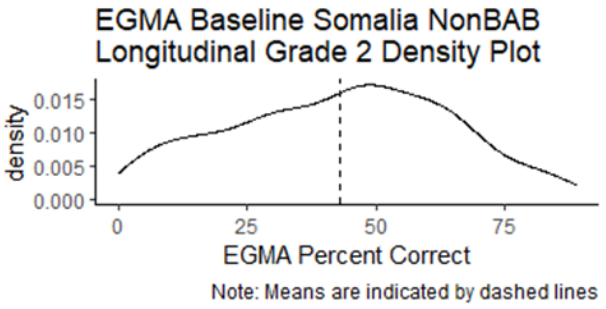
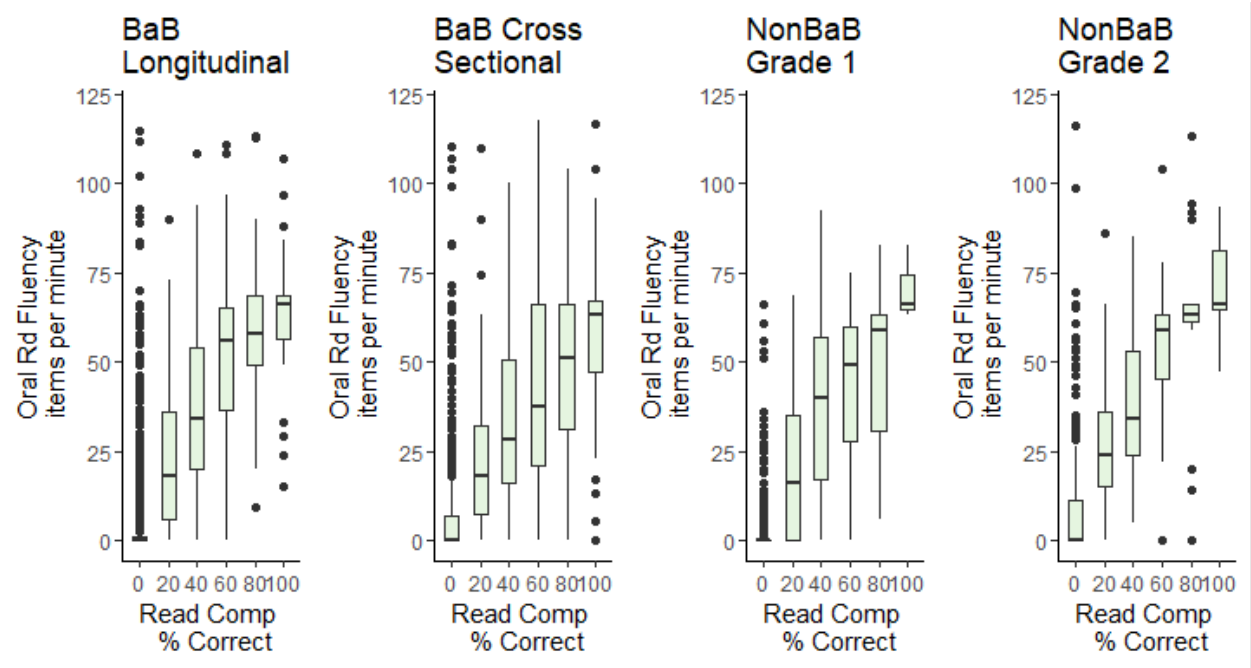


Figure 58: Effect of Sampling Cohort on Literacy Proficiency Scores

Baseline Longitudinal and Cross-Sectional Oral Reading Fluency vs. Reading Comprehension on EGRA



One of the biggest differences between the longitudinal and cross-sectional samples is learner age, with the longitudinal sample including a larger number of younger learners (Figure 59). This difference is impactful, as age is the variable most strongly correlated with academic achievement in our analysis. Figures 60 and 61 show EGRA and EGMA scores by learner age at baseline for both cohorts (both including age 7-17 learners only according to the cross-sectional data preference).

While age has the strongest correlation with learner achievement, learner SES, maternal literacy, and quality of life indicators are also significantly correlated in our analysis. As the longitudinal cohort over-represents rural learners, who score lower on these constructs in our analysis, it is not surprising that longitudinal baseline scores at all age levels are lower than the cross-sectional scores at the same age. These relationships are shown in the lollipop graphs in Figure 62 and 63.

Figure 59: Student Age and Frequency for Longitudinal vs Cross-section Cohorts

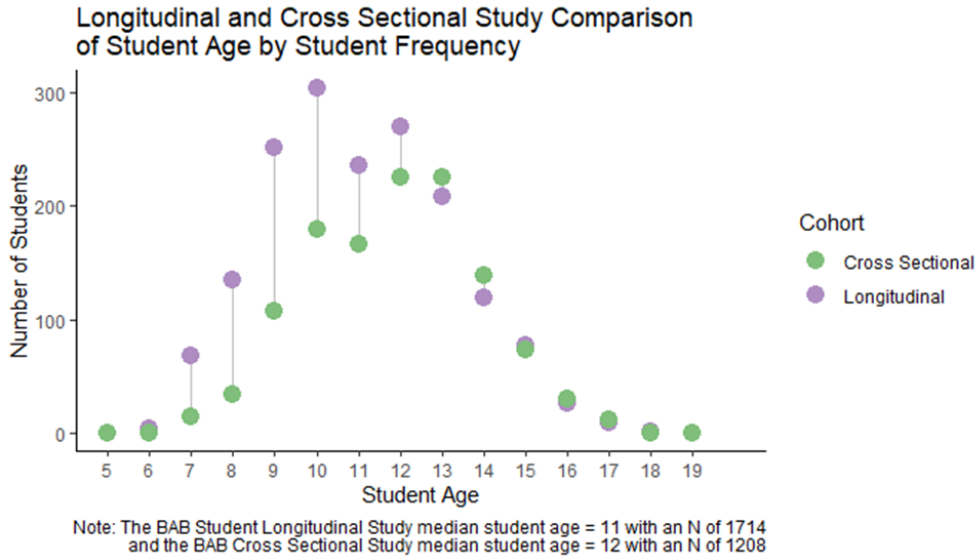


Figure 60: Longitudinal Baseline EGRA and EGMA Scores by Age

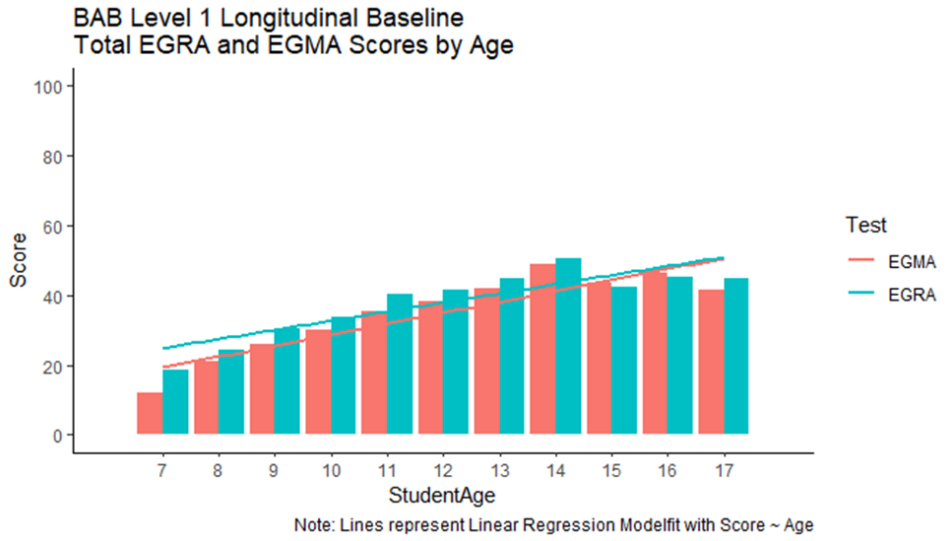


Figure 61: Cross-sectional Baseline EGRA and EGMA Scores by Age

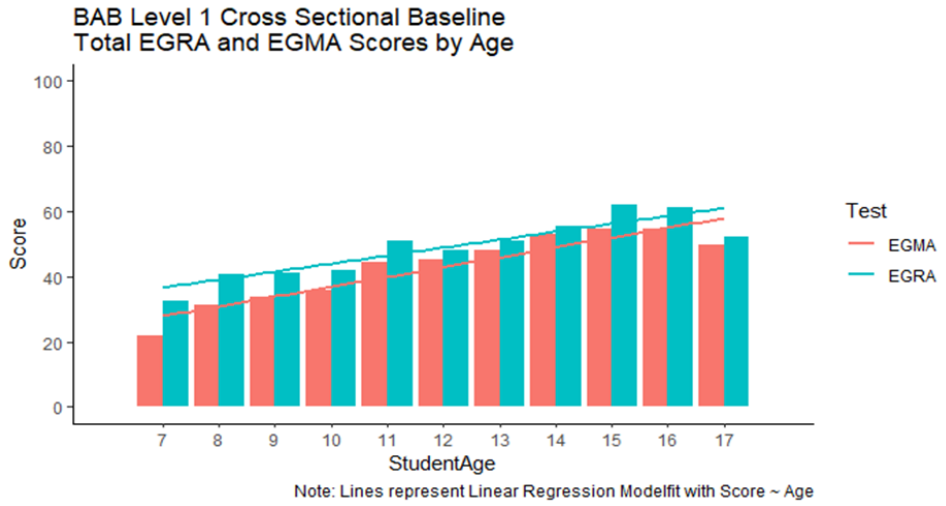


Figure 62: Longitudinal EGRA and EGMA Score Comparison by Age

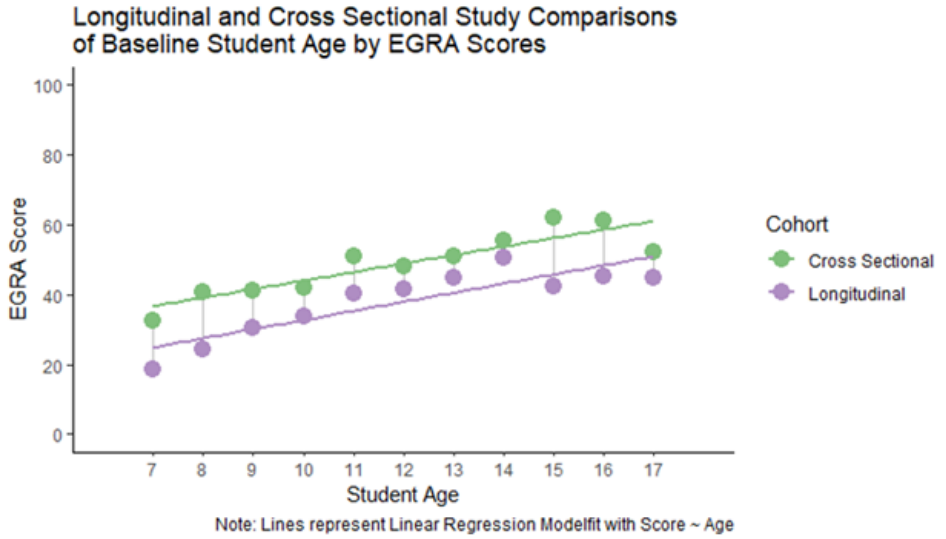
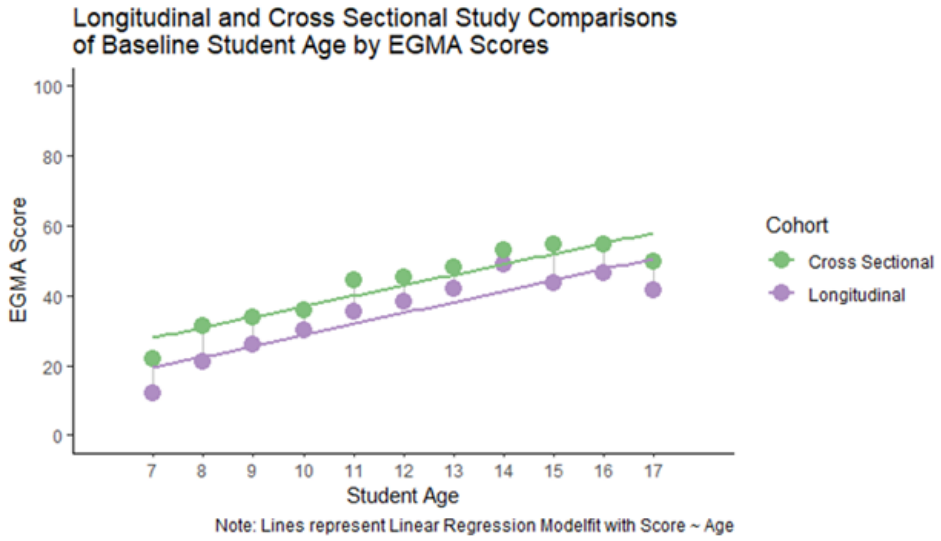


Figure 63: Cross-Sectional EGRA and EGMA Score Comparison by Age



7.1 Cost Analysis

This section provides a preliminary analysis of the cost data provided by Creative Associates and outlines the cost analysis plan for the project. All the calculations are based on the expenditure data provided by Creative Associates from April 2020 to September 2021.

Cost Analysis Questions

To answer the cost analysis questions the evaluation team needs expenditure data, cost of activity components, contributions, beneficiary opportunity cost, and intervention details (duration of teacher training, geographic coverage of intervention, dosage, etc.). Table 37 displays the cost analysis questions and the associated cost analysis methods, and data requirements. The cost analysis questions were collaboratively developed by USAID, and the evaluation and implementation (Creative) teams.

Table 37: Cost Analysis Questions

No.	Question	Method	Data requirement
1.	What is the cost per learner of increasing one proficiency level in reading and math skills (e.g.: non-learners, basic learners, emergent learners, etc.)?	Cost Effectiveness	1) Estimates of the impact of SEL intervention on reading and math skills
2	A. What is the cost per learner of expanding the BAB reading, math, and SEL intervention package in intervention areas? This will be estimated from USAID	Cost Efficiency Prospective Analysis	1) Information on output. 2) Information and data about the new areas where the

	and the Government of Somalia's perspective. B. What is the cost per learner of replicating the BAB reading, math, and SEL intervention package in new areas? This will be estimated from USAID and the Government of Somalia's perspective.		intervention is to be implemented. 3) Information from USAID and Government of Somalia's perspective
3	What is the total cost per learner of SEL instruction, disaggregated by levels (L0, L1, L1-L2)?	Cost Efficiency	1) Information on output. 2) Information and data required for stakeholder analysis (Government, USAID, Private/non-profit organizations)
4	What are the start-up costs for the BAB program?	Cost Economy	1) Cost categories and activities are related to the start-up cost? 2) Start and end dates for these activities 3) Local price database 4) Output data
5	What is the development cost per learner of teaching and learning material?	Cost Economy	1) Cost categories, contributions, activities are related to the development cost of TLM 2) Start and end dates for these activities 3) Local price database 4) Output data
6	What is the procurement cost per learner of teaching and learning material?	Cost Economy	1) Local price database 2) Output data
7	What is the implementation (production and distribution) cost for teacher training, per teacher?	Cost Efficiency	1) Information on output

8	A. What is the estimated unit cost of teacher training when scaled up in intervention areas? B. What is the estimated unit cost of teacher training when replicated in new areas?	Cost Efficiency Prospective Analysis	1) Information on output. 2) Information and data about the new areas where the intervention is to be implemented. 3) Information from USAID and Government of Somalia's perspective
9	What is the cost of classroom/teaching infrastructure, per learner?	Cost Economy	1) Local price database 2) Output data
10	What is the cost of community engagement, per learner?	Cost Economy	1) Local price database 2) Output data 2) Information and data required for stakeholder analysis (Government, USAID, Private/non-profit organizations)

Note: All the questions require the expenditure and contributions reports disaggregated by start-up cost categories and ingredients. Column 3 notes the additional information required for answering the questions.

Cost Expenditures

Table 38 shows the expenditure data using cost categories for the intervention. The expenditure data was provided from April 2020 to September 2021. The highest expenditure was in the general management and operations category (36%) followed by the non-ingredient data and fixed fee (complementary activities; 35%). Block grants accounted for 11% of the expenditure and development and implementation of teaching and learning material accounted for 6.5% of the total expenditure. Safe school and infrastructure (0.3%) and community engagement (0.6%) categories amounted for the lowest expenditure (see Figure 64).

Table 38: BAB expenditure data using cost categories reported from April 2020 to September 2021

Summary Expenditure Data	**Amount (USD)**	**Percent**
Cost Category 1: General Management and Operations	3,167,375	35.96
Cost Category 2: M&E and Reporting	424,735	4.82
Cost Category 4: Teacher Training	161,291	1.83

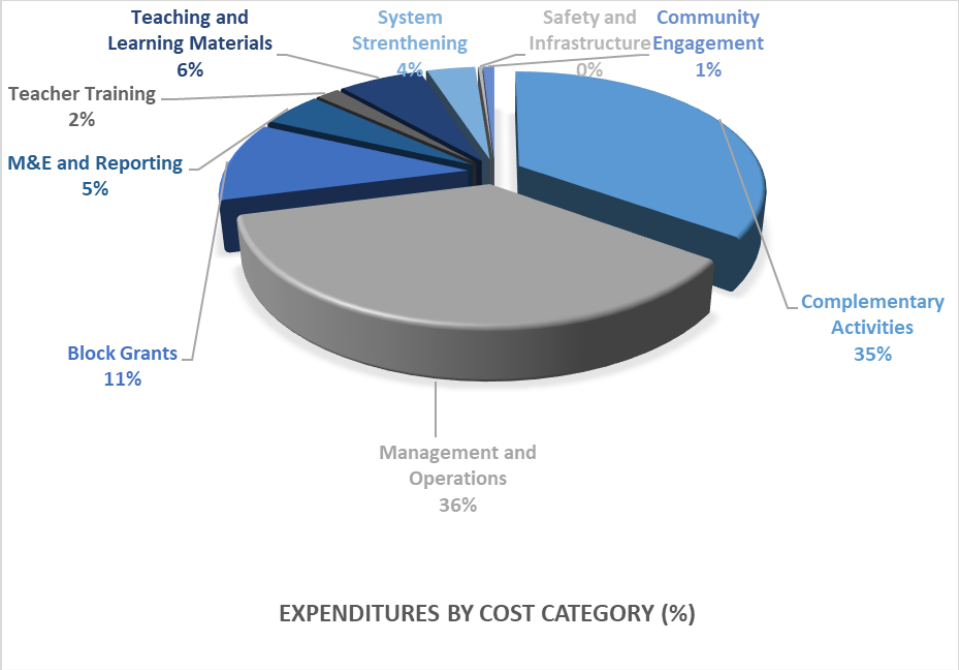
Subcategory 4: Development	27,205	0.30
Subcategory 4: Implementation	134,086	1.52
Cost Category 5: Teaching and Learning Materials	576,384	6.54
Subcategory 5: TLM Development	386,3.8	4.38
Subcategory 5: Production and Distribution	190,076	2.15
Cost Category 6: System Strengthening	332,386	3.77
Cost Category 8: Parents/Community Engagement	81,663	0.92
Subcategory 8: Intervention Development	53,255	0.60
Subcategory 8: Intervention Implementation	28,408	0.32
Cost Category 9: Safe Schools and Infrastructure	31,519	0.3
Cost Category 11: Block Grants	967,922	10.99
Cost Category II: Complementary Activities	3,063,332	34.76
Total	8,806,606	100

Note: The calculations for the categories include expenditure data from April 2020 to September 2021, except for the complementary activities category. The complementary activities cost category is the sum of non-ingredient data and fixed fee and includes expenditure pre-April 2020.

Cost Drivers

To better understand the main cost drivers for the expenditure, Table 39 shows the top ten cost drivers. Non-ingredient data is the largest cost driver accounting for 30% of the expenditure. Contract grants to governmental organizations are the second highest cost driver (8%). The labor cost for General Operations (6%) and General Management (3.7%) cost categories are also the top cost drivers. Same is the case for the indirect cost for the General Operations (6.6%) and General Management (3.5%) cost categories.

Figure 64: Expenditures by Category



7.1.1.1 Implementation cost per teacher of teacher training

We conducted a preliminary cost-efficiency analysis to calculate the per teacher implementation cost for teacher training. The expenditure for teacher training is noted in Q3 and Q4 of year 2021. In total 127 teachers were trained in year 2021 (Reference: Contributions Dosage Report Q1Y3). To calculate the implementation cost, we also considered the management and logistic costs for the entire duration available.

Our current preliminary estimate of implementation cost per teacher of teacher training is $1,436.

Assumptions and required information:

1) We assumed that the management and logistic cost for all the five quarters is applicable for the teacher training. The cost would differ if the management and logistic cost were considered only for the duration of teacher training. However, a calendar of activities related to teacher training is required to precisely estimate this cost.

2) Similarly, information about teacher training per-diems and any other related expenses (such as food, donations) will lead to a more accurate estimate.

Table 39: Top Ten Cost Drivers

No.	Cost Driver	Amount (USD)	Percent
1	Non-ingredient Data	2,602,158	29.5%
2	Grants Under Contract (Block Grants to Governmental Organizations)	697,082	7.91%
3	Indirect Cost (General Operations)	588,652	6.68%
4	Labor (General Operations)	529,290	6.01%
5	Other Direct Costs (General Operations)	483,587	5.49%
6	Fixed Fee	461,174	5.23%
7	Grants Under Contract (Block Grants to Non-governmental Organizations)	413,062	4.69%
8	Labor (General Management)	329,577	3.74%
9	Indirect Cost (General Management)	308,800	3.50%
10	Other Direct Costs (Learning Assessment and Evaluation)	222,725	2.52%

Note: The calculations are based on the expenditure data reported from April 2020 to September 2021. Non-ingredient data and Fixed Fee categories belong to the Complementary Activities cost category in Table 39.

Preliminary Cost Analysis

7.1.1.2 Start-up cost for BAB program

We conducted a preliminary cost-economy analysis to calculate the start-up cost for BAB program. The expenditure in all five quarters was used in calculating this cost.

The current preliminary estimate for start-up costs is $4,018,232.

Assumptions and required information:

1) We are using all implementation and development data from all the quarters available.

2) To estimate the start-up cost we need a calendar of activities related to set-up and implementation for the BAB program

3) Similarly, local price database, contributions from teacher training etc. will help in obtaining accurate estimates.

7.1.1.3 Cost per learner of increasing one proficiency level in reading and math skills

We will conduct a cost-effectiveness analysis to answer this question. We will use all implementation and development data from all the quarters available. Monetized contributions will also be included in the calculations. The estimates of the impact of ABE intervention on reading and math skills will be used in the analysis. The proficiency levels for the learners will be defined using existing benchmarks from BAB and ES1-48 (2020 Compendium of Standard PIRS for Education Programming) and Somali national benchmarks as they become available.

Assumptions and required information:

1) Contributions for the intervention and the associated cost for estimating monetized contributions.

2) Estimates of the impact of ABE intervention on reading and math skills from the impact evaluation calculations.

From the current available data, the total expenditure on increasing proficiency for all the learners is $384,088 (implementation cost). The cost, including management and logistic costs, is $440,750.

7.1.1.4 Development cost of teaching and learning material

We will conduct a cost economy analysis to answer this question. The preliminary estimate of total development cost for the teaching and learning material is $525,248.

To calculate the development cost per learner for teaching and learning material, a total number of learners affected by the teaching and learning material is required. To calculate the development cost of teaching and learning material by level, we will need the development cost disaggregated by material dedicated for each level and the number of learners at each level benefiting from the material.

If it is not possible to provide expenditure data by levels, this information might be obtained from other sources available to Creative Associates.

7.1.1.5 Procurement cost of teaching and learning material

The expenditure related to the procurement cost of teaching and learning is not captured in the quarterly expenditure reports. This procurement cost for the teaching and learning material will have to be accounted for answering this question.

7.1.1.6 Total cost of classroom/teaching infrastructure

The preliminary total cost of classroom infrastructure (Safe schools and Infrastructure) is $31,529.

To calculate the classroom infrastructure cost per learner, the total number of learners affected by the Safe Schools and Infrastructure expense is required.

7.1.1.7 Cost per learner of ABE instruction per learner

We will conduct a cost economy analysis that will also include a stakeholder analysis. Detailed information about stakeholder contributions and costs associated with those contributions will be required to obtain an accurate estimate.

To calculate this cost per learner, the total number of learners benefiting from this instruction is required.

7.1.1.8 Cost of community engagement per learner

We will conduct a cost economy analysis that will also include a stakeholder analysis. Detailed information about stakeholder contributions and costs associated with those contributions will be required to obtain an accurate estimate.

The preliminary total cost of community engagement (implementation and development) is $81,663.

To calculate the community engagement cost per learner, a total number of learners affected by community engagement expense is required.

7.1.1.9 Cost of expanding and replicating the BAB program and cost of scaling and replicating teacher training in new areas (Prospective Analysis)

We will conduct a cost efficiency prospective analysis to answer questions related to expanding the BAB program in the intervention areas and replicating the program in new areas and scaling and replicating teacher training in new areas.

To conduct prospective analysis, we will need the following detailed information:

- Perspective (e.g.: USAID, Government of Somalia) from which the scaling, expansion, or replication of the program will be conducted.
- Geographical location and characteristics of the expansion areas
- Characteristics of the target beneficiaries from the program
- Number of teachers to be trained
- Number of learners to be trained at each level
- Who will be conducting the expansion, scaling, or replication

Challenges and Limitations

The goal is to ensure that the cost analysis questions are answered as best as we can, given the available data and information. However, there can be certain challenges and limitations. The prospective cost efficiency analysis will depend on several factors that are not included in the expenditure data. Information from stakeholders about the replications and scaling up of the ABE, such as changes when scaling up, population characteristics where replicating, etc., is important for the successful estimation of these costs. Stakeholders can decide on how precise the estimates for the prospective analysis are required. Similarly, provision and accuracy of the information that is not available through the expenditure data will be crucial in the precise cost estimation. This information includes cost-share contributions, local price database, intervention details, volunteer contributions, beneficiary, and volunteer opportunity cost, etc.

8 CONCLUSIONS AND RECOMMENDATIONS

This external baseline evaluation focused on: 1) describing learner, teacher, and head teacher characteristics for both BAB and non-BAB components of a longitudinal cohort – we will follow this cohort for 2 years; 2) testing the quality and performance of data collection tools; and 3) examining expected correlational associations between variables of interest to test the validity of the measurement model.

8.1 POPULATIONS SERVED BY BAB AND NON-BAB SCHOOLS

The evaluation noted little variation in characteristics between BAB and Non-BAB learners at baseline -- including age distribution, gender distribution, and prior school attendance. Both the BAB learners and learners attending non-accelerated schools varied in age from below 5 to above 19 years of age with median ages of 10-11 for all groups. Gender appeared relatively balanced for both the BAB and non-BAB cohorts overall, but differences were noted by school funding type. Cohort learners from all areas enjoyed high levels of prior educational experience, with Qur'anic school being most common. Examining the baseline data for differences across school type (community, public, private, and BAB) revealed distinct patterns suggesting that different populations of Somalia children and youth are served by the various school types. SES indicator scores were lower for the BAB cohort than for the combined non-BAB cohort, suggesting that BAB is serving financially marginalized students. Although the BAB project aspires to serve OOSCY ages 9-16, their actual learner population spans a much larger age range. As age is strongly correlated with baseline skills and competencies, younger learners will be disadvantaged and, given the compressed nature of the accelerated curriculum, may not be well served by this learning environment. **Future cohort recruitment should prioritize target age ranges.**

8.2 STUDENT PERCEPTIONS OF EQUITY, SAFETY, AND PSYCHOSOCIAL INDICATORS

Longitudinal cohort students, regardless of location, school type, or demographic characteristics expressed moderately positive perceptions of safety, engagement, and quality of life. However, variations were noted by location, with rural students scoring lower on most measures than students in IDP or urban settings. Student perceptions also varied based on school type. Students attending community schools had consistently lower scores than private school students with public and BAB students intermediate. **These findings point to a need for school policies and procedures, including teacher preparation and curricula, that consider the specific needs of their learners and adjust to target areas of greatest need for growth.**

8.3 GEOGRAPHIC LOCATION

SES, psychosocial, and learning outcomes all varied by state. Interestingly, although students in Hirshabelle scored significantly higher on SES indicators than students living in Benadir, Jubaland, or Southwest, they had consistently lower literacy and numeracy scores and lower scores on measures of equity, safety, and psychosocial indicators at baseline. Environmental factors, including flooding and instability, during data collection may have impacted student scores. Data collection and analysis

at midline should try to capture the role of external factors on student outcomes. **Teacher and head teacher training should include information, recommendations, and approaches for supporting students during times of uncertainty, stress, or trauma.**

8.4 Baseline Academic Measures

Learning outcomes varied by cohort group, with BAB learners scoring between grade 1 and grade 2 learners on all measures. Differences were also noted in student baseline skills and competencies based on location, gender, and school type. Correlation and multiple regression analysis confirmed positive correlations between student demographics (SES, maternal literacy, and age), psychosocial measures (engagement and quality of life) and baseline EGRA and EGMA scores. **The correlation between psychosocial indicators and academic achievement corroborates the need for intentional efforts to provide training, curricula, and supports that promote positive student feelings of engagement, self-esteem, and other psychosocial measures, with particular emphasis on girls, students with disabilities, and other marginalized populations**

Floor and ceiling effects on the EGRA and EGMA for the longitudinal sample at baseline were minimal. To assess and improve the reliability of all surveys, the evaluation conducted item analyses at various levels; factor analysis, reliability analysis, item difficulty and discrimination analysis (for EGRA and EGMA). **The assessments performed well overall, but some questions on the EGMA had negative discrimination indices that pose a threat to instrument validity. These items should be modified or replaced for future administrations to ensure data quality.**

8.5 Quality of Instruction

Teacher baseline surveys revealed a high-quality teacher workforce dominated by men. Cohort teachers were generally young, educated, and experienced. Male teachers scored high to moderately high on indicators associated with teacher effectiveness, including perceptions of safety, self-efficacy, and knowledge, preparation, and skills. Female teachers and teachers that only teach BAB classes typically scored lower on these measures. Rural teachers expressed lower perceptions of resource availability than urban or IDP colleagues. **Teacher recruitment should continue to prioritize identifying, developing, and supporting female teachers. Teacher training and mentoring plans should consider the specific needs of teachers in lower scoring groups and develop and implement proper supports.**

Head teacher survey responses suggest commitment to equitable teaching of all learners. Further head teachers reported high levels of community engagement, as evidenced by the presence of an active community education committee.

Overall, the baseline evaluation revealed learner, teacher, and community assets that are foundational to educational success.

8.6 Sampling Effects

This evaluation examined the impact of two different sampling methodologies – the evaluation longitudinal and the BAB cross-sectional approach on cohort characteristics. Although both

methodologies used a random sampling approach, they differed based on the purposes of the sample. The resulting cohort samples were statistically different in every demographic aspect except gender. Most notably, the longitudinal sample had higher components of lower age learners, rural learners, and IDP learners. The differences in demographics between the two samples is reflected in EGRA and EGMA baseline assessments. **These comparisons highlight the importance of considering sample composition when developing educational standards or benchmarks to ensure equitable representation of target groups.**

9 APPENDIX 1: EVALUATION QUESTIONS, INDICATORS, AND DATA COLLECTION AND ANALYSIS TOOLS AND METHODS

ABBREVIATIONS:

ABE	Accelerated Basic Education
AEP	Accelerated Education Programs
BAB	Bar ama Baro - "Teach or Learn"
CA	Community Assessment
CEC	Community Education Committee
CMA	Center Management Assessment
CPE	Community and Parent Engagement
EGMA	Early Grades Math Assessment
EGRA	Early Grades Reading Assessment
ISELA	International Social and Emotional Learning Assessment
LFI	Learning Facilitator Interviews
LFS	Learning Facilitator Survey
LKFAP	Learning Facilitator Knowledge-Attitudes and Practices
LP	Learning Profiles and Reading, Math, and SEL Assessments
LS	Learning Survey
OOSCY	Out of School Children and Youth
PGCI	Parent/Caregiver Interviews
PGCS	Parent/Caregiver Survey
RLA	Rapid Learning Assessments
SLEC	Learner Learning in Emergency Checklist

Question	Indicator	Quantitative Measures	Qualitative Measures	Analysis
Evaluation Question 1: To what extent is BAB effective in improving access to quality education for Somali out-of-school children and youth ages 9-16?				

1.1 Increased Enrollment				
Access				
1.1a What are the key barriers to learner access? Do barriers differ based on learner and/or community characteristics?	AE Center locations, programs (flexibility), and practices meet learner and community needs	LP LFS LS PCGS	PCGI CA, LFI Case Studies	Content, Descriptive, Factor, Regression, Mixed-effects models Subgroup, Thematic analysis
1.1b To what extent did BAB succeed in reducing barriers for all learners and for vulnerable* learners, in particular? (where vulnerable includes socially marginalized such as Digil-Mirifle clan-family and Somali Bantu, etc.)	BAB enrolls a pool of learners that reflect community (AMEP Indicator 1)	LP LFS LS PCGS	PCGI CA LFI Case Studies	Content, Descriptive, Thematic analysis
Retention				
1.1c To what extent was BAB effective in retaining targeted learners during a school year and promoting learners between educational levels? Retaining and promoting the most vulnerable learners, particularly girls? Did retention/dropout vary based on implementation context and/or learner characteristics?	Learners are retained for at least one Academic Year (AY) Learners are promoted from L1 to L2 Learners complete 2 years of instruction	Project Metrics, LP LFS CMA (attendance)	LFI CA	Content, Descriptive, Time-series, Survival, Trend, Subgroup Note: Time-series, survival, trend, and subgroup analysis require detailed attendance records.
1.1d How did family and community engagement, attitudes and beliefs, and other factors influence learner retention/dropout?	Parent/CG support and prioritize attendance of all learners; CECs are actively engaged in ABE site	LP LS PCGS CPE	PCGI CA LFI Case Studies	Content, Descriptive, Thematic analysis Factor Correlations, Regression, Mixed-Effects models

1.2 Increased Safety				
1.2a To what extent did BAB facilities and practices meet physical and emotional needs of all learners? Were some learners better served than others?	Learners perceive ABE site and program as safe	LP LS LFS CPE	PCGI LFI CA Case Studies	Content, Descriptive, Factor, Regression, Subgroup Thematic analysis
1.2b To what extent did parents/caregivers perceive their child's participation in BAB as safe and secure? Did perceptions vary based on learner characteristics (e.g., girls vs boys)?	Parent/Caregivers perceive the BAB site and program as safe for their children	CPE	PCGI Case studies	Content, Descriptive, Subgroup, Thematic analysis
1.2c To what extent is the community engaged in enhancing safety and security?	CECs (Community Education Committee) active in developing risk mitigation plans	CPE	CA	Content, Descriptive
1.2d Learning Facilitators/head teachers perceive BAB site as safe and secure	Learning Facilitators and Head Teachers feel supported and perceive ABE site as safe and secure for themselves and the learners	LFS	LFI	Content, Descriptive
1.3 Increase Learning Outcomes				
High Quality Instruction				
1.3a To what extent was BAB effective in recruiting and retaining teachers from implementation areas? What challenges/barriers did BAB encounter?	Quality Learning Facilitators identified, recruited	Program Records, LFS	BAB program staff interviews	Content, Descriptive, Subgroup
1.3b To what extent was BAB effective in training and supervising LFs (Learning Facilitators)?	LFs have appropriate knowledge of content and effective and inclusive instructional practices, are confident in their ability to teach effectively, and feel supported. Center Head Facilitators create a positive environment and support effective teaching practices.	LFS LFKAP CMA	LFI	Content, Descriptive, Factor, T-tests, Regression Mixed-effects models

Access to appropriate Curriculum, and needed materials, supplies				
1.3c To what extent is BAB curriculum able to meet the needs of all learners?	Curriculum is sensitive to the needs of learners across the full array of age, gender, context, and other demographic variables. LFs and learners have access to support, materials, and supplies needed for effective teaching and learning	LS LFS CMA CPE	LFI PCGI Case studies	Content, Descriptive, Thematic analysis
Learning				
1.3d What impact did BAB have on learning outcomes (math, reading, socio-emotional learning) for participating out of school children and youth (OOSCY)?	Learners demonstrate increased proficiencies in reading, math, and SEL assessments	LS LP RLA EGRA, EGMA, ISELA, SLEC End of Year exams	LFI	Content, Descriptive, Regression, Mixed-effects models Meta-analysis
1.3e Are there subgroups of learners that benefited to a lesser extent from the intervention? If so, who are they and why did they benefit less?	Learning outcomes differ based on learner or community characteristics	LP LFS LS CPE	LFI Case studies	Content, Descriptive, Regression, Mixed-effects models Subgroup, Meta-analysis, Thematic analysis
1.3f How successful was BAB in creating a safe, secure, and inclusive learning environment?	All learners feel safe, secure, and a sense of belonging	LS LP LFS CPE	LFI Case Studies	Content, Descriptive, Thematic analysis
Evaluation Question 2: How do learning outcomes of diverse learners differ across formal (public, community, and private), and non-formal education options and interventions in the CDCS focal zones? What is the impact of contextual and demographic indicators on learning outcomes?				

2.1 Enrollment				
Access				
2.1a To what extent do all learners have access to non-formal and formal schools (community, public, and private) in target communities	Demographics of learners enrolled in formal (community, private, and public) and non-formal schools (e.g., gender, age, disability, minority) reflect community demographics.	school enrollment records demographic questions on standard assessments (EGRA/EGMA/ISELA) Community Surveys		Descriptive, Factor, Regression, Subgroup
Retention				
2.1b To what extent are non-formal and formal school learners retained and promoted between education levels? How do educational options, learner characteristics, and context (e.g., urban, rural, or IDP) impact retention and dropout rates?	Learners are retained for at least one AY. Learners are promoted from one grade to the next.	Attendance records and other school records. Participation in pre and post test		Descriptive Time-series, Survival, Trend, Subgroup Note: Time-series, survival, trend, and subgroup analysis require detailed attendance records.
2.2 Safety				
2.2a To what extent do formal and non-formal schools meet the physical and emotional needs of all learners? Were some learners better served than others?	Learners perceive school sites and programs as safe	ISELA LS		Descriptive, Factor, Regression, Mixed-effects models Subgroup
2.2b Teachers perceive school as safe and secure.	Teachers feel the school is safe and secure for themselves and the learners	LFS		Descriptive
2.3 Learning Outcomes				
Learning				
2.3a What impact did instruction have on learning outcomes (math, reading, socio-	Learners demonstrate increased proficiencies in	EGRA, EGMA, ISELA		Descriptive,

emotional learning) for participating learners?	reading, math, and SEL assessments			Regression, Mixed-effects models Meta-analysis
2.3b Are there subgroups of learners that benefited to a lesser extent from the instruction? If so, who are they and why did they benefit less?	Learning outcomes differ based on learner or community characteristics	LP LFS LS CPE		Descriptive, Regression, Mixed-effects models Subgroup, Meta-analysis
Evaluation Question 3: What can we learn from the BAB implementation to inform decision-making for scale-up and sustainability?				
Program Costs				
3.1a What is the cost per learner of increasing one proficiency level in reading and math skills (e.g.: non-learners, basic learners, emergent learners, etc.)? The proficiency levels will be defined using existing benchmarks from BAB and ES1-48 (2020 Compendium of Standard PIRS (Performance Indicator Reference Sheets) for Education Programming).	Cost per participant	BAB cost and program data		Cost analysis
3.1b.1 What is the cost per learner of expanding the BAB reading, math, and SEL intervention package in intervention areas? This will be estimated from USAID and the Government of Somalia's perspective. 3.1b.2 What is the cost per learner of replicating the BAB reading, math, and SEL intervention package in new areas? This will be estimated from USAID and the Government of Somalia's perspective.	Cost per learner	BAB cost and program data		Cost analysis
3.1.c What is the total cost per learner of ABE instruction,				

disaggregated by levels (L0,L1,L1-L2)?				
Benefits by Participant Group/Context				
3.2 What types of learners/and communities would be best served by scaling the BAB ABE model? What types of learners/community would be left out when scaling the BAB ABE model?	What groups/subgroups of participants benefited the most/least from the intervention? What characteristics of learners and their communities affected learner outcomes?	LP RLA LS LFS	CA LFI Case studies	Content, Descriptive, Regression, Mixed-effects models Meta-analysis Subgroup Gap Trend Thematic analysis
Impacts of Program Elements on learners				
3.3a What teacher, teacher training, and teacher support characteristics drive positive changes in access, retention, and learning outcomes? Does this vary by learner demographic and/or implementation context?	Correlations between teacher background, preparation, training and support and learner outcomes. Variations in observed outcomes by learner characteristics and context.	LP LFS LFKAP CMA	LFI Case studies	Content, Descriptive, Correlation, Regression, Mixed-effects models Meta-analysis Subgroup Mixed-effects Models Thematic analysis
3.3b What elements of the BAB implementation (e.g. curriculum and instruction; infrastructure; materials and supplies; safety and security; etc.) appeared to drive positive changes in learner feelings of safety and well-being? Does this vary by demographic and/or implementation context?	Correlations between program elements and learner feelings of safety and well-being. Variations in observed outcomes by learner characteristics and context.	LP LS C	LFI Case studies	Content, Descriptive, Correlation, Regression, Mixed-effects models Meta-analysis Thematic analysis
3.3c What is the relationship between learner feelings of	Correlations between indicators of well-being and	LP	Case studies	Descriptive,

safety and well-being and learning outcomes?	learning outcomes. Variations in observed outcomes by learner characteristics and context.	LS EGRA, EGMA, ISELA, SLEC RLA End of year exams		Correlation, Regression, Mixed-effects models Thematic analysis
Impact of Family and Community Factors on Outcomes				
3.4 What community or family level factors drive positive changes in access, retention, and learning outcomes? Does this vary by demographic and/or implementation context?	How do parent and community attitudes and beliefs, support, and engagement affect learner outcomes?	PCGS LP LS RLA LFS	PCGI CA LFI Case studies	Content, Descriptive, Factor Correlation, Regression, Mixed-effects models Meta-analysis, Thematic analysis
Civic Engagement				
3.5 To what extent does youth civic engagement contribute to increased access, retention, and learning outcomes? Does this vary by demographic and/or implementation context?	Correlations between measures of youth engagement and learning outcomes, including SEL and measures of well-being and self-efficacy. How do findings vary based on learner demographics and community context?	Program records LP LS	CA PCGI LFI	Content, Descriptive, Correlation, Regression, Mixed-effects models Meta-analysis
Unintended Consequences				
3.6 Is there evidence that the BAB implementation model caused harm or has the potential to worsen inequality of conflict if implemented at scale?	Consider the possibility that socially marginalized groups, such as the Digil-Mirifle clan-family, Somali Bantu, or other socially marginalized, have been excluded.		CA LFI PCGI Case studies	Content, Thematic analysis

APPENDIX 2: LONGITUDINAL COHORT SAMPLING SCHEME

BAB LONGITUDINAL COHORT

The evaluation's selection protocol for identifying a random sample of BAB learners included a two-stage sampling approach since learners were nested within BAB sites. Thus, the evaluation first drew the sites to be sampled; we then randomly chose a class within each selected site. Specifics of the sample selection process are described below.

BAB enrollment data for the academic year starting August 2021 reported a total of 29,901 level 1 learners in 598 classes at 197 school sites distributed across 11 districts, 5 regions, and 4 states. School sites were distributed across location types, as follows: 41 IDP sites (4,849 learners), 29 rural sites (4,101 learners), and 127 urban sites (20,951 learners). The number of learners per classroom ranged from a minimum of 23 to a maximum of 100 BAB learners, with an average of 50 learners across all sites. Sites in the three contexts (urban, rural, and IDP) were selected independently of each other.

The evaluation used stratified sampling to conduct accurate sampling along with a proper representation across the population. We first selected sites based on the site-based sampling scheme with the number of sample classrooms selected for each context IDP/rural/urban based on the number of sites in each category. The evaluation power calculations call for a BAB sample size of approximately 2,000 learners. We selected learners in a classroom, as opposed to an individual basis, to simplify data collection logistics and to decrease potential sampling bias. With an average of 50 learners per classroom, the evaluation targeted 40 classrooms for the BAB longitudinal cohort sample (one classroom per site). Distributing the 40 sites following the ratio IDP: Rural: Urban as 41:29:127, results in a sample including 8 IDP, 6 rural and 26 urban sites. The evaluation changed this sample scheme, as our SORDI partners advise that data collection in rural and IDP sites might offer more challenge due to population mobility, access issues, and other considerations, making it desirable to oversample IDP and rural areas, while slightly under-sampling urban sites. We used the following approach to determine a suitable number of sample sites per category with these challenges in mind.

We calculated a representative sample of the 20,951 learners from urban sites, considering 3%-5% margin of error for most education projects, as approximately 1,000 urban learners. With an average of 50 learners in a classroom, the random sample will include 20 urban classrooms This sample size was estimated based on the formula below; where N is the number of learners in the study population, p is estimated variance in the population; e is the precision desired (3%-5% for most education projects); and z is based on confidence level (1.96 for 95% confidence).

$$\text{Sample size} = \frac{\frac{z^2 \times p\,(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p\,(1-p)}{e^2 N}\right)}$$

To maximize the geographic reach of the data collection, the evaluation planned to select a level one classroom within each selected site. Thus, the evaluation's urban sample calls for 20 classrooms – one each from 20 different sites. To determine the number of IDP and rural classes in the sample, we distributed the remaining 20 classrooms using the proportion of IDP to rural sites (41: 29) -- resulting in 11 IDP sites and 9 rural sites. Thus, the evaluation's sampling target at the site (classroom) level is 11:9:20 (IDP:Rural:Urban).

Next, the evaluation drew sites proportional to the number of BAB site numbers in the respective districts according to the table below. Considering the evaluation will randomly select a first-grade classroom from each non-BAB school co-located with the selected BAB sites to form the non-BAB cohort, we also tried to maximize the variety of school types selected in each district. For example, there is only one private school among the Baidoa IDP sites. Thus, we selected this school first, then randomly selected the other 6 sites from Baidoa based on the ratio of community schools vs. public schools. Because some sites, particularly in IDP and rural areas, may become inaccessible to enumerators due to a host of factors, 3 backup sites were selected for each context. These backup sites will be used to replace original sites, should they become inaccessible.

Higher levels of safety concerns were reported in Balcad and Jowhar district, resulting in a request from SORDI for additional backup sites in these areas. Additional backup sites (2 for each context) in each district were randomly selected by Purdue researchers right before data collection activities began on 10//10/2021. They are noted in the site selection spreadsheet.

Context	**District**	**# Selected**	**# Backups**	**# Additional Backups**
IDP	Baidoa	7		
	Deynile	3	1	
	Kahada	1	2	
Rural	Balcad	3		2
	Barawe	1		
	Jowhar	2	1	2
	Kismayo	2	1	
	Walanweyn	1	1	
Urban	Baidoa	2		
	Balcad	1		2
	Barawe	1		
	Deynile	1		

	Diinsor	1	1	
	Hamarwayne	1		
	Jowhar	2		2
	Kahada	3		
	Kismayo	5		
	Shibis	1	2	
	Walanweyn	2		

Finally, the evaluation randomly selected a level one classroom within each selected site (for a total of at least 40 classrooms). The procedure for selecting the class was random sampling. Most BAB classes are afternoon classes (13:30-17:20). We have paid attention to including enough morning classes (08:00-12:30). The randomly selected classes turned out to have a balance of morning classes and afternoon classes according to the original proportion.

NON-BAB LONGITUDINAL COHORT

The evaluation's selection of the non-BAB longitudinal cohort sought to include non-BAB classes that were co-located in or near a BAB cohort classroom. Because the BAB level 1 longitudinal cohort will cover grades 1 and 2 in a single year, we included both first and second grade classrooms in the non-BAB longitudinal cohort sample. Using this expanded strategy allows for 26 non-BAB classes co-located with BAB cohort classes (13 1st grade and 13 2nd grade). The following table shows the co-located sites, # of available 1st and 2nd grade classes, and sampling scheme for the co-located sites. We supplemented these sites by recruiting another 5 classes from IDP and 5 classes from rural areas to round out our sample. Because BAB is not sampling any nonBAB classes, we could randomly select co-located classrooms without concern about double sampling. We sought USAID's input regarding selection approaches that minimize impacts on data collection timelines and budget, while minimizing sampling bias.

Sampling scheme for non-BAB classes co-located with BAB Longitudinal Cohort Classes

School Type	State	Region	District	School	Non-BAB	# Grade 1	# Grade 2	Grade 1 Class Selected	Grade 2 Class Selected
IDP/public	South West State	Bay	Baidoa	Mustaqbal	Yes	4	3	3	2
Rural/Public	Hirshabelle State	Middle Shabelle	Balcad	Arofag School	Yes	1	1	1	1

Urban/Private	Hirshabelle State	Middle Shabelle	Balcad	Ifiye School	Yes	1	2	1	1
Urban/Private	South West State	Lower Shabelle	Barawe	Zeynul Abidiin	Yes	2	1	2	1
Urban/Public	Benadir	Benadir	Hamarwayne	Moalim Jama	Yes		2		1
Urban/Public	Hirshabelle State	Middle Shabelle	Jowhar	Horseed Jowhar	Yes	1	1	1	1
Urban/Community	Hirshabelle State	Middle Shabelle	Jowhar	Jaahweyn two Primary	Yes	2	0	2	
Urban/Private	Hirshabelle State	Middle Shabelle	Jowhar	Jahweyn 1 primary	Yes	1	0	1	
Urban/Private	Benadir	Benadir	Kahada	Osama Bin Zaid3	Yes	1	1	1	1
Urban/public	Jubaland	Lower Juba	Kismayo	Ahmed Gurey	Yes	1	1	1	1
Urban/Private	Jubaland	Lower Juba	Kismayo	Kismayo Pri&Sec School	Yes	0	1		1
Urban/Private	Jubaland	Lower Juba	Kismayo	Nasiib Bundo Pri School	Yes	2	2	2	1
Urban/Public	Jubaland	Lower Juba	Kismayo	Wadajir Pri School	Yes	2	1	1	1
Rural/Public	Jubaland	Lower Juba	Kismayo	Yontoy Pri School	Yes	1	2	1	2

Urban/Private	Benadir	Benadir	Shibis	Dar Altarabiya	Yes	0	2		1
Rural/Community	South West State	Lower Shabelle	Walanweyn	Danwadag	Yes	1	0	1	

SORDI's senior researchers conducted a survey for existing non-BAB co-located classes; in schools with multiple grades one and two, the team chose classes randomly. For instance, school X could have 2 classes for grade one and 3 classes for grade two of non-BAB co-located classes; the teams randomized the classes and recruited the randomly selected classes.

In Baidoa, Sheikh Ashiro has non-BAB classes, and SORDI randomly chose available classes for grade 1&2. Also, Diinsoor has two schools with BAB and Non-BAB classes (Waberi and Hawlwadag), SORDI randomly selected the Waberi School in Dinsoor for collecting non-BAB data in place of the sole selected BAB school, Yaqshid, which does not have non-BAB classes.

For areas where SORDI researchers were unable to identify non-BAB classes, SORDI field coordinators and supervisors reached out to district education officers (DEOs) and the state/region level education authority to provide a list of schools with similar socio-economic characteristics with both BAB and non-BAB classes. SORDI randomly selected non-BAB schools and classes from the list of schools and contact information provided.

APPENDIX 3: ITEM ANALYSIS, SURVEY RELIABILITY AND VALIDITY

INSTRUMENT VALIDITY AND RELIABILITY

The evaluation's Longitudinal Learner, Teacher, and Head Teacher surveys were developed and tested to ensure effective construction and adequate psychometric quality using a process (depicted in Figure A3-1) that is based on the core principles and best practices of the scale construction process outlined by Furr. Inner ovals in the model diagram depict the primary tasks of scale development and testing, while rectangles describe the evaluation's approach to each task. The process began with a desk review to understand the status of primary grade education in Somalia and a literature review to explore educational theory and current research linking key constructs identified in the project's theory of change to learning outcomes. Next, we scanned the literature for scales measuring these constructs that have been validated in Somalia (ideally), East Africa, or other places. When no previously validated scale was found, we selected individual items or groups of items from existing instruments (e.g., BAB Learner Survey) or created original items. Learner scales and items were tested for face validity with a small group of target-age children and with colleagues in Somalia. All surveys, including Learner, Teacher and Head Teacher surveys were reviewed by SORDI colleagues for contextual appropriateness. Based on the feedback from the mini pilot (the process through which we assessed face validity), some items on the surveys were revised; items that were originally presented in statement format were converted to question format, since the participants had trouble responding to statement formats. All instruments were then translated into the official Somali language by SORDI colleagues and further piloted and modified during enumerator training. After data collection was completed, SORDI researchers translated any free-response items back to English for analysis.

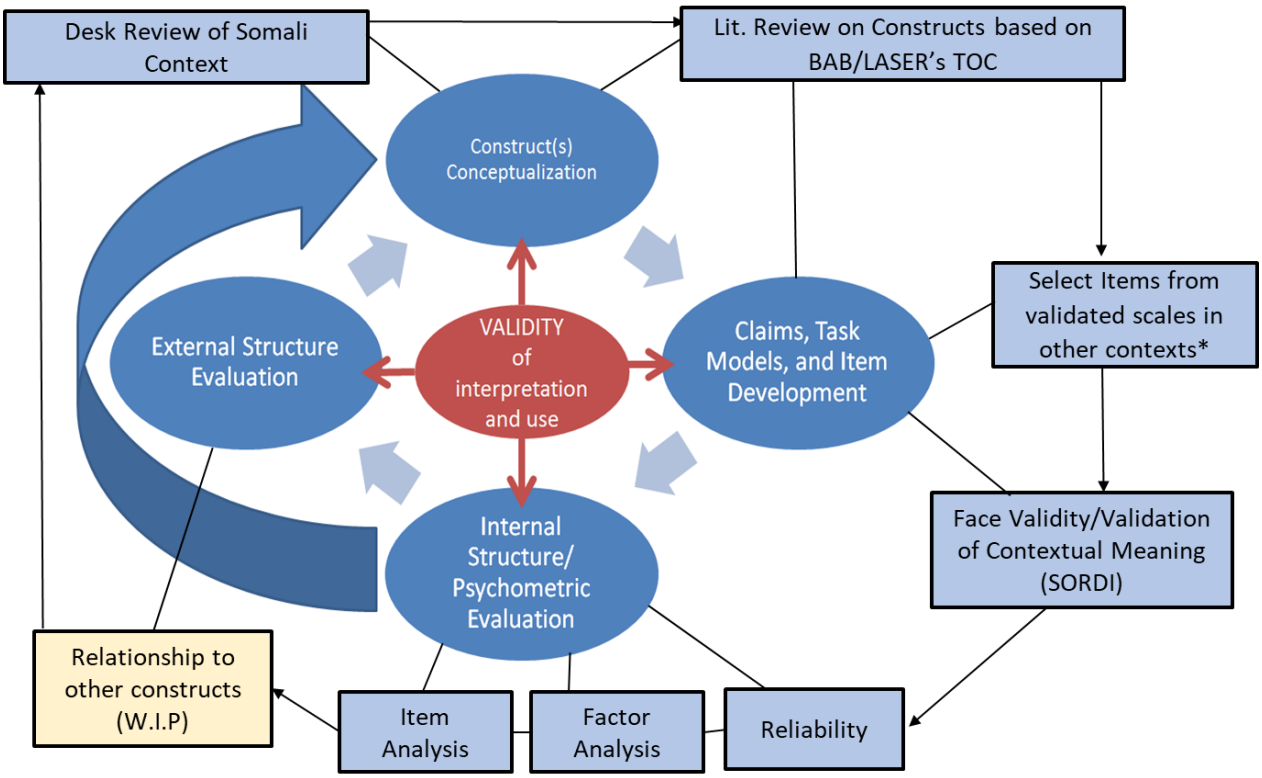
To assess and improve the reliability of all surveys, the evaluation conducted item analyses at various levels; factor analysis, reliability analysis, item difficulty and discrimination analysis (for EGRA and EGMA)

Factor Analysis

We conducted factor analysis to measure the internal consistency and dimensionality of various constructs in our surveys (Furr 2011). For items extracted from previously validated scales, we conducted Confirmatory Factor Analysis (CFA) to confirm that hypothesized relationships among those items hold true for our data (see Figure A5-2 for CFA process). For those items that had not been previously validated, we first conducted Exploratory Factor Analysis (EFA) to explore the relationships among those items and form hypotheses that were later confirmed with a CFA. For example, out of seven (7) items we extracted from the BAB Learner Survey that were intended to measure socio-economic status (SES), only four (4) of those items hung together in the EFA. Therefore, only those items were used in the final CFA model as SES indicators. Results from our final CFA model indicate that the hypothesized relationships among the items that have been previously validated and the new ones developed based on our EFA results hold true for our data

(CFI = 0.972, TLI = 0.966, RMSEA = 0.065, SRMR = 0.071). This satisfies the cut off for good model fit prescribed by (Hu & Bentler, 1999).

Figure A3-1: External Evaluation Tool Development and Validation Process[5]



For the Teacher's Survey, we ran a separate CFA model for each of the constructs of interest because of sample size limitations (n = 54). Factor analysis typically requires a large sample size for stable results. Just like the learner survey, some items from the Teacher's Survey did not favor the hypothesized model in our sample and were therefore excluded from the CFA model. Notwithstanding, each CFA model for the teacher data met the satisfactory model fit criteria.

Reliability

In addition to factor analysis, we estimated Alpha reliability for each of the scales/subscales measuring our constructs of interest. This process is critical, as it allows us to examine the level of correlation among the items that are expected to measure the same thing. Higher correlations result

[5] TOC = Theory of Change, W.I.P = work in progress, * = Some items/subscales have not been previously validated in other contexts.

in a higher reliability coefficient and vice versa, and this is referred to as a measure of internal consistency. Two items that showed poor correlation with other items on their respective scales were kept, since they did not significantly affect the CFA model fit to our data. These items are the third item on the safety scale and the fourth item on the friendship subscale. Apart from the scale and subscale mentioned, all other scales/subscales maintained acceptable Alpha coefficient greater than 0.7, indicating evidence of internal consistency.

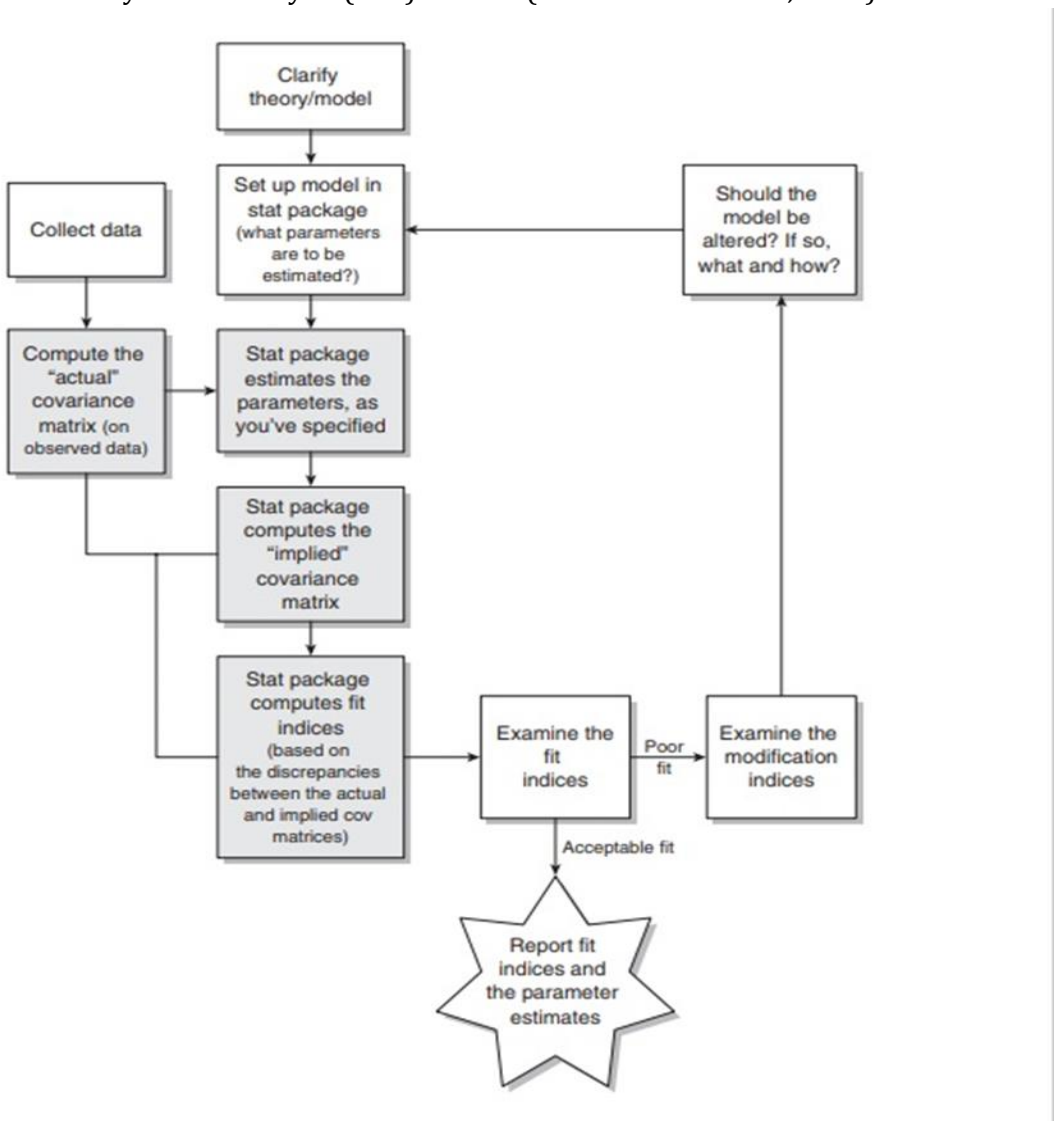
Item Analysis (Difficulty and Discrimination)

Assessments like EGRA and EGMA are designed to measure learners' ability to perform certain predefined tasks. A good assessment should be able to place learners in a continuum that provides information about what each learner can and cannot do. Therefore, it should contain items with a range of difficulty (e.g., very easy, easy, moderately difficult, and very difficult, as the case may be). That way, learners are expected to perform well on the tasks that they have mastered and poorly on tasks they have not mastered. Additionally, students who score higher on the overall assessment are expected to have a higher probability of getting each item correct. However, this is not always the case with some items. Therefore, it is important to conduct item difficulty and discrimination analysis to ascertain the quality of the assessment.

We conducted item difficulty and item discrimination analysis to ascertain the level of difficulty and discrimination of the items on the EGRA and EGMA assessments. This was done using the Classical Test Theory (CTT) framework. This framework was chosen because it has been widely used in similar studies and is easy to understand. Here, the item difficulty index is defined as the proportion of learners who answered an item correctly. It is counterintuitive in nature, in that an item with a higher difficulty index (with values ranging from 0 to 1) implies an easier item and vice versa. The item discrimination index on the other hand is a measure of how well an item can distinguish between learners who have mastered a given task or the material they are being assessed on and those who have not. It is the correlation between responses to a particular item and the overall score on the assessment. Item discrimination is typically affected by two factors: item difficulty and guessing. Items that are either too difficult -- such that only a few learners can provide the correct response -- or too easy -- such that most of the learners can provide a correct response -- usually have low discrimination indices. Additionally, items in which most of the learners who supplied the correct response guess the answer would also have poor discrimination. Complete item analysis results are included below.

The results from the item difficulty and discrimination analyses help us determine if we need to add some easier items or more difficult items in later rounds of assessment to combat any ceiling and/or floor effects; and if we need to modify or remove poor items in terms of discrimination. It is important to note that items with a negative discrimination index pose a great threat to the construct validity of a test/assessment. This is because it implies that learners who performed poorly on the overall assessment are more likely to supply a correct response to such items, which is a sign of construct-irrelevant variance. That is, it is something other than being knowledgeable or mastering the assigned tasks that contributes to a correct response, e.g., guess work. Therefore, such items would either need to be modified or completely removed from future administration(s) of the assessment.

Figure A3-2: Confirmatory Factor Analysis (CFA) Process (Extracted from Furr, 2011)



Overall, items on the EGRA and EGMA were of high quality; the item difficulty index for the EGRA ranged from 0.05 to 0.83, indicating a range of exceedingly difficult items to moderately easy/difficult items. Similarly, the EGMA had difficulty indices ranging from 0.05 to 0.84, showing a similar difficulty range. Item discrimination on the EGRA was very satisfactory, only three (3) out of two hundred and ninety-four (294) items had discrimination indices less than the acceptable threshold of 0.3 (item 5 on the reading comprehension subtask, items 47 and 49 on the invent word subtask, with discrimination indices 0.24, 0.20, and 0.25, respectively. This is not concerning since they were exceedingly difficult items. Meanwhile, the EGMA assessment, which has one hundred and eight (108) items, has twelve (12) items performing below the acceptable threshold of 0.3 or greater. Of these 12 items, only three (3) are considered problematic, since they have negative discrimination indices (items 1 and 8 on the number identification subtask, and the

first strategy item on the word problems subtask with discrimination indices of –0.38, -0.02, and –0.06, respectively. This is concerning because the results show that learners who performed poorly on the overall assessment were more likely to answer these questions correctly. Therefore, it poses a threat to validity. Although the items are included in the results at baseline, they should not be included when it is time to make decisions about promotion and graduation from the BAB program. We will also investigate why these items performed this way prior to collecting midline data to figure out if they should be included in that round of data collection or if there are modifications to be made.

Table A3-1: CFA Results for Learner Survey: Parameter Estimates & Fit Indices

	SES	Safety	Engagement	Wellbeing	Self-Esteem	Friendship	Cronbach's Alpha
SES_04	0.732						0.565
SES_05	0.584						
SES_06	0.798						
SES_07	0.584						
Safety_01		0.753					0.449
Safety_02		0.855					
Safety_03 Rcd		0.241					
Engagement_01			0.788				0.783
Engagement_02			0.776				
Engagement_03			0.761				
Wellbeing_02Rcd				0.648			0.705
Wellbeing_03Rcd				0.772			
Wellbeing_04Rcd				0.800			
Self-Esteem_01					0.727		0.729
Self-Esteem_02					0.698		
Self-Esteem_03					0.805		
Self-Esteem_04					0.568		
Frienship_01						0.745	0.386
Frienship_02						0.712	
Frienship_03						0.769	
Frienship_04						0.183	

Note: CFI = 0.972, RMSEA = 0.065, SRMR = 0.071

Table A3-2: CFA Results for Teachers' Self Efficacy Scale

Item	Instructional Self Efficacy	Disciplinary Self Efficacy	Cronbach's Alpha

TInstrSE_02	0.427		0.756
TInstrSE_03	0.528		
TInstrSE_05	0.427		
TInstrSE_06	0.544		
TInstrSE_07	0.691		
TInstrSE_09	0.557		
TInstrSE_10	0.667		
TDiscSE_01		0.753	0.538
TDiscSE_02		0.457	
TDiscSE_03		0.379	

Note: CFI = 0.960, RMSEA = 0.048, SRMR = 0.079, Overall scale Alpha = 0.805

Table A3-3: CFA Results for the Engaged Teachers' Scale

Item	Emotional Engagement	Social Engagement with Learner	Cognitive Engagement	Cronbach's Alpha
TEmtnlEngm_01	0.604			0.622
TEmtnlEngm_02	0.539			
TEmtnlEngm_03	0.710			
TSclEngmnt_01		0.572		0.692
TSclEngmnt_02		0.846		
TSclEngmnt_04		0.578		
TCgntvEnggm_01			0.511	0.730
TCgntvEnggm_02			0.702	
TCgntvEnggm_03			0.722	
TCgntvEnggm_04			0.620	

Note: CFI = 0.931, RMSEA = 0.079, SRMR = 0.075, Overall scale Alpha = 0.847

Table A3-4: CFA Results for Teachers' Perceptions of Support Scale

Item	Internal	External	Cronbach's Alpha
TSupport_03	0.571		0.567
TSupport_04	0.509		
TSupport_06	0.567		
TSupport_01		0.831	0.833
TSupport_02		0.780	
TSupport_05		0.721	

TSupport_07	0.656

Note: CFI = 0.997, RMSEA = 0.023, SRMR = 0.062, Overall scale Alpha = 0.821

Table A3-5: CFA Results for the Teachers' Perceptions of Safety Scale

Item	Safety	Cronbach's Alpha
TSafety_01	0.861	0.795
TSafety_02	0.727	
TSafety_04	0.658	
TSafety_06	0.590	

Note: CFI = 0.989, RMSEA = 0.083, SRMR = 0.035

Table A3-6: Item Analysis Results for EGRA (Longitudinal Cohort)

Item	Mean (Difficulty Index)	Variance	Discrimination Index
InventWord_48	0.05	0.05	0.31
ReadComp_05	0.05	0.05	0.24
InventWord_47	0.06	0.06	0.2
InventWord_49	0.06	0.06	0.25
InventWord_50	0.06	0.05	0.32
LetSound_100	0.07	0.07	0.5
InventWord_46	0.07	0.06	0.42
ReadComp_04	0.07	0.06	0.41
LetSound_91	0.08	0.08	0.46
LetSound_92	0.08	0.08	0.52
LetSound_93	0.08	0.08	0.48
LetSound_94	0.08	0.07	0.46
LetSound_95	0.08	0.07	0.46

LetSound_96	0.08	0.07	0.47
LetSound_97	0.08	0.07	0.5
LetSound_98	0.08	0.07	0.57
LetSound_99	0.08	0.07	0.63
InventWord_43	0.08	0.08	0.69
InventWord_44	0.08	0.07	0.67
InventWord_45	0.08	0.07	0.68
RealWords_47	0.08	0.08	0.7
RealWords_48	0.08	0.07	0.67
RealWords_49	0.08	0.08	0.66
RealWords_50	0.08	0.07	0.68
LetSound_87	0.09	0.08	0.7
LetSound_88	0.09	0.08	0.68
LetSound_89	0.09	0.08	0.72
LetSound_90	0.09	0.08	0.72
InventWord_41	0.09	0.08	0.69
InventWord_42	0.09	0.08	0.73
RealWords_45	0.09	0.09	0.72
RealWords_46	0.09	0.08	0.73
LetSound_84	0.1	0.09	0.72
LetSound_85	0.1	0.09	0.72
LetSound_86	0.1	0.09	0.74
InventWord_40	0.1	0.09	0.75
RealWords_42	0.1	0.09	0.75
RealWords_43	0.1	0.09	0.76
RealWords_44	0.1	0.09	0.77

ReadComp_03	0.1	0.09	0.75
LetSound_81	0.11	0.1	0.77
LetSound_82	0.11	0.1	0.69
LetSound_83	0.11	0.09	0.77
InventWord_38	0.11	0.1	0.78
InventWord_39	0.11	0.1	0.78
RealWords_40	0.11	0.1	0.78
RealWords_41	0.11	0.1	0.78
LetSound_78	0.12	0.11	0.78
LetSound_79	0.12	0.11	0.78
LetSound_80	0.12	0.1	0.77
InventWord_35	0.12	0.11	0.77
InventWord_36	0.12	0.1	0.77
InventWord_37	0.12	0.1	0.77
RealWords_38	0.12	0.11	0.77
RealWords_39	0.12	0.1	0.77
OralRdFlncy_57	0.12	0.11	0.77
OralRdFlncy_59	0.12	0.11	0.77
OralRdFlncy_60	0.12	0.11	0.77
OralRdFlncy_61	0.12	0.1	0.77
OralRdFlncy_62	0.12	0.1	0.77
OralRdFlncy_63	0.12	0.1	0.77
OralRdFlncy_64	0.12	0.1	0.76
OralRdFlncy_65	0.12	0.1	0.77
RealWords_36	0.13	0.12	0.75
RealWords_37	0.13	0.11	0.75

OralRdFlncy_54	0.13	0.12	0.74
OralRdFlncy_55	0.13	0.11	0.74
OralRdFlncy_56	0.13	0.11	0.67
OralRdFlncy_58	0.13	0.11	0.66
OralRdFlncy_66	0.13	0.11	0.72
ReadComp_02	0.13	0.11	0.68
LetSound_73	0.14	0.12	0.7
LetSound_75	0.14	0.12	0.7
LetSound_76	0.14	0.12	0.66
LetSound_77	0.14	0.12	0.67
InventWord_31	0.14	0.12	0.66
InventWord_33	0.14	0.12	0.65
InventWord_34	0.14	0.12	0.64
RealWords_33	0.14	0.12	0.63
RealWords_34	0.14	0.12	0.63
RealWords_35	0.14	0.12	0.62
OralRdFlncy_51	0.14	0.12	0.61
OralRdFlncy_52	0.14	0.12	0.58
OralRdFlncy_53	0.14	0.12	0.58
LetSound_74	0.15	0.12	0.58
InventWord_32	0.15	0.13	0.58
RealWords_32	0.15	0.13	0.57
OralRdFlncy_42	0.15	0.12	0.54
OralRdFlncy_43	0.15	0.13	0.55
OralRdFlncy_44	0.15	0.13	0.53
OralRdFlncy_45	0.15	0.13	0.53

OralRdFlncy_46	0.15	0.12	0.53
OralRdFlncy_47	0.15	0.13	0.5
OralRdFlncy_48	0.15	0.13	0.5
OralRdFlncy_49	0.15	0.13	0.5
OralRdFlncy_50	0.15	0.12	0.49
ReadComp_01	0.15	0.13	0.48
LetSound_72	0.16	0.13	0.48
InventWord_30	0.16	0.14	0.47
RealWords_31	0.16	0.13	0.46
OralRdFlncy_37	0.16	0.14	0.46
OralRdFlncy_40	0.16	0.13	0.43
OralRdFlncy_41	0.16	0.13	0.45
LetSound_68	0.17	0.14	0.44
LetSound_69	0.17	0.14	0.44
LetSound_70	0.17	0.14	0.42
LetSound_71	0.17	0.14	0.43
RealWords_27	0.17	0.14	0.43
RealWords_29	0.17	0.14	0.42
RealWords_30	0.17	0.14	0.42
OralRdFlncy_38	0.17	0.14	0.41
OralRdFlncy_39	0.17	0.14	0.41
InventWord_28	0.18	0.15	0.41
InventWord_29	0.18	0.14	0.4
RealWords_28	0.18	0.15	0.4
OralRdFlncy_34	0.18	0.15	0.64
OralRdFlncy_35	0.18	0.15	0.72

OralRdFlncy_36	0.18	0.15	0.72
InventWord_25	0.19	0.16	0.7
InventWord_26	0.19	0.16	0.73
InventWord_27	0.19	0.15	0.68
RealWords_26	0.19	0.15	0.7
OralRdFlncy_31	0.19	0.16	0.72
OralRdFlncy_32	0.19	0.16	0.74
OralRdFlncy_33	0.19	0.15	0.71
LetSound_66	0.2	0.16	0.68
LetSound_67	0.2	0.16	0.71
InventWord_23	0.2	0.16	0.71
RealWords_25	0.2	0.16	0.73
OralRdFlncy_28	0.2	0.16	0.7
OralRdFlncy_30	0.2	0.16	0.7
LetSound_64	0.21	0.17	0.71
LetSound_65	0.21	0.16	0.71
InventWord_22	0.21	0.17	0.71
InventWord_24	0.21	0.17	0.71
RealWords_24	0.21	0.17	0.69
OralRdFlncy_25	0.21	0.17	0.66
OralRdFlncy_26	0.21	0.16	0.65
OralRdFlncy_27	0.21	0.17	0.67
OralRdFlncy_29	0.21	0.16	0.64
LetSound_62	0.22	0.17	0.65
LetSound_63	0.22	0.17	0.65
RealWords_20	0.22	0.17	0.62

RealWords_22	0.22	0.17	0.63
RealWords_23	0.22	0.17	0.61
OralRdFlncy_23	0.22	0.17	0.57
LetSound_61	0.23	0.18	0.59
InventWord_21	0.23	0.18	0.57
RealWords_13	0.23	0.18	0.56
RealWords_21	0.23	0.18	0.53
OralRdFlncy_24	0.23	0.17	0.52
LetSound_59	0.24	0.18	0.53
OralRdFlncy_21	0.24	0.18	0.5
OralRdFlncy_22	0.24	0.18	0.51
LetSound_60	0.25	0.19	0.49
InventWord_16	0.25	0.19	0.47
InventWord_17	0.25	0.19	0.46
InventWord_18	0.25	0.19	0.44
InventWord_19	0.25	0.19	0.44
InventWord_20	0.25	0.19	0.43
RealWords_17	0.25	0.19	0.39
RealWords_18	0.25	0.19	0.37
RealWords_19	0.25	0.19	0.35
OralRdFlncy_08	0.25	0.19	0.37
OralRdFlncy_15	0.25	0.19	0.36
OralRdFlncy_16	0.25	0.19	0.69
LetSound_54	0.26	0.19	0.74
LetSound_56	0.26	0.19	0.72
InventWord_11	0.26	0.19	0.74

InventWord_15	0.26	0.19	0.71
RealWords_15	0.26	0.19	0.68
RealWords_16	0.26	0.19	0.73
OralRdFlncy_07	0.26	0.19	0.75
OralRdFlncy_17	0.26	0.19	0.76
OralRdFlncy_18	0.26	0.19	0.74
OralRdFlncy_19	0.26	0.19	0.75
LetSound_53	0.27	0.2	0.74
LetSound_58	0.27	0.2	0.66
InventWord_12	0.27	0.2	0.73
OralRdFlncy_14	0.27	0.2	0.71
OralRdFlncy_20	0.27	0.19	0.72
Dictation_02_02	0.27	0.2	0.71
LetSound_57	0.28	0.2	0.71
InventWord_06	0.28	0.2	0.71
InventWord_07	0.28	0.2	0.68
InventWord_13	0.28	0.2	0.7
InventWord_14	0.28	0.2	0.68
RealWords_06	0.28	0.2	0.69
RealWords_14	0.28	0.2	0.68
OralRdFlncy_04	0.28	0.2	0.66
InventWord_10	0.29	0.21	0.64
RealWords_05	0.29	0.21	0.61
OralRdFlncy_10	0.29	0.21	0.64
OralRdFlncy_11	0.29	0.2	0.63
OralRdFlncy_12	0.29	0.21	0.62

OralRdFlncy_13	0.29	0.21	0.6
Dictation_02_01	0.29	0.21	0.59
InventWord_04	0.3	0.21	0.58
RealWords_12	0.3	0.21	0.58
OralRdFlncy_09	0.3	0.21	0.57
LetSound_55	0.31	0.21	0.57
InventWord_08	0.31	0.21	0.56
RealWords_10	0.31	0.21	0.55
OralRdFlncy_03	0.31	0.21	0.53
OralRdFlncy_05	0.31	0.21	0.51
OralRdFlncy_06	0.31	0.21	0.53
InventWord_03	0.32	0.22	0.5
InventWord_09	0.32	0.22	0.5
RealWords_07	0.32	0.22	0.5
RealWords_11	0.32	0.22	0.48
Dictation_02_03	0.32	0.22	0.47
LetSound_52	0.33	0.22	0.45
InventWord_05	0.33	0.22	0.43
RealWords_08	0.33	0.22	0.45
OralRdFlncy_02	0.33	0.22	0.44
LetSound_51	0.34	0.22	0.58
InventWord_02	0.34	0.22	0.72
LetSound_49	0.35	0.23	0.73
LetSound_50	0.35	0.23	0.71
RealWords_03	0.35	0.23	0.72
RealWords_09	0.35	0.23	0.72

RealWords_02	0.36	0.23	0.68
RealWords_04	0.36	0.23	0.68
LetSound_47	0.37	0.23	0.74
LetSound_48	0.37	0.23	0.73
InventWord_01	0.37	0.23	0.71
LetSound_44	0.38	0.24	0.73
LetSound_46	0.38	0.24	0.74
LetSound_43	0.39	0.24	0.72
LetSound_45	0.39	0.24	0.69
LetSound_40	0.4	0.24	0.66
LetSound_41	0.4	0.24	0.68
LetSound_37	0.41	0.24	0.7
LetSound_39	0.41	0.24	0.7
LetSound_42	0.41	0.24	0.7
OralRdFlncy_01	0.42	0.24	0.63
LetSound_34	0.43	0.25	0.66
LetSound_35	0.43	0.25	0.64
LetSound_38	0.43	0.25	0.65
LetSound_27	0.44	0.25	0.62
RealWords_01	0.44	0.25	0.62
LetSound_36	0.45	0.25	0.62
LetSound_03	0.47	0.25	0.61
LetSound_33	0.47	0.25	0.62
LetSound_25	0.48	0.25	0.6
LetSound_32	0.48	0.25	0.59
PhonAware_07	0.49	0.25	0.59

LetSound_23	0.49	0.25	0.58
LetSound_29	0.49	0.25	0.58
LetSound_31	0.49	0.25	0.58
LetSound_30	0.5	0.25	0.57
LetSound_18	0.51	0.25	0.54
LetSound_20	0.51	0.25	0.55
LetSound_12	0.52	0.25	0.55
LetSound_28	0.52	0.25	0.53
LetSound_13	0.53	0.25	0.53
LetSound_14	0.53	0.25	0.5
Dictation_01_02	0.53	0.25	0.52
LetSound_10	0.54	0.25	0.52
LetSound_26	0.54	0.25	0.52
LetSound_05	0.55	0.25	0.5
LetSound_16	0.55	0.25	0.51
LetSound_24	0.55	0.25	0.51
LetSound_22	0.57	0.25	0.51
LetSound_21	0.58	0.24	0.5
ListeningComp_05	0.59	0.24	0.49
LetSound_04	0.59	0.24	0.48
LetSound_06	0.59	0.24	0.49
LetSound_15	0.59	0.24	0.47
LetSound_19	0.59	0.24	0.45
LetSound_17	0.61	0.24	0.45
PhonAware_05	0.63	0.23	0.43
LetSound_09	0.63	0.23	0.45

Dictation_01_04	0.65	0.23	0.44
PhonAware_04	0.66	0.23	0.44
LetSound_08	0.68	0.22	0.42
LetSound_11	0.68	0.22	0.42
Dictation_01_05	0.68	0.22	0.42
LetSound_07	0.69	0.22	0.42
ListeningComp_04	0.7	0.21	0.42
PhonAware_06	0.7	0.21	0.38
PhonAware_08	0.71	0.21	0.5
PhonAware_10	0.71	0.2	0.46
PhonAware_02	0.72	0.2	0.41
PhonAware_01	0.73	0.2	0.33
LetSound_02	0.73	0.2	0.3
Dictation_01_01	0.75	0.19	0.5
PhonAware_09	0.76	0.18	0.62
PhonAware_03	0.78	0.17	0.42
LetSound_01	0.78	0.17	0.6
ListeningComp_01	0.79	0.17	0.56
ListeningComp_02	0.79	0.17	0.6
ListeningComp_03	0.83	0.14	0.58
Dictation_01_03	0.83	0.14	0.62

Table A3-7: Item Analysis Results for EGMA (Longitudinal Cohort)

Item	**Mean (Difficulty Index**	**Variance**	**Discrimination Index**

AddLvl2Strat4	0.05	0.05	0.53
SubLvl2Strat3	0.05	0.05	0.57
SubLvl2Strat4	0.05	0.04	0.6
AddLvl2Strat3	0.06	0.05	0.68
SubLvl1_20	0.06	0.06	0.68
SubLvl1_19	0.07	0.07	0.7
SubLvl1_18	0.08	0.07	0.7
AddLvl1_20	0.09	0.08	0.7
SubLvl1_17	0.09	0.08	0.7
AddLvl1_19	0.1	0.09	0.72
SubLvl1_16	0.1	0.09	0.7
SubLvl2_04	0.1	0.09	0.71
SubLvl2_05	0.1	0.09	0.7
WordProbStrat3	0.1	0.09	0.71
AddLvl1_18	0.11	0.1	0.72
MissNum_08	0.12	0.1	0.69
MissNum_10	0.13	0.11	0.67
AddLvl2_04	0.13	0.12	0.69
SubLvl1_15	0.13	0.11	0.71
SubLvl2_03	0.13	0.11	0.68
MissNum_05	0.14	0.12	0.65
MissNum_07	0.14	0.12	0.67
MissNum_09	0.14	0.12	0.69
AddLvl1_17	0.14	0.12	0.69
AddLvl2_05	0.14	0.12	0.67
WordProbStrat4	0.14	0.12	0.67

SubLvl1_14	0.15	0.13	0.62
MissNum_03	0.16	0.13	0.53
AddLvl2_03	0.17	0.14	0.59
AddLvl1_16	0.18	0.14	0.58
SubLvl1_13	0.19	0.16	0.58
MissNum_04	0.22	0.17	0.64
AddLvl1_15	0.22	0.17	0.37
SubLvl1_11	0.22	0.17	0.45
SubLvl1_12	0.22	0.17	0.33
SubLvl2Strat1	0.24	0.18	0.58
AddLvl1_14	0.25	0.19	0.26
SubLvl1_10	0.25	0.19	0.21
SubLvl1_09	0.26	0.19	0.26
SubLvl2Strat2	0.28	0.2	0.2
SubLvl2_02	0.3	0.21	0.67
AddLvl2Strat1	0.31	0.22	0.73
SubLvl1_05	0.31	0.21	0.75
AddLvl1_13	0.32	0.22	0.77
SubLvl1_08	0.33	0.22	0.77
SubLvl2_01	0.33	0.22	0.78
SubLvl1_04	0.34	0.22	0.78
SubLvl1_06	0.34	0.23	0.78
AddLvl1_11	0.35	0.23	0.75
AddLvl1_12	0.35	0.23	0.74
SubLvl1_07	0.35	0.23	0.72
AddLvl2Strat2	0.37	0.23	0.72

WordProb_05	0.37	0.23	0.7
AddLvl1_10	0.38	0.24	0.65
SubLvl1_01	0.38	0.24	0.63
MissNum_06	0.39	0.24	0.57
SubLvl1_03	0.39	0.24	0.52
NumIdentify_20	0.4	0.24	0.47
AddLvl2_02	0.4	0.24	0.45
SubLvl1_02	0.4	0.24	0.42
NumIdentify_17	0.41	0.24	0.71
NumIdentify_18	0.41	0.24	0.69
NumDisc_08	0.41	0.24	0.49
WordProb_06	0.41	0.24	0.46
WordProb_03	0.42	0.24	0.4
NumIdentify_19	0.43	0.24	0.5
AddLvl1_09	0.43	0.24	0.53
AddLvl2_01	0.43	0.25	0.22
WordProbStrat1	0.43	0.24	-0.06
NumIdentify_16	0.45	0.25	0.55
AddLvl1_04	0.48	0.25	0.72
AddLvl1_05	0.48	0.25	0.74
AddLvl1_08	0.49	0.25	0.7
WordProb_04	0.49	0.25	0.69
AddLvl1_07	0.5	0.25	0.72
NumDisc_09	0.51	0.25	0.74
AddLvl1_06	0.51	0.25	0.73
AddLvl1_03	0.52	0.25	0.69

NumDisc_07	0.53	0.25	0.67
NumDisc_10	0.54	0.25	0.65
AddLvl1_01	0.54	0.25	0.65
AddLvl1_02	0.55	0.25	0.62
WordProb_02	0.57	0.25	0.55
WordProbStrat2	0.6	0.24	0.52
NumDisc_06	0.61	0.24	0.46
NumIdentify_14	0.64	0.23	0.42
NumDisc_03	0.64	0.23	0.4
NumIdentify_09	0.65	0.23	0.38
NumIdentify_12	0.65	0.23	0.35
NumIdentify_13	0.65	0.23	0.69
NumDisc_05	0.65	0.23	0.66
NumIdentify_11	0.66	0.23	0.46
NumIdentify_15	0.66	0.22	0.38
MissNum_02	0.66	0.22	0.37
NumIdentify_10	0.67	0.22	0.51
NumDisc_04	0.68	0.22	0.55
NumIdentify_07	0.69	0.21	0.23
NumIdentify_08	0.69	0.21	-0.02
NumIdentify_06	0.71	0.21	0.4
NumIdentify_04	0.72	0.2	0.6
NumIdentify_05	0.72	0.2	0.55
NumDisc_02	0.73	0.2	0.58
MissNum_01	0.76	0.18	0.47
NumDisc_01	0.77	0.18	0.56

WordProb_01	0.77	0.18	0.17
NumIdentify_03	0.8	0.16	0.22
NumIdentify_02	0.83	0.14	0.13
NumIdentify_01	0.84	0.13	-0.38

Table A3-8: Floor & Ceiling Effects (Longitudinal Cohort)

Assessment	Learner Scores (%)	Flooring	Ceiling
EGRA	0	42 (1.4%)	
	0-10	276 (9.5%)	
	90-100		65 (2.2%)
	100		4 (0.1%)
EGMA	0	137 (4.7%)	
	0-10	**604 (20.7%) ***	
	90-100		0 (0%)
	100		0 (0%)

Notes: *Ceiling/flooring effects greater than 15% are considered significant (Terwee et. al., 2007). However, we are not concerned about this flooring effect, since less than 5% of our sample scored a zero (0) and the assessment couldn't have been easier. So, there is room for growth.

Table A3-9: Item Analysis Results for EGRA (Cross-Sectional Cohort)

Item	Mean (Difficulty Index)	Variance	Discrimination Index
InventWord_49	0.07	0.07	0.2
InventWord_48	0.08	0.08	0.17
InventWord_50	0.08	0.07	0.11
InventWord_47	0.09	0.08	0.14
InventWord_46	0.1	0.09	0.18
RealWords_48	0.11	0.1	0.37

RealWords_49	0.11	0.1	0.3
RealWords_50	0.11	0.1	0.29
LetSound_94	0.12	0.11	0.35
LetSound_95	0.12	0.11	0.42
LetSound_96	0.12	0.1	0.33
LetSound_97	0.12	0.1	0.4
LetSound_98	0.12	0.1	0.37
LetSound_99	0.12	0.1	0.29
LetSound_100	0.12	0.1	0.33
InventWord_45	0.12	0.11	0.51
RealWords_46	0.12	0.1	0.54
RealWords_47	0.12	0.1	0.65
LetSound_92	0.13	0.12	0.64
LetSound_93	0.13	0.11	0.69
InventWord_44	0.13	0.11	0.64
RealWords_44	0.13	0.12	0.57
RealWords_45	0.13	0.11	0.57
LetSound_91	0.14	0.12	0.63
InventWord_40	0.14	0.12	0.67
InventWord_41	0.14	0.12	0.55
InventWord_42	0.14	0.12	0.69
InventWord_43	0.14	0.12	0.68
RealWords_42	0.14	0.12	0.63
RealWords_43	0.14	0.12	0.65
LetSound_89	0.15	0.13	0.64
LetSound_90	0.15	0.13	0.6

RealWords_40	0.16	0.13	0.72
LetSound_86	0.17	0.14	0.65
LetSound_87	0.17	0.14	0.72
LetSound_88	0.17	0.14	0.66
InventWord_35	0.17	0.14	0.65
InventWord_36	0.17	0.14	0.72
InventWord_38	0.17	0.14	0.65
InventWord_39	0.17	0.14	0.7
RealWords_39	0.17	0.14	0.69
RealWords_41	0.17	0.14	0.64
OralRdFlncy_57	0.17	0.14	0.69
OralRdFlncy_61	0.17	0.14	0.69
OralRdFlncy_62	0.17	0.14	0.7
OralRdFlncy_64	0.17	0.14	0.71
LetSound_85	0.18	0.15	0.71
RealWords_38	0.18	0.15	0.71
OralRdFlncy_58	0.18	0.15	0.72
OralRdFlncy_59	0.18	0.15	0.69
OralRdFlncy_63	0.18	0.15	0.72
ReadComp_04	0.18	0.15	0.72
ReadComp_05	0.18	0.15	0.71
LetSound_83	0.19	0.15	0.74
LetSound_84	0.19	0.15	0.75
InventWord_37	0.19	0.15	0.74
OralRdFlncy_56	0.19	0.15	0.73
OralRdFlncy_65	0.19	0.15	0.75

RealWords_34	0.2	0.16	0.74
RealWords_36	0.2	0.16	0.74
RealWords_37	0.2	0.16	0.75
OralRdFlncy_52	0.2	0.16	0.76
OralRdFlncy_53	0.2	0.16	0.74
OralRdFlncy_54	0.2	0.16	0.76
OralRdFlncy_60	0.2	0.16	0.76
LetSound_81	0.21	0.16	0.74
LetSound_82	0.21	0.16	0.76
InventWord_31	0.21	0.17	0.7
InventWord_34	0.21	0.16	0.7
RealWords_33	0.21	0.17	0.75
RealWords_35	0.21	0.17	0.71
OralRdFlncy_42	0.21	0.17	0.75
OralRdFlncy_51	0.21	0.17	0.74
OralRdFlncy_55	0.21	0.16	0.7
LetSound_78	0.22	0.17	0.73
LetSound_80	0.22	0.17	0.72
InventWord_33	0.22	0.17	0.72
RealWords_32	0.22	0.17	0.72
OralRdFlncy_46	0.22	0.17	0.72
LetSound_79	0.23	0.18	0.72
OralRdFlncy_44	0.23	0.18	0.71
OralRdFlncy_47	0.23	0.18	0.7
OralRdFlncy_48	0.23	0.18	0.65
OralRdFlncy_49	0.23	0.17	0.65

ReadComp_03	0.23	0.18	0.66
OralRdFlncy_43	0.24	0.18	0.65
OralRdFlncy_50	0.24	0.18	0.65
LetSound_77	0.25	0.19	0.61
InventWord_30	0.25	0.19	0.63
RealWords_27	0.25	0.19	0.61
RealWords_29	0.25	0.19	0.61
RealWords_30	0.25	0.19	0.6
RealWords_31	0.25	0.19	0.56
OralRdFlncy_34	0.25	0.19	0.58
OralRdFlncy_37	0.25	0.19	0.57
LetSound_73	0.26	0.19	0.55
LetSound_75	0.26	0.19	0.55
LetSound_76	0.26	0.19	0.54
InventWord_32	0.26	0.19	0.52
OralRdFlncy_33	0.26	0.19	0.5
OralRdFlncy_38	0.26	0.19	0.5
OralRdFlncy_39	0.26	0.19	0.49
OralRdFlncy_41	0.26	0.19	0.48
OralRdFlncy_45	0.26	0.19	0.45
OralRdFlncy_66	0.26	0.19	0.45
LetSound_74	0.27	0.2	0.43
RealWords_25	0.27	0.2	0.42
RealWords_26	0.27	0.2	0.42
RealWords_28	0.28	0.2	0.42
OralRdFlncy_36	0.28	0.2	0.4

OralRdFlncy_40	0.28	0.2	0.4
ReadComp_02	0.28	0.2	0.39
LetSound_69	0.29	0.21	0.39
LetSound_71	0.29	0.21	0.39
LetSound_72	0.29	0.21	0.38
InventWord_28	0.29	0.21	0.69
RealWords_13	0.29	0.21	0.73
OralRdFlncy_32	0.29	0.21	0.74
LetSound_68	0.3	0.21	0.7
LetSound_70	0.3	0.21	0.74
InventWord_25	0.3	0.21	0.68
InventWord_29	0.3	0.21	0.7
RealWords_20	0.3	0.21	0.71
RealWords_22	0.3	0.21	0.73
RealWords_24	0.3	0.21	0.69
OralRdFlncy_26	0.3	0.21	0.68
OralRdFlncy_31	0.3	0.21	0.7
InventWord_22	0.31	0.21	0.73
InventWord_23	0.31	0.21	0.73
InventWord_26	0.31	0.21	0.75
InventWord_27	0.31	0.21	0.71
RealWords_23	0.31	0.21	0.74
OralRdFlncy_21	0.31	0.21	0.76
OralRdFlncy_23	0.31	0.21	0.73
OralRdFlncy_28	0.31	0.21	0.74
OralRdFlncy_29	0.31	0.21	0.74

OralRdFlncy_35	0.31	0.22	0.7
OralRdFlncy_16	0.32	0.22	0.7
OralRdFlncy_27	0.32	0.22	0.74
OralRdFlncy_30	0.32	0.22	0.72
ReadComp_01	0.32	0.22	0.73
InventWord_11	0.33	0.22	0.73
InventWord_24	0.33	0.22	0.71
RealWords_06	0.33	0.22	0.73
RealWords_15	0.33	0.22	0.67
OralRdFlncy_08	0.33	0.22	0.61
Dictation_02_02	0.33	0.22	0.7
InventWord_10	0.34	0.22	0.62
RealWords_17	0.34	0.22	0.63
RealWords_21	0.34	0.22	0.58
OralRdFlncy_24	0.34	0.23	0.59
OralRdFlncy_25	0.34	0.22	0.62
LetSound_65	0.35	0.23	0.6
LetSound_66	0.35	0.23	0.59
LetSound_67	0.35	0.23	0.55
InventWord_16	0.35	0.23	0.55
InventWord_19	0.35	0.23	0.55
InventWord_20	0.35	0.23	0.54
InventWord_21	0.35	0.23	0.52
RealWords_19	0.35	0.23	0.51
OralRdFlncy_07	0.35	0.23	0.46
OralRdFlncy_22	0.35	0.23	0.42

LetSound_64	0.36	0.23	0.41
InventWord_12	0.36	0.23	0.39
InventWord_15	0.36	0.23	0.41
InventWord_17	0.36	0.23	0.71
RealWords_18	0.36	0.23	0.74
LetSound_59	0.37	0.23	0.72
LetSound_62	0.37	0.23	0.76
LetSound_63	0.37	0.23	0.72
InventWord_04	0.37	0.23	0.67
InventWord_13	0.37	0.23	0.72
RealWords_14	0.37	0.23	0.75
OralRdFlncy_18	0.37	0.23	0.73
Dictation_02_01	0.37	0.23	0.76
LetSound_56	0.38	0.24	0.74
LetSound_61	0.38	0.23	0.76
InventWord_07	0.38	0.23	0.64
InventWord_18	0.38	0.24	0.74
RealWords_05	0.38	0.24	0.7
RealWords_16	0.38	0.24	0.75
OralRdFlncy_14	0.38	0.23	0.73
OralRdFlncy_17	0.38	0.24	0.74
LetSound_54	0.39	0.24	0.75
InventWord_14	0.39	0.24	0.7
OralRdFlncy_04	0.39	0.24	0.76
OralRdFlncy_19	0.39	0.24	0.72
Dictation_02_03	0.39	0.24	0.72

LetSound_53	0.4	0.24	0.73
LetSound_60	0.4	0.24	0.69
InventWord_06	0.4	0.24	0.69
InventWord_08	0.4	0.24	0.67
OralRdFlncy_11	0.4	0.24	0.73
InventWord_05	0.41	0.24	0.69
InventWord_09	0.41	0.24	0.69
RealWords_07	0.41	0.24	0.68
RealWords_10	0.41	0.24	0.67
RealWords_11	0.41	0.24	0.65
RealWords_12	0.41	0.24	0.64
OralRdFlncy_15	0.41	0.24	0.66
LetSound_58	0.42	0.24	0.64
RealWords_08	0.42	0.24	0.63
LetSound_57	0.43	0.25	0.62
OralRdFlncy_13	0.43	0.25	0.61
OralRdFlncy_20	0.43	0.25	0.59
LetSound_55	0.44	0.25	0.6
InventWord_03	0.44	0.25	0.56
OralRdFlncy_03	0.44	0.25	0.57
OralRdFlncy_06	0.44	0.25	0.55
OralRdFlncy_09	0.44	0.25	0.54
InventWord_01	0.45	0.25	0.51
InventWord_02	0.45	0.25	0.51
RealWords_09	0.45	0.25	0.49
OralRdFlncy_12	0.45	0.25	0.49

LetSound_51	0.47	0.25	0.48
LetSound_52	0.47	0.25	0.71
RealWords_04	0.47	0.25	0.72
OralRdFlncy_01	0.47	0.25	0.73
LetSound_49	0.48	0.25	0.72
LetSound_50	0.48	0.25	0.66
RealWords_02	0.48	0.25	0.72
RealWords_03	0.49	0.25	0.68
OralRdFlncy_02	0.49	0.25	0.69
OralRdFlncy_10	0.49	0.25	0.73
LetSound_48	0.5	0.25	0.67
LetSound_44	0.51	0.25	0.7
LetSound_46	0.51	0.25	0.74
LetSound_47	0.51	0.25	0.76
OralRdFlncy_05	0.51	0.25	0.74
LetSound_45	0.52	0.25	0.6
LetSound_40	0.53	0.25	0.66
LetSound_43	0.53	0.25	0.74
LetSound_41	0.54	0.25	0.74
LetSound_42	0.54	0.25	0.74
LetSound_35	0.55	0.25	0.68
LetSound_37	0.55	0.25	0.65
LetSound_39	0.55	0.25	0.71
LetSound_03	0.56	0.25	0.67
LetSound_34	0.56	0.25	0.71
RealWords_01	0.56	0.25	0.58

LetSound_27	0.57	0.25	0.66
LetSound_38	0.57	0.25	0.71
PhonAware_07	0.58	0.24	0.69
LetSound_12	0.59	0.24	0.69
LetSound_36	0.59	0.24	0.6
LetSound_18	0.6	0.24	0.67
LetSound_33	0.6	0.24	0.68
LetSound_23	0.61	0.24	0.64
LetSound_13	0.62	0.24	0.63
LetSound_25	0.62	0.24	0.59
LetSound_29	0.62	0.24	0.65
LetSound_31	0.62	0.23	0.6
LetSound_32	0.62	0.24	0.63
LetSound_05	0.63	0.23	0.63
LetSound_14	0.63	0.23	0.55
LetSound_20	0.63	0.23	0.62
LetSound_30	0.63	0.23	0.56
LetSound_10	0.64	0.23	0.6
LetSound_28	0.64	0.23	0.58
Dictation_01_02	0.65	0.23	0.52
LetSound_04	0.67	0.22	0.56
LetSound_16	0.67	0.22	0.58
LetSound_24	0.67	0.22	0.58
LetSound_26	0.67	0.22	0.57
LetSound_22	0.69	0.22	0.51
LetSound_06	0.7	0.21	0.55

LetSound_21	0.7	0.21	0.53
LetSound_15	0.71	0.21	0.54
LetSound_19	0.71	0.21	0.53
PhonAware_05	0.72	0.2	0.46
LetSound_09	0.72	0.2	0.52
ListeningComp_05	0.73	0.2	0.48
LetSound_17	0.73	0.2	0.5
LetSound_08	0.77	0.18	0.49
LetSound_11	0.77	0.18	0.44
Dictation_01_04	0.77	0.18	0.42
PhonAware_04	0.78	0.17	0.46
LetSound_07	0.79	0.16	0.47
ListeningComp_04	0.8	0.16	0.46
Dictation_01_01	0.8	0.16	0.39
PhonAware_06	0.81	0.15	0.18
PhonAware_08	0.81	0.16	0.63
PhonAware_01	0.82	0.15	0.6
PhonAware_02	0.82	0.15	0.49
PhonAware_10	0.82	0.15	0.47
LetSound_02	0.82	0.15	0.46
Dictation_01_05	0.82	0.15	0.39
PhonAware_09	0.83	0.14	0.46
LetSound_01	0.84	0.13	0.26
ListeningComp_02	0.86	0.12	0.45
PhonAware_03	0.86	0.12	0.46
Dictation_01_03	0.87	0.11	0.6

ListeningComp_01	0.88	0.11	0.59
ListeningComp_03	0.9	0.09	0.63
Dictation_02_04	0.99	0.01	0.07

Table A3-10: Item Analysis Results for EGRA (Cross-Sectinal Cohort)

Item	Mean (Difficulty Index)	Variance	Discrimination Index
SubLvl2Strat4	0.02	0.02	0.41
AddLvl2Strat4	0.03	0.03	0.42
AddLvl2Strat3	0.04	0.04	0.48
SubLvl2Strat3	0.04	0.04	0.56
WordProbStrat4	0.04	0.04	0.55
WordProbStrat3	0.08	0.08	0.55
SubLvl2_04	0.14	0.12	0.55
SubLvl2_05	0.15	0.13	0.57
MissNum_08	0.17	0.14	0.57
SubLvl1_20	0.17	0.14	0.6
AddLvl2_05	0.18	0.15	0.58
AddLvl1_20	0.19	0.15	0.58
AddLvl2_04	0.19	0.15	0.6
SubLvl1_18	0.19	0.15	0.61
SubLvl1_19	0.19	0.15	0.59
MissNum_09	0.2	0.16	0.59
SubLvl1_17	0.2	0.16	0.61

MissNum_05	0.21	0.16	0.64
AddLvl2_03	0.21	0.17	0.65
SubLvl2_03	0.21	0.16	0.6
MissNum_03	0.22	0.17	0.35
MissNum_07	0.23	0.18	0.44
AddLvl1_19	0.23	0.18	0.48
AddLvl1_18	0.24	0.18	0.51
SubLvl1_16	0.24	0.18	0.47
MissNum_10	0.25	0.19	0.49
MissNum_04	0.27	0.2	0.52
AddLvl1_17	0.27	0.2	0.5
SubLvl1_15	0.28	0.2	0.48
SubLvl1_14	0.31	0.21	0.5
AddLvl1_16	0.34	0.23	0.29
SubLvl2Strat2	0.34	0.23	0.46
SubLvl1_13	0.35	0.23	0.36
SubLvl2Strat1	0.36	0.23	0.37
AddLvl2Strat2	0.39	0.24	0.34
SubLvl1_12	0.39	0.24	0.42
AddLvl2Strat1	0.4	0.24	0.36
SubLvl1_11	0.4	0.24	0.25
AddLvl1_15	0.41	0.24	0.32
AddLvl1_14	0.42	0.24	0.32
SubLvl1_10	0.42	0.24	0.62
SubLvl1_09	0.43	0.25	0.65
SubLvl2_02	0.43	0.25	0.68

SubLvl2_01	0.45	0.25	0.68
WordProbStrat1	0.45	0.25	0.68
SubLvl1_05	0.48	0.25	0.68
MissNum_06	0.49	0.25	0.69
WordProb_05	0.49	0.25	0.7
AddLvl1_13	0.51	0.25	0.7
AddLvl2_02	0.52	0.25	0.71
SubLvl1_08	0.52	0.25	0.71
SubLvl1_04	0.53	0.25	0.73
AddLvl1_11	0.54	0.25	0.74
SubLvl1_01	0.54	0.25	0.71
SubLvl1_07	0.54	0.25	0.71
NumIdentify_18	0.55	0.25	0.65
NumIdentify_20	0.55	0.25	0.58
NumDisc_08	0.55	0.25	0.59
AddLvl2_01	0.55	0.25	0.56
SubLvl1_06	0.55	0.25	0.49
NumIdentify_17	0.56	0.25	0.6
AddLvl1_12	0.56	0.25	0.63
AddLvl1_10	0.57	0.25	0.49
SubLvl1_03	0.57	0.24	0.46
NumIdentify_16	0.58	0.24	0.44
NumIdentify_19	0.58	0.24	0.36
WordProbStrat2	0.59	0.24	0.19
SubLvl1_02	0.6	0.24	0.06
AddLvl1_09	0.61	0.24	-0.1

WordProb_03	0.61	0.24	0.61
WordProb_06	0.61	0.24	0.69
AddLvl1_05	0.63	0.23	0.72
AddLvl1_01	0.64	0.23	0.7
AddLvl1_04	0.64	0.23	0.69
NumDisc_09	0.65	0.23	0.73
WordProb_04	0.66	0.22	0.73
NumDisc_10	0.67	0.22	0.74
AddLvl1_08	0.67	0.22	0.73
NumDisc_07	0.68	0.22	0.73
AddLvl1_03	0.68	0.22	0.72
AddLvl1_07	0.68	0.22	0.74
AddLvl1_06	0.69	0.21	0.72
AddLvl1_02	0.71	0.21	0.66
WordProb_02	0.72	0.2	0.63
NumDisc_06	0.74	0.19	0.59
NumIdentify_14	0.77	0.18	0.53
NumIdentify_09	0.78	0.17	0.51
NumIdentify_13	0.78	0.17	0.49
MissNum_02	0.78	0.17	0.45
NumIdentify_07	0.79	0.17	0.64
NumIdentify_11	0.79	0.17	0.68
NumIdentify_15	0.79	0.16	0.52
NumDisc_03	0.79	0.17	0.43
NumDisc_04	0.79	0.17	0.43
NumDisc_05	0.79	0.17	0.39

NumIdentify_10	0.8	0.16	0.28
NumIdentify_12	0.8	0.16	0.11
NumIdentify_06	0.82	0.15	-0.1
NumIdentify_08	0.82	0.15	0.52
NumIdentify_05	0.83	0.14	0.52
NumIdentify_04	0.84	0.14	0.48
NumDisc_02	0.87	0.11	0.46
NumIdentify_03	0.88	0.11	0.5
MissNum_01	0.88	0.1	0.15
NumIdentify_02	0.89	0.1	-0.04
NumIdentify_01	0.9	0.09	0
NumDisc_01	0.9	0.09	-0.21

Table A3-11: Floor & Ceiling Effects (Cross-Sectional Cohort)

Assessment	Learner Scores (%)	Floor	Ceiling
EGRA	0	0 (0%)	
	0-10	26 (2.1%)	
	90-100		65 (5.4%)
	100		1 (0.08%)
EGMA	0	21 (1.7%)	
	0-10	102 (8.4%)	
	90-100		0 (0%)
	100		0 (0%)

APPENDIX 4: ENUMERATOR TRAINING

SORDI researchers employed a combination of tools, methods, and communication styles to conduct an extensive enumerator training, which included a 7-day workshop, followed by one-day pilot, for all teams at the main meeting site in Mogadishu. Team members from Kismayo and Baidoa (Baydhabo) who were unable to travel to Mogadishu took part in the theoretical and presentation parts of the 7-day meeting virtually through Google Meet. In addition, SORDI trainers traveled to Kismayo and Baidoa sites to conduct five consecutive days of practice and piloting the tools and protocols using the Tangerine™ platform with the remote participants. All trainees received the following:

- Overview of each survey instrument and the Tangerine™ platform before beginning the detailed administration of any tool. Facilitators used PowerPoint Slides to present information and shared the slides with the trainees at the end of each training day for study and reflections. The facilitators also used simulation tools to enhance practice and understanding.
- Practice sessions for each instrument allowed trainees to get hands-on experience. Practice sessions included two steps.
    - Step one: Trainees practiced using each tool to gain familiarity with its functionality and components and to raise any areas of concern encountered along the way.
    - Step Two: Randomly paired trainee teams practiced using the tools together from the learner/teacher/ head teacher and assessor's perspective. This supplied additional practice for trainees, while also helping facilitators pinpoint any problem areas shared among most trainees.

Enumerator Training Program – Baidoa

Training Title:	**Evaluation/BAB Enumerators' Training**	
Start: October 06, 2021 (Five days)		End: October 11, 2021
Training Hours:	40 Contact Hours	
Female: 1	Male: 13	**Total: 14**
Location: Haldoor Hotel		
District: Baidoa		State: Southwest Somalia
Objectives:		

1. To introduce and inform the MOECHE officials about the LASER-BAB evaluation purposes, methods, and teams 2. To train the enumerators to accurately and effectively administer EGRA, EGMA, Learner Survey, and other survey questionnaires, in electronic format. 3. To enhance the knowledge of the enumerators on ethical and quality aspects of data collection and ethical issues around children's participation in research activity. 4. To identify skilled assessors to serve as enumerators for the data collection for the baseline evaluation survey. 5. To pilot evaluations tools and platforms and check errors and enumerator consistency before deployment		
Tea & Refreshment:	☐ Working	☒ Non-working/Full break
Meals/Lunch:	☒ Yes	☐ No

Time	Session		Baidoa Site
Day 1 (October 06, 2021)			
08:00 – 08:30 a.m.	Registration		Abdullahi Ankar
08:30 – 08:45 a.m.	Participants' Introduction		Mohamed Amin
08:45 – 09:00 a.m.	Opening Remarks		Mohamed Amin Jamal and Abdullahi Ankar
	Review of the Agenda and setting ground rules		Mohamed Amin Jamal and Abdullahi Ankar
09:00 – 09:35 a.m.	Overview of the Evaluation • About LASER • Evaluation Background • Evaluation objectives • Evaluation sites • Methods		Mohamed Amin Jamal and Abdullahi Ankar

Time	Session		Baidoa Site
09:35 – 09:45 a.m.	Tea Break		
09:45 – 10:15 a.m.	Overview: • Early Grade Reading • Early Grade Mathematics • Learner Survey • Teacher and Head-teacher Survey		Mohamed Amin Jamal and Abdullahi Ankar
10:15 – 10:35 a.m.	Tea Break		
10:35 – 11:10 a.m.	Overview of Early Grade Reading Assessment (EGRA) (Purpose, Instrument Content, Results use)		Mohamed Amin Jamal and Abdullahi Ankar
11:10 – 12:30 p.m.	Early Grade Reading Assessment Tasks		Mohamed Amin Jamal and Abdullahi Ankar
12:30 – 01:30 p.m.	Prayers and Lunch Break		
01:30 – 02:00 p.m.	Using Tablets for EGRA and protocols		Mohamed Amin Jamal and Abdullahi Ankar
02:00 – 03:00 p.m.	Practice EGRA Tasks: • Listening Comprehension • Letter Sound Identification • Phonemic Awareness		Mohamed Amin Jamal
03:00 – 03:30 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal
03:30 – 04:00 p.m.	Prayers and tea break		
04:00 – 05:00 p.m.	Practice EGRA Tasks:		Mohamed Amin Jamal

Time	Session		Baidoa Site
	• Invented Words (non-words) • Familiar Words		
05:00 – 05:15 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal
05:15 p.m.	**End of Day-1**		
	Day 2 (October 07, 2021)		
08:30 – 08:45 a.m.	Day-1 Review		Live Session (Kismayo and Baidoa Teams)
08:45 – 10:00 a.m.	Practice EGRA Tasks: • Oral Reading Fluency (ORF) • Reading Comprehension • Writing/Dictation		Mohamed Amin Jamal
10:00 – 10:15 a.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal Abdullahi Ankar
10:15 – 10:45 a.m.	Tea Break		
10:45 – 12:15 p.m.	Complete EGRA Practice-1 (All Tasks)		Mohamed Amin Jamal Abdullahi Ankar
12:15 – 12:30 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal Abdullahi Ankar
12:30 – 01:30 p.m.	Prayers and Lunch Break		
01:30 – 02:30 p.m.	Complete EGRA Practice-2 (All Tasks)		Mohamed Amin Jamal Abdullahi Ankar

Time	Session		Baidoa Site
02:30 – 02:45 p.m.	Reflections By Facilitators		Mohamed Amin
05:15 p.m.	End of Day-2		
Day 3 (October 08, 2021)			
08:30 – 08:45 a.m.	Day-2 Review		
08:45 – 10:00 a.m.	Practice EGMA Tasks: • Missing Numbers • Addition Level-1 • Addition Level-2		Mohamed Amin Jamal Abdullahi Ankar
10:00 – 10:15 a.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal Abdullahi Ankar
10:15 – 10:35 a.m.	Tea Break		
10:35 – 12:00 noon	Practice EGMA Tasks: • Subtraction Level 1 • Subtraction Level-2 • Word Problem		Mohamed Amin Jamal Abdullahi Ankar
12:00 – 12:15 p.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal Abdullahi Ankar
12:15 – 01:15 p.m.	Prayers and Lunch Break		
01:15 – 02:30 p.m.	Complete EGMA Practice-1 (All Tasks)		Mohamed Amin Jamal Abdullahi Ankar

Time	Session		Baidoa Site
02:30 – 02:40 p.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Mohamed Amin Jamal Abdullahi Ankar
02:40 – 03:30 p.m.	Complete EGMA Practice-2 (All Tasks)		Mohamed Amin Jamal Abdullahi Ankar
03:30 – 04:00 p.m.	Prayers and Tea Break		
04:00 – 05:00 p.m.	Practice EGRA & EGMA Combined		Mohamed Amin Jamal Abdullahi Ankar
05:00 – 05:15 p.m.	Reflections By Facilitators		
05:15 p.m.	End of Day-3		
Day 4 (October 09, 2021)			
08:30 – 08:45 a.m.	Practice Longitudinal Learner Survey		Mohamed Amin Jamal Abdullahi Ankar
08:45 – 09:15 a.m.	Reflections on Longitudinal Learner Survey		Mohamed Amin Jamal Abdullahi Ankar
08:45 – 09:15 a.m.	Practice Questionnaires		Mohamed Amin Jamal Abdullahi Ankar
09:15 – 10:15 a.m.	Tea Break		
10:15 – 10:35 a.m. 10:35 – 10:50 a.m.	Reflections on Teacher/Head-Teacher Questionnaires		Mohamed Amin Jamal Abdullahi Ankar
01:30 – 03:30 p.m.	Full Assessment Practice-1		Mohamed Amin Jamal Abdullahi Ankar

Time	Session		Baidoa Site
03:30 – 04:00 p.m.	Prayers and tea break		
04:00 – 05:00 p.m.	Full Assessment Practice-2 (Group Practice)		Mohamed Amin Jamal Abdullahi Ankar
05:00 – 05:15 p.m.	Reflections by Facilitators		Mohamed Amin Jamal Abdullahi Ankar
05:15 p.m.	End of Day-5		
Day 5 (October 10, 2021)			
08:30 – 08:45 a.m.	Day-4 Review		Mohamed Amin Jamal
08:45 – 10:00 a.m.	Group Practice		Abdullahi Ankar
10:00 – 10:15 a.m.	Tea Break		Mohamed Amin Jamal
10:15 – 11:45 a.m.	Interrater Reliability Test-1		Abdullahi Ankar
11:45 – 12:15 p.m.	Reflections on Interrater Reliability Test-1		Mohamed Amin Jamal
12:15 – 01:15 p.m.	Prayers and Lunch Break		
01:15 – 02:15 p.m.	Interrater Reliability Test-2		Mohamed Amin Jamal
02:15 – 02:45 pm.	Logistics: • Review of Protocols (repeat) • Roles of team • Using and following up teaming plan • Coordination • Reporting		Abdullahi Ankar

Time	Session		Baidoa Site
2:45– 3:00 noon	Logistics Handover: • Stationery • Tablets • Stimuli • Other assessment materials		Mohamed Amin Jamal
3:00 – 3:30 p.m.	Break and Prayers		
03:30 – 04:30 p.m.	Recap Assessment: • Process • Ethical concerns • Data quality		Mohamed Amin Jamal Abdullahi Ankar
04:30 – 04:45 p.m.	Closing Remarks		

Additions

- **Covid-19 mitigation measures**
- **Pretest and debrief from Day 8 and 9**

Enumerator Training Program – Kismayo

Training Title:	**Evaluation/BAB Enumerators' Training**	
Start: October 06, 2021 (Five days)		End: October 11, 2021
Training Hours:	40 Contact Hours	
Female: 2	Male: 12	**Total: 14**

Location: Madina Hotel		
District: Kismayo	State: Jubaland	
Objectives: 1. To introduce and inform the MOECHE officials about the LASER-BAB evaluation purposes, methods, and teams 2. To train the enumerators to accurately and effectively administer EGRA, EGMA, Learner Survey, and other survey questionnaires, in electronic format. 3. To enhance the knowledge of the enumerators on ethical and quality aspects of data collection and ethical issues around children's participation in research activity. 4. To identify skilled assessors to serve as enumerators for the data collection for the baseline evaluation survey. 5. To pilot longitudinal evaluation tools and platforms and check errors and enumerator consistency before deployment		
Tea & Refreshment:	☐ Working	☒ Non-working/Full break
Meals/Lunch:	☒ Yes	☐ No

Time	Session		Baidoa Site
Day 1 (October 06, 2021)			
08:00 – 08:30 a.m.	Registration		Abdirahman Sufi
08:30 – 08:45 a.m.	Participants' Introduction		Abdirahman Sufi
08:45 – 09:00 a.m.	Opening Remarks		Omar Abdullahi and Abdirahman Sufi
	Review of the Agenda and setting ground rules		Omar Abdullahi and Abdirahman Sufi
09:00 – 09:35 a.m.	Overview of the Evaluation • About LASER • Evaluation Background • Evaluation objectives		Omar Abdullahi and Abdirahman Sufi

Time	Session		Baidoa Site
	• Evaluation sites • Methods		
09:35 – 09:45 a.m.	Tea Break		
09:45 – 10:15 a.m.	Overview: • Early Grade Reading • Early Grade Mathematics • Learner Survey • Teacher and Head-teacher Survey		Omar Abdullahi and Abdirahman Sufi
10:15 – 10:35 a.m.	Tea Break		
10:35 – 11:10 a.m.	Overview of Early Grade Reading Assessment (EGRA) (Purpose, Instrument Content, Results use)		Omar Abdullahi and Abdirahman Sufi
11:10 – 12:30 p.m.	Early Grade Reading Assessment Tasks		Omar Abdullahi and Abdirahman Sufi
12:30 – 01:30 p.m.	Prayers and Lunch Break		
01:30 – 02:00 p.m.	Using Tablets for EGRA and protocols		Omar Abdullahi and Abdirahman Sufi
02:00 – 03:00 p.m.	Practice EGRA Tasks: • Listening Comprehension • Letter Sound Identification • Phonemic Awareness		Omar Abdullahi
03:00 – 03:30 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Omar Abdullahi

Time	Session		Baidoa Site
03:30 – 04:00 p.m.	Prayers and tea break		
04:00 – 05:00 p.m.	Practice EGRA Tasks: • Invented Words (non-words) • Familiar Words		Omar Abdullahi
05:00 – 05:15 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Omar Abdullahi
05:15 p.m.	**End of Day-1**		
Day 2 (October 07, 2021)			
08:30 – 08:45 a.m.	Day-1 Review		Abdirahman Sufi
08:45 – 10:00 a.m.	Practice EGRA Tasks: • Oral Reading Fluency (ORF) • Reading Comprehension • Writing/Dictation		Omar Abdullahi
10:00 – 10:15 a.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Omar Abdullahi Abdirahman Sufi
10:15 – 10:45 a.m.	Tea Break		
10:45 – 12:15 p.m.	Complete EGRA Practice-1 (All Tasks)		Omar Abdullahi Abdirahman Sufi
12:15 – 12:30 p.m.	Reflections EGRA Tasks: • Participants' reflections • Challenges		Omar Abdullahi Abdirahman Sufi

Time	Session		Baidoa Site
12:30 – 01:30 p.m.	Prayers and Lunch Break		
01:30 – 02:30 p.m.	Complete EGRA Practice-2 (All Tasks)		Omar Abdullahi Abdirahman Sufi
02:30 – 02:45 p.m.	Reflections By Facilitators		Mohamed Amin
05:15 p.m.	End of Day-2		
Day 3 (October 08, 2021)			
08:30 – 08:45 a.m.	Day-2 Review		
08:45 – 10:00 a.m.	Practice EGMA Tasks: • Missing Numbers • Addition Level-1 • Addition Level-2		Omar Abdullahi Abdirahman Sufi
10:00 – 10:15 a.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Omar Abdullahi Abdirahman Sufi
10:15 – 10:35 a.m.	Tea Break		
10:35 – 12:00 noon	Practice EGMA Tasks: • Subtraction Level 1 • Subtraction Level-2 • Word Problem		Omar Abdullahi Abdirahman Sufi
12:00 – 12:15 p.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Omar Abdullahi Abdirahman Sufi

Time	Session		Baidoa Site
12:15 – 01:15 p.m.	Prayers and Lunch Break		
01:15 – 02:30 p.m.	Complete EGMA Practice-1 (All Tasks)		Omar Abdullahi Abdirahman Sufi
02:30 – 02:40 p.m.	Reflections EGMA Tasks: • Participants' reflections • Challenges		Omar Abdullahi Abdirahman Sufi
02:40 – 03:30 p.m.	Complete EGMA Practice-2 (All Tasks)		Omar Abdullahi Abdirahman Sufi
03:30 – 04:00 p.m.	Prayers and Tea Break		
04:00 – 05:00 p.m.	Practice EGRA & EGMA Combined		Omar Abdullahi Abdirahman Sufi
05:00 – 05:15 p.m.	Reflections By Facilitators		
05:15 p.m.	End of Day-3		
Day 4 (October 09, 2021)			
08:30 – 08:45 a.m.	Practice Longitudinal Learner Survey		Omar Abdullahi Abdirahman Sufi
08:45 – 09:15 a.m.	Reflections on Longitudinal Learner Survey		Omar Abdullahi Abdirahman Sufi
08:45 – 09:15 a.m.	Practice Questionnaires		Omar Abdullahi Abdirahman Sufi
09:15 – 10:15 a.m.	Tea Break		
10:15 – 10:35 a.m.	Reflections on Teacher/Head-Teacher Questionnaires		Omar Abdullahi Abdirahman Sufi

Time	Session		Baidoa Site
10:35 – 10:50 a.m.			
01:30 – 03:30 p.m.	Full Assessment Practice-1		Omar Abdullahi Abdirahman Sufi
03:30 – 04:00 p.m.	Prayers and tea break		
04:00 – 05:00 p.m.	Full Assessment Practice-2 (Group Practice)		Omar Abdullahi Abdirahman Sufi
05:00 – 05:15 p.m.	Reflections by Facilitators		Omar Abdullahi Abdirahman Sufi
05:15 p.m.	End of Day-5		
Day 5 (October 10, 2021)			
08:30 – 08:45 a.m.	Day-4 Review		Omar Abdullahi
08:45 – 10:00 a.m.	Group Practice		Abdirahman Sufi
10:00 – 10:15 a.m.	Tea Break		Omar Abdullahi
10:15 – 11:45 a.m.	Interrater Reliability Test-1		Omar Abdullahi Abdirahman Sufi
11:45 – 12:15 p.m.	Reflections on Interrater Reliability Test-1		Omar Abdullahi
12:15 – 01:15 p.m.	Prayers and Lunch Break		
01:15 – 02:15 p.m.	Interrater Reliability Test-2		Omar Abdullahi
02:15 – 02:45 pm.	Logistics: • Review of Protocols (repeat)		Omar Abdullahi Abdirahman Sufi

Time	Session		Baidoa Site
	• Roles of team • Using and following up teaming plan • Coordination • Reporting		
2:45– 3:00 noon	Logistics Handover: • Stationery • Tablets • Stimuli • Other assessment materials		Omar Abdullahi Abdirahman Sufi
3:00 – 3:30 p.m.	Break and Prayers		
03:30 – 04:30 p.m.	Recap Assessment: • Process • Ethical concerns • Data quality		Omar Abdulahi
04:30 – 04:45 p.m.	Closing Remarks		

Additions

- **Covid-19 mitigation measures**
- **Pretest and debrief from Day 8 and 9**

APPENDIX 5: SURVEY INSTRUMENTS

LONGITUDINAL LEARNER SURVEY

LONGITUDINAL TEACHER SURVEY

LONGITUDINAL HEAD TEACHER SURVEY

APPENDIX 6: FIELD NOTES

SORDI Field Report

DATA ANNEX

REFERENCES

AEWG (Accelerated Education Working Group) (2020). *Accelerated Education Programme Monitoring & Evaluation Toolkit.* Geneva: UNHCR. Retrieved from https://inee.org/resources/accelerated-education-programme-monitoring-evaluation-toolkit

BAB. "BAB Fy21 Annual Report." 2021.

Bandura, A. 1977. "Self-efficacy: Towards a unifying theory of behavioral change." *Psychological Review* 84 (2): 192-215.

Bedard, K., and P J Kuhn. 2006. "Where class size really matters: Class size and student ratings of instructor effectiveness." *Economics of Education Review* 27: 253-265.

Bhaumik DK, Roy A, Aryal S, Hur K, Duan N, et al. (2009). Sample size determination for studies with repeated continuous outcomes. *Psychiatric Ann*, 38,765–71.

Blozis, S.A., & Harring, J.R. (2021). Fitting nonlinear mixed-effects models with alternative residual covariance structures. *Sociological Methods & Research. 50(2), 531-566.*

Boisvert, K., Flemming, J. & Shah, R. (2017). *AEWG Guide to the Accelerated Education Principles*. Education in Crisis and Conflict Network. 4. Retrieved from https://scholarworks.umass.edu/cie_eccn/4

Dee, T., and M West. 2011. "The non-cognitive returns to class size." *Educational Evaluation and Pollicy Analysis* 33 (1): 23-46.

Federal Government of Somalia (2017). *Federal Government of Somalia Ministry of Higher Education, Culture and Higher Education -- Education Sector Strategic Plan 2018-2020.* Ministry of Education, Culture and Higher Education, Federal Government of Somalia.

Flemming, Thomas, Tara Toutant, and Helen Raptis. 2002. "Class Size and Effects: A Review." *Fastback 497.*

Fredricks, Jennifer A., Phyllis Blumenfeld, Jeanne Friedel, and Alison Paris. 2005. *School Engagement.* Vol. 3, in *What Do Children Need to Flourish? The Search Institute Series on Developmentally Attentive Community and Society*, edited by Lippman L.H. Moore K.A. Boston, MA: Springer. https://doi-org.ezproxy.lib.purdue.edu/10.1007/0-387-23823-9_19.

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage. [Crossref], [Google Scholar]

Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, 30(22), 2696–2707.

Harter, S. (1988). *Manual for the Self Perception Profile for Adolescents.* Denver: University of Denver.

Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. Structural Equation Modeling, 6, 1-55. http://dx.doi.org/10.1080/10705519909540118

Jackson, Margot, Kathleen Kiernan, and Sara McLanahan. 2017. "Maternal Education, Changing Family Circumstances, and Children's Skoll Development in the United States and UK." *Annals of the American Academy of Political and Social Sciences* 674 (1): 59-84.

Jerald, C,D,. 2007. *Believing and achieving.* Issue brief, Washington, DC: Center for Comprehensive School Reform and Improvement.

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & de Boor, H. (2020). The relationships between school belonging and learners' motivational, social-emotional, behavioural, and academic outcomes in secondary education: a meta-analytic review. *Research Papers in Education, 35*(6), 641-680.

Ladd, G. W., & Dinella, L. M. (2009, February 1). Continuity and Change in Early School Engagement: Predictive of Children's Achievement Trajectories from First to Eighth Grade? *Journal of Educational Psychology*, 190-206.

Miettunen, J., Machova, Z., & Peterson, B.D. "Adolescent Girls' Education in Somalia – Baseline Evaluation." 2020.

Mullis, Ina VS, Michael O Martin, Pierre Foy, and Alka Arora. *Timss 2011 International Results in Mathematics*. ERIC, 2012.

OECD. 2012. "Equity and Quality in Education: Supporting Disadvantaged Students and Schools." https://www.oecd.org/education/school/equityandqualityineducation-supportingdisadvantagedstudentsandschools.htm.

Ravens-Sieberer, U., & Bullinger, M. (1998). News form the KINDL-Questionnair - A new version for adolescents. *Quality of Life Research, 7*, 653.

Roy A, Bhaumik DK, Aryal S, Gibbons RD. (2007) Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*,63, 699–707.

RTI International. 2017. *Setting and Using Benchmarks for Reading Performance.* 02 16. Accessed 03 14, 2022. https://ierc-publicfiles.s3.amazonaws.com/public/resources/10746_Benchmarks_brief_02-16-17_r6.pdf.

RTI International. 2014. *Early Grade Mathematics Assessment (EGMA) Toolkit.* Washington, DC: United States Agency for International Development.

RTI International. 2015. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition.* Washington, D.C.: United States Agency for International Development.

Rubie, C., Townsend, M. A., & Moore, D. W. (2004). Motivational and academic effects of cultural experiences for indigenous minority students in New Zealand. *Educational Psychology, 24*, 143-160.

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research, 75*(3), 417-453.

Stufflebeam, D. L., & Shinkfield, A. J. (1984). Systematic evaluation: a self-instructional guide to theory and practice: Springer

Somalia Education Cluster (2017). *Annual Report 2016.* UNICEF.

Terwee C. B., Bot S. D. M., de Boer M. R., van der Windt DAWM, Knol D. L., et al Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60(1):34–42. - PubMed

UNESCO. 2014. "Quality education needs qualified teachers." *UNESCO.* 3 10. Accessed 3 23, 2022. https://en.unesco.org/news/quality-education-needs-qualified-teachers.

USAID. "Alternative Basic Education for Pastoralists (Abe) Fact Sheet." https://www.usaid.gov/somalia/fact-sheets/somalia-alternative-basic-education-pastoralists-abe.

USAID. 2016. *Draft New Standard Indicators for PIRS for SLE.* United States Agency for International Development.

Wafula, Charles, and Godfrey Mulongo. "Are Children in South and Central Somalia Accessing Education, and Are They Learning? Baseline Information." Social Sciences & Humanities Open 2, no. 1 (2020): 100011.

Wallace Foundation. 2011. "The School Principal as Leader: Guiding Schools to Better Teaching and Learning. https://www.wallacefoundation.org/knowledge-center/Documents/The-School-Principal-as-Leader-Guiding-Schools-to-Better-Teaching-and-Learning-2nd-Ed.pdf