RTI End Line Assessment Report

Core Education Skills for Liberian Youth (CESLY)

Analytical Report

CESLY

Contract No. EDH-I-00-05-00031-00
Task Order No. EDH-I-05-05-00031-00
October 2011

This publication was produced for review by the United States Agency for International Development (USAID) and by Education Development Center, Inc. (EDC). It was prepared by RTI International for the USAID/Core Education Skills for Liberian Youth (CESLY) Project, under EDC Subcontract No. 05379-RTI-00.

Core Education Skills for Liberian Youth (CESLY)

Final Report

Prepared for
USAID/Liberia
and the Ministry of Education, Republic of Liberia
with partial funding from the Presidential Initiative for Expanding Education

and

Education Development Center, Inc.
IDD Division
55 Chapel Street
Newton, Massachusetts 02458-1060

Prepared by
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

RTI International is a trade name of Research Triangle Institute.

Table of Contents

List of Figures v
List of Tables vii
Abbreviations ix
1. Executive Summary 1
2. Introduction: Early Level Intervention in Reading—CESLY 6
2.1 Introduction 6
2.2 Lesson that CESLY Offers for Future Education Projects 8
2.2.1 Curriculum Resource Development 8
2.2.2 Instructional Materials 9
2.2.3 Teacher Training and School-Based Support 10
2.2.4 Community Outreach and Field Implementation 10
2.2.5 System Improvements 11
3. Sustainability and Scale-Up 12
4. Research Design 13
4.1 Assessment Framework 13
4.2 Sample (ALP, NFE) 14
4.2.1 ALP Sample: Baseline, Mid-term, and Final 14
4.2.2 NFE Sample: Baseline and Final 17
4.3 Assessment Tools 18
4.3.1 EGRA Learner Instrument 19
4.3.2 EGMA Learner Instrument 20
4.3.3 Learner Context Interview 20
4.3.4 Principal Instrument 21
4.3.5 Teacher Instrument 21
4.3.6 Classroom Observation 21
4.3.7 School Observation 21
4.3.8 Focus Groups 22
4.3.9 NFE Instrument 22
4.4 Identification, Training, and Selection of Data Collection Assessors 22
4.4.1 Identification 22
4.4.2 Training and Selection of Assessors 22
4.5 Field Deployment and Supervision 23
4.5.1 ALP Data Collection 23
4.5.2 NFE Data Collection 23
4.6 Data Entry 24
5. EGRA and EGMA Reliability Analysis 25
6. Calibration of Mathematics Tasks 27
7. CESLY Program Impact 29

7.1 ALP Program Impact by Group (Regular or Youth) and Sex 29
7.3 NFE Program Impact by Sex 32
7.4 ALP Program Impact Comparing Baseline and Final; Baseline to Midterm to Final Assessments 34
7.5 NFE Program Impact Comparing Baseline and Final 43
8. Liberia Comparisons and Benchmarks 47
8.1 Comparisons with International Benchmarks: DIBELS 47
8.2 Comparisons with Guyana 48
9. EGRA Impact Analysis 50
9.1 Results of Regression Analysis: ALP 50
9.2 Results of Regression Analysis: NFE 51
9.3 Zero Scores 51
10. EGMA Impact Analysis 55
10.1 General Findings, Descriptive Analysis: ALP 55
10.2 General Findings, Descriptive Analysis: NFE 60
11. Principal Questionnaire Results for ALP and NFE 69
11.1 ALP Principals 69
11.2 NFE Principals 71
12. Teacher Interviews: ALP and NFE 74
12.1 ALP Teachers 74
12.2 NFE Teachers 75
13. Student, Teacher, and Principal Optimism 77
14. Perceptions of Safety and Violence 85
14.1 ALP Respondents 85
14.2 NFE Respondents 87
15. Recommendations 91

List of Figures

Figure 1: ALP Average Reading Achievement Scores, by Group Type (Regular and Youth), Final Assessment 30

Figure 2: ALP Average Reading Achievement Scores, by Group Type and Sex, Final Assessment 32

Figure 3: NFE Program Impact on Reading Achievement, by Sex 34

Figure 4: ALP Program Impact on Reading Achievement, Comparing Baseline and Final (Youth and Regular) 36

Figure 5: ALP Program Impact on Reading Achievement, Comparing Baseline to Midterm to Final (Youth and Regular) 37

Figure 6: ALP Average Reading Achievement, Reading Connected Text Score, By Group and Sex, Baseline and Final 39

Figure 7: ALP Average Reading Comprehension Score, by Group and Sex, Baseline and Final Assessments 40

Figure 8: ALP Average Reading Achievement Scores, by Group and Sex: Baseline, Midterm, and Final Assessments 41

Figure 9: ALP Average Reading Comprehension Score by Group and Sex, Baseline to Midterm to Final Assessments (% of students answering 80% or higher correctly) 42

Figure 10: NFE Program Impact on Reading Achievement, Comparing Baseline and Final—Reading and Comprehension 44

Figure 11: NFE Program Impact on Reading Achievement, Comparing Baseline and Final, by Sex—Reading and Comprehension 46

Figure 12: Oral Reading Fluency Scores Compared to International Benchmarks 48

Figure 13: Oral Reading Fluency Scores in Liberia Compared to Other Interventions 49

Figure 14: Percentages of ALP Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline, Midterm, and Final Assessments 53

Figure 15: Percentages of NFE Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline and Final Assessment Scores 54

Figure 16: ALP Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Score, by Subtask 57

Figure 17: Percentages of ALP Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline, Midterm and Final 60

Figure 18: NFE Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores 62

Figure 19: NFE General EGMA Findings 63

Figure 20: NFE Program Impact on Mathematics, Part 1, by Sex: Average Scores 66

Figure 21: NFE Program Impact on Mathematics, Part 2, by Sex: Average Scores 67

Figure 22: Percentages of NFE Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline and Final 68
Figure 23: ALP Principals' Levels of Concern about Safety: Baseline, Midterm, and Final Assessments 71
Figure 24: NFE Principals' Expectations about When NFE Learners Should Become Fluent Readers 72
Figure 25: NFE Principals' Levels of Concern about Safety: Baseline, Midterm, and Final Assessments 73
Figure 26: Question for ALP Teachers: "In the past two years, have things generally improved in your community?" 74
Figure 27: Question for NFE Teachers: "In the past two years, have things generally improved in your community?" 76
Figure 28: Comparison of ALP Principals, Teachers, and Students: "Have things in your community improved in general?" 77
Figure 29: Comparison of ALP Principals, Teachers, and Students: "Has the life of your family and friends improved?" 78
Figure 30: Comparison of NFE Principals, Teachers, and Students: "Have things in your community improved in general?" 79
Figure 31: Comparison of NFE Principals, Teachers, and Students: "Has the life of your family and friends improved?" 79
Figure 32: ALP Principals' Views on What Might Have Caused Any Changes in Their Lives 81
Figure 33: NFE Principals' Views on What Might Have Caused Any Changes in Their Lives 81
Figure 34: ALP Teachers' Views on What Might Have Caused Any Changes in Their Lives 82
Figure 35: NFE Teachers' Views on What Might Have Caused Any Changes in Their Lives 83
Figure 36: ALP Students' Views on What Might Have Caused Any Changes in Their Lives 83
Figure 37: NFE Students' Views on What Might Have Caused Any Changes in Their Lives 84
Figure 38: ALP Principal, Teacher, and Student Views on Safety in their Community (% "yes") 85
Figure 39: ALP Principals' and Teachers' Views on Safety in School (% "yes") 86
Figure 40: ALP Principal, Teacher, and Student Views on Harassment (% "yes") 87
Figure 41: NFE Principal, Teacher, and Student Views on Safety in Their Community (% "yes") 88
Figure 42: NFE Principal and Teacher Views on Safety in School (% "yes") 89
Figure 43: NFE Principal, Teacher, and Student Views on Harassment (% "yes") 90

List of Tables

Table 1: CESLY Assessment Framework 7
Table 2: CESLY ALP and NFE Assessment Framework 7
Table 3: CESLY ALP and NFE Assessment Framework 13
Table 4: Intended versus Actual Sample Size for All Three ALP Assessments 15
Table 5: ALP: Districts Chosen for Sample and Number of Schools per District—ALP Baseline, Mid-term, and Final Assessments 16
Table 6: Intended versus Actual Sample Size for All NFE Baseline and Final Assessments 17
Table 7: NFE: Districts Chosen for Sample and Number of Schools per District: Baseline, Mid-term, and Final Assessments 18
Table 8: Pearson Correlations for EGRA Subtasks 25
Table 9: Cronbach's Alpha Scores for EGRA 26
Table 10: Pearson Correlations for EGMA Subtasks 27
Table 11: Cronbach's Alpha Scores for EGMA 28
Table 12: ALP Average Reading Achievement Scores by Group Type (Regular and Youth), Final Assessment 29
Table 13: ALP Average Reading Achievement Scores, by Group Type and Sex, Final Assessment 31
Table 14: NFE Program Impact on Reading Achievement, by Sex, Final Assessment 33
Table 15: ALP Program Impact on Reading Achievement, Comparing Baseline and Final (Youth and Regular) 35
Table 16: ALP Average Reading Achievement Scores, by Group Type and Sex, Baseline and Final Assessments 38
Table 17: ALP Average Reading Comprehension Score, by Group and Sex, Baseline and Final Assessments 40
Table 18: ALP Average Reading Achievement Scores, by Group and Sex: Baseline, Midterm, and Final Assessments 41
Table 19: ALP Average Reading Comprehension Score, by Group and Sex: Baseline to Midterm to Final Assessments (% of students answering 80% or higher correctly) 42
Table 20: NFE Program Impact on Reading Achievement, Comparing Baseline and Final 43
Table 21: NFE Program Impact on Reading, by Sex, Baseline and Final 45
Table 22: NFE Program Impact on Reading Achievement, Comparing Baseline and Final, by Sex—Reading and Comprehension 45
Table 23: Percentages of ALP Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline, Midterm, and Final 52

Table 24: Percentages of NFE Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline and Final Assessment Scores 53
Table 25: Interpreting Mathematics Average Scores as Assessed by EGMA: ALP Program 55
Table 26: ALP Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores 56
Table 27: ALP Program Impact on Mathematics, Comparing Baseline and Final: Proportion of Students Answering Mathematics Subtasks Correctly 58
Table 28: Percentages of ALP Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline, Midterm, and Final 59
Table 29: Interpreting Mathematics Average Scores as Assessed by EGMA: NFE Program 60
Table 30: NFE Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores 61
Table 31: NFE General EGMA Findings 62
Table 32: NFE Program Impact on Mathematics Scores, by Sex 64
Table 33: Percentages of NFE Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline and Final 67

Abbreviations

AA	Associate of Arts
ALP	Accelerated Leaning Program
Avg	average = mean score
BA	Bachelor of Arts
CESLY	Core Education Skills for Liberia Youth
CIASES	El Centro de Investigación y Acción Educativa Social
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
EDC	Education Development Center, Inc.
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
GOL	Government of Liberia
IDD	International Development Division
M&E	Monitoring and Evaluation
MOE	Ministry of Education
N	sample size
NFE	Non-Formal Education
PMP	Performance Monitoring Plan
PTA	Parent-Teacher Association
RTI	Research Triangle Institute
SD	standard deviation
SSME	Snapshot of School Management Effectiveness
USAID	United States Agency for International Development

1. Executive Summary

In a move to improve reading and math skills among learners who are significantly older than standard, primary school-age, the United States Agency for International Development (USAID) Liberia and the Liberian Ministry of Education (MOE) have collaborated to implement the Core Education Skills for Liberian Youth (CESLY) program. The CESLY program is tasked with providing high-quality education to youth and adults who have not had a chance to attend to their primary education. CESLY is composed of two complementary interventions: (1) the Accelerated Learning Program (ALP) and (2) the Non-Formal Education (NFE) program, coupled with components for development of teachers, instructional materials, the work force, and policy.

CESLY's interventions strive to foster student learning through improved teaching, by providing focused pedagogical training and mentorship to teachers. These initiatives aim to encourage teachers to use learner-centered methods, which in turn will enable learners to learn and improve their performance.

Accelerated Learning Program (ALP). ALP was developed by the Government of Liberia (GOL) as a response to a large number of youth and adults who, because of the long civil war in Liberia, have never attended school or dropped out and thus never had a chance to continue their studies. ALP represents a condensed primary education curriculum for Grades 1 through 6, and is organized into three levels: Level 1 (equivalent of Grades 1 and 2 in regular primary schools), Level 2 (equivalent of Grades 3 and 4 in regular primary schools), and Level 3 (equivalent of Grades 5 and 6 in regular primary schools).

CESLY ALP learners attend school in two types of program groups: either ALP Regular or ALP Youth programs. ALP Regular groups serve learners who are slightly older than the regular school age (targeting youth aged 10–18 years), but who have been delayed from school entry because of the variety of constraining conditions that Liberia has endured. The ALP Youth group serves learners who are significantly older than primary school age (targeting youth aged 18–35 years). Overall, an ALP learner graduating from Level 3 would be expected to have the equivalent of a Grade 6 education, on par with other Liberian students.

Non-Formal Education Program (NFE). In September 2010, CESLY commenced the piloting of the new NFE curriculum. NFE targeted youth aged 15–35 years, but without a strict cap on the upper age limit. The intervention targeted Liberians with zero or little literacy background, but did not limit participation to those who had a little higher content knowledge and wanted to participate. The enrollment of women in NFE was much higher at 73%, than compared to 43% females in ALP Regular and 61% females in ALP Youth. Note that according to the CESLY program goals, NFE learners exiting from Level 1 would be expected to have very rudimentary knowledge of basic literacy and numeracy.

Assessment approach

The CESLY partnership elected to use the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) methodology to investigate the learners' basic literacy and numeracy skills.[1] The program design and two-year program duration demanded a complex assessment framework, described as follows:

- A baseline assessment was conducted in November 2009, at the beginning of the project, in ALP schools for both group types—Youth and Regular—and for Levels 2 and 3.
- At the end of intervention Year 1 (May 2010), a mid-term assessment was conducted with the goal of determining the impact of the CESLY program on learners' reading and mathematics achievement. At this stage, ALP Level 3 graduated in May 2010, and ALP Level 2 became Level 3 in September 2011. This Level 3 was now the last ALP cohort. At this time, CESLY introduced a new NFE curriculum, with the new students that constituted NFE Level 1. For this new NFE cohort, an initial assessment was conducted in January 2011, and then a final assessment was conducted in May 2011.
- A final assessment was conducted at the end of Year 2 (May 2011), to determine the impact of CESLY interventions in both ALP and NFE classrooms.

Key Findings

Reading. The results of the baseline, midterm, and final assessments show that participation in both ALP and NFE interventions had a positive impact on reading achievement. Average performance scores on most reading subtasks that examined skills, including letter identification, oral reading fluency, and reading comprehension, increased from the baseline to the final assessments. The results of descriptive analyses showed ALP learners posting higher average reading achievement than their NFE counterparts (within the context of ALP students working on the Grade 6-level curriculum and NFE students working on the entry-level literacy curriculum). Despite these apparent differences, the results of t-tests showed that the differences between school types were not statistically significant.

Differences in average reading performance also appeared within program types for ALP students. The results of descriptive analyses of the ALP schools showed that Regular group students, who were younger in age, tended to post higher average reading scores than older Youth group students. In the final phase of data collection, the average reading achievement of ALP Regular students was higher than that of ALP Youth students on nearly all (4 of 6) subtasks. However, these differences were small and not statistically significant.

[1] Independently of EGRA and EGMA tools, the CESLY program developed a Rapid Reading Assessment tool and then conducted an initial Rapid Reading Assessment in September 2010 that was used to place learners in the correct level and assess basic reading needs.

In contrast, differences by student sex were statistically significant for both school types. ALP and NFE female students consistently posted lower average scores on each reading achievement subtask in comparison to their male peers in the baseline assessment. This pattern was also noted with both ALP and NFE students in the final assessment. However, when looking at the overall gains, among ALP students, males outperformed females on nearly all subtasks and across both group types (Regular and Youth) in the final assessment (females both scored lower and made smaller improvements between baseline and final than their male counterparts). This finding suggests that the ALP program may have made a greater impact on male achievement than on female achievement.

Given that NFE and ALP students are older than typical primary school age children who are learning to read, it is difficult to make cross-country comparisons with other adolescents and/or adults in terms of levels of key early reading skills. However, when comparing the oral reading fluency scores of NFE and ALP students with those of primary school age students in other countries, it appears that ALP and NFE students continue to lag behind, albeit to different degrees. The ALP groups, both Regular and Youth, approached the oral reading fluency scores found among other Liberian students. Both groups performed considerably lower than U.S. oral reading fluency benchmarks. Undoubtedly, these comparisons between adults and children, who are learning at very different paces, are not ideal, but they provide some insight into the abilities of NFE and ALP students.

Mathematics. A descriptive analysis of EGMA data among ALP participants suggests that participation in the intervention may have positively influenced average reading performance in some mathematical content knowledge. However, the NFE learners are rather behind in terms of their performance on all of the skills tested, and in some cases, their performance was not as strong as at the time of the initial assessment.

Recommendations

The CESLY program showed a positive impact on reading achievement. At the same time, it is not possible to determine the impact of the program as currently conceived in comparison to other types of reading interventions or standard instruction. Further, the NFE program was assessed after a short period of implementation, with three and a half months between the two testing periods, as well as final assessment occurring after only eight and a half months of class operation. Thus, NFE is not directly comparable to the ALP program, which represents a cumulative investment of three years in reading or language instruction. With these limitations in mind, the following recommendations are proposed:

- **Consider sustaining and scaling up the CESLY program.** In May 2011, the ALP program has been phased out and replaced with the NFE program. However,

both ALP and NFE programs showed student learning increases in reading. Given the brief implementation period of NFE in particular, it stands to reason that a longer intervention would result in more reading gains. It is strongly recommended that the CESLY program focus its support toward teachers and schools, so that NFE learners acquire all of the basic skills, as tested by EGRA and EGMA, before the end of Level 1. That is, to learn how to read and perform basic mathematical functions before learners transition to the next level.

- **Improve female reading achievement.** The findings suggest that males outperformed females at baseline, and that gaps in achievement by sex increased over the course of the intervention. For the both programs, additional emphasis should be placed on raising female achievement. During this project period, CESLY has already hired a Gender Specialist, launched retention campaigns for females, conducted female club strengthening, produced reading materials specifically oriented to females, and reoriented the curriculum to meet females' needs (e.g., by placing prevention of pregnancy and gender violence at the start of the lessons before females drop out). However, these and other efforts to strengthen female achievement should be continued and strengthened.
- **Focus on reading comprehension.** Students appeared to make marked improvements in several early literacy skills, such as letter identification. However, the percentage of students who successfully answered 80% of reading comprehension questions remained low. In the CESLY program, basic components of reading were emphasized in Year 1, while Year 2 focused on reading comprehension and practice of writing. In future programming, reading comprehension should be a focus of instruction from program inception, and it should be increasingly emphasized as basic reading skills are acquired. It is recommended that the CESLY program review the EGRA Plus: Liberia reading intervention, which provides clear scope and sequence as to how to approach reading. This reading intervention has proved to be successful in ensuring that students learn how to read in less than one year[2] and has been adjusted for Grades 1–3. It is currently being implemented through the Liberia Teacher Training Program: Phase 2.
- **Focus on mathematics.** Some improvements have been made on the numeracy front, but these are rather low, and more emphasis and focus needs to be given to the mathematics area. The NFE curriculum and support provided to teachers need to be more intense so that all learners acquire basic numeracy skills, as assessed by the EGMA tool, before the end of Level 1. The assessment results indicate that students can perform some tasks (number identification, simple additions and subtractions), but not all basic mathematical functions (multiplication, quantity discrimination, divisions) that are key foundational skills for complex mathematical functions that will come in subsequent y**e**ars and curriculum.
- **Calibrate the curriculum.** Even though at the end of Level 1, NFE learners are expected to have rudimentary literacy knowledge according to the CESLY program, the assessment after only three months of intervention showed that

[2] The effectiveness of the reading interventions developed by RTI are not only limited to Liberia, but also have proven successful in other countries like Kenya and South Africa.

important gains are possible. Reading and mathematics skills that were assessed using EGRA and EGMA are the most basic skills that anyone can learn, and learning them quickly is critical for future learning. It is recommended that the NFE curriculum and teacher training material be calibrated so that these skills are acquired before NFE learners transition to Level 2.

- **Conduct a quasi-experimental study of CESLY.** To fully gauge the impact of the ALP and NFE programs, a quasi-experimental study would be an ideal scenario. The lack of a control group means that it is not possible to accurately and entirely attribute the learning gains seen in CESLY participants to the program itself. Creation of a control group was considered at the beginning of the project; however, it was determined that it would not be possible to create one given the sampling population. For instance, all of the ALP schools had been receiving support for a few years before the CESLY program started, and they were to continue receiving support through the CESLY program. In the context of randomized control trials, this type of scenario leads to the "contamination" of the control group in case one would have been selected. In other words, no true control group would exist because there would be no schools that would not be receiving the intervention. The same issue was noted for NFE schools. All of the NFE schools were to receive the intervention, resulting in the inability to create control schools. After reviewing different possibilities, it was agreed in the end that measuring progress over baseline would be sufficient to gauge the impact of the project.
- **Consider further investigating the implementation of the mathematics intervention.** On more than half of all mathematics tasks, the percentage of zero scores (in other words, the number of students who could not complete the tasks) increased from baseline to final assessment, and those findings were statistically significant. On the others, zero scores decreased, and this, too, was statistically significant in nearly all cases. This finding suggests a certain amount of disequilibrium in the quality of the math intervention by subtask; explanations for this variation should be investigated, so that appropriate steps can be taken to improve mathematics achievement across the board. At the very least, the CESLY program could investigate the following two possibilities. First, for the NFE program, it may be the case that the curriculum is calibrated in such a way that some of the EGMA-assessed tasks are taught in later levels, e.g., Level 2. This would not apply to the ALP students, because they should have already learned this material. Second, it would be very important to determine if teachers' content knowledge of some of these skills are at the level desired and needed.

2. Introduction: Early Level Intervention in Reading—CESLY

2.1 Introduction

Project description: CESLY is a USAID-funded program tasked with providing high-quality education to youth and adults in Liberia and comprises two complementary interventions: ALP and NFE. The CESLY program has been tasked with graduating the final two ALP cohorts over a period of two years, and while doing so, developing and piloting a new NFE curriculum that will cater to Youth adults aged 18 and older.

CESLY's interventions strive to foster student learning through improved teaching, by providing focused pedagogical training and mentorship to teachers. These initiatives aim to encourage teachers to use learner-centered methods, which will enable learners to advance and to improve their performance. CESLY's interventions are directed at promoting effective teacher support and addressing professional challenges.

Accelerated Learning Program (ALP). ALP was developed as a response to a large number of youth and adults who, because of the long civil war in Liberia, have never attended school, dropped out, and never had a chance to continue their studies. There are two ALP program groups—Regular and Youth. The ALP Regular group serves learners who are *slightly* older (targeting, but not limited to, 10- to18-year-olds) than the regular school age. The ALP Youth group consists of learners who are *significantly* older (18 to 35 years of age) than primary school age. The ALP program is basically a condensed primary education curriculum for Grades 1 through 6 and is organized into three levels: Level 1 (equivalent of Grades 1 and 2 in regular primary schools), Level 2 (equivalent of Grades 3 and 4 in regular primary schools) and Level 3 (equivalent of Grades 5 and 6 in regular primary schools).

Non-Formal Education Program (NFE). In September 2010, CESLY commenced the piloting of the new NFE curriculum. NFE targeted youth aged 15–35 years, but without a strict cap on the upper age limit. The enrollment of women in NFE was much higher at 73%, than compared to 43% females in ALP Regular and 61% females in ALP Youth. The NFE curriculum is designed for Liberians with zero or little literacy background, but did not limit participation to those who had a little higher content knowledge and wanted to attend. Note that according to the CESLY program goals, NFE learners exiting from Level 1 would be expected to have very rudimentary knowledge of basic literacy and numeracy.

The CESLY partnership elected to use the EGRA and EGMA methodology to investigate the learners' basic literacy and numeracy skills. Independently of the EGRA and EGMA tools that were applied by RTI International, the CESLY program developed a Rapid Reading Assessment tool and then conducted an initial Rapid Reading Assessment in

September 2010 that was used to place learners in the correct level and assess basic reading needs.

Assessment approach: Using the EGRA and EGMA methodology agreed upon at the outset of the program to investigate learners' basic literacy and numeracy skills, the original assessment framework is presented in ***Table 1.*** For each assessment, equivalent forms assess reading and math performance of students, regardless of whether they belong to the ALP or the NFE program. The rationale behind this approach is that both EGRA and EGMA are testing foundational skills in literacy and numeracy as well as multiple levels of students who are not dependent on any curriculum.

Table 1: CESLY Assessment Framework

Program	Year 1 /Baseline (May 2009) and Mid-term (May 20110)	Year 2 Final Assessment (May 2011)
ALP Regular	Level 2, Level 3	Level 3, Non-Formal literacy/math curriculum
ALP Youth	Level 2, Level 3	Level 3, Non-Formal literacy/math curriculum

However, in light of the MOE mandate to create an entirely new program with an entirely new cohort of students, in early December 2010, RTI International and Education Development Center, Inc. (EDC) discussed the possibility of conducting a quick snapshot of NFE student performance in January 2011, that would serve as an initial measurement point and help detect improvement changes between then and the end of the field testing of Level 1 in June 2011, of this newly introduced curriculum. See ***Table 2***.

Table 2: CESLY ALP and NFE Assessment Framework

Program	Year 1 /Baseline (May 2009) and Mid-term (May 20110)	Year 2 / NFE Initial Assessment (Jan 2011)	NFE and ALP Final Assessment (May 2011)
ALP Regular	Level 2, Level 3	Level 2, Level 3	Level 3 (last year's Level 2)
ALP Youth	Level 2, Level 3	Level 2, Level 3	Level 3 (last year's Level 2)
NFE Level 1	N/A	NFE Level 1	NFE Level 1

This report presents comparisons between CESLY outcomes identified at the baseline, midterm, and final assessments. The increases and the project impact will be presented as "improvements over baseline," since the Project Monitoring Plan (PMP) dictates that the project is evaluated by detecting changes between baseline and final assessments. For the baseline to mid-term step, it is critical to note that because the baseline assessment was undertaken at the beginning of the academic year and the midterm at the end of the year, the assumption is that significant improvements will be seen as a result of typical intra-

level learning. Unlike comparing learning gains in a population with treatment and control schools, comparisons from baseline to midterm had to be carefully examined to understand whether they were distinguishable from the typical learning gains that would have accrued if learners were in non-CESLY schools.

For the final assessment, it is possible to compare end-of-level to end-of-level data for Level 3 students, because this was the one remaining cohort of ALP students. We will be able to determine how much these students learned from Level 2 to Level 3, but also be able to compare their performance to the Level 3 students from the previous year, who had already graduated in the past academic year.

For the NFE students, it is important to note that the initial assessment was conducted in late January 2011, thus, five months after the intervention started and after two major in-service professional development interventions. Therefore, the initial assessment does not measure the base knowledge with which learners started the program. Moreover, the final assessment took place in the middle of May 2011, which means only three and a half months after the baseline assessment was conducted.[3]

The sections that follow discuss both the ALP and NFE interventions' impact on student performance in reading and mathematics.

2.2 Lesson that CESLY Offers for Future Education Projects

The CESLY project has conducted an internal impact assessment of teaching and learning, analysis of field testing documentation, measurement of change in academic scores, and performance assessment in non-formal education content. From these, plus extensive monitoring and other inputs, the project has gleaned the following learning to be applied to future programming for youth.

2.2.1 Curriculum Resource Development

- It appeared that the lesson plans focused more on vocabulary than on some other reading skills initially. The focus on learning the knowledge of sounds and manipulation of sounds is important, as is learning how to blend and decode sounds for reading. Coupled with this, teachers must teach comprehension from the very beginning, so the right balance between pre-reading skills to reading skills and then language must be found to maximize student learning. Initially, more focus should be put on blending and decoding of sounds, but as students learn this skill, increasing emphasis needs to be put on comprehension. EGRA Plus: Liberia has developed a scope (what) and sequence (how) for teaching

[3] Depending on the design of the intervention, it is possible to detect improvements in learner performance. A similar experience was recorded during the USAID-funded EGRA Plus: Liberia project, where students in the first year were assessed at the beginning of the academic year, but the intervention was in effect only three and a half months before the mid-term assessment was conducted. Despite such a short time, significant improvements in student performance were noted.

reading in Liberia that proved that it is possible for regular school-age students to learn how to read in less than one year.

- Our experience in Liberia indicates that the teachers prefer more scripted lesson plans so that they can get through lessons faster. At the same time, teachers need to learn how to assess students as they go, through built-in mastery checks, and catch up those students who are lagging behind.
- The skills that were assessed through EGRA and EGMA need to be acquired by these students as soon as possible, even before they move to Level 2 of the NFE Program.
- Writing and actual practice needs to be emphasized. In the context of reading and moving to higher language concepts, it is important that students learn the sound-to-print relationship as soon as possible.
- The use of student workbooks needs to be emphasized. Liberia is one of the poorest countries in the world, and there are almost no books or workbooks provided to students. For all future projects, it is important to mention that system improvements will have limited impact on student performance. If learners do not have books or workbooks, it will be more difficult for learning to take place.

2.2.2 Instructional Materials

- Visual aids can greatly help in the teaching and learning process, but it is most effective to provide plenty of learning materials so that learners have items to read that are relevant to their lives as youth.
- Achieving a ratio of 1:1 for student to learning materials wherever possible has a significant impact on reading.
- Where sufficient materials prove too costly to procure for all learners, mobile libraries can be an effective means for ensuring that materials reach classroom level. Projects and governments should engage more in creating public-private partnerships at the global level to bring books to Liberia. It is possible to find books that are culturally appropriate for a short-period of time. For example, a charity, Books for Africa, provides such books. At the same time, projects and the GOL need to establish the context in-country where local publishing should be encouraged. Liberia is probably one of the most resource-constrained countries in the world, and it is not a surprise to see that student performance is lagging behind.
- The project has learned that the teachers' skills to plan lessons and implement these can vary and lack greatly. Some teachers are able to apply newly acquired knowledge with some support, and other teachers require much more support. It is recommended that another look is given to the way teacher manuals are developed, and if possible, for at least the early and foundational skills in reading and mathematics, some of these lessons be scripted. This is beneficial for at least three reasons. First, it provides clear and direct instructions to teachers on what to teach, how to teach it, and how to measure progress, and as such it ensures that a scope and sequence for teaching either reading or mathematics is followed. This process will lead to faster acquisition of these skills by students. Second, it

"forces" teachers to adopt and learn new pedagogical practices that are effective. Third, learners will gain the necessary foundational skills faster, enabling them to start reading and learning on their own.

2.2.3 Teacher Training and School-Based Support

- Continue emphasis on basic core content knowledge in teacher training sessions. Of critical importance is not to overwhelm teachers initially. Experience has taught us on other projects in Liberia that if daily lesson plans are structured in a similar fashion, then teaching teachers how to implement one or two weeks' worth of lessons is often sufficient for them to grasp the approach necessary for realizing these and future lesson plans. At the same time, we need to be cognizant that teachers are acquiring new pedagogy as they go, and more often than not lack content knowledge. More support for both of these areas (approach and content knowledge) needs to be considered during the next phase of the CESLY program.
- Consider developing and piloting a pre-service training in alternative basic education so that the new teachers receive a more solid foundation in literacy and numeracy and are better prepared to teach adults.
- Budget for adequate monitoring to allow facilitators to receive constant, continual on-site coaching to help them master letter sounds, sounding out words, and the content of teaching reading, to better help their learners master literacy and reading. Given the long-term disruption of education services provision and support to teachers in the past decade or so, it is of utmost importance that teachers are supported at the school level, including mentorship provided at the school using the structured approach, and coaching needs to be leveraged so that teachers can start working and learning together. If possible, a master teacher should be selected at the school level and put in charge of ensuring that all teachers learn, regardless of whether it is asked of them or not. It is also important that the principals are oriented about the new practices and that they are supportive of this program. Without the strong and committed leadership at the school level, very little can be expected.

2.2.4 Community Outreach and Field Implementation

- Continued acquisition of reading and mathematical skills cannot succeed without constant vigorous effort to keep youth enrolled. Enrollment campaigns, targeted efforts to reach girls, and high level of community engagement are important strategies to keep youth interested in learning. As youth get older, they will be increasingly dependent on having their own sources of income. Thus, linking education to its practical application in their lives, such finding jobs that require literacy and numeracy, is important.
- Pilot better functioning solar options to provide light at night that is more conducive to learning. This is the biggest problem in Liberia overall, lack of electricity. Just as it was highly recommended that student materials are provided in large quantities, this is another one of those critical elements of improved learning. Most NFE classes are held at night, and visibility is limited. Research shows that any light, be it candle or solar light, is good for reading and that low

visibility does not have any impact on one's eyesight. When designing new interventions, this particular issue—sources of light for learning at night—must be adequately addressed.

- Plan timing and scheduling of classes so that youth and older learners can maximize the time that they are available to participate in class. This has been discussed previously, but lessons need to be so well planned that there is no "downtime" and that every minute of students' time in school leads them to being one step closer to becoming literate.

2.2.5 System Improvements

The CESLY program has helped the MOE develop and pass an alternative, basic education policy. The program has also helped the ministry develop an out-of-school youth component as part of the five-year medium term plan. To sustain gains for youth, these policies and strategies need to be supported and put into practice, both by the ministry and the implementing partners that support the ministry. The learning that RTI has supplied to the USAID/CESLY program at the close of implementation feed into these lesson learned and help the program focus on the strategies that it has developed to improve reading scores in the future.

Overall, the MOE and CESLY staff have learned how to conduct EGRA and EGMA assessments. Over the period of two years and longer, if including other projects, dozens of MOE staff have been trained and deployed for data collection. Nevertheless, this knowledge is now harbored with individual MOE staffers who may or may not be working for the MOE next year. Unless the use of early grade learning assessment is incorporated into the MOE's official assessment framework, it is unlikely that data collected by applying these assessments, and assessment tools themselves, will be used in future as a policy and decision-making tool. Use of early grade learning assessment, or in the context of youth this could be early stage learning assessments, will remain a tool used by individual projects.

3. Sustainability and Scale-Up

EDC has gleaned that in the next phase of youth programming, the alternative basic education strategy needs to be rolled out to all Liberian counties. Newly formed county school boards need to be integrally involved in implementation. Civil society implementers need to be provided with the curriculum and given the freedom to implement. The ministry officials, at the highest levels, need to provide leadership in advocating for the advancement of youth. In all of these efforts, reading needs to be emphasized as a key priority foundation for sustaining higher gains in education. This focus on reading can occur through interventions in curriculum development, instructional materials development, instructional materials design, teacher training, teacher support, assessment focus, and other activity interventions areas.

Focusing attention on the needs of youth, and specifically on their need to develop sound foundational skills in reading, serves as a foundation for advanced development of the nation. In comparing Liberia's reading scores with the global equivalent, it is noted that great advocacy and attention needs to be placed on significant support to increase reading gains. The ability to read needs to be recognized as foundational for success of older learners in the workforce. Even the private sector, which has a need for a competent workforce, has a role in leveraging support for reading education for youth. When government, donors, civil society, the private sector, and teachers and learners themselves come to work together with focused effort to increase reading scores, significant gains can be accomplished.

4. Research Design

4.1 Assessment Framework

With the increasing focus on being able to demonstrate the project impact on improved student learning, USAID/Liberia requested that as part of the monitoring and evaluation (M&E) efforts, the CESLY project also include quantitative assessments of student learning in the areas of reading and mathematics. In collaboration with EDC, RTI suggested that reading and math assessments are conducted, along with asking additional questions in response to the needs of the PMP. Given that the ALP curriculum also includes other subjects—science, social studies, and life-skills—EDC requested that tests for these three subjects be developed as well. At the outset of the project, it was agreed that developing these tests well would require significant time and resources, and as a result, EDC and RTI agreed to "inherit" the tests that were developed as part of the CESLY predecessor program—the USAID-funded ALP.

The quantitative and qualitative assessments that were, in the end, conducted in a sample of CESLY schools during the life of the project provided needed information for measurement of reading, mathematics, and other areas of non-formal education. Quantitative assessments focused on reading and mathematics, including several test items for other ALP skills. Qualitative assessment focused on discerning time-on-task spent by teachers on ALP subjects (classroom observation), and the impact of ALP curriculum on violence reduction and other challenges that students are facing (student, teacher, and principal interviews). The impact of the CESLY project was measured for both ALP students and NFE students as presented in ***Table 3*** below.

Table 3: CESLY ALP and NFE Assessment Framework

Program	Year 1 /Baseline (May 2009) and Mid-term (May 2010)	Year 2 / NFE Baseline (Jan 2011) NFE and ALP Final Assessment (May 2011)
ALP Regular	Level 2, Level 3	Level 3 (last year's Level 2)
ALP Youth	Level 2, Level 3	Level 3 (last year's Level 2)
NFE Level 1	N/A	NFE Level 1

With this assessment approach, the learners were assessed in three waves for ALP intervention (baseline, mid-term, and final assessments) and in two waves for NFE interventions (initial and final assessments). For each application, the research design was intended to ensure a sample large enough to be able to note the project impact with statistical significance. Data collectors also recorded information from classroom and school observations (as well as interviews with principals and teachers) to help determine the school context and its ability to support learning, in terms of the availability of

resources and of the quality of pedagogical supports, as well as of prevailing teacher practices in classrooms.

Note for future projects. On several occasions, it was noted that conducting quantitative assessments of this scale takes time, because these assessments usually take place at the end of a given academic year. Therefore, it is important to keep in mind for future projects that the data coming from these assessments will most likely come in too late for the next year's programming. In this context, it is recommended that projects conduct smaller, less formal assessments of reading and math performance to be able to determine if the intervention is unfolding as planned. Larger studies with statistical significance could be conducted as impact studies, while smaller, more informal studies would be more informative for guiding the project's strategy development.

More details about the sampling approach, instruments, and findings are presented in the following sections.

4.2 Sample (ALP, NFE)

4.2.1 ALP Sample: Baseline, Mid-term, and Final

To estimate the levels of reading and mathematics achievement in sampled CESLY schools, as well as the overall impact of the program on learner achievement in these areas, the CESLY research design was to implement the EGRA and EGMA tools to assess learners in a representative sample of ALP schools, with proportional sampling across ALP Regular and ALP Youth programs—although the collected data varied from the research design in specific areas explained below.

A total of three assessments have been conducted: baseline (November 2009), midterm (May 2010), and final (May 2011). The assessment of students in math and reading skills in ALP schools is reported by the following four levels for baseline and midterm: ALP Regular, ALP Youth, Level 2, and Level 3. For the final assessment, given that there was only one cohort—Level 3—the data was collected across ALP Regular and Youth groups for this level alone.

To ensure representation of the sample for assessment of reading and math skills in ALP Youth and ALP Regular groups for Levels 2 and 3 in Year 1 and 2 of the project, RTI employed a clustered two-stage proportional-to-population sampling approach. The first level of sampling was to choose districts that are proportional to population in ALP schools—thus sampling at the district level. The second or next level of sampling was to choose a fixed number of schools in each district. ***Table 4*** provides an overview of the intended and actual samples, while ***Table 5*** provides an overview of chosen districts and number of schools per chosen districts for all three assessments.

In each school, RTI randomly sampled 10 students from Level 2 and 10 students from Level 3, in the end resulting in a total sample of more or less 2,000 students for baseline and mid-term assessments. For the final assessment, RTI aimed at choosing 20 students

from ALP Youth and 20 students from ALP Regular groups, resulting in a sample of 1,200 students. The intended and actual samples for each assessment are as follows.

Table 4: Intended versus Actual Sample Size for All Three ALP Assessments

Assessment/Sample Size	Baseline: Nov 2011	Mid-Term: May 2010	Final: May 2011
Intended number of schools/ students	100 schools / 2,000 students	100 schools / 2,000 students	60 schools / 1,200 students
Actual number of schools/ students	95 schools / 1,703 students	84 schools / 1,637 students	60 schools / 667students

For each of the assessments, we oversampled to make sure that we have a sufficient number of observations for the analysis. For the final assessment, we oversampled by more than usual because we knew that the enrollment in the last year of ALP program may be problematic. This assumption turned out to be true, and in the end, we received just enough data to be able to perform the analysis. Our standard error will be higher than we would like, but not significant to the point that data is not going to be reliable.

Table 5: ALP: Districts Chosen for Sample and Number of Schools per District—ALP Baseline, Mid-term, and Final Assessments

Baseline: November 2009			Mid-Term: May 2011			Final Assessment: May 2011		
County	**District**	**# of schools**	**County**	**District**	**# of schools**	**County**	**District**	**# of schools**
Bong	Fuamah	3	Bong	Fuamah	4	Bong	Fuamah	2
Bong	Gbarnga	2	Bong	Panta Kpaii	4	Bong	Suakoko	5
Bong	Kokoyah	4	Bong	Salala	5	Bong	Kokoyah	1
Bong	Konobo	6	Bong	Suakoko	5	Bong	Panta Kpaii	2
Bong	Panta Kpaii	6	Grand Gedeh	Konobo	4	Bong	Salala	3
Bong	Salala	5	Grand Gedeh	Putu	4	Bong	Zota	2
Bong	Zota	3	Grand Gedeh	Tchien	5	Grand Gedeh	Konobo	2
Grand Gedeh	Gbarzon	4	Lofa	Kolahun	9	Grand Gedeh	Gbarzon	4
Grand Gedeh	Putu	4	Lofa	Voinjama	6	Grand Gedeh	Putu	2
Lofa	Foya	3	Lofa	Zorzor	5	Lofa	Salayea	3
Lofa	Salayea	5	Maryland	Barrobo	5	Lofa	Zorzor	5
Lofa	Voinjama	4	Maryland	Karluway 2	5	Maryland	Sodoken	2
Lofa	Zorzor	6	Maryland	Pleebo	4	Maryland	Pleebo	2
Maryland	Barrobo	3	Monrovia	Careysburg	5	Monteserrado	Careysburg	3
Monteserrado	Left Bank	9	Monrovia	Greater Monrovia I	3	Monteserrado	Left Bank	9
Maryland	Karluway # 2	5	Monrovia	Greater Monrovia ll	1	Nimba	Bain-Garr	2
Maryland	Pleebo	8	Monrovia	St. Paul River Left Bank	10	Nimba	Saclepea # 2	3
Monteserrado	Careysburg	4	Monrovia	St. Paul River Right Bank	4	Nimba	Sanniquellie- Mah	6
Monteserrado	Greater Monrovia I	3	Nimba	Saclepea 1	5	Nimba	Zoe-Geh	2
Nimba	Bain-Garr	5	Nimba	Twah River	2	**Total schools for final**		**60**
Nimba	Saclepea # 2	3	Nimba	Sanniquellie-Mah	5			
Nimba	Sanniquellie- Mah	5						
Total schools for baseline		**100**	**Total schools for mid-term**		**100**			

4.2.2 NFE Sample: Baseline and Final

As part of CESLY's mandate, a total of 89 NFE sites were operational in six CESLY counties. This is the first cohort of students of NFE—Level 1. The NFE program is open to all who wish to continue their education and the program is targeted to learners in the age range of 15 to 35 years, while actual learners' ages ranged from 18 to 60+ years.

A sample of around 400 student observations was determined to be sufficient to make reliable statistical inferences about the performance of NFE Level 1 students. With respect to the measurement tools, the CESLY technical team advised that the ALP mid-term assessment tools were used for the NFE assessment, with some important, yet not too-demanding adjustments. The NFE sites posed more challenges for assessment than regular or ALP schools for the following reasons: they only operate in the evening hours between 6:30 and 9:30 p.m., and they operate only on certain days of the week that is not necessarily common for all sites. Some sites are in operation Monday through Wednesday, some sites are operating Friday through Sunday, and some Monday, Tuesday, and Friday.

To ensure that a sufficient number of student observations was available, RTI oversampled to account for possible challenges that we might encounter, such as low enrollment in some sites, bad road conditions, for example. As a result, we drew a sample of 30 NFE sites, with the goal of assessing 20 students per site, which would lead to 600 student observations. This approach was supposed to provide us with a buffer that would ensure that we reached the sample size of 400 students. In the end, the sample realized was 433 students, which is slightly more than what we had hoped to have. The sample was drawn to be proportional to enrollment in NFE sites, thus being representative of all sites across all six counties. See ***Table 6*** for the intended sample versus actual sample, and ***Table 7*** for an overview of sites chosen for NFE baseline and mid-term assessments.

Table 6: Intended versus Actual Sample Size for All NFE Baseline and Final Assessments

Assessment/Sample Size	Baseline: Nov 2011	Final: May 2011
Intended number of schools/students	30 sites schools / 600 students	30 sites schools / 600 students
Actual number of schools/students	23 schools / 366 students	30 schools / 433 students

Table 7: NFE: Districts Chosen for Sample and Number of Schools per District: Baseline, Mid-term, and Final Assessments

Baseline - November 2009			Final Assessment: May 2011		
County	**District**	**# of schools**	**County**	**District**	**# of schools**
Bong	Zota	1	Bong	Kokoyah	3
Bong	Panta Kpaii	2	Bong	Panta Kpaii	1
Grand Gedeh	Gbarzon	3	Bong	Suakoko	2
Grand Gedeh	Konobo	2	Bong	Zota	1
Grand Gedeh	Putu	1	Grand Gedeh	Gbarzon	2
Lofa	Kolahun	1	Grand Gedeh	Konobo	1
Lofa	Salayea	2	Grand Gedeh	Putu	1
Lofa	Zorzor	1	Grand Gedeh	Tchien	1
Maryland	Harper #2	1	Lofa	Kpakio	1
Maryland	Karluway # 1	1	Lofa	Vezala	1
Maryland	Karluway # 2	2	Lofa	Voinjama	1
Maryland	Pleebo	1	Lofa	Zorzor	1
Monteserrado	Careysburg	1	Maryland	Harper #1	1
Monteserrado	Greater Monrovia	1	Maryland	Karluway # 2	2
Monteserrado	Left Bank	3	Maryland	Pleebo	1
Maryland	Right Bank	1	Monteserrado	Greater Mon. II	3
Nimba	Bain-Garr	1	Monteserrado	Left Bank	1
Nimba	Gbehlay-Geh	2	Maryland	Right Bank	2
Nimba	Saclepea 2	1	Nimba	Bain-Garr	2
Nimba	Yarpea-Mah	1	Nimba	Twah River	2
Nimba	Sanniquellie- Mah	1	Total schools for final assessment		30
Total schools for baseline		30			

4.3 Assessment Tools

Reading is the most foundational skill that we acquire during our time spent in school. It is used as a foundation for our introduction to content and knowledge that is located in more complicated texts, and thus important to learn how to read so that we can read to learn. EGRA is an orally administered assessment that is targeted at measuring the pre-reading and reading skills that are foundational to later reading (and academic) success. EGRA takes approximately 15 minutes to administer and is often combined with a questionnaire that measures a several student background variables to assist in explaining some of the reading outcome findings.

Given the increased importance of mathematics to the knowledge economy and the ability of learners in Sub-Saharan Africa to find gainful and appropriate employment, as well as the relationship between early mathematics acquisition and future academic success, there has been increased interest in using the findings and experience of EGRA-

based reading assessments and the resultant reforms in several countries to inform assessment and associated interventions in the area of mathematics.

The EGRA and EGMA tools were adapted to the Liberian context and specifically for CESLY, because Liberia had previously developed EGRA tools for the USAID-funded EGRA Plus: Liberia program; mini-pilot assessments were carried out to ensure that the tools were appropriately matched to learners' prior knowledge. This process was led by RTI's reading and mathematics experts, with strong input from local stakeholders, MOE staff, and assessor candidates.

EGRA has been used in assessing students in Grades 1–6, while the EGMA has been used for assessing students in Grades 4 and 5. However, neither of these two instruments had ever been used for assessing adult learners. Therefore, we carefully calibrated the instruments by applying several different methods to ensure their reliability and validity:

- First, RTI's reading (Dr. Marcia Davidson) and mathematics (Dr. Linda Platas) experts reviewed the ALP Teacher Manuals and the content taught to learners. Based on this review, the first draft tools both for reading and for mathematics were crafted.
- Next, we organized a week-long workshop to discuss the reading and math assessments with Liberian stakeholders as a way of "ground truthing" their validity. The reading and mathematics assessment tools were then piloted in two ALP schools.
- Finally, all instruments went through a series of iterative revisions and additions with the CESLY team, including the EDC core team and affiliated experts.

In addition to the reading and mathematics tools, the following presents a list of tools developed and used through CESLY, each discussed in turn and in detail in forthcoming sections: Learner Instruments (EGRA, EGMA, and Interview), Teacher Instrument, Principal Instrument, Classroom and School Observation Tools, Focus Group Discussion Tool, and NFE Instrument.

4.3.1 EGRA Learner Instrument

EGRA assessment is an orally administered tool and when timed, it takes on average of 15 minutes to assess. The following are the tasks that were included in the reading assessment:

1. *Letter-naming fluency*: ability to read the letters of the alphabet without hesitation and naturally. This is a timed task that assesses automaticity and fluency of letter recognition. It is timed to 1 minute, which saves time and also prevents the child from having to spend time on something with which they are having difficulty.
2. *Non-familiar or nonsense word oral reading fluency*: ability to process words that could exist in the language in question, but do not, or are likely to be very unfamiliar. The non-words used for EGRA are truly made-up words. They assess the child's ability to "decode" words fluently. This task is timed to 1 minute.

3. *Connected text oral reading fluency*: ability to read a passage, about 60 words long, that tells a story. It is timed to 1 minute.
4. *Comprehension in connected text*: ability to answer five questions based on the passage read.
5. *Listening comprehension*: being able to follow and understand a simple oral story. This assesses the child's ability to concentrate and focus to understand a very simple story of three sentences with simple non-inferential (factual) questions. It is considered a pre-reading skill.
6. *Dictation*: ability to hear, comprehend, and write a basic sentence with content familiar to the assessed child, using appropriate spelling, punctuation, and capitalization.

4.3.2 EGMA Learner Instrument

EGMA is also orally administered and takes about 15 minutes. The following are the tasks that are included into the mathematics assessment:

1. *Number identification*: learners were asked to identify particular numbers of varying difficulty levels but appropriate for Grades 1–3 learners according to the curriculum.
2. *Quantity discrimination*: learners were asked which of two numbers were bigger, testing place value and number sense. This task was timed.
3. *Missing number*: given a list of three of four numbers, one of which is missing, the child is asked to identify the missing number.
4. *Addition:* a list of common and simple addition facts were presented to the learners, who were asked to solve them as quickly as possible. There were two versions of this addition task, with the second presenting slightly more computational problems.
5. *Subtraction:* similar to the addition task above, learners were presented with simple subtraction problems and asked to solve them. There were two versions of this subtraction task, with the second one slightly more difficult. This was timed.
6. *Multiplication:* learners were presented with a set of multiplication problems and asked to solve them. This was not timed.
7. *Division:* for this task, learners were asked to solve a set of division problems. Similar to the multiplication task, this task was not timed.
8. *Shape recognition:* given several shapes on a page, the learners were asked to identify the circles, squares, rectangles, and triangles.

4.3.3 Learner Context Interview

As per the commitment to EDC, we also included several test items for assessing student knowledge in life skills, social studies, and science. These were assessed in group settings using paper and pencil. The support was provided by an assessor who read the questions

and multiple choice answers aloud. A total of 38 questions were asked across the three mentioned areas, and this enabled the researchers to discuss each of the subjects. No analyses on these questions will be presented in this report.

We have also included a number of student context interview questions that will provide (a) information on the socio-economic status of learners interviewed, (b) information on the learning environment, and (c) inputs on changes in perception towards various issues that communities and schools face.

4.3.4 Principal Instrument

The questions in the principal instrument were designed to aim at assessing (a) the experience of principals in their current position, (b) the opportunities for in-service teacher training, (c) the overall learning environment at the school level and principals' support to teachers, (d) the role and authority of parent-teacher associations, (e) the support received from education officers, and (f) inputs on changes in perception towards various issues that communities and schools face.

4.3.5 Teacher Instrument

The questions in the teacher instrument were designed in collaboration with the CESLY technical team and assess the following aspects: (a) practices used to teach reading, (b) practices used to teach math, (c) support received from the principal and other education officers, (d) general teaching practices and collaboration with other teachers, and (e) inputs on changes in perception towards various issues that communities and schools face.

4.3.6 Classroom Observation

The classroom observation tool consists of two parts: (1) recording information on availability of teaching and learning materials, and (2) recording pedagogical practices used by teachers in transferring knowledge. The tool also provides the time on task. The content knowledge of teachers is not discerned through this tool, but rather the most prevalent teaching style and time spent teaching. The content knowledge of teachers is easily understood, based on student scores on reading and math and other subjects, as discussed above. No analysis will be presented in this report.

4.3.7 School Observation

The data collected using this tool provides information about (a) enrollment per level of interest (broken down by sex), (b) state of school infrastructure, (c) availability of learning and teaching materials, and (d) whether teachers and students are on task while classes are in session.

4.3.8 Focus Groups

For the baseline assessment, we conducted focus group discussions in 12 schools. These discussions were intended to enrich quantitative data and provide needed information for programming purposes for the CESLY technical team. Nevertheless, the data collected through focus group discussions was not as varied in the participant response, and painted a rather positive picture of the issues at hand. This is in line with our experiences in other Sub-Saharan African countries, and we suggest that the focus group discussions no longer be administered. No analysis will be presented in this report.

4.3.9 NFE Instrument

The CESLY technical team suggested that the mid-term instruments for ALP be used for the NFE baseline assessment. The first round of adjustments included removing some of the sections from the student instrument, such as social studies, science, and life skills questions. It was also agreed that no classroom or school observation would be conducted for NFE. In the end, the student instrument included reading and math sections, with some background questions about the student. No changes to teacher and principal instruments were made. The second round of adjustments took place during the assessor training for the baseline assessment. Every question was reviewed and adjusted for the new NFE context. The NFE instruments can be provided upon request.

4.4 Identification, Training, and Selection of Data Collection Assessors

4.4.1 Identification

For CESLY assessments, a total of 27 assessors were required to collect data. From the data quality point of view, the project draws on already trained assessors from both the MOE and the pool of those assessors who are not employed by the MOE but have been previously trained through other projects (e.g., EGRA Plus: Liberia).

4.4.2 Training and Selection of Assessors

Training. The training of assessors for each of the assessments took place over a period of two weeks, as follows:

- Week 1: the experts presented literature on reading and math, as well as explained reading and mathematics tools that we would be using, introduced draft instruments for CESLY, and commenced their training in administration of instruments. By the end of this week, the instruments were finalized and piloted.
- Week 2: The participants were invited to continue training to master their assessment skills. The assessors learned how to become more accurate in marking student performance and how to conduct themselves in the field, draw a sample of students at the school level, and handle and store assessment tools. During this week, the trainers selected the best assessors from the participants.

Selection. The "interrater reliability" is defined as follows: a trainer selects several tasks from the EGRA assessment instrument and develops a modified version of each task that

contains mistakes. The trainer, posing as a student, then reads these new subtasks aloud to the assessors, who are supposed to have marked the same mistakes. For instance, if the trainer made six mistakes on the letter-knowledge task, then the enumerators should have marked the same six mistakes. The candidates completed two interrater task exercises: one unannounced task, to reduce pre-test anxiety, and one announced task. The scores of both announced and unannounced tasks were added together and used to rank assessors. This exercise also allowed the trainers to pinpoint the struggles that assessors were experiencing. EDC found the assessor training to be one of the most valuable exercises. The assessor training provided a significant contribution to the sector and helped meet the project's indicators in the context of education management and information systems targets.

4.5 Field Deployment and Supervision

4.5.1 ALP Data Collection

Baseline Assessment. Data collection commenced on November 23, 2009, and ended on December 18, 2009. This allowed for four weeks of data collection in 100 schools. Ten teams, consisting of three members each, spent two days per school.

Midterm Assessment. Data collection commenced on May 17, 2010, and ended on June 11, 2010. This allowed for four weeks of data collection in 100 schools. Ten teams, consisting of three members each, spent two days per school.

Final ALP Assessment. Data collection commenced on May 20, 2011, and ended on June27, 2011. A total of six teams, each consisting of three members were deployed. Each team spent two days per school, for a total of 60 schools.

As a means of supervising the assessment activity and also to give in-the-field support to the assessors, two to three separate field trips for each assessment were organized. Most of the schools visited were receptive of assessors and provided their full cooperation. A few schools had to be changed because of worsened road conditions, and some schools had to be visited more than twice.

4.5.2 NFE Data Collection

Baseline Assessment. Data collection took place between January 27 and February 10, 2011. There were six teams, each consisting of two members. Each team assessed five sites, resulting in a total of 30 NFE sites assessed.

Final Assessment. Between May 27 and June15, 2011, a total of six teams, each consisting of two members, spent between 8 to 10 days in the field assessing a total of 30 NFE sites.

4.6 Data Entry

The first data entry in Liberia, as part of the EGRA Plus program, had been conducted in Microsoft Access or Excel, and lead to significant numbers of errors and long delays in data cleaning. RTI provided technical services to adapt the Visual Basic EGRA Data Entry System to reflect the CESLY instruments for Liberia, as well as to train data entry managers and staff. Given the complexity and length of instruments used for CESLY, it took twice as much time to develop the data entry system. For each of the assessments conducted under CESLY, the data entry program was successfully used. For each of these assessments, the MOE supervised the data entry. Data entry was performed by externally hired data clerks, who were trained by our assessment coordinators.

5. EGRA and EGMA Reliability Analysis

To verify whether EGRA would adequately estimate the reading skills of the participants in the sample, we measured its reliability using Pearson's correlations and Cronbach's alpha statistic.

Table 8 shows the Pearson correlations for the EGRA subtasks. The results indicate that fluency in letter naming correlated with reading comprehension ($r = 0.40$) and ability to read connected words ($r = 0.50$), but much less with listening comprehension, or the ability to respond correctly to questions asked orally (as opposed to written text; $r = 0.06$). Oral reading fluency (reading text aloud) was highly correlated with reading comprehension ($r = 0.69$), but not with listening comprehension ($r = 0.21$). The correlations were not very high overall, with especially low correlation ($r = 0.09$) between listening comprehension and unfamiliar word fluency. This shows that although some of the subtasks are measuring related ideas in students, others are measuring other areas, allowing for EGRA to assess a wide variety of important details of early reading competence.

Table 8: Pearson Correlations for EGRA Subtasks

Item (Subtask)	Letter Naming Fluency	Unfamiliar Word Fluency	Oral Reading Fluency	Reading Compre-hension	Listening Compre-hension	Dictation
Letter naming fluency	1					
Unfamiliar word fluency	0.24	1				
Oral reading fluency	0.50	0.41	1			
Reading comprehension	0.40	0.23	0.69	1		
Listening comprehension	0.06	0.09	0.21	0.31	1	
Dictation	0.47	0.30	0.61	0.54	0.24	1

Analysis using Cronbach's alpha scores indicated ***(see Table 9)*** an overall average of 0.56..

Table 9: Cronbach's Alpha Scores for EGRA

Item (Subtask)	Item-Test Correlation	Item-Rest Correlation	Alpha
Letter naming fluency	0.86	0.57	0.31
Unfamiliar word fluency	0.47	0.39	0.53
Oral reading fluency	0.88	0.55	0.25
Reading comprehension	0.65	0.63	0.56
Listening comprehension	0.22	0.20	0.58
Dictation	0.71	0.68	0.54
Overall test			**0.56**

6. Calibration of Mathematics Tasks

As with the EGRA instrument, the project team tested the reliability of the EGMA instrument, investigating the relationships between various assessment subtasks by using Pearson's correlation analysis ***(Table 10).*** Results of this analysis showed that number identification was weakly correlated with quantity discrimination ($r = 0.31$) and least correlated with the division task ($r = 0.07$). Quantity discrimination also appears not to have been correlated with any other subtasks. As expected, addition and subtraction tasks are highly correlated with each other, as are the division and multiplication subtasks. As ***Table 10*** shows, the EGMA subtasks were able to assess various components of mathematics skill.

Table 10: Pearson Correlations for EGMA Subtasks

Subtask	Number Identification	Quantity Discrimination	Missing Number	Addition 1	Addition 2	Subtraction 1	Subtraction 2	Multiplication	Division
Number identification (%)	1								
Quantity discrimination (%)	0.31	1							
Missing number (%)	0.10	0.17	1						
Addition task 1 (%)	0.49	0.34	0.16	1					
Addition task 2 (%)	0.46	0.35	0.14	0.59	1				
Subtraction task 1 (%)	0.37	0.27	0.11	0.59	0.48	1			
Subtraction task 2 (%)	0.35	0.24	0.08	0.44	0.51	0.56	1		
Multiplication (%)	0.12	0.40	0.24	0.24	0.21	0.21	0.14	1	
Division (%)	0.07	0.27	0.24	0.17	0.14	0.21	0.09	0.62	1

Table 11 below presents the Cronbach's alpha scores for the entire set of EGMA subtasks. The alpha scores for every subtask were over 0.70, and the entire test's alpha score was 0.79. As at baseline, shape recognition had the lowest item-rest correlation at 0.08, with the missing number subtask not far behind at 0.29. This shows that these two subtasks are measuring slightly different portions of math, with shapes focused on geometric knowledge and the missing number focused on numerical pattern recognition information.

Table 11: Cronbach's Alpha Scores for EGMA

Item (Subtask)	Item-Test Correlation	Item-Rest Correlation	Alpha
Number identification (%)	0.55	0.43	0.75
Quantity discrimination (%)	0.64	0.50	0.74
Missing number (%)	0.43	0.29	0.77
Addition task 1 (%)	0.75	0.65	0.72
Addition task 2 (%)	0.67	0.56	0.73
Subtraction task 1 (%)	0.72	0.60	0.72
Subtraction task 2 (%)	0.59	0.46	0.74
Multiplication (%)	0.57	0.45	0.74
Division (%)	0.47	0.36	0.75
Shape recognition (%)	0.37	0.08	0.80
Overall task			**0.77**

7. CESLY Program Impact

7.1 ALP Program Impact by Group (Regular or Youth) and Sex

In the final phase of data collection, the average reading achievement of ALP Regular students in Level 3 was higher than that of ALP Youth students on nearly all (4 of 6) subtasks. However, these differences were small. For example, on average, ALP Regular students correctly identified 79.92 letters per minute in comparison to 75.23 letters by their ALP Youth peers. Also, the standard deviations for the average achievement scores of each subtask were similar; that is, one group did not appear to demonstrate noticeably greater variation in scores than another. Perhaps most importantly, students in both groups continued to demonstrate inadequate reading skills in key areas such as reading connected text and comprehension.

In ***Table 12***, more details on individual tasks have been presented. Here we discuss only the two most critical skills—reading of connected text and reading comprehension. When compared to baseline, students in Regular schools read 46.95 correct words per minute. The learners in these same schools did not do as well on reading comprehension. Only 9.74% of students were able to respond to 80% and more of the questions asked. This can be explained if they are still reading slowly, which is indicated by their reading of about 47 words per minute, which is insufficient for translating what is being read into meaning. In other words, the transfer from short-term to long-term memory of what is read is not taking place, given that for this to occur, a certain level of speed and accuracy is required. They should be reading at least 60 words per minute or 1 word per second to make the transfer. The performance on the listening comprehension task is much better, where almost all of the students are answering all of the questions correctly. This reinforces the fact that adult learners bring a much richer knowledge base to the classroom than regular school age students, and because of this, adult learners can be expected to learn how to read much faster than regular school age students.

Table 12: ALP Average Reading Achievement Scores by Group Type (Regular and Youth), Final Assessment

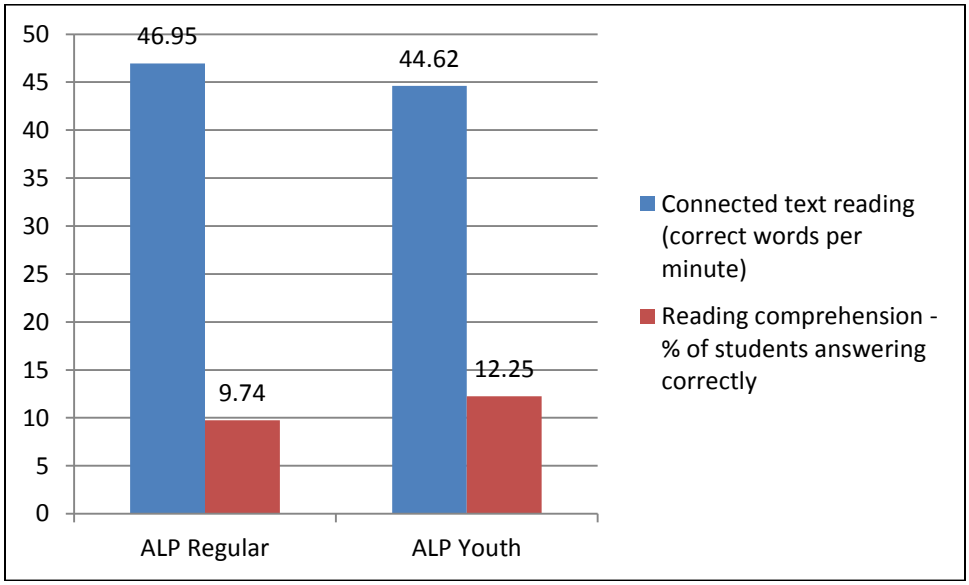
	Final Assessment				
Task	**Program**	***N***	**Average**	**SD**	**Gains**
Letter naming fluency	Regular	346.	79.92	29.61	-2.80
Letter naming fluency	Youth	248	75.23	33.81	-4.84
Invented word fluency	Regular	341	3.62	5.77	1.43
Invented word fluency	Youth	248	3.47	6.13	1.76
Connected text fluency	Regular	344	46.95	33.12	20.52
Connected text fluency	Youth	247	44.62	35.24	24.15
Reading comprehension 80% proportion*	Regular	349	9.74%	0.30	0.02
Reading comprehension 80% proportion	Youth	253	12.25%	0.33	0.07
Listening comprehension score	Regular	336	2.18	0.94	-0.17

Task	Final Assessment				
	Program	*N*	Average	SD	Gains
Listening comprehension score	Youth	231	2.26	0.95	-0.13
Dictation score	Regular	324	4.69	2.74	1.70
Dictation score	Youth	228	4.31	2.97	1.33

N=Sample Size, Average = Mean Score; SD=Standard Deviation.
* Proportion or percentage of students answering 80% or higher correctly.
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

Figure 1, below is a visual representation of performance on reading of connected text and reading comprehension by students in Regular and Youth group schools. There is no significant difference on reading and comprehension between two school group types. Given that the learners in Level 3 are supposed to be graduating the ALP program, which is equivalent to six grades of regular primary education, these scores are very low.

Figure 1: ALP Average Reading Achievement Scores, by Group Type (Regular and Youth), Final Assessment



The sex of the student appeared to continue to play a role in influencing reading achievement in the final assessment among both groups ***(see Table 13)***. For example, male Youth students correctly identified 81.62 letters in one minute in comparison to 69.04 for their female Youth counterparts. Similarly, male Regular students correctly identified 86.45 words per minute in comparison to their female counterparts, who identified just 73.08 words. Only on the listening comprehension subtask did female Youth students post higher average achievement than males, and that difference was small (2.29 versus 2.20).

The performance for female students continues a pattern of gross underachievement when compared to male students on the baseline. It is worthwhile to investigate the cause

of this huge jump further. Is it related to the implementation focus on female students' achievement in Year 2, as it was recommended by our midterm report? Is it related to the targeted girls' retention campaign, which kept girls in school and increased classroom contact hours? Is it related to the effect of girls' achievement being so low that they responded well to any intervention that increased their input? The drastic shift is worth further investigation as the program continues.

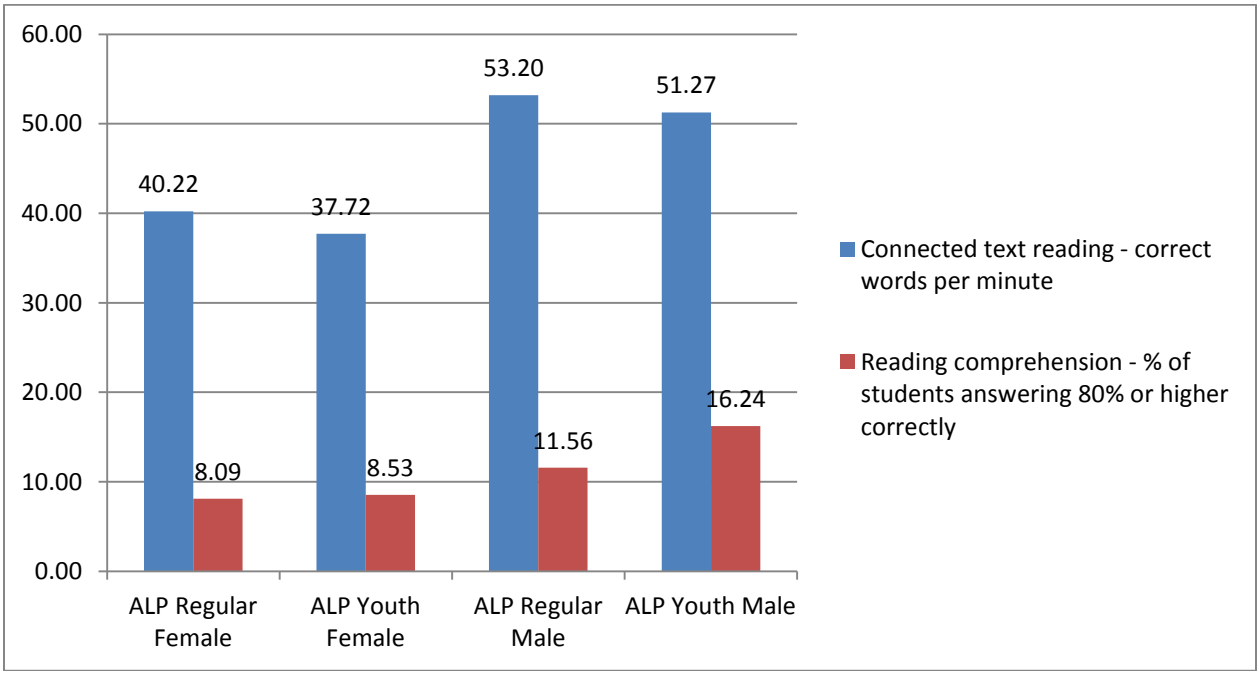
Table 13: ALP Average Reading Achievement Scores, by Group Type and Sex, Final Assessment

			Final Assessment		
Task	**Program**	**Sex**	***N***	**Avg**	**SD**
Letter naming fluency	Regular	Male	173	86.45	26.70
Letter naming fluency	Regular	Female	170	73.08	31.13
Letter naming fluency	Youth	Male	116	81.62	34.04
Letter naming fluency	Youth	Female	127	69.04	33.11
Invented word fluency	Regular	Male	170	4.52	5.87
Invented word fluency	Regular	Female	168	2.74	5.57
Invented word fluency	Youth	Male	116	4.38	6.51
Invented word fluency	Youth	Female	127	2.45	5.09
Connected text fluency	Regular	Male	172	53.20	31.53
Connected text fluency	Regular	Female	169	40.22	33.60
Connected text fluency	Youth	Male	115	51.27	33.41
Connected text fluency	Youth	Female	127	37.72	35.80
Reading comprehension 80% proportion*	Regular	Male	173	11.56%	0.32
Reading comprehension 80% proportion	Regular	Female	173	8.09%	0.27
Reading comprehension 80% proportion	Youth	Male	117	16.24%	0.37
Reading comprehension 80% proportion	Youth	Female	129	8.53%	0.28
Listening comprehension score	Regular	Male	168	2.26	0.95
Listening comprehension score	Regular	Female	165	2.12	0.93
Listening comprehension score	Youth	Male	105	2.20	1.00
Listening comprehension score	Youth	Female	122	2.29	0.91
Dictation score	Regular	Male	161	5.36	2.70
Dictation score	Regular	Female	160	4.03	2.65
Dictation score	Youth	Male	103	4.98	2.85
Dictation score	Youth	Female	120	3.71	2.89

N=sample size; Avg = mean score; SD=standard deviation.
* Proportion or percent of students answering 80% or higher correctly
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

If we present the two critical variables visually, in ***Figure 2*** below, we will see the difference between female and male performance that warrants further investigation. Also shown are that neither male nor female students, regardless of how many words they can read, are understanding very much. More emphasis on comprehension skills need to be incorporated into the next phase of the program.

Figure 2: ALP Average Reading Achievement Scores, by Group Type and Sex, Final Assessment



7.3 NFE Program Impact by Sex

NFE female students consistently posted lower average scores on each reading achievement task in comparison to their male peers in the final assessment (***See Table 14***). For females, the average final achievement score in identifying letters in one minute was 47.50, in comparison to 61.85 for males. On some tasks, such as invented word fluency, connected text fluency, and proportion reading at 80% comprehension, females posted scores that were less than half that of their male counterparts. Although scores for both males and females remained low overall, the continuing gender disparities in reading achievement found at the final assessment suggest further improvements could be made in the NFE program to narrow these gaps.

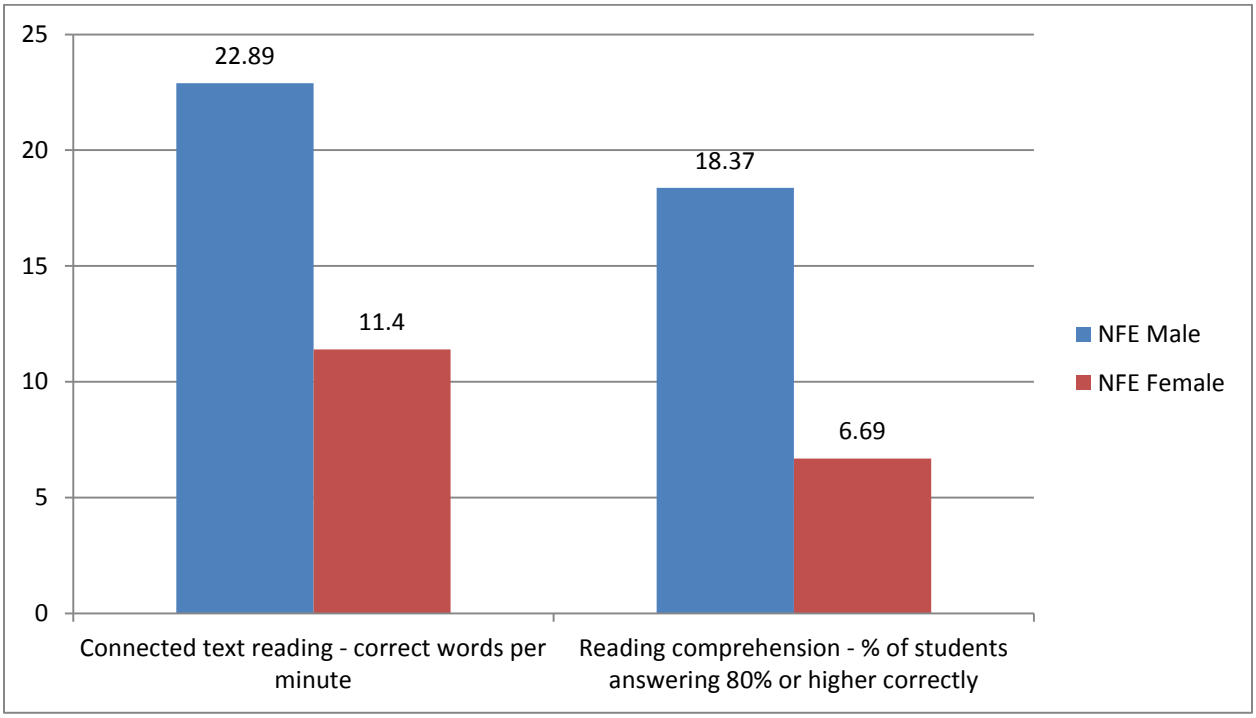
Table 14: NFE Program Impact on Reading Achievement, by Sex, Final Assessment

Task	Final assessment			
	Sex	*N*	Average	Standard Deviation
Letter naming fluency	Male	98	61.85	28.89
Letter naming fluency	Female	314	47.50	31.45
Invented word fluency	Male	98	5.95	12.49
Invented word fluency	Female	314	2.06	4.07
Connected text fluency	Male	98	22.89	26.03
Connected text fluency	Female	313	11.40	17.93
Reading comprehension 80% proportion*	Male	98	18.37 %	0.39
Reading comprehension 80% proportion	Female	314	6.69%	0.25
Listening comprehension score	Male	93	2.27	0.84
Listening comprehension score	Female	292	1.96	0.98
Dictation score	Male	82	3.37	3.37
Dictation score	Female	237	1.72	2.49

N=sample size; Average=mean score; SD=standard deviation.
* Proportion or percent of students answering 80% or higher correctly
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

In ***Figure 3*** below, we can see that, similar to ALP schools, in NFE schools male students are outperforming female students on reading and comprehension. This calls for more focus on female learners as the program continues. We noted a similar situation with ALP schools, where male students were doing better than their female counterparts. Overall, both male and female students are performing at alarmingly low levels on these two tasks. If structured and implemented well, the reading intervention should be effective in less than a year. In the Recommendations Section of this report, we provide several suggestions as to how this could be accomplished.

Figure 3: NFE Program Impact on Reading Achievement, by Sex



7.4 ALP Program Impact Comparing Baseline and Final; Baseline to Midterm to Final Assessments

The analysis of ALP reading achievement scores by learners in Level 3 from baseline to final assessment suggest that participation in the program may be related to improvements in reading achievement. The absence of the control group is preventing us to entirely attribute these changes to the ALP program. However, given that the CESLY program and its predecessor ALP program were tasked with working in ALP schools, we can assume that these results are accomplished due to those interventions.

At baseline, the average reading achievement of Regular students was higher on nearly all (5 out of 6) subtasks than that of Youth students on the baseline assessment. Youth students performed higher only on the listening comprehension subtasks and the difference in the average score was small (2.36 versus 2.38). In the final phase of data collection, the average reading achievement of Regular students continued to be higher than that of Youth students on nearly all (4 of 6) subtasks.

Despite this trend of lower average performance among Youth students from baseline to final assessment, Youth students showed larger percentage gains in four of six reading subtasks—letter naming fluency, invented word fluency, connected text fluency, and reading comprehension—than their regular counterparts. The percentage increase on these subtasks was higher for Youth than for Regular ALP students, suggesting that they made more rapid gains than their counterparts.

The detailed overview of the results and changes from baseline to the final assessment is presented in ***Table 15***.

Table 15: ALP Program Impact on Reading Achievement, Comparing Baseline and Final (Youth and Regular)

		Baseline			Final			Impact			
Task	**Program**	***N***	**Avg**	**SD**	***N***	**Avg**	**SD**	**Gains**	**Percent increase**	**Pooled SD**	**Effect size (SD)**
Letter naming fluency	Regular	376.	82.72	24.82	346	79.92	29.61	-2.80	-3%	27.18	-0.10
Letter naming fluency	Youth	330	80.07	25.82	248	75.23	33.81	-4.84	-6%	29.46	-0.16
Invented word fluency	Regular	407	2.19	6.73	341	3.62	5.77	1.43	65%	6.30	0.23
Invented word fluency	Youth	355	1.71	6.03	248	3.47	6.13	1.76	103%	6.07	0.29
Connected text fluency	Regular	361	26.43	22.82	344	46.95	33.12	20.52	78%	28.27	0.73
Connected text fluency	Youth	316	20.47	19.71	247	44.62	35.24	24.15	118%	27.57	0.88
Reading comprehension 80% proportion*	Regular	439	7.28%	0.26	349	9.74%	0.30	0.02	34%	0.28	0.09
Reading comprehension 80% proportion*	Youth	364	5.49%	0.23	253	12.25 %	0.33	0.07	123%	0.27	0.25
Listening comprehension score	Regular	424	2.36	0.82	336	2.18	0.94	-0.17	-7%	0.87	-0.20
Listening comprehension score	Youth	357	2.38	0.76	231	2.26	0.95	-0.13	-5%	0.84	-0.15
Dictation score	Regular	439	2.99	2.41	324	4.69	2.74	1.70	57%	2.55	0.67
Dictation score	Youth	364	2.98	2.27	228	4.31	2.97	1.33	45%	2.56	0.52

N=sample size, Avg =mean score; SD=standard deviation.
* proportion or percent of students answering 80% or higher correctly.
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

In ***Figure 4*** we can see that significant improvements were made between baseline and final assessment on reading connected text; however, the improvements on comprehension are lagging. The figure that follows this one reveals interesting programmatic implications.

Figure 4: ALP Program Impact on Reading Achievement, Comparing Baseline and Final (Youth and Regular)

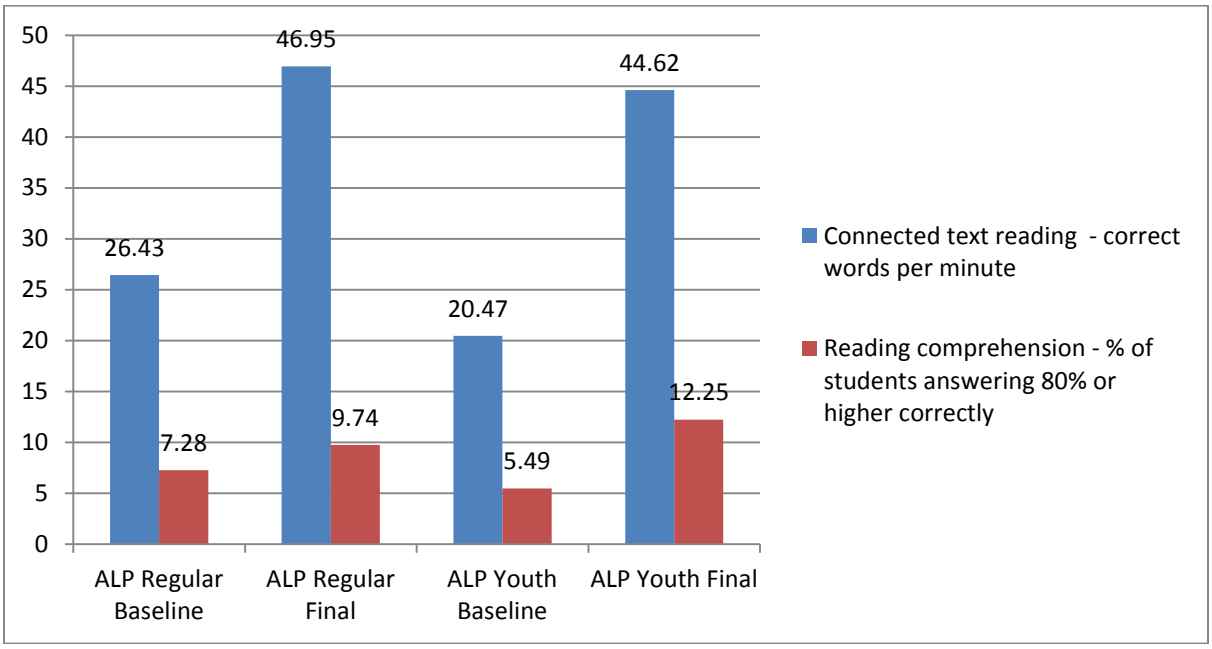
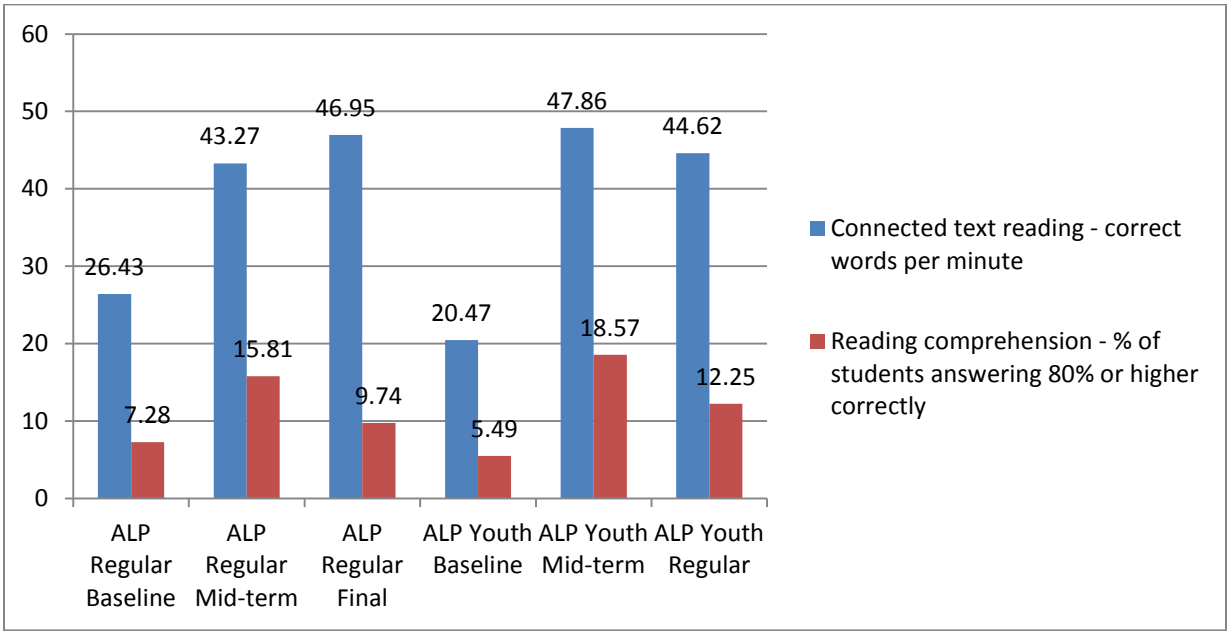


Figure 5 represents the comparisons between baseline, to midterm to final assessment for ALP Regular and Youth students. As it can be seen, significant improvements were made between baseline and midterm assessment. But if we compare the change between midterm to final assessment, one can note that there has been leveling off with respect to performance on reading connected text. It appears as if teachers have reached their maximum when it comes to accelerating reading performance. This could be tied to many different reasons, but one that is certain is best explained by the lack of improved reading comprehension. Teachers appear to be focused too much on the early reading skills, such as letter naming and decoding and blending, instead of moving from that, once mastered, into more intensive work on reading comprehension. We are not suggesting that teachers make a departure from phonics and initial reading skills, but rather that they learn how to move on from these skills once mastered and focus more on comprehension. This shift would improve the fluency and enrich vocabulary, but most importantly, would force students to read more and faster, which would lead to both improved reading and comprehension scores. Overall, the program has had significant impact when compared over baseline, but it appears as if it is reaching the ceiling and the program needs to take this into account as it continues. It seems more support to teachers needs to be provided in terms of enriching their arsenal of comprehension strategies and improved speed and accuracy. The way EGRA Plus: Liberia structured its lesson plans, with the focus on both early grade reading skills and comprehension, start from day one of teaching. It is important to do all at once at the outset, but as the time goes on and children master sounds, decoding, and blending, more time can be dedicated to comprehension strategies.

Figure 5: ALP Program Impact on Reading Achievement, Comparing Baseline to Midterm to Final (Youth and Regular)



Similar to the consistent baseline differences between ALP Youth and Regular students, female students on the baseline assessment performed at consistently lower average levels on reading subtasks than their male peers. ***See Table 16.*** Regardless of group type (Regular or Youth), males posted higher reading achievement baseline scores than females on nearly all subtasks; on only two subtasks—reading comprehension for Regular students and listening comprehension for Youth students—did females perform higher than males. Unlike the continued trend of lower average performance among Youth students at the final assessment, female students actually caught up with and surpassed male students. Males outperformed females on nearly all subtasks and across both groups types; only on the listening comprehension subtask did youth male students post lower average achievement than females, and that difference was small (2.20 versus 2.29). This finding suggests that the ALP program may have made a greater impact on male achievement than on female achievement, possibly due to the higher starting point males had at baseline.

Table 16: ALP Average Reading Achievement Scores, by Group Type and Sex, Baseline and Final Assessments

Task	Program	Sex	Baseline			Final		
			N	Avg	SD	*N*	Avg	SD
Letter naming fluency	Regular	Male	164	79.36	26.26	173	86.45	26.70
Letter naming fluency	Regular	Female	202	86.23	23.13	170	73.08	31.13
Letter naming fluency	Youth	Male	153	79.47	28.70	116	81.62	34.04
Letter naming fluency	Youth	Female	169	80.21	22.91	127	69.04	33.11
Invented word fluency	Regular	Male	186	1.39	5.80	170	4.52	5.87
Invented word fluency	Regular	Female	210	2.99	7.53	168	2.74	5.57
Invented word fluency	Youth	Male	163	0.80	2.96	116	4.38	6.51
Invented word fluency	Youth	Female	184	2.44	7.72	127	2.45	5.09
Connected text fluency	Regular	Male	158	22.22	20.13	172	53.20	31.53
Connected text fluency	Regular	Female	192	30.34	24.59	169	40.22	33.60
Connected text fluency	Youth	Male	149	19.27	20.53	115	51.27	33.41
Connected text fluency	Youth	Female	161	21.38	18.49	127	37.72	35.80
Reading comprehension 80% proportion**	Regular	Male	197	5.08%	0.22	173	11.56%	0.32
Reading comprehension 80% proportion	Regular	Female	228	9.21%	0.29	173	8.09%	0.27
Reading comprehension 80% proportion	Youth	Male	168	5.95%	0.24	117	16.24%	0.37
Reading comprehension 80% proportion	Youth	Female	188	5.32%	0.23	129	8.53%	0.28
Listening comprehension score	Regular	Male	188	2.36	0.87	168	2.26	0.95
Listening comprehension score	Regular	Female	223	2.35	0.77	165	2.12	0.93
Listening comprehension score	Youth	Male	165	2.24	0.80	105	2.20	1.00
Listening comprehension score	Youth	Female	184	2.51	0.69	122	2.29	0.91
Dictation score	Regular	Male	197	2.64	2.36	161	5.36	2.70
Dictation score	Regular	Female	228	3.28	2.42	160	4.03	2.65
Dictation score	Youth	Male	168	2.81	2.18	103	4.98	2.85
Dictation score	Youth	Female	188	3.19	2.35	120	3.71	2.89

N=sample size; Avg=mean score; SD=standard deviation.
** proportion or percent of students answering 80% or higher correctly
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

Figure 6 shows the changes between baseline and final by group type and sex for reading connected text. It is clear that males have done much better than female students on reading the connected text. The improvement made by males is rather impressive. These are not just percent increases, which in case of a lower starting base for students would mean that change is larger, given that they made a sizeable leap. Rather, the figure below provides changes in absolute terms—number of correct words per minute—and as such confirms that learning has occurred within program type and also within gender.

Figure 6: ALP Average Reading Achievement, Reading Connected Text Score, By Group and Sex, Baseline and Final

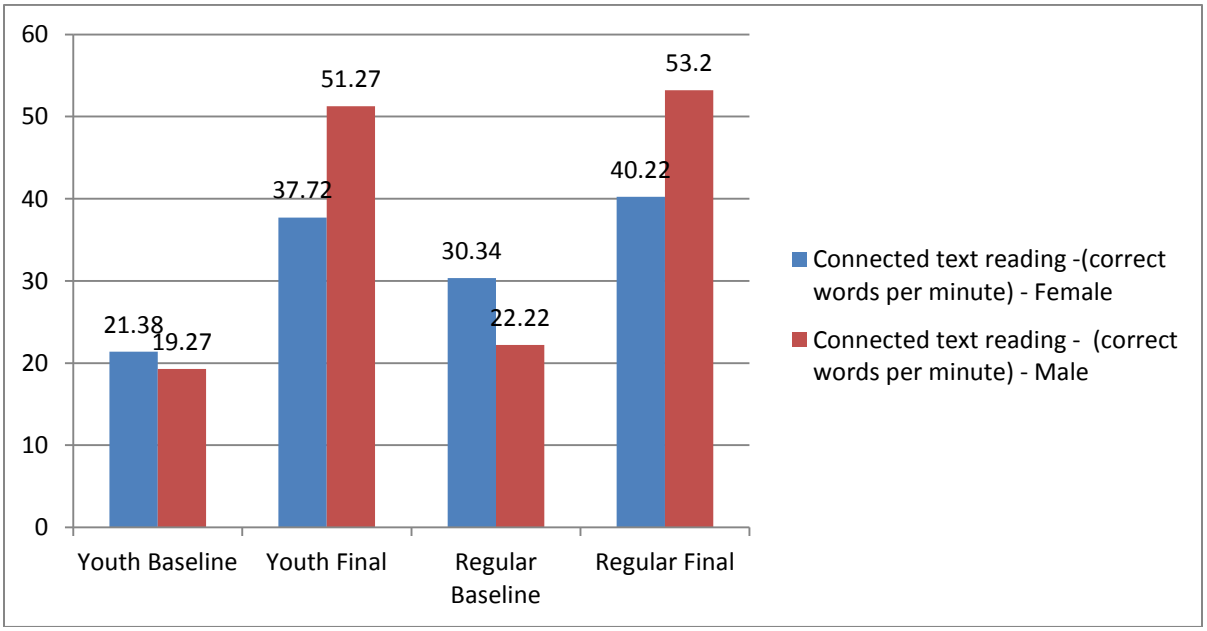


Table 17 provides information on reading comprehension scores as compared from baseline to the final assessment, and disaggregated by school type. ***Figure 7*** provides an overview of the differences in comprehension by group type and sex. The trend we discussed in the previous figure still holds—male students are outperforming female students on reading comprehension.

Table 17: ALP Average Reading Comprehension Score, by Group and Sex, Baseline and Final Assessments

Task	Youth		Regular	
	Baseline	Final	Baseline	Final
Reading comprehension – % of students answering 80% or higher correctly – Female	5.31%	8.52%	9.21%	8.09%
Reading comprehension – % of students answering 80% or higher correctly – Male	5.95%	16.23%	5.07%	11.56%

Figure 7: ALP Average Reading Comprehension Score, by Group and Sex, Baseline and Final Assessments

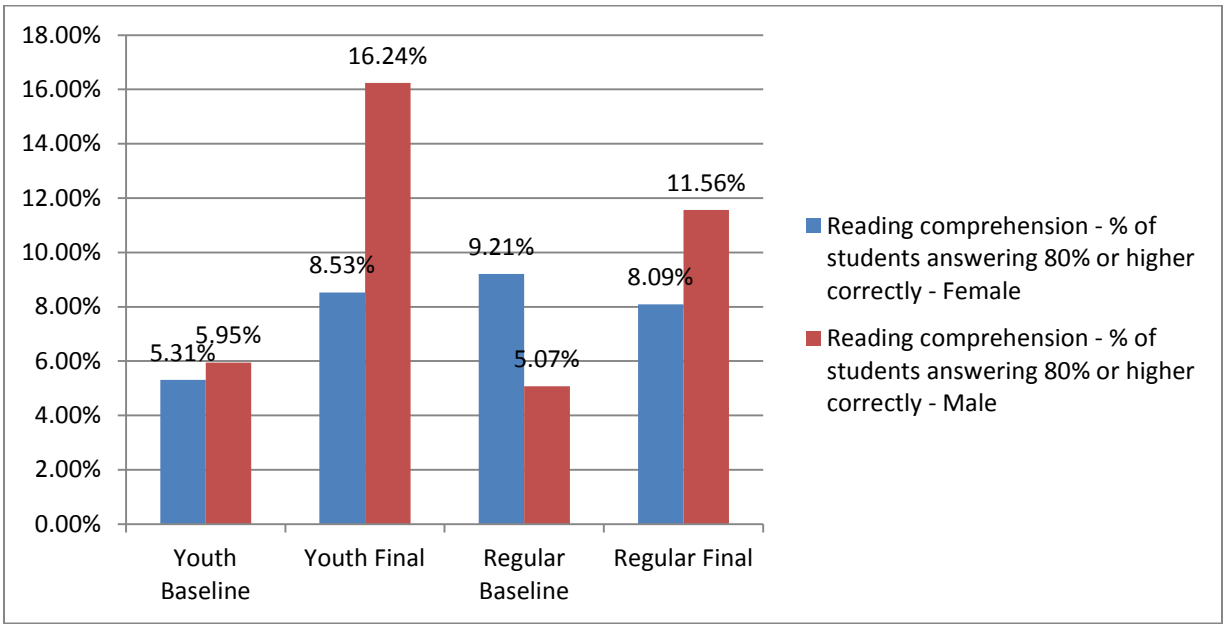


Table 18 and ***Figure 8*** reveal an important finding when we compare performance of ALP learners from baseline to midterm to final students. While improvements were made from baseline to midterm assessment, there has been a decrease of performance from midterm to the final assessment. This could mean that teachers have reached their maximum with respect to accelerating student performance at this point. Given that the ALP program has been completed, this finding would be important to keep in mind for NFE learners. Both teaching and curriculum need to focus initially on early reading skills, with appropriate level of comprehension strategies, but once children learn how to read, they need not re-learn sounds, but rather spend their time learning new comprehension strategies and just simply reading more.

Table 18: ALP Average Reading Achievement Scores, by Group and Sex: Baseline, Midterm, and Final Assessments

	Youth			Regular		
	Baseline	Midterm	Final	Baseline	Midterm	Final
Connected text reading – Female	21.38	38.77	37.72	30.34	41.10	40.22
Connected text reading – Male	19.27	60.49	51.27	22.22	44.90	53.2

Figure 8: ALP Average Reading Achievement Scores, by Group and Sex: Baseline, Midterm, and Final Assessments

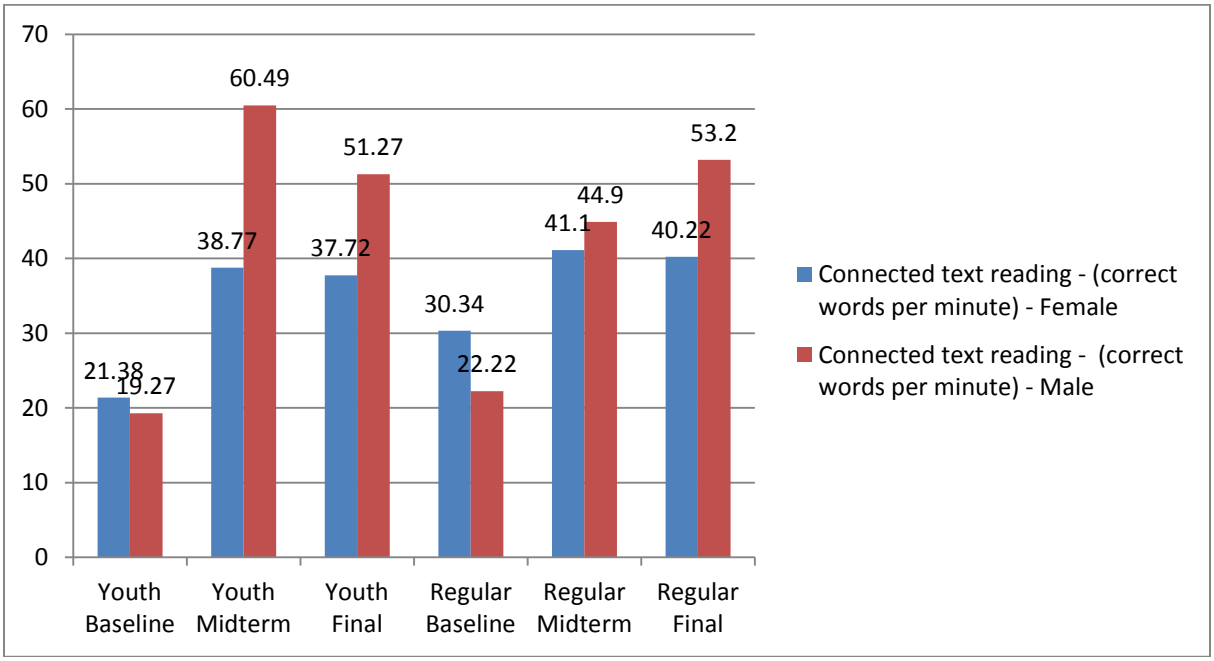
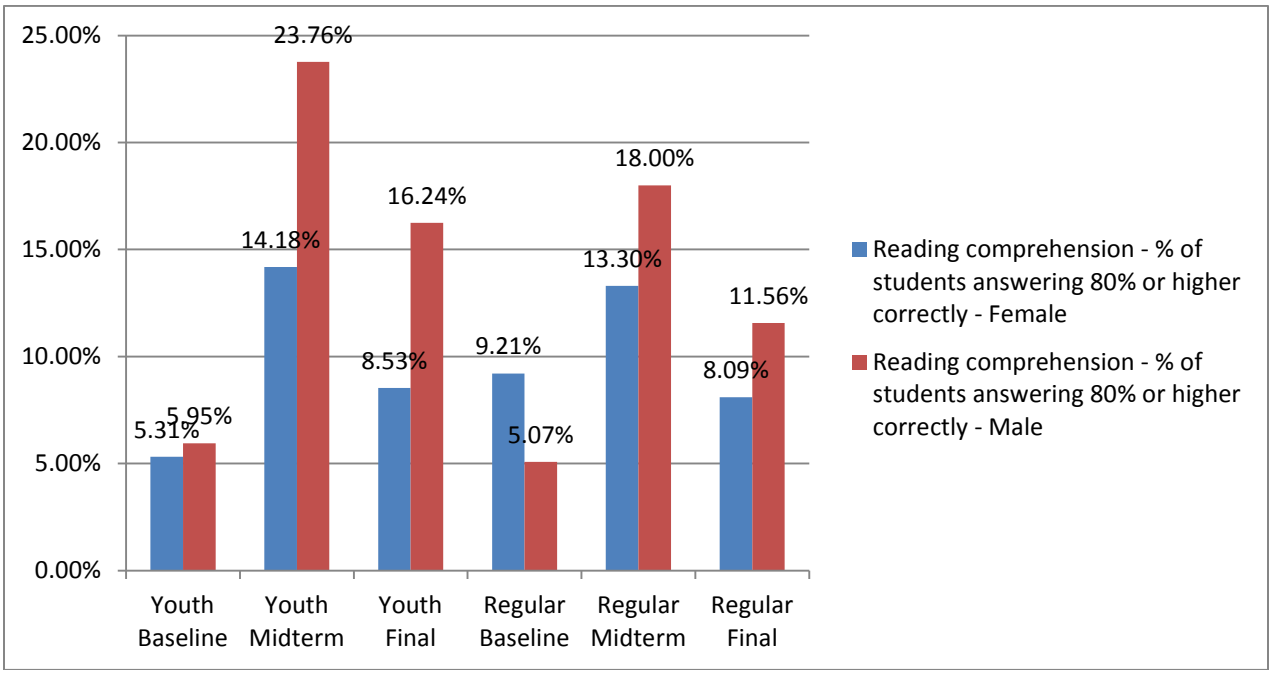


Table 19 and ***Figure 9*** present the achievement on reading comprehension. It is presented as a percentage of students who answered at least 80% of questions correctly. Similar to that of reading comprehension, students made improvements from baseline to midterm assessment, but for the final assessment, comprehension scores decreased. This can be related to the fact that both students and teachers realized that this is the end of the program, so they did not pay much attention to the comprehension scores. If we look at the changes over time, it is again clear that male students are doing better than female students.

Table 19: ALP Average Reading Comprehension Score, by Group and Sex: Baseline to Midterm to Final Assessments (% of students answering 80% or higher correctly)

Task	Youth			Regular		
	Baseline	Midterm	Final	Baseline	Midterm	Final
Reading comprehension – % of students answering 80% or higher correctly – Female	5.31%	14.18%	8.52%	9.21%	13.30%	8.09%
Reading comprehension – % of students answering 80% or higher correctly – Male	5.95%	23.76%	16.23%	5.07%	18.00%	11.56%

Figure 9: ALP Average Reading Comprehension Score by Group and Sex, Baseline to Midterm to Final Assessments (% of students answering 80% or higher correctly)



7.5 NFE Program Impact Comparing Baseline and Final

Based on the comparison of average reading achievement from baseline to final assessment, it appears that the NFE program likely contributed to improvements in reading achievement scores.[4] The average reading achievement score in all subtasks, except reading comprehension, increased from the baseline to the final sample (See ***Table 20***). For example, at baseline, the average number of letters correctly named per minute was 41.21; at final, that number had increased to 50.41 letters per minute. The largest increase occurred on the invented word fluency subtask. At baseline, the average number of invented words correctly read per minute was 0.69; at final, the average number had increased to 2.93. Although the gain was large on this subtask, the average still remained low, with fewer than 3 invented words correctly read per minute.

Table 20: NFE Program Impact on Reading Achievement, Comparing Baseline and Final

	Baseline			Final			Impact			
Task	*N*	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled (SD) Variation	Effect size (SD)
Letter naming fluency	362	41.21	32.04	432	50.41	31.31	9.21	22%	31.60	0.29
Invented word fluency	361	0.69	3.48	432	2.93	7.09	2.24	323%	5.73	0.39
Connected text fluency	361	9.62	18.36	431	14.06	20.68	4.44	46%	19.63	0.23
Reading comprehension 80% proportion*	367	4.63%	21	433	9.48%	0.29	0.04	104%	0.26	0.18
Listening comprehension score	367	1.74	1.18	400	2.03	0.95	0.29	17%	1.06	0.27
Dictation score	335	1.55	2.60	331	2.10	2.81	0.55	36%	2.70	0.20

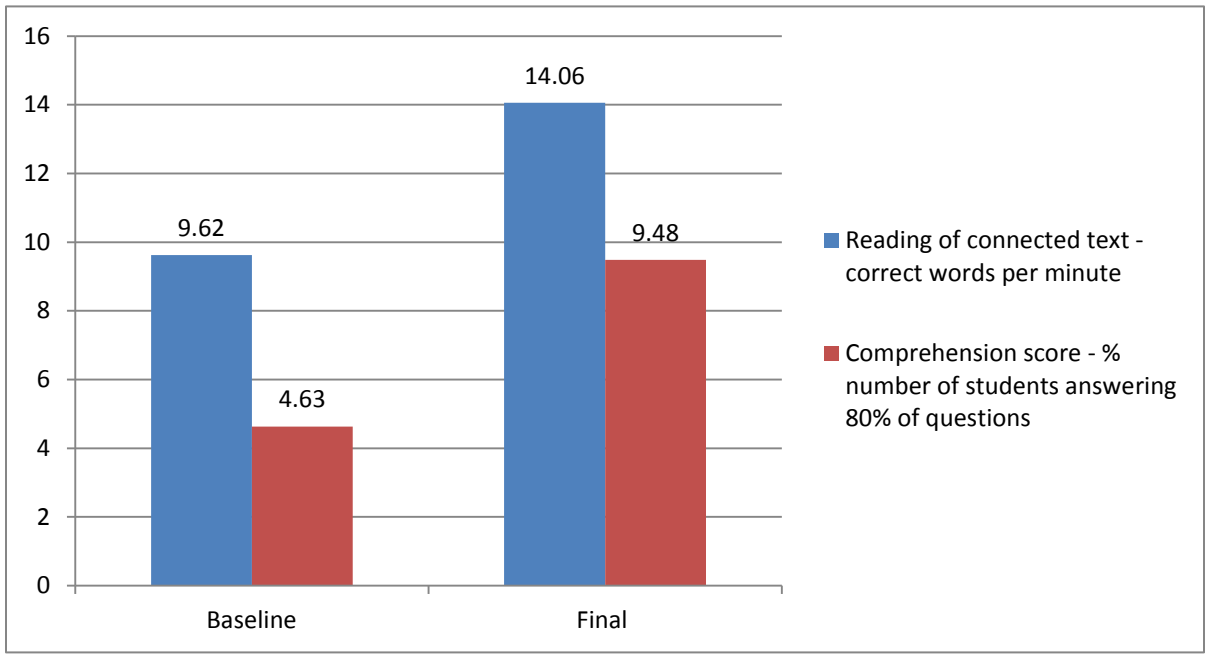
N=sample size; Avg =mean score; SD=standard deviation.
*Proportion or percent of students answering at 80% questions correctly
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

Figure 10 below represents the improvements in gains on reading of a connected text and a comprehension score. On baseline, learners read 9.62 correct words, and only 12% of them were able to answer 80% or higher of questions correctly. The reason why the average for correct words per minute is so low, is simply answered by the fact that there are many learners who could not read at all, and thus pulled the average down. Going from baseline to final assessment in three and a half months, learners have shown gains that are rather statistically significant. **On the final assessment, they scored 14.06 correct words per minute, but their comprehension increased by 104%.**

[4] A quasi-experimental analysis that analyzes the impact of the program in comparison to no intervention could confirm this supposition.

Regardless of significant improvements between baseline and final assessments, what needs to be noted is that the averages are very low and insufficient for reading with fluency. The task of the NFE program is difficult, given that these are the learners who have not had any previous education. Nevertheless, the program needs to tighten its approach and provide more intense support to teachers in the next academic year. Learning to read for adults should take place at a faster pace than that of regular students, given the wealth of knowledge that these adult learners are bringing with them to the classrooms.

Figure 10: NFE Program Impact on Reading Achievement, Comparing Baseline and Final—Reading and Comprehension



Similar to ALP students where male students consistently outperformed female counterparts from baseline, to midterm, to final assessment, female NFE students on the baseline assessment performed at consistently lower average levels on reading subtasks than their male peers. Males posted higher reading achievement baseline scores than females on all subtasks. We had hoped that Among the ALP & NFE intervention participants, female students would catch up with male students by the final assessment. This finding did not bear out for either intervention. We strongly recommend testing females after nine or more months of reading intervention to test whether this trend is maintained across the programs. At the final assessment, female students continued to perform at consistently lower average levels than their male peers. ***See Table 21.***

Table 21: NFE Program Impact on Reading, by Sex, Baseline and Final

Task	Sex	Baseline			Final			Impact					
		N	Avg.	SD	*N*	Avg.	SD	Gains	Percent increase	Pooled SD	Effect size (SD)	Female gains (SD)	Female effect size (SD)
Letter naming fluency	M	85	51.95	31.99	98	61.85	28.89	9.90	19%	30.20	0.33		
Letter naming fluency	F	265	37.70	31.68	314	47.50	31.45	9.79	26%	31.50	0.31	-0.11	-0.02
Invented word fluency	M	85	1.82	4.62	98	5.95	12.49	4.13	226%	9.62	0.43		
Invented word fluency	F	264	0.34	3.03	314	2.06	4.07	1.72	505%	3.62	0.47	-2.41	0.05
Connected text fluency	M	85	17.96	23.63	98	22.89	26.03	4.92	27%	24.81	0.20		
Connected text fluency	F	264	6.90	15.64	313	11.40	17.93	4.50	65%	16.89	0.27	-0.42	0.07
Reading comprehension 80% proportion*	M	85	9.41%	0.29	98	18.37%	0.39	0.09	95.15%	0.35	0.26		
Reading comprehension 80% proportion*	F	270	3.33%	0.18	314	6.69%	0.25	0.03	100.64%	0.22	0.15	-0.06	-0.11
Listening comprehension score	M	85	1.91	1.04	93	2.27	0.84	0.36	19%	0.93	0.39		
Listening comprehension score	F	270	1.67	1.21	292	1.96	0.98	0.29	17%	1.10	0.26	-0.07	-0.12
Dictation score	M	80	2.79	3.17	82	3.37	3.37	0.58	21%	3.25	0.18		
Dictation score	F	243	1.17	2.27	237	1.72	2.49	0.55	47%	2.38	0.23	-0.03	0.05

N=sample size; Avg =mean score; SD=standard deviation.
*Proportion or percent of students answering at 80% questions correctly
**Percent increase has been calculated as follows: "Percent Increase= (FinalMean-BaselineMean)/BaselineMean."

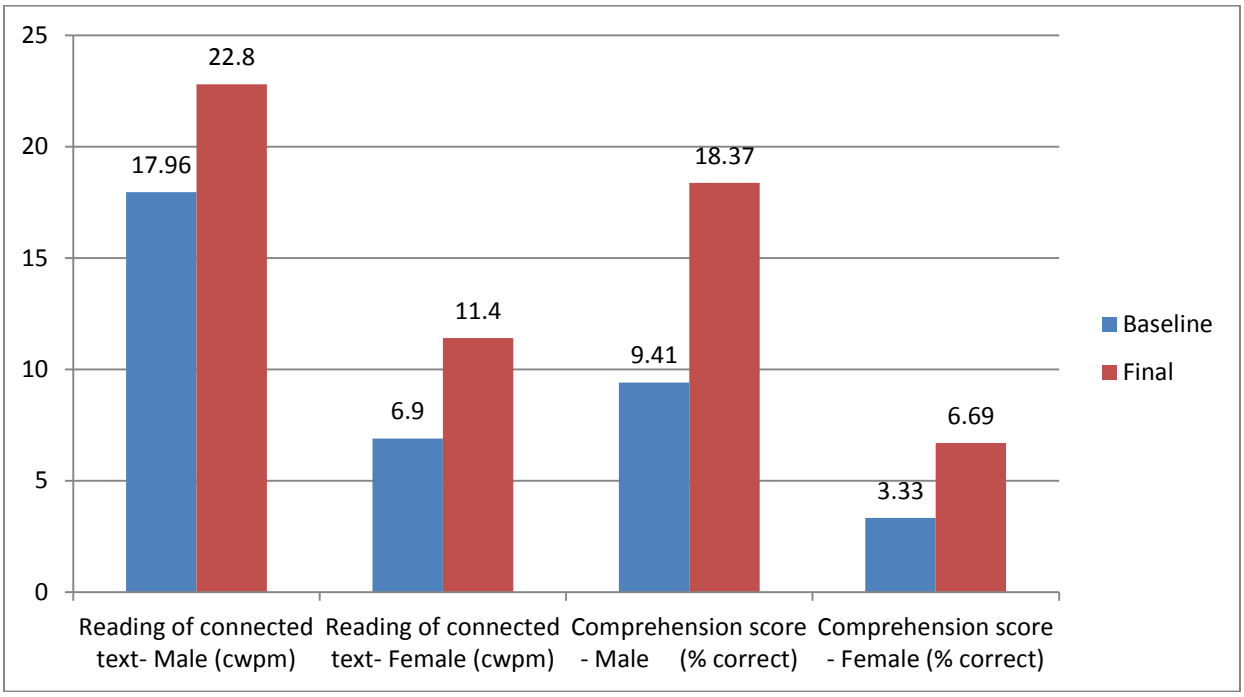
Table 22 provides an overview of changes between initial and final assessment on reading and comprehension by females and males. ***Figure 11*** that follows represents this same data visually. Overall, both reading and comprehension scores are very low and we have noted decreases between baseline and final assessment.

Table 22: NFE Program Impact on Reading Achievement, Comparing Baseline and Final, by Sex—Reading and Comprehension

Task/NFE	Baseline	Final
Reading of connected text – male	17.96	22.8
Reading of connected text – female	6.9	11.4
Comprehension score – male*	9.41%	18.37%
Comprehension score – female*	3.33%	6.69%

*proportion of students answering correctly at least 80% of questions

Figure 11: NFE Program Impact on Reading Achievement, Comparing Baseline and Final, by Sex—Reading and Comprehension



*cwpm = correct words per minute.

8. Liberia Comparisons and Benchmarks

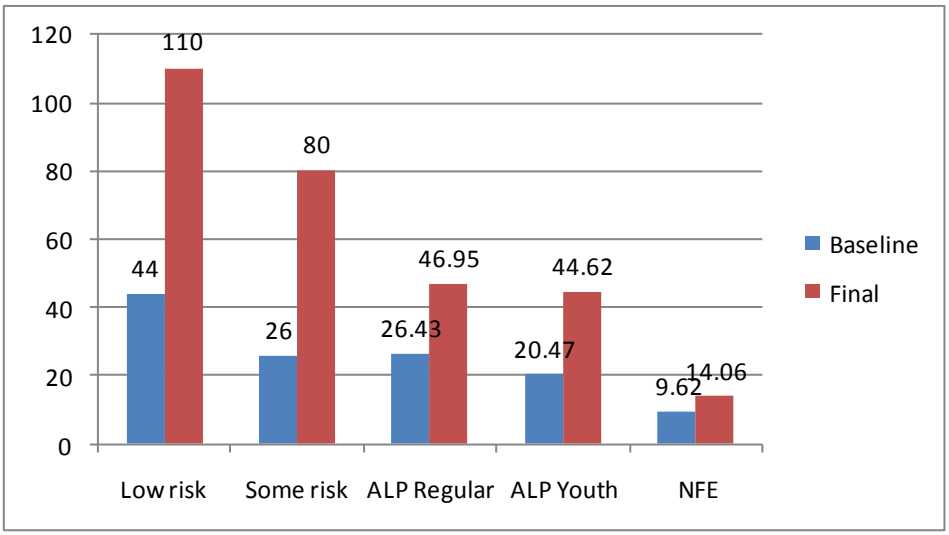
8.1 Comparisons with International Benchmarks: DIBELS

The findings in this report suggest that the ALP and NFE interventions had a positive impact on basic reading and math skills. This section compares some aspects of oral reading fluency in ALP and NFE schools and the DIBELS benchmarks for oral reading fluency.[5] It is very important to note here that these comparisons are illustrative, and that comparing two different populations on measures such as these generally is not at all recommended. Nevertheless, in the case of reading, which is a skill, teachers and programs should strive to have students reach and then pass the goal of reading 60 words per minute correctly, and achieve 100-percent accuracy on comprehension questions. While some may learn reading skills faster than others, for a variety of reasons, there is a scientifically proven approach for learning to read, regardless of which student population is of interest. Thus, here we contrast some of the observed performance results to benchmarks for the United States and a few of other countries.

In ***Figure 12,*** the blue bar shows the DIBELS "some risk" benchmark for oral reading fluency, while the red bar shows the DIBELS "low risk" benchmark. The comparisons for ALP and NFE schools are from the baseline and final assessments for ALP (Regular and Youth) and NFE intervention schools. At baseline, students in ALP schools were closer to the low-risk benchmarks from DIBELS than were students from NFE schools. However, at the final assessment, neither the ALP nor the NFE students approached the oral reading fluency scores of either the some risk or low risk DIBELS measures. This suggests that, although students in both groups made some progress, the average oral reading fluency scores remained much lower than the international benchmarks.

[5] DIBELS stands for Dynamic Indicators of Basic Early Literacy Skills. It is the assessment format upon which much of EGRA is based. DIBELS comparisons are useful, because while DIBELS is specific to the United States context, it has well-developed benchmarks for oral reading fluency scores for students who are deemed to be either at some risk or low risk of experiencing reading difficulties. More information can be found at https://dibels.uoregon.edu/.

Figure 12: Oral Reading Fluency Scores Compared to International Benchmarks

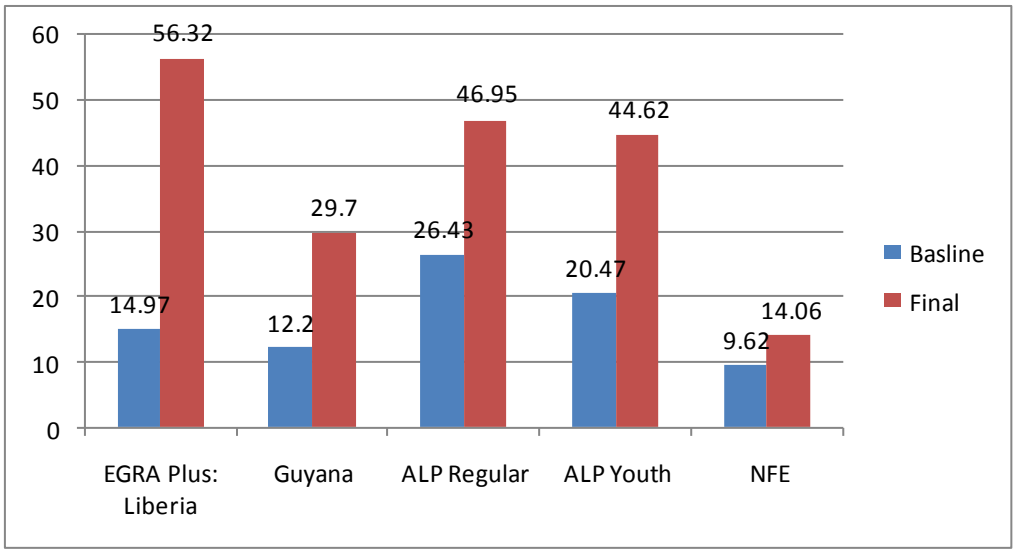


8.2 Comparisons with Guyana

While the discussion above is noteworthy, and there is some value in comparing Liberian students' results to what is found in the United States DIBELS benchmarks, it is more appropriate and valuable to compare Liberia's scores to the oral reading fluency scores from other in-country samples, namely the USAID EGRA Plus: Liberia intervention; and to those from other low-income countries, in this case, Guyana. Note, however, that even this type of comparison is fraught with problems given the language differences and the local adaptation of EGRA in each country. Even in countries where English is assessed, the assessments can be quite different since each EGRA oral reading fluency story is locally created. That said, it is still worth taking a look at the comparisons between students in different countries.

The comparison in ***Figure 13*** shows that the ALP baseline scores were very similar to or even higher than the baseline scores found in EGRA Plus (Liberia) and Guyana. In contrast, the scores among NFE students were lower at baseline than all other comparison groups. However, at the final assessment, the EGRA Plus scores were noticeably higher than those of the other comparison groups. The ALP groups, both Regular and Youth, approached the EGRA Plus scores, but the NFE scores showed little comparative progress. These students performed below EGRA Plus: Liberia and Guyana scores. However, the NFE group represents a population of learners who are largely illiterate or neoliterate adults. EGRA has as yet not been administered to other adult populations, so little comparative data exist.

Figure 13: Oral Reading Fluency Scores in Liberia Compared to Other Interventions



Regardless of this, the findings of ALP and NFE assessments show some improvements, but these improvements are not sufficient and the program needs to refocus its intervention so that the learners learn how to read in less than a year. A similar experience was noted during EGRA Plus: Liberia and several other projects, where midterm assessments showed improvements (in case of EGRA Plus: Liberia, they were significant), but they were still not at the desired level. It was only in the second year in which the true impact was achieved. This was because in the first year of the project, both teachers and program staff were learning the new curriculum, its delivery, and support. As time progressed, the scores improved. However, the critical step must be to focus the intervention so that the goal is for learners to achieve at least the goal of reading accurately 60 words per minute (and more is desired), and answering correctly all questions asked about a given story.

9. EGRA Impact Analysis

As the preceding sections have shown, *descriptive* analyses of achievement data, collected to examine the influence of participation in ALP and NFE interventions on student achievement, showed positive trends. That is, the descriptive analyses suggested that participation in the interventions positively impacted student achievement. In this section, we present the results of regression models to determine whether, in fact, the program had an impact on student achievement. Regression models allow for testing to estimate whether or not an individual predictor (sex or grade, for example) has a statistically significant impact on a particular outcome, such as oral reading fluency or letter identification.

9.1 Results of Regression Analysis: ALP

Overall, the results of regression analyses of ALP data found that school type was not a significant predictor of reading outcomes. Alternatively, student sex *was* a significant predictor of reading outcomes. Despite the progress of female students from the baseline to final assessment, female students performed at lower levels on the EGRA subtasks for letter identification, invented word reading, oral reading fluency, and reading comprehension. The descriptive analyses of the final assessment showed that the average reading achievement of the Regular group was higher on nearly all subtasks than that of the Youth group. This finding was also consistent with the findings of the baseline assessment, on which students in the Regular group also tended to perform higher than the Youth students, on average. Despite these trends, the regression analysis examining the relationship between school type (Regular vs. Youth) and multiple reading subtask outcomes did not find statistical significance. Thus, participation in either school type did not predict a statistically significant difference in reading outcome score.

Similar to the consistent baseline differences between ALP Youth and Regular students, female students on the baseline assessment performed at consistently lower average levels on reading subtasks than their male peers. The results of the regression analysis found that there was a relationship between sex and oral reading fluency. The coefficient for the relationship between being female and performance on oral reading fluency was -13.15542; that is, for each unit increase in sex (indicating that a student was female), there was a 13.15542 words-per-minute decrease in the oral reading fluency score. The *p*-value for this relationship was 0.000, indicating that the relationship was statistically significant. A similar relationship was found between sex and reading comprehension. The coefficient for the relationship between being female and performance on reading comprehension was -1.843431; that is, for each unit increase in sex (indicating that a student was female), there was a corresponding 1.843431 words-per-minute decrease in reading comprehension. The *p*-value for this relationship was also 0.000, indicating that the relationship was statistically significant.

9.2 Results of Regression Analysis: NFE

For students participating in the NFE program, regression analyses were conducted by two key variables: age[6] and sex. Overall, the results of regression analyses of NFE data found that student age did not significantly impact reading achievement scores. However, similar to the analyses of ALP data, sex did influence reading achievement scores. Female NFE students performed lower on average than male students and this difference was statistically significant for two measures of reading.

The average reading achievement score on all subtasks, except for reading comprehension, increased from the baseline to the final sample. For example, at baseline, the average number of letters correctly named per minute was 41.21; at final, that number had increased to 50.41 letters per minute. The largest increase occurred on the invented-word fluency subtask. At baseline, the average number of invented words correctly read per minute was 0.69; at final, the average number had increased to 2.93. Although the gain was large on this subtask, the average still remained low, with fewer than three invented words correctly read per minute.

There was a relationship between sex and oral reading fluency. The coefficient for the relationship between being female and performance on oral reading fluency was -1.4875; that is, for each unit increase in sex (indicating that a student was female), there was a corresponding 11.4875 words-per-minute decrease in oral reading fluency. The *p*-value for this relationship was 0.000, indicating that the relationship was statistically significant.

A similar relationship was found between sex and reading comprehension. The coefficient for the relationship between being female and performance on reading comprehension was -0.1167945; that is, for each unit increase in sex (indicating that a student was female), there was a corresponding 0.1167945 words-per-minute decrease in reading comprehension. The *p*-value for this relationship was 0.001, indicating that the relationship was statistically significant.

9.3 Zero Scores

Table 23 presents the percentage of ALP students who scored zero on the various reading subtasks in the final assessment. On most EGRA subtasks, the percentage of NFE students who scored zero was *lower* at the final assessment in comparison to the baseline assessment. The most impressive accomplishment was on the invented-word fluency subtask where the percentage of students scoring zero decreased from 83.51% at baseline to 53.31% at final assessment. With the exception of letter naming and listening comprehension, the number of zero scores decreased from the baseline to the final assessment.

[6] Unlike ALP data, which were analyzed by program type based on age groupings (regular vs. youth), NFE data were analyzed on a continuum of age, as reported by participants.

Table 23: Percentages of ALP Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline, Midterm, and Final

Subtask	Baseline	Midterm	Final	T-test
Letter naming fluency (zero letters per minute)	0.84%	1.03%	1.01%	Not significant
Unfamiliar word fluency (zero words per minute)	83.51%	70.97%	53.31%	t=12.7839; p=0.0000
Oral reading fluency (zero words per minute)	18.54%	8.63%	12.84%	t=2.7853; p=0.0054
Reading comprehension (zero correct)	25.53%	10.37%	18.05%	t=2.8860; p=0.004
Listening comprehension (zero correct)	3.04%	2.49%	7.94%	t=-4.0671; p=0.0001
Dictation (zero correct)	20.72%	16.71%	12.14%	t=4.1380; p=0.0000

Figure 14 indicates that the number of learners who could not identify any letters was quite low at all three assessment periods. However, except for listening comprehension, we can note the downward trend in terms of the number of students who could not perform on any of the EGRA subtasks. This means that the ALP program was successful in moving learners from low levels of literacy onto the path to literacy.

Figure 14: Percentages of ALP Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline, Midterm, and Final Assessments

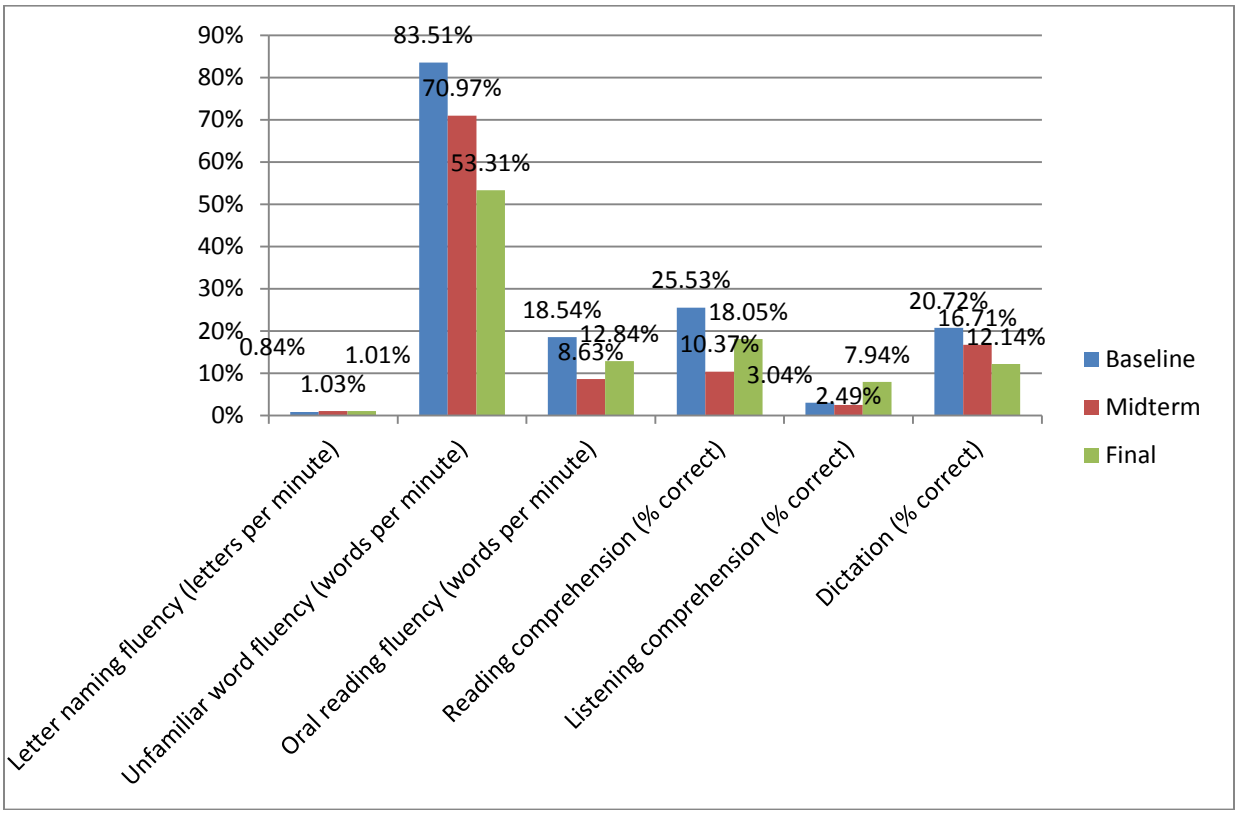


Table 24 presents the percentages of NFE students who scored zero on the various reading subtasks in the final assessment, across each treatment group. As it can be seen, NFE program has made impressive accomplishment by reducing the number of students who could not respond to any of the task. All of the results are significant with huge improvements from the baseline to the final assessment. For example, on the reading comprehension task, slightly more than 73% of students could not answer any questions at the baseline assessment. This number decreased to 17% on the final assessment.

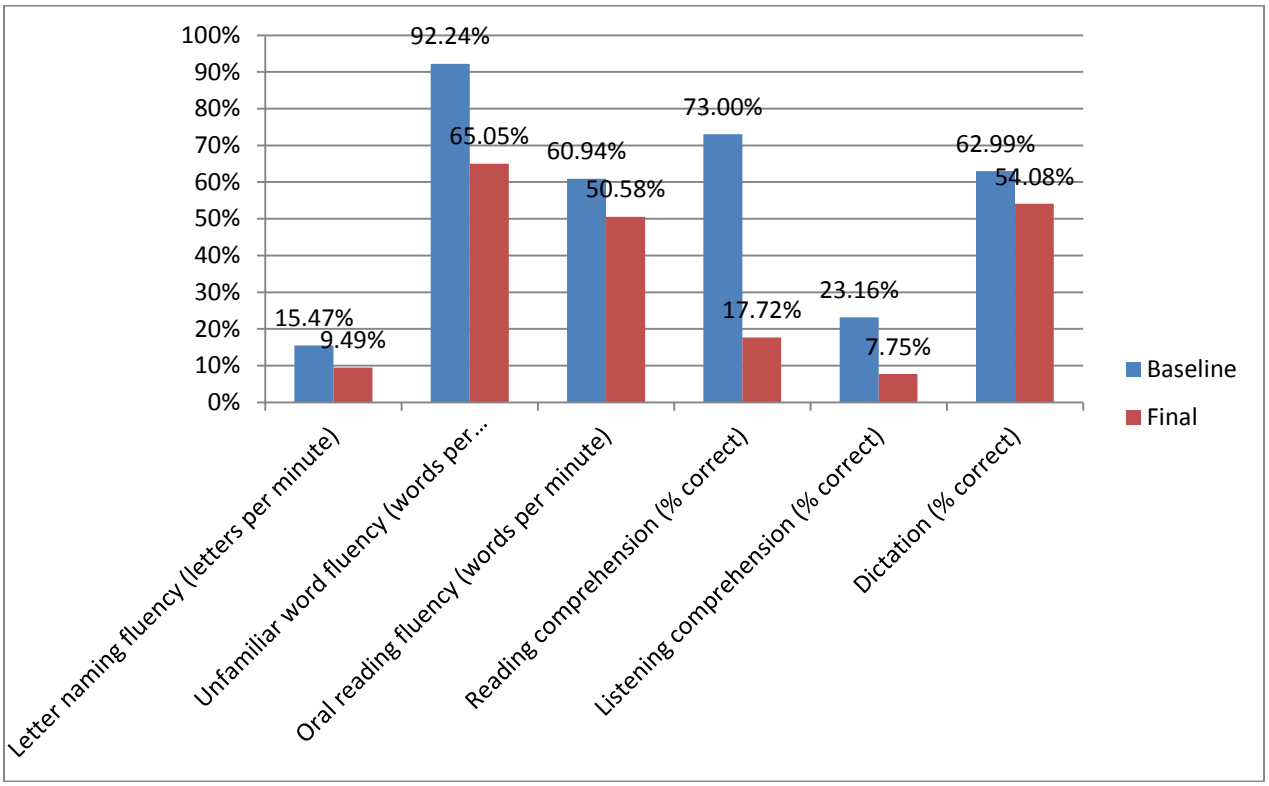
Table 24: Percentages of NFE Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline and Final Assessment Scores

Subtask	Baseline	Final	T-test
Letter naming fluency (letters per minute)	15.47%	9.49%	t=2.5697; p=0.0104
Unfamiliar word fluency (words per minute)	92.24%	65.05%	t=9.6312; p=0.0000
Oral reading fluency (words per minute)	60.94%	50.58%	t=2.9333; p=0.0035
Reading comprehension (% correct)	73.41%	17.72%	t=13,7265; p=0.000
Listening comprehension (% correct)	23.16%	7.75%	t=6.0850; p=0.0000

Subtask	Baseline	Final	T-test
Dictation (% correct)	62.99%	54.08%	t=2.3390; p=0.0196

Figure 15 below represents these data visually; we can see a positive trend in terms of lowering the number of students who could not read at all at the time of initial assessment. This trend needs to be continued until zero students are scoring zero on EGRA subtasks.

Figure 15: Percentages of NFE Learners with Zero Scores on EGRA Subtasks: Comparison of Baseline and Final Assessment Scores



10. EGMA Impact Analysis

10.1 General Findings, Descriptive Analysis: ALP

For this impact evaluation report, we report on average mathematics scores by learners on individual subtasks as well as presenting the proportion of students who were able to answer all of the subtasks correctly. This approach allows us to compare changes across assessments but also to present important information in terms of the levels at which students were performing. ***Table 25*** shows what the maximum score would be on mathematics subtasks as assessed by EGMA. When discussed in the forthcoming text, the average student performance is to be contrasted to the number of total possible answers. For example, if the average performance on the number identification for ALP students on the final assessment was 14.64, this would mean that ALP students on average had 14.64 out of 20 possible answers correct. The same approach to interpreting results is applied to all other subtasks.

Table 25: Interpreting Mathematics Average Scores as Assessed by EGMA: ALP Program

Subtask	Number of total possible correct answers
Number identification	20
Quantity discrimination	20
Missing number	10
Addition score (level 1)	10
Addition score (level 2)	9
Subtraction score (level 1)	10
Subtraction score (level 2)	9
Multiplication score	6
Division score	6
Shape identification score	4

Overall, descriptive analyses of EGMA data among ALP participants suggested that participation in the intervention might have positively influenced average reading performance in some mathematical content knowledge. The average math achievement score on most subtasks increased from the baseline to the final sample (see ***Table 26***). For example, at baseline, the average number identification score was 16.78 items identified correctly; at final, that number had decreased to 16.77. with the exception of quantitative comprehension fluency, the average mathematics achievement scores for the sample decreased from baseline to final on all tasks.

Table 26: ALP Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores

		Baseline			Midterm			Final			
Task	Program	N	Mean	SD	N	Mean	SD	N	Mean	SD	Percent Increase
Number identification score	Regular	421	16.78	4.20	488	17.78	2.97	298	16.77	3.76	-0.06%
Number identification score	Youth	351	17.32	3.19	219	17.67	3.32	211	16.25	4.57	-6.18%
Quantitative comparison fluency	Regular	275	0.43	0.92	525	10.79	6.61	344	8.35	6.04	1824.96%
Quantitative comparison fluency	Youth	232	0.44	0.89	248	11.45	6.94	246	7.57	6.54	1609.86%
Quantitative comparison score	Regular	306	0.39	0.73	526	10.46	6.11	344	8.12	5.69	1970.25%
Quantitative comparison score	Youth	256	0.43	0.74	248	10.93	6.11	246	7.41	6.23	1640.86%
Missing digit score	Regular	425	3.72	2.55	524	2.68	3.25	344	1.45	2.33	-61.13%
Missing digit score	Youth	350	3.15	2.27	248	2.63	3.42	246	1.39	2.38	-55.89%
Addition score (level 1)	Regular	424	7.91	2.53	521	6.71	2.74	345	6.57	2.58	-16.89%
Addition score (level 1)	Youth	350	7.60	2.61	245	7.20	2.61	246	6.08	2.76	-20.00%
Addition score (level 2)	Regular	417	6.53	2.48	520	5.01	2.67	315	5.15	2.14	-21.11%
Addition score (level 2)	Youth	345	6.05	2.59	245	5.35	2.74	210	5.10	2.11	-15.69%
Subtraction score (level 1)	Regular	420	6.15	3.09	522	5.19	2.85	344	4.88	2.75	-20.64%
Subtraction score (level 1)	Youth	347	6.17	2.98	247	6.05	2.88	245	4.85	2.79	-21.32%
Subtraction score (level 2)	Regular	417	4.62	2.59	518	3.51	2.92	274	4.19	2.24	-9.25%
Subtraction score (level 2)	Youth	348	4.34	2.49	244	4.16	3.09	186	4.35	2.29	0.24%
Multiplication score	Regular	413	2.23	1.51	503	1.43	1.78	339	0.86	1.41	-61.42%
Multiplication score	Youth	346	2.11	1.32	241	1.32	1.89	243	0.74	1.27	-65.04%
Division score	Regular	404	1.34	1.48	497	1.00	1.57	341	0.43	1.08	-67.71%
Division score	Youth	335	1.47	1.42	236	0.94	1.70	242	0.48	1.15	-67.36%
Shape identification score	Regular	423	1.05	1.02	419	1.53	1.00	210	1.11	1.68	6.16%
Shape identification score	Youth	345	1.00	0.92	216	1.55	1.05	172	0.83	1.54	-16.62%

Figure 16 represents average performance from ALP baseline assessment to final assessment on the mathematics subtasks measured by EGMA. As described above, for the most part, some improvements occurred, with the exception of a few subtasks. For mathematics, however—unlike in reading—it is not possible to create one composite score.

Figure 16: ALP Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Score, by Subtask

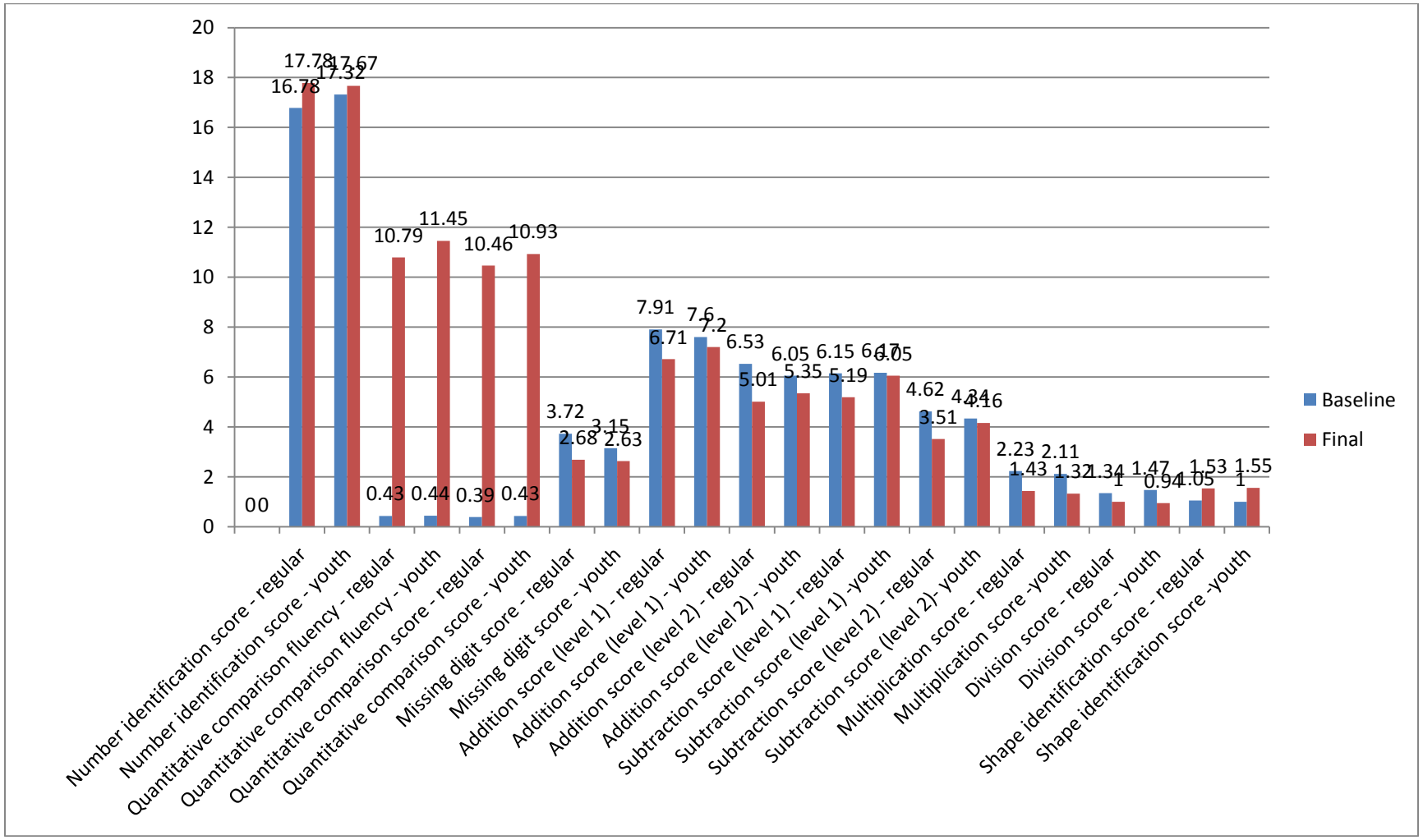


Table 27 is an overview of the proportion of Regular students who were able to answer all of the questions correctly. For example, on the number identification subtask, at baseline, 84% of students were able to answer all of the questions correctly. The analysis of baseline mathematics results for ALP students by group (Youth vs. Regular) shows that any differences between average mathematics achievement scores by Youth vs. Regular group were few and small. For example, the aforementioned 84% baseline average for Regular students in number identification compared to 87% for the Youth group.

At the midterm analysis, the average scores for both Regular and Youth groups across most subtasks had increased. On some subtasks, the increase was small: Average scores on number identification for Regular students changed from 84% to 89%, and for Youth students from 87% to 88%. On other subtasks, the increase was more noticeable: Average quantity discrimination scores for Regular students rose from 2% to 52%. For Youth students, the same scores rose slightly more, from 2% to 55%. Similar to the baseline, the differences in average mathematics subtask scores between groups were few and small.

At the final assessment, few percentage increases in average mathematics subtask scores were shown. In fact, most subtasks showed a decrease in the average mathematics score. These trends were similar across the Regular and Youth students.

Table 27: ALP Program Impact on Mathematics, Comparing Baseline and Final: Proportion of Students Answering Mathematics Subtasks Correctly

		Baseline			Midterm			Final			Impact			
Subtask	Program	*N*	Avg	SD	*N*	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled SD	Effect size (SD)
Number identification percent	Regular	421	84%	0.21	488	89%	0.15	298	84%	0.19	0.00	0%	0.20	0.00
Number identification percent	Youth	351	87%	0.16	219	88%	0.17	211	81%	0.23	-0.05	-6%	0.19	-0.28
Quantity discrimination comparison percent	Regular	306	2%	0.04	526	52%	0.31	344	41%	0.28	0.39	197%	0.21	1.86
Quantity discrimination comparison percent	Youth	256	2%	0.04	248	55%	0.31	246	37%	0.31	0.35	164%	0.22	1.59
Missing digit percent	Regular	425	37%	0.26	524	27%	0.32	344	14%	0.23	-0.23	-61%	0.25	-0.93
Missing digit percent	Youth	350	32%	0.23	248	26%	0.34	246	14%	0.24	-0.18	-56%	0.23	-0.76
Addition score (level 1) percent	Regular	424	79%	0.25	521	67%	0.27	345	66%	0.26	-0.13	-17%	0.25	-0.52
Addition score (level 1) percent	Youth	350	76%	0.26	245	72%	0.26	246	61%	0.28	-0.15	-20%	0.27	-0.57
Addition score (level 2) percent	Regular	417	73%	0.28	520	56%	0.30	315	57%	0.24	-0.15	-21%	0.26	-0.59
Addition score (level 2) percent	Youth	345	67%	0.29	245	59%	0.30	210	57%	0.23	-0.11	-16%	0.27	-0.39
Subtraction score (level 1) percent	Regular	420	62%	0.31	522	52%	0.28	344	49%	0.28	-0.13	-21%	0.29	-0.43
Subtraction score (level 1) percent	Youth	347	62%	0.30	247	61%	0.29	245	49%	0.28	-0.13	-21%	0.29	-0.45
Subtraction score (level 2) percent	Regular	417	51%	0.29	518	39%	0.32	274	47%	0.25	-0.05	-09%	0.27	-0.17

Subtask	Program	Baseline			Midterm			Final			Impact			
		N	Avg	SD	*N*	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled SD	Effect size (SD)
Subtraction score (level 2) percent	Youth	348	48%	0.28	244	46%	0.34	186	48%	0.25	0.00	0%	0.27	0.00
Multiplication percent	Regular	413	37%	0.25	503	24%	0.30	339	14%	0.23	-0.23	-61%	0.24	-0.93
Multiplication percent	Youth	346	35%	0.22	241	22%	0.31	243	12%	0.21	-0.23	-65%	0.22	-1.06
Division percent	Regular	404	22%	0.25	497	17%	0.26	341	7%	0.18	-0.15	-68%	0.22	-0.70
Division percent	Youth	335	24%	0.24	236	16%	0.28	242	8%	0.19	-0.16	-67%	0.22	-0.76
Shape identification percent	Regular	423	26%	0.25	419	38%	0.25	210	28%	0.42	0.02	6%	0.32	0.05
Shape identification percent	Youth	345	25%	0.23	216	39%	0.26	172	21%	0.38	-0.04	-17%	0.29	-0.14

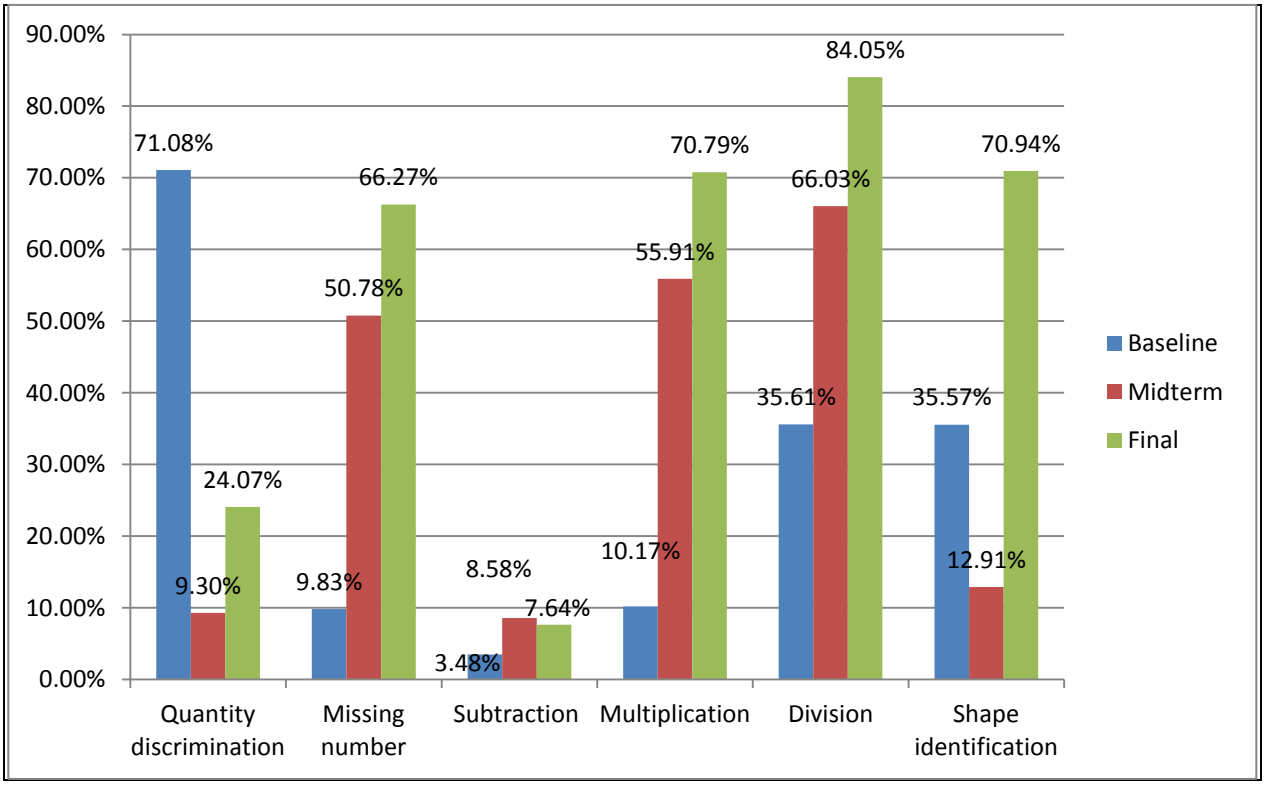
N=sample size, Avg=mean; SD=standard deviation.

When we compare ALP learners' zero scores on EGMA at baseline, midterm, and final assessment, the results are disappointing (see ***Table 28*** and ***Figure 17***). For all mathematics subtasks, the percentage of zero scores *increased* from baseline to final assessment. In most instances, these changes were dramatic. For example, the percentage of ALP learners with zero scores on multiplication rose from 10.17% at baseline to 70.79% at final. Such dramatic increases in students scoring zero on mathematics subtasks seem unlikely, if the intervention was faithfully applied; further investigation into the appropriate administration of EGMA or the data entry process itself seems warranted.

Table 28: Percentages of ALP Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline, Midterm, and Final

Subtask	Baseline	Midterm	Final	T-test
Quantity discrimination	71.08%	9.30%	24.07%	t=18.1349; p=0.0000
Missing number	9.83%	50.78%	66.27%	t=-27.0163; p=0.0000
Subtraction	3.48%	8.58%	7.64%	t=-3.4123; p=0.0007
Multiplication	10.17%	55.91%	70.79%	t=-29.3300; p=0.0000
Division	35.61%	66.03%	84.05%	t=-20.2244; p=0.0000
Shape identification	35.57%	12.91%	70.94%	t=-12.0137; p=0.0000

Figure 17: Percentages of ALP Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline, Midterm and Final



10.2 General Findings, Descriptive Analysis: NFE

Identical to the interpretation of mathematics performance for ALP students, for NFE learners, we report on average mathematics score by learners on individual subtasks as well as presenting the proportion of students who were able to answer all of the subtasks correctly. This approach allows us to compare changes across assessments but also to present important information in terms of the levels at which student were performing. For ease of reference, ***Table 29*** repeats the maximum possible scores on mathematics subtasks as assessed by EGMA. As with the ALP groups, when discussed in the forthcoming text, the average student performance is to be contrasted to the number of total possible answers. For example, if the average performance on the number identification for NFE students on the final assessment was 12.81, this would mean that NFE students on average had 12.81 out of 20 possible answers correct. The same approach to interpreting results is applied to all other subtasks.

Table 29: Interpreting Mathematics Average Scores as Assessed by EGMA: NFE Program

Subtask	Number of total possible correct answers
Number identification	20
Quantity discrimination	20
Missing number	10

Subtask	Number of total possible correct answers
Addition score (level 1)	10
Addition score (level 2)	9
Subtraction score (level 1)	10
Subtraction score (level 2)	9
Multiplication score	6
Division score	6
Shape identification score	4

Unlike for ALP, based on the comparison of average mathematics achievement from baseline to final assessment, it appears that the NFE program likely contributed to improvements in mathematics achievement scores;[7] however, these results were not as consistent across subtasks as they were in reading. Positive changes in average mathematics achievement from baseline to final assessment occurred in just half of the subtasks reported here. See ***Table 30*** and ***Figure 18*** for the overall performance of NFE learners on EGMA.

Table 30: NFE Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores

	Baseline			Final			Impact			
Subtask	*N*	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled variation	Effect size (SD)
Number identification score	275	12.81	6.54	298	14.64	5.74	1.83	14%	6.13	0.29
Quantity I discrimination comparison fluency	358	6.13	5.64	430	6.08	5.86	-0.04	0%	5.75	-0.00
Quantity discrimination score	358	6.01	5.15	430	6.07	5.85	0.06	1%	5.53	0.01
Missing digit score	354	0.93	2.06	429	0.52	1.70	-0.41	43%	1.87	-0.21
Addition score (level 1)	351	5.06	3.54	428	5.86	3.60	0.79	15%	3.57	0.22
Addition score (level 2)	347	3.85	3.26	346	5.66	2.53	1.81	47%	2.91	0.62
Subtraction score (level 1)	352	4.33	3.36	427	4.65	3.19	0.31	7%	3.26	0.09
Subtraction score (level 2)	351	2.55	2.75	298	3.73	2.04	1.178	46%	2.45	0.48
Multiplication score	349	0.55	1.19	423	0.48	1.16	-0.07	-12%	1.17	-0.06
Division score	349	0.41	1.09	424	0.27	0.92	-0.13	-33%	1.00	-0.13
Shape identification score	127	1.09	1.01	213	0.57	1.30	-0.48	-45%	1.20	-0.40

N=sample size, Avg=mean; SD=standard deviation.

[7] A quasi-experimental analysis that analyzed the impact of the program in comparison to no intervention could confirm this supposition.

Figure 18: NFE Program Impact on Mathematics, Comparing Baseline and Final Assessments: Average Student Scores

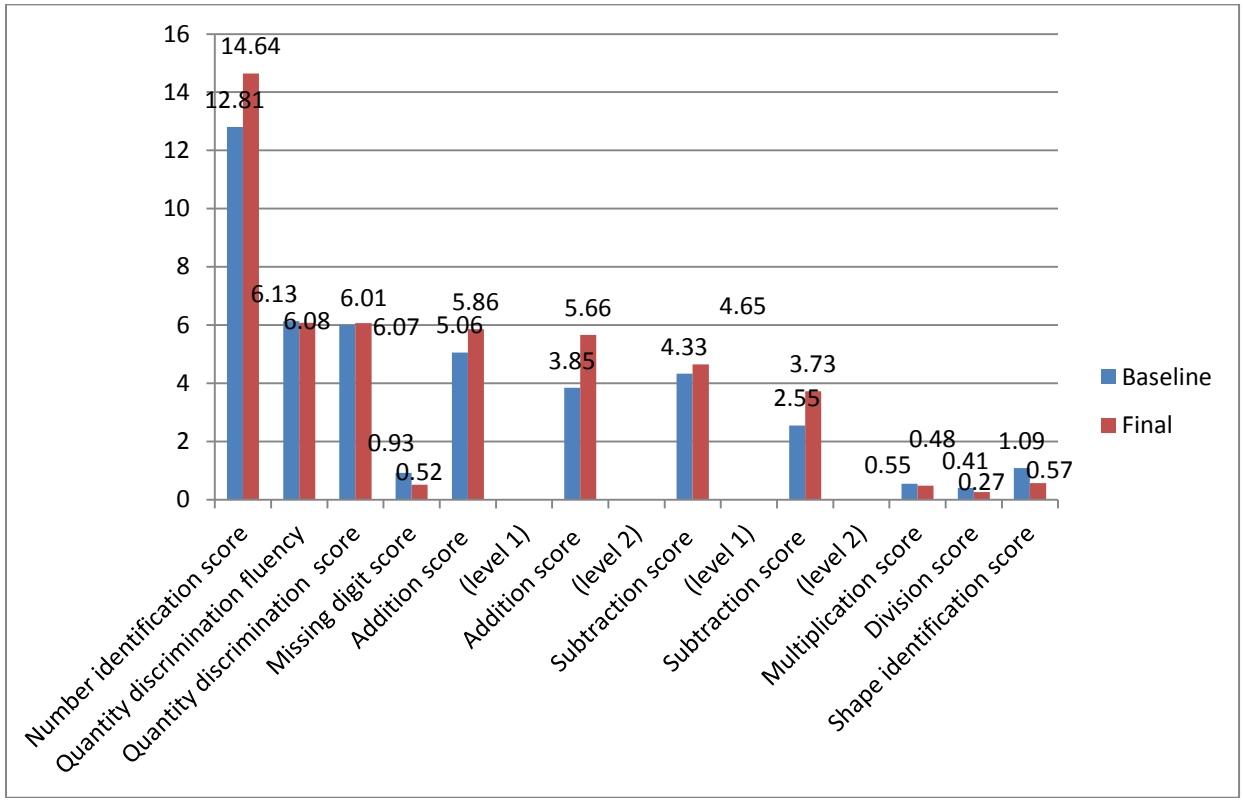


Table 31 and ***Figure 19*** present the proportion of students who were able to answer all of the items correctly. The average mathematics achievement score in number identification, addition (level 1), addition (level 2), subtraction (level 1), and subtraction (level 2), increased from the baseline to the final sample. For example, at baseline, the proportion of students who answered all items correctly was 64%; at final, that average had increased to 73%. The largest percentage increase occurred on the addition score (level 1). At baseline, the average addition score for addition (level 2) was 43%; at final, we seen an increase to 63%.

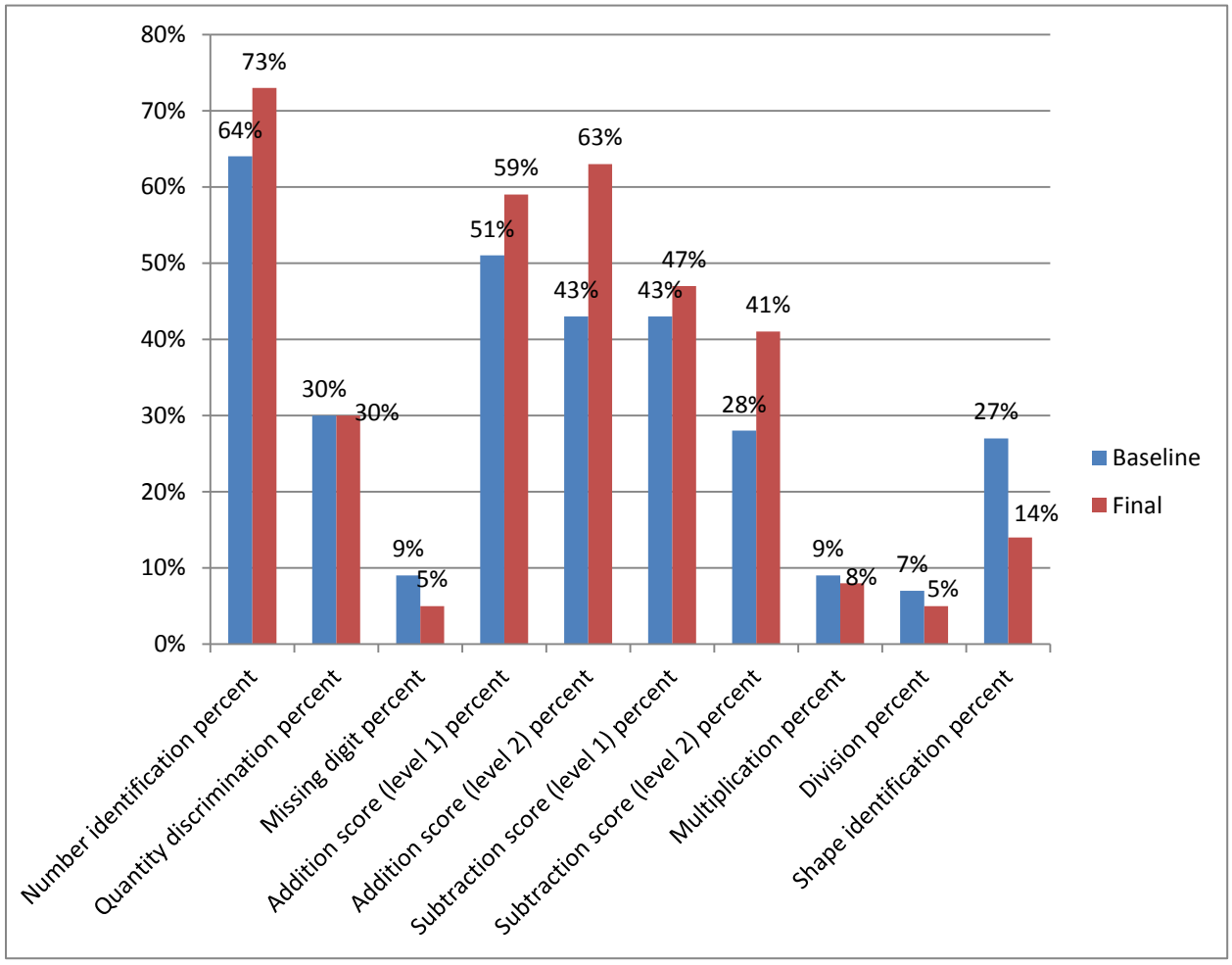
Table 31: NFE General EGMA Findings

	Baseline			Final			Impact			
Subtask	*N*	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled variation	Effect size
Number identification percent	275	64%	0.33	298	73%	0.29	0.09	14%	0.31	0.30
Quantity discrimination percent	358	30%	0.26	430	30%	0.29	0.00	1%	0.28	0.01
Missing digit percent	354	9%	0.21	429	5%	0.17	-0.04	-44%	0.19	-0.22
Addition score (level 1) percent	351	51%	0.35	428	59%	0.36	0.08	16%	0.36	0.22
Addition score (level 2)	347	43%	0.36	346	63%	0.28	0.20	47%	0.32	0.62

Subtask	Baseline			Final			Impact			
	N	Avg	SD	*N*	Avg	SD	Gains	Percent increase	Pooled variation	Effect size
percent										
Subtraction score (level 1) percent	352	43%	0.34	427	47%	0.32	0.03	7%	0.33	0.10
Subtraction score (level 2) percent	351	28%	0.31	298	41%	0.23	0.13	46%	0.27	0.48
Multiplication percent	349	9%	0.20	423	8%	0.19	-0.01	-13%	0.20	-0.06
Division percent	349	7%	0.18	424	5%	0.15	-0.02	-34%	0.17	-0.14
Shape identification percent	127	27%	0.25	213	14%	0.33	-0.12	-46%	0.30	-0.40

N=sample size; Avg=mean; SD=standard deviation.

Figure 19: NFE General EGMA Findings



Female NFE students on the baseline assessment performed at consistently lower average levels on mathematics subtasks than their male peers (see ***Table 32*** and ***Figures 20 and 21***). Males posted higher mathematics achievement baseline scores on all subtasks than

females. For example, the average quantity discrimination fluency result for male students at baseline was 9.58; in comparison, the average quantitative comparison fluency score of female students was approximately half that at 5.07 problems correct per minute. At the final assessment, female students continued to perform at consistently lower average levels than their male peers. Moreover, on four subtasks, average female scores not only did not increase, but decreased from the baseline to the final assessment.

At the same time, on the remaining six subtasks, female students showed percentage increases in average mathematics achievement scores from the baseline to the final assessment. In those instances, their percentage increases were larger than those of male students. For example, the percentage increase of the average addition score (level 2) of female students was 63% in contrast to 14% for male students. This finding suggests that participation in the NFE intervention may have positively influenced female student mathematics outcomes to a greater degree than male outcomes, at least on some subtasks. Of course, this does not change the fact that the overall average scores of male students on mathematics subtasks were higher than those of female students on the final assessment.

Table 32: NFE Program Impact on Mathematics Scores, by Sex

		Baseline			Final			Impact					
Subtask	Sex	N	Avg	SD	N	Avg	SD	Gains	% increase	Pooled variation	Effect size	Female gains	Female effect size
Number identification score	M	72	16.76	4.44	89	17.13	4.34	0.371	2.21%	4.36	0.09		
Number identification score	F	194	11.49	6.66	193	13.59	5.92	2.096	18.24%	6.29	0.33	1.72	0.25
Quantity discrimination fluency	M	84	9.58	7.00	98	7.82	7.10	-1.76	-18%	7.01	-0.25		
Quantity discrimination fluency	F	262	5.07	4.70	312	5.48	5.33	0.40	8%	5.05	0.08	2.17	0.33
Quantity discrimination score	M	84	9.17	5.60	98	7.81	7.08	-1.36	-15%	6.40	-0.21		
Quantitative discrimination score	F	262	5.04	4.63	312	5.46	5.31	0.43	08%	5.00	0.09	1.78	0.30
Missing digit score	M	84	1.57	2.94	97	0.78	2.18	-0.79	-50%	2.55	-0.31		
Missing digit score	F	258	0.71	1.64	312	0.41	1.49	-0.31	-43%	1.55	-0.20	0.48	0.11
Addition score (level 1)	M	84	6.93	3.15	98	7.10	3.37	0.17	2%	3.25	0.05		
Addition score (level 1)	F	256	4.47	3.46	310	5.44	3.57	0.98	22%	3.51	0.28	0.81	0.23
Addition score (level 2)	M	82	5.57	3.11	86	6.37	2.36	0.80	14%	2.74	0.29		
Addition score (level 2)	F	253	3.29	3.11	243	5.34	2.52	2.06	63%	2.83	0.73	1.25	0.43

Subtask	Sex	Baseline			Final			Impact					
		N	Avg	SD	N	Avg	SD	Gains	% increase	Pooled variation	Effect size	Female gains	Female effect size
Subtraction score (level 1)	M	83	5.92	3.35	98	5.94	3.20	0.02	0%	3.25	0.01		
Subtraction score (level 1)	F	257	3.80	3.23	309	4.19	3.08	0.38	10%	3.15	0.12	0.36	0.12
Subtraction score (level 2)	M	83	3.43	2.97	76	4.29	2.28	0.86	25%	2.65	0.32		
Subtraction score (level 2)	F	256	2.26	2.65	206	3.53	1.90	1.27	56%	2.34	0.54	0.41	0.22
Multiplication score	M	83	0.99	1.57	96	0.97	1.66	-0.02	-2%	1.61	-0.01		
Multiplication score	F	254	0.41	0.98	308	0.33	0.92	-0.07	-18%	0.94	-0.08	-0.05	-0.06
Division score	M	83	0.57	1.40	98	0.57	1.36	0.01	1%	1.37	0.00		
Division score	F	254	0.35	0.97	307	0.18	0.73	-0.16	-47%	0.84	-0.19	-0.17	-0.20
Shape identification score	M	42	1.26	1.11	60	0.90	1.55	-0.36	-29%	1.37	-0.26		
Shape identification score	F	81	0.98	0.97	142	0.43	1.14	-0.55	-56%	1.08	-0.50	-0.18	-0.24

N=sample size; Avg=mean; SD=standard deviation.

Figure 20: NFE Program Impact on Mathematics, Part 1, by Sex: Average Scores

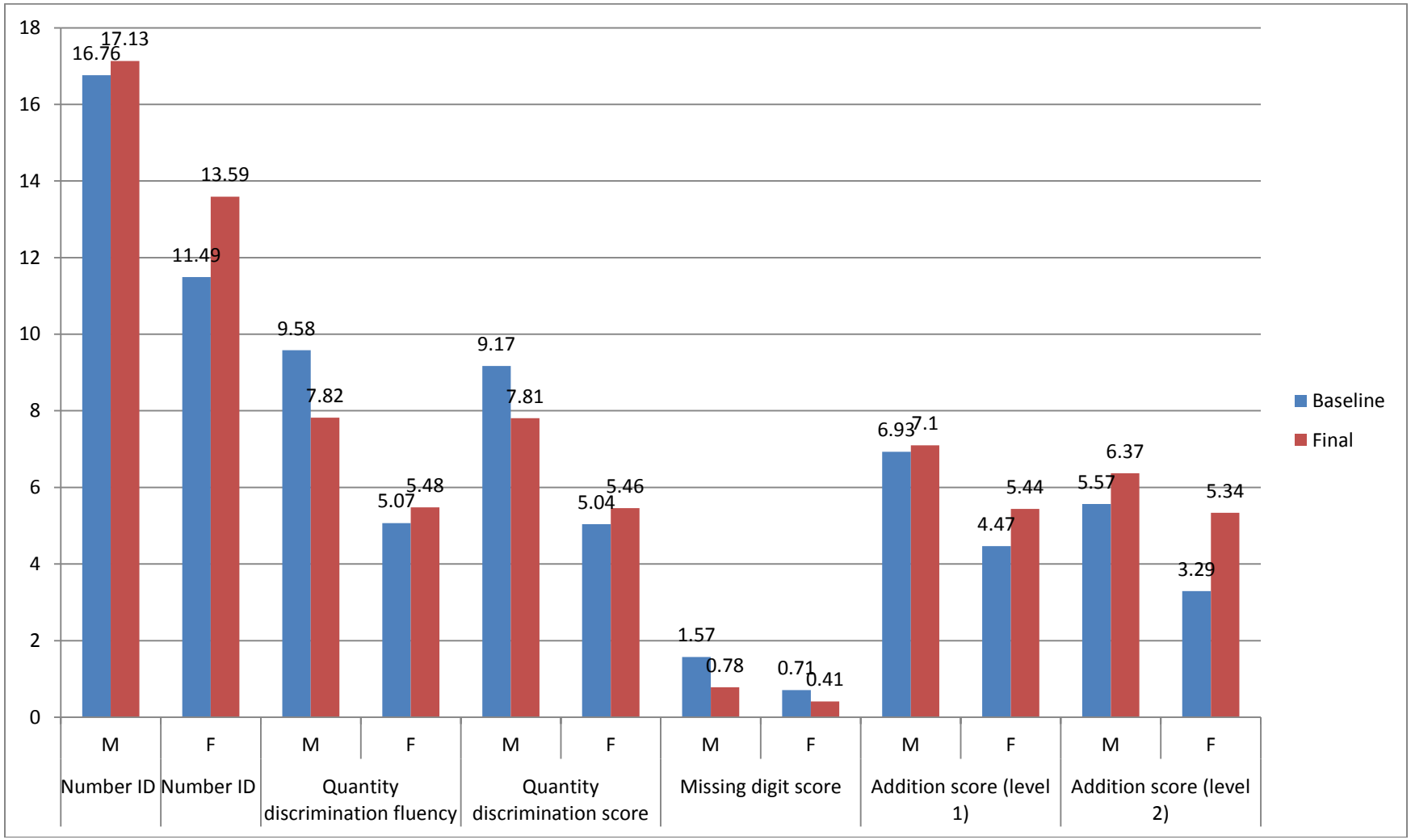
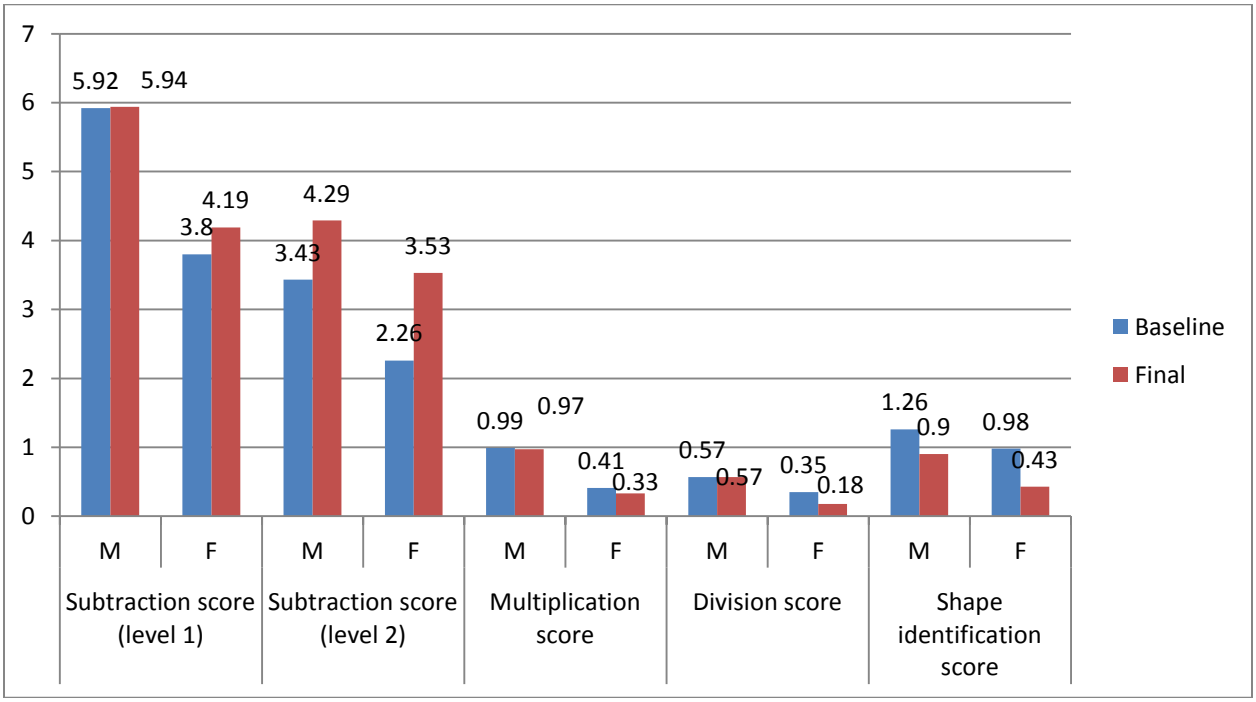


Figure 21: NFE Program Impact on Mathematics, Part 2, by Sex: Average Scores



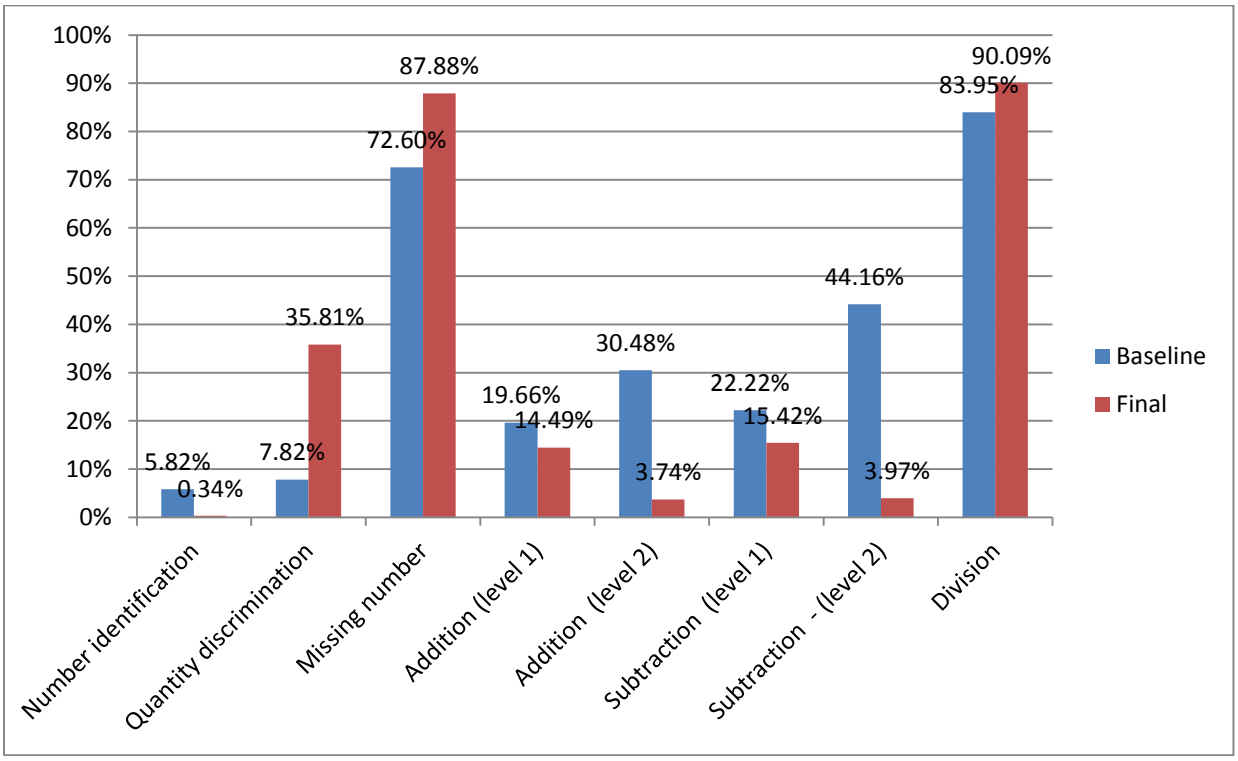
When we compare NFE learners' zero scores on EGMA at baseline and final, the results are mixed ***(Table 33*** and ***Figure 22).*** That is, on five of ten subtasks, participants showed fewer zero scores on the final assessment than they did on the baseline assessment. For all but one of these subtasks (addition, level 1), this difference was statistically significant. However, for the remaining subtasks, the percentage of zero scores increased between baseline and final assessment. Moreover, these increases in zero scores were statistically significant. This finding suggests that there may have been differences in the quality of intervention by subtask; perhaps instructors themselves showed greater mastery of the material in certain subtasks than others and these strengths were manifested in their instruction.

Table 33: Percentages of NFE Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline and Final

Subtask	Baseline	Final	T-test
Number identification	5.82%	0.34%	t=3.9089; p=0.0001
Quantity discrimination	7.82%	35.81%	t=-9.8245; p=0.0000
Missing number	72.60%	87.88%	t=-5.5188; p=0.0000
Addition (level 1)	19.66%	14.49%	t=1.9225; p=0.0549
Addition (level 2)	30.48%	3.74%	t= 10.9264; p=0.0000

Subtask	Baseline	Final	T-test
Subtraction (level 1)	22.22%	15.42%	t= 2.4394; p=0.0149
Subtraction (level 2)	44.16%	3.97%	t= 15.3373; p=0.0000
Division	83.95%	90.09%	t=--2.5608; p=0.0106

Figure 22: Percentages of NFE Learners with Zero Scores on EGMA Subtasks: Comparison of Baseline and Final



11. Principal Questionnaire Results for ALP and NFE

The principal questionnaire was designed to gather detailed quantitative information about the school principals' views of the relationships among teacher and school characteristics, home background, and learner outcomes. The same questions were asked of both ALP and NFE principals.

11.1 ALP Principals

Similar to the results at baseline, at midterm we found that ALP school principals still had low expectations of their CESLY learners, with only 1.72% (compared with 3.6% at midterm and 3.1% at baseline) expecting learners to read fluently by the end of Level 2.

Characteristics of the ALP principals themselves can be summarized as follows:

- *Experience*—Grouping the respondents by years of experience, the largest number of principals in the sample (15.52%) had only three years of experience as head of their current school, nearly the same percentage as reported in the midterm analysis.
- *Teaching duties*—48.28% of principals also were teaching ALP classes, slightly less than was reported in the midterm analysis. More than 36% of these were Level 3 teachers. By the same token, 24.14% of the principals in the sample reported teaching in the regular primary school where the ALP program was held. A full 75.86% reported also serving as the principal of the regular primary school.
- *Qualifications*—The qualifications of the sample of principals were divided among senior high school graduates (13.79%), associate's degree (5.17%), C certificate (56.9%), B certificate (15.52%), AA certificate (1.72%), and bachelor's degree (3.45%). Notably, 87.93% of principals reported receiving special training or taking courses in school management.

Reading resources. As far as school facilities conducive to reading are concerned, just 32.76% of principals reported having a library at their school; this percentage was lower than that reported at the midterm (36%) and higher than reported at baseline (25%). Another notable change from the midterm assessment was the percentage of principals reporting that learners were allowed to take library books to read at home; the final percentage was just 39.66, compared to 59.4% of principals at midterm and 36.6% at baseline. The dip in the percentage indicates that school principals still need continued support and training in managing school libraries by allowing learners to borrow more books even if, for example, students return books in less-than-new condition.

External monitoring and support. When interviewed about the number of visits by school inspectors or education officers, only two school principals indicated that their schools did not receive a single annual visit, down from six at midterm. The majority of principals, however, indicated that their schools had been visited once every two to three months during the past year. Also, 19 principals reported their schools were visited by inspectors once a week. In previous waves of data collection, 21 principals (midterm) and 33 principals (baseline) reported the same. However, given how difficult it is to reach

some schools, it is uncertain whether these data are trustworthy or resulted from principals' inadequate comprehension of the question. Studies show that school inspection in Liberia is a daunting task. As noted by Mulkeen (2010),[8] district and county education officers are responsible for inspecting schools. But due to transportation and other logistical problems, some schools are visited as infrequently as once every five years.

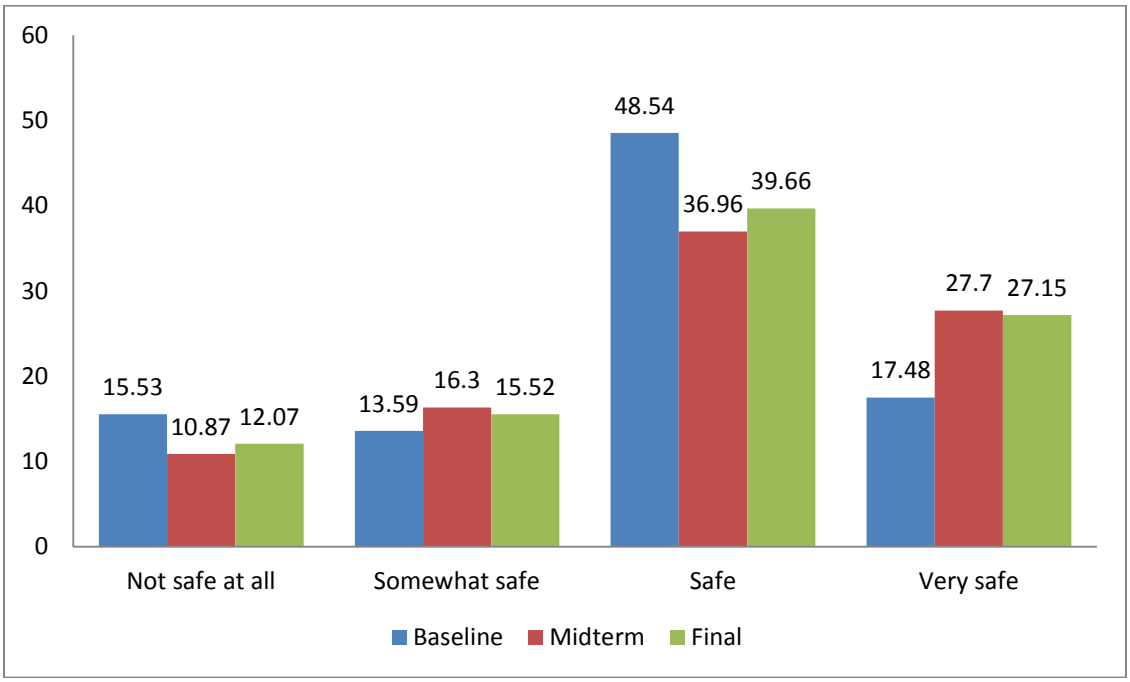
For pedagogical support activities, it is evident from the principals' midterm responses that CESLY interventions are enhancing teaching methodologies from the principals' perspective:

- 84.48% of principals indicated that they checked for learner attendance using registers. This was a decrease from 95.1% at the midterm, but an improvement from 83.2% at baseline.
- 86.21% of principals indicated that they examined pupils' progress records, an increase over the midterm (83.9%) and baseline (82.1%) assessments.
- 84.48% indicated that they observed classrooms in session. This was a decrease from the 86.6% recorded at midterm, but an increase over the 75.8% that was recorded at the baseline.
- 70.69% of interviewed school principals said they examined recent learner assessments and evaluations. This was a notable decrease from the midterm assessment, in which 96.3% of principals reported the same. Surprisingly, this percentage was even lower than the 72.0% reported at baseline.
- 93.10% of principals indicated that they gave advice on learner discipline, just below that reported at midterm and baseline (97.6%), although still high.
- 93.10% indicated that that they advised on teaching techniques. This was below the percentages reported at the baseline and midterm assessments (95.1%).
- 89.66% indicated that they provided information to teachers on curriculum innovations. This percentage increased from 82.9% at midterm but was still lower than the 92.6% reported at baseline.
- 89.66% indicated that they shared information on potential professional development opportunities, just above the percentage reported at baseline and midterm (88.9%).

Figure 23 presents information about safety and violence concerns expressed by the school principals, since these are ancillary goals of the ALP programs supported by CESLY (more on similar questions appears in the section that follows). The results show that most principals felt that they were themselves safe—see the "safe" and "very safe" categories in Figure 23.

[8] Mulkeen, A. (2010). *Teachers in anglophone Africa: Issues in teacher supply, training, and management*. Washington, DC: International Bank for Reconstruction and Development/World Bank.

Figure 23: ALP Principals' Levels of Concern about Safety: Baseline, Midterm, and Final Assessments



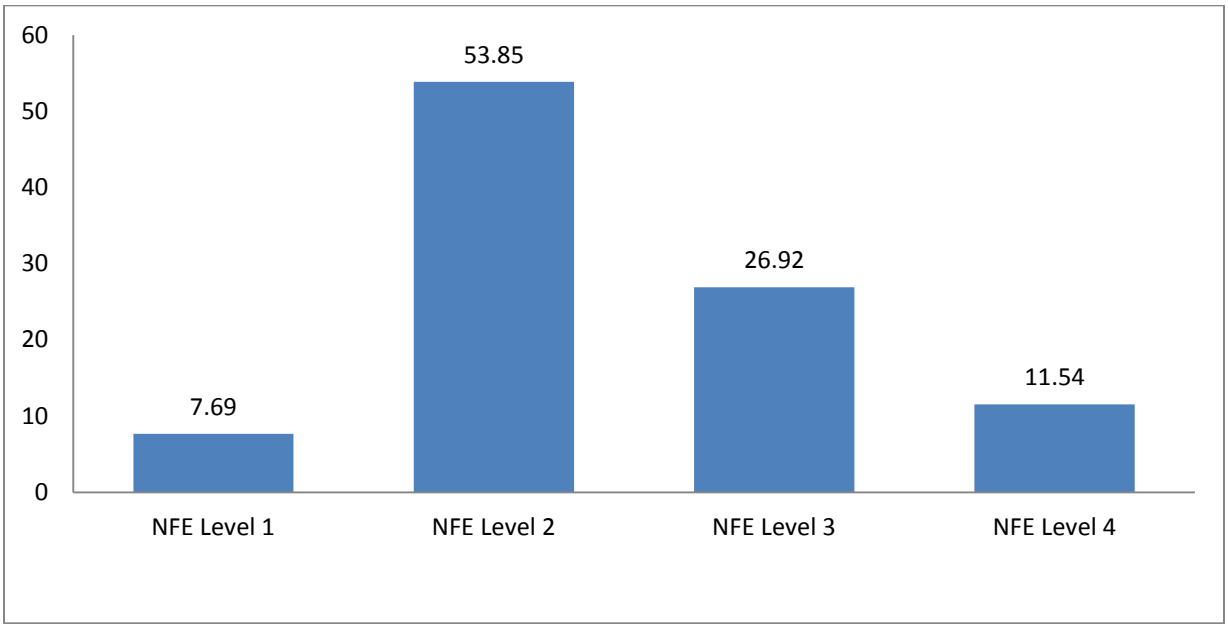
11.2 NFE Principals

About 30% of NFE principals also said that they were teaching NFE classes, which poses a serious burden on their time to perform their administrative tasks. In addition, a similar proportion of NFE principals reported that they also were teaching either regular or ALP classes. Given that regular classes take place in the mornings and NFE classes take place at night, this would mean long working hours in a given day. The CESLY program should think of ways to support these principals. When asked about their qualifications, about 8% of them said that they had elementary education; 23% had secondary education; 38% reported having a C certificate; 20% reported having a B certificate; and around 10% reported having either an AA certificate or a bachelor's degree. Interestingly, 100% of the respondents indicated that they received some training in school management, and that 80% of this training was provided by a nongovernment entity.

As was the case with the ALP principals, the majority of NFE principals had low expectations of how quickly their learners should learn how to read (see ***Figure 24***). The CESLY program should focus on changing expectations so that NFE principals, as well as teachers and students, expect the students to learn how to read in their first year of school. Although this problem also was discussed in the previous sections, we would like to draw some serious attention to this issue. Moreover, even though adult learners do indeed learn differently from regular-age students, adult learners bring a wealth of knowledge and a richer vocabulary, and can learn to read just as fast if not faster. The CESLY program should focus on ensuring that majority of NFE learners learn how to

read in Level 1 so that they can spend the rest of the three years of their education reading to learn.

Figure 24: NFE Principals' Expectations about When NFE Learners Should Become Fluent Readers



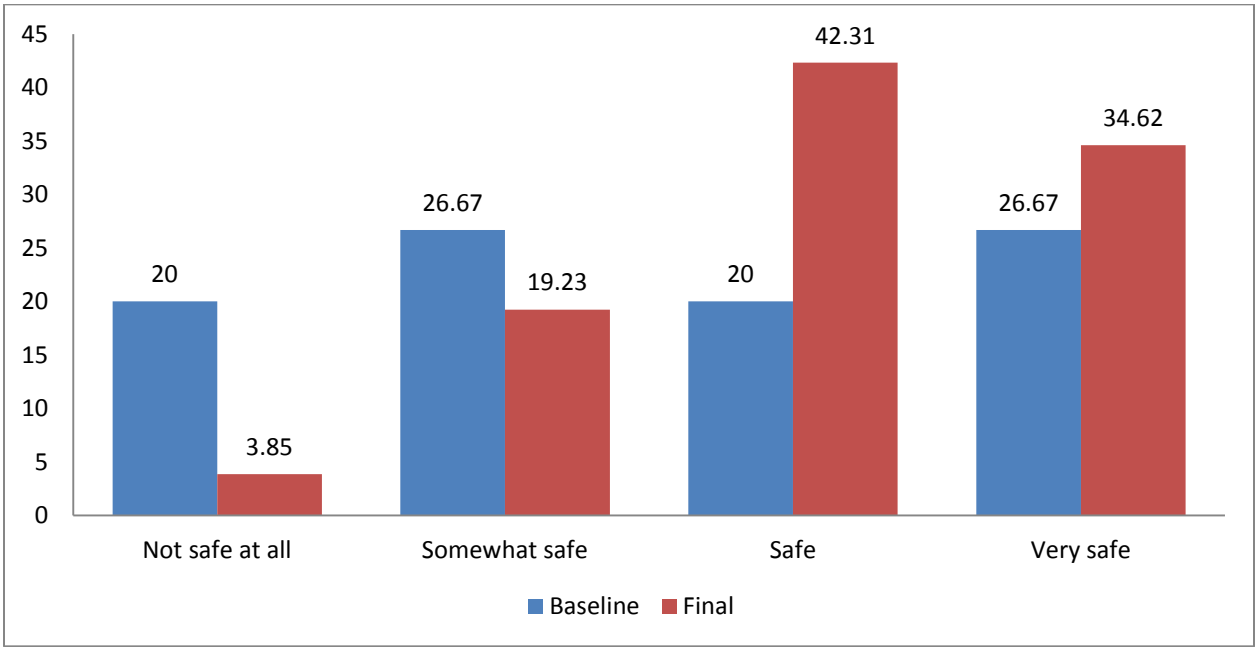
The majority of NFE principals—80.77%—indicated that there were no libraries at their schools. The majority of NFE principals also responded that they held PTA meetings on a regular basis. The CESLY program should strive to use these PTA meetings to raise awareness about the problems of lack of reading materials as well as low literacy and numeracy among students and in their communities overall. Moreover, about 60% of principals indicated that they made it a practice to discuss the safety of learners at the school, and 65% of principals also said they discussed student learning outcomes. These meetings, then, are venues at which some of the important aspects of school development and improved learning outcomes are being discussed, but they need to be utilized even more. For example, RTI's previous research in Liberia (See *Piper, B; Korda, M.: Early Grade Reading Assessment (EGRA): Plus Liberia Program Evaluation – www.eddataglobal.org)* indicates that if learners read about 20 minutes at home every day, this will lead to a whole grade of improvement on its own. So if adult learners, and guardians or parents in the case of younger learners, can enforce this daily habit of reading at home, reading outcomes will improve greatly.

Interestingly, the majority of NFE principals—65%—indicated that they received a school inspection visit from an education officer. Our experience and visits to schools indicate that this is most likely far from reality, unless some change has been introduced in the past year or so, which is highly doubtful. However, it would be important to determine whether the principals considered visits by nongovernment agencies to be a

formal means of inspection. If this were this case, their responses could be seen as more realistic.

When asked about their concerns regarding safety and violence in their community, NFE principals' responses were similar to those of ALP principals (see ***Figure 25***). Very few said they felt "not safe at all"; a significant majority felt either safe or very safe, but a significant percentage—almost 20%—indicated feeling somewhat safe. However, great decreases occurred in the number of those who responded "not safe at all" and "somewhat safe" between baseline and final assessment; and there was a significant increase in the number of those who said they felt "safe" or "very safe." Overall, the responses to this question are showing a positive trend.

Figure 25: NFE Principals' Levels of Concern about Safety: Baseline, Midterm, and Final Assessments



12. Teacher Interviews: ALP and NFE

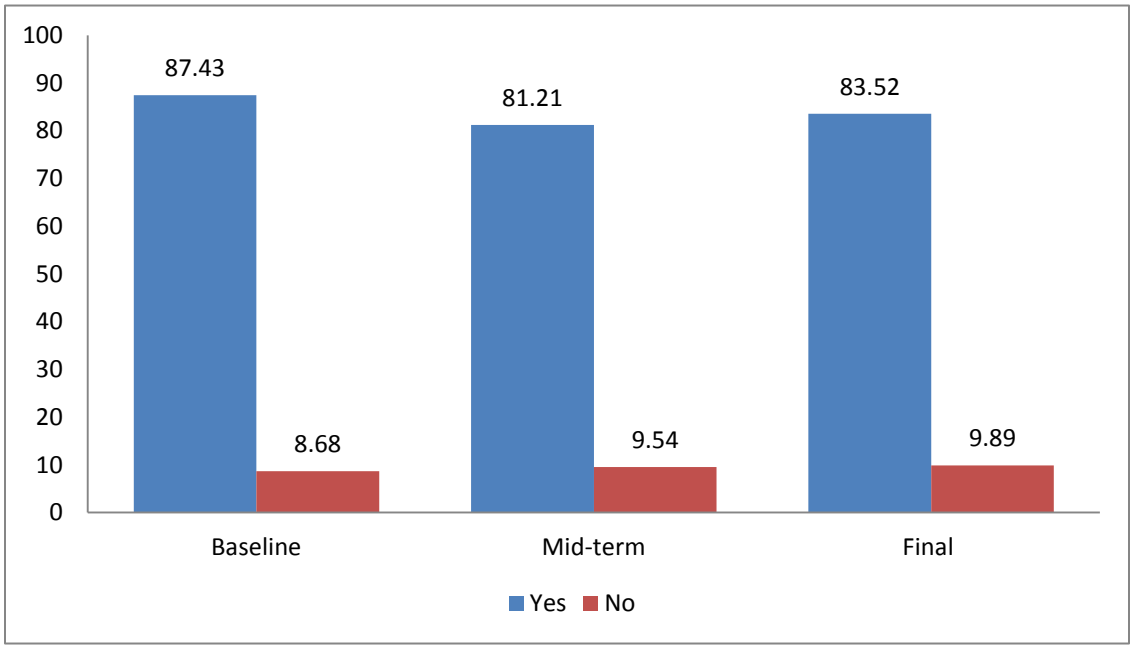
12.1 ALP Teachers

For ALP teacher interviews, we discuss only a few critical issues. In Section 12.2 we provide more detail about the NFE teachers.

About 85% of the ALP teachers indicated that they also were teaching in regular schools, and about 86% indicated that they held either a C certificate (33%) or high school degree (53%). Very few teachers said they had an elementary degree or an associate's or bachelor's degree. A significant 90% of these teachers indicated that they were using ALP manuals, and about 73% said that they found them useful. About 90% of teachers indicated that they had received a visit from an education officer. Again, as with the principals' responses, this is a rather high number and warrants further investigation, given that the number of education officers in Liberia is limited enough that we would not expect visits to happen with this frequency. It is possible that teachers are confusing nongovernmental organizations with government officers.

With respect to teachers' concerns about safety and violence in their community, the majority of ALP teachers indicated that they felt safe, a positive finding that was present from the beginning of the program (see ***Figure 26***).

Figure 26: Question for ALP Teachers: "In the past two years, have things generally improved in your community?"



12.2 NFE Teachers

As noted above, part of the final assessment involved a qualitative analysis via interviews with CESLY teachers in addition to the principals. In the NFE program, 54 teachers were interviewed. Of the NFE teachers interviewed, 96.30 said they taught in Level 1. (For comparison, in the ALP program, 91 teachers were interviewed, and 93.41% said they taught in ALP Level 3.)

The NFE teachers reported on professional development in reading. Of the total group, 77.78% reported that they had received training on how to teach reading. This figure was a decrease from 88.9% at midterm, but still an increase over the 48.7% at baseline.

The NFE teachers also reported on their pedagogical techniques. Positively, 96.3% of the teachers reported asking learners to read aloud, a figure similar to that found at the midterm assessment (98.4%), and an increase over the baseline (90.5%). The percentage of teachers who reported reading aloud to learners increased to 96.3, from 95.2% at midterm and 94.5% at the baseline.

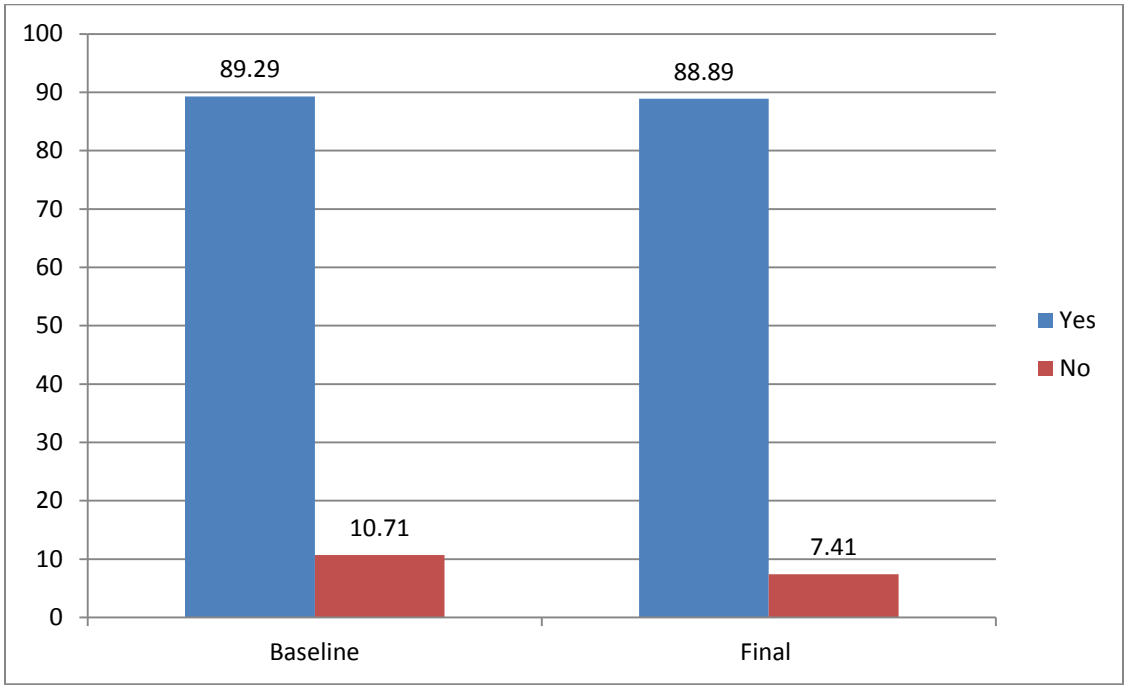
NFE teachers also responded to questions about the enabling environment surrounding the classroom. Their answers showed that 31.48% did not receive the CESLY stipend on time, down from 48.6% at midterm and 44% at baseline. A noticeable improvement was reported in the length of delays in CESLY stipend payments: Just 3.7% of teachers reported a delay of at least two months, as compared to 9.3% of teachers at midterm and 34% at the baseline. Finally, 42.6% of teachers reported having other incomes beyond teaching, the same percentage as reported at midterm and higher than the percentage reported at baseline (36.8%). To allow NFE teachers to focus exclusively on teaching, it appears that additional monetary resources may need to be supplied.

Besides teaching, these NFE interviewees indicated that they had significant responsibility in taking care of family members. A full 44.4% of teachers reported they had to support as many as six dependents, down from 59.5% at midterm. In continuing to implement CESLY, a key consideration is whether this personal burden may be related to teacher absenteeism, which could make the reforms intrinsic to ALP and CESLY difficult to implement.

The NFE teacher questionnaire was also a way to find out whether the teachers were using improved pedagogical methods. When asked if they if they often assigned reading for their learners to do at home, 88.89% of sample teachers reported that they did, or fewer than the 95.2% of sampled teachers who responded to the question at midterm. A notable change in pedagogical practices from midterm to final assessment was in whether teachers reported that they practiced mathematics problems with their students; at midterm, 95.2% of teachers said they practiced mathematics problems with their learners, but on the final assessment, just 51.85% reported the same. As this is considered an effective pedagogical practice, the decrease in reported use suggests that NFE teachers may be reverting to less effective methods.

As for the NFE teacher concerns about safety and violence in their communities, the answers reveal a trend similar to that of ALP, where no significant changes were detected (see ***Figure 27***). Overall, NFE teachers said they do feel safe in their communities, which has been the case from the beginning of the program.

Figure 27: Question for NFE Teachers: "In the past two years, have things generally improved in your community?"



13. Student, Teacher, and Principal Optimism

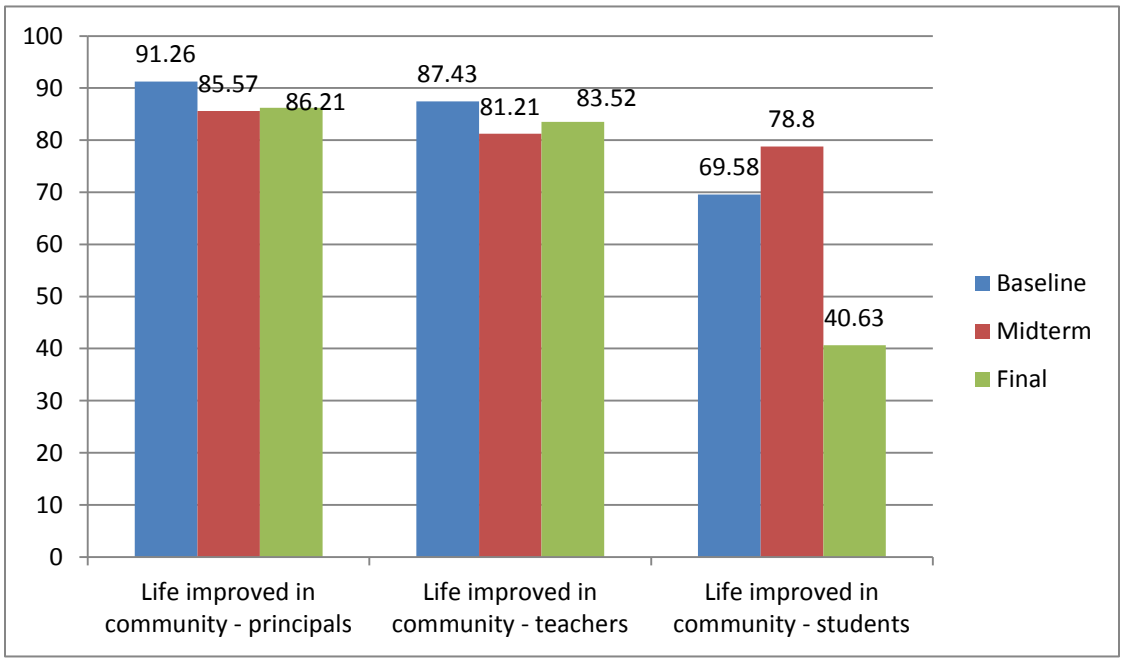
This section analyzes the responses to the "optimism" questions posed to three groups: learners (at the group interviews), teachers, and principals. The three groups were asked three questions in common:

1. "In the past two years, have things generally improved in your community?"
2. "In the past two years, have the lives of your friends and family gotten better, or "stayed the same?" and

3. "What do you think caused this change?"

The level of optimism with respect to the first question—has life in your community improved overall—went down across all of the ALP respondents between baseline and midterm and between the midterm and final assessment. The most significant decrease was in the case of students, whose "yes" responses to this question decreased from 78.8% at midterm to 40.63% on the final assessment. Overall, if we look at the percentages, the opinions of teachers and principals are positive, with nonsignificant decreases from baseline to midterm to final assessment (see ***Figure 28***).

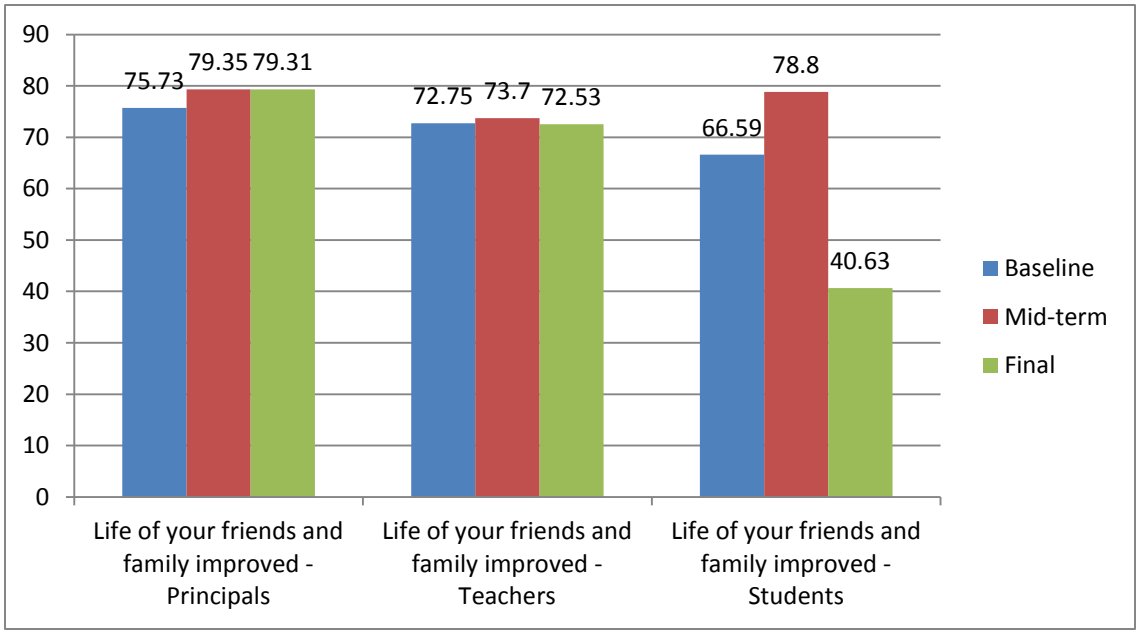
Figure 28: Comparison of ALP Principals, Teachers, and Students: "Have things in your community improved in general?"



The second question—Has the life of your family and friends gotten better or stayed the same—yielded a similar set of responses. Principals and teachers remained fairly positive about this issue, with "yes" responses in the range of 75% to 80% across all three assessments, indicating that they felt their lives "have gotten better." However, the

learners were not feeling as positive about this issue at the time of the final assessment, as indicated by a significant decrease from 78.8% answering "yes" on the midterm assessment to only 40.63% answering positively on the final assessment (see ***Figure 29***).

Figure 29: Comparison of ALP Principals, Teachers, and Students: "Has the life of your family and friends improved?"



As for the NFE respondents, the answer to the first question was more consistent across three respondent groups. All principals, teachers, and students felt more optimistic about their lives and situation in their communities. ***Figures 30 and 31*** below indicate generally high levels of optimism about improvements in the respondents' communities and their own lives from baseline to final assessment. Students felt a bit less optimistic than teachers and principals, however (Figure 30).

A similar trend was recorded on the question about improvements in the lives of the respondents' friends and families (Figure 31).

Figure 30: Comparison of NFE Principals, Teachers, and Students: "Have things in your community improved in general?"

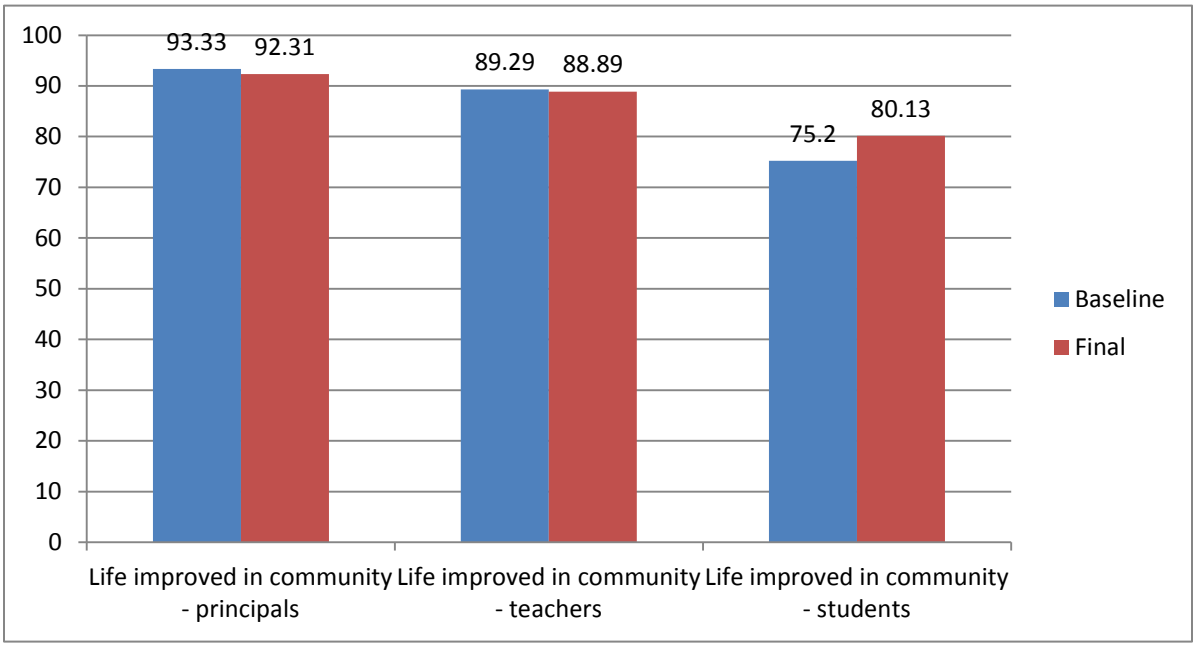
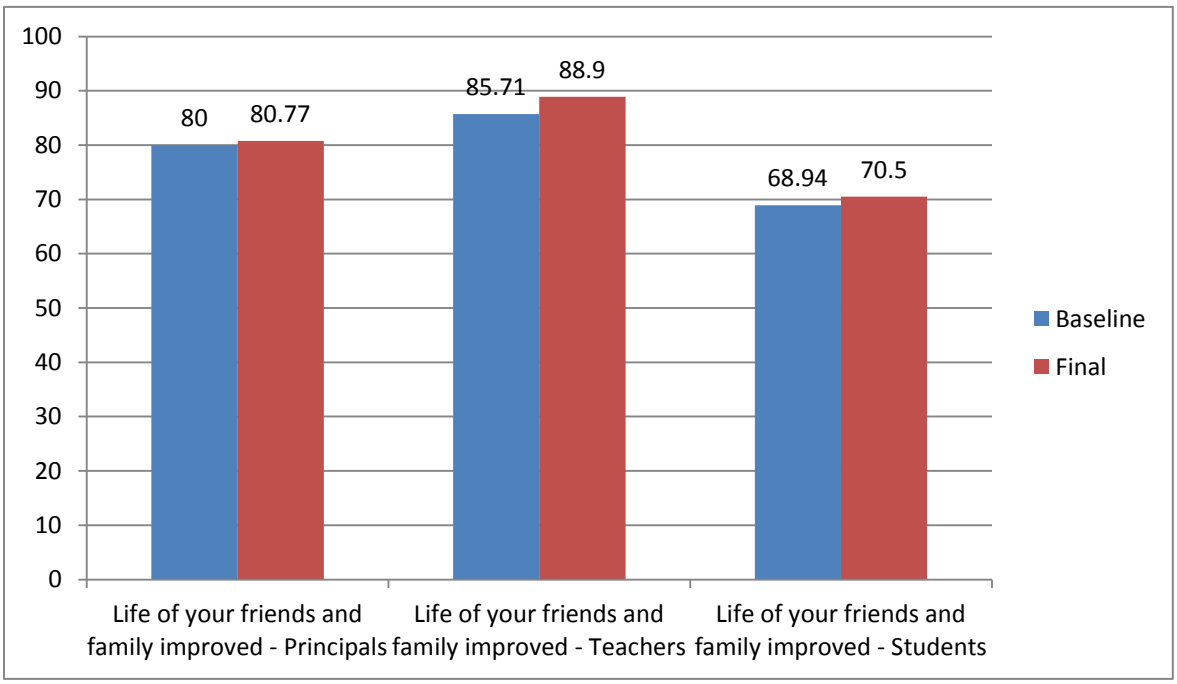


Figure 31: Comparison of NFE Principals, Teachers, and Students: "Has the life of your family and friends improved?"



When the researchers probed further to find out what might have caused these changes, the ALP and NFE respondents provided answers for a number of different categories (see ***Figures 32 through 37***). For example, we asked how many principals thought that the

change in the life of their community had something to do with the government intervention, some change in their health status, etc. Very few respondents thought that any of the categories we inquired about had anything to do with the changes in their lives. Similar responses were provided by ALP teachers and ALP students, as well as all of the NFE respondents (principals, teachers, and students).

However, it is worth noting that "participation in ALP" was chosen as the most frequent answer for improved well-being, even if not expressed by the majority of the respondents. This was interesting because ALP, like other educational programs, is seen as a program designed to benefit learners; however, principals and teachers appear to believe it is benefiting them as well. This may be because of varying definitions of the "benefit" of the program: ALP provides consistent income for principals and teachers. Student responses were also interesting, as one supposes that they may have been successful in starting their own businesses at least in part through their participation in the ALP program. The relationship between these two variables is one avenue for future investigation.

Among NFE participants, explanations for optimism varied across the three participant groups. Among principals, more than half attributed their optimism to participation in the NFE program. This finding was consistent with the reports of respondents in the ALP program. Likewise, more than half of NFE teachers attributed their optimism to participating in the NFE program, with another 40% attributing it to economic conditions. Finally, although the percentage of students' explanations of optimism were not especially high in any of the areas, about 10% attributed their optimism to participating in NFE.

Figure 32: ALP Principals' Views on What Might Have Caused Any Changes in Their Lives

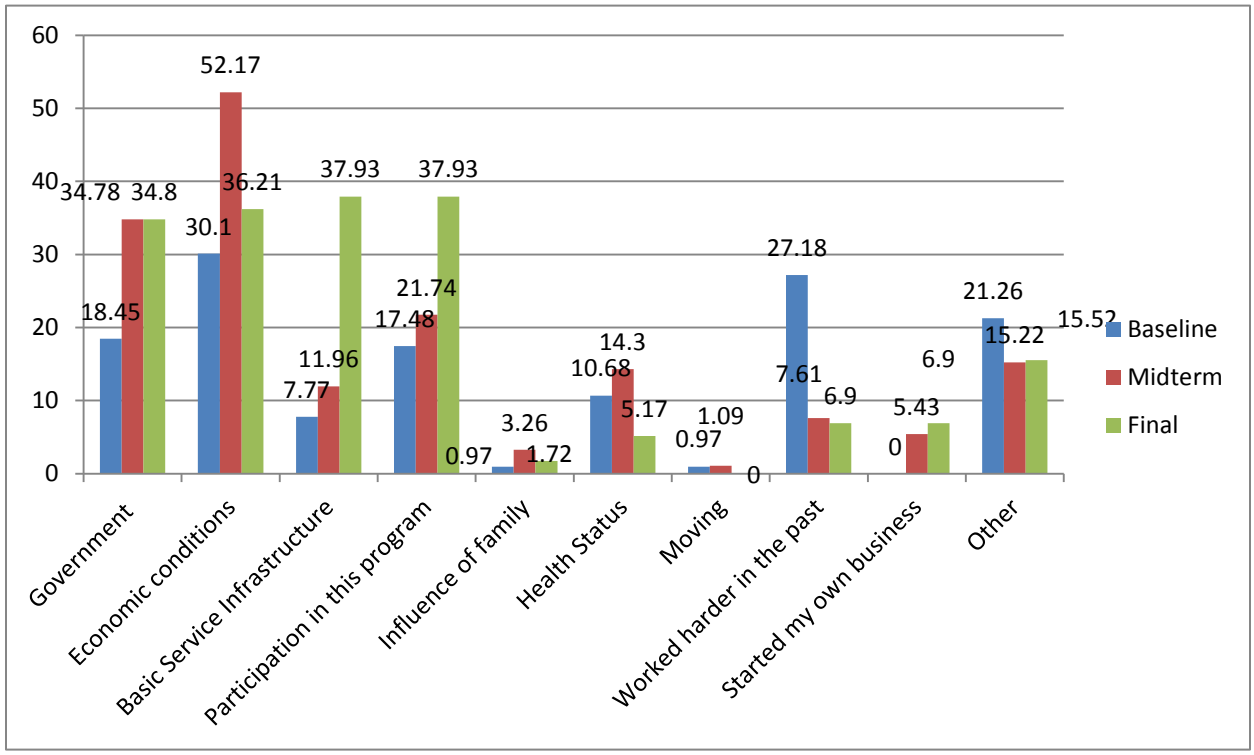


Figure 33: NFE Principals' Views on What Might Have Caused Any Changes in Their Lives

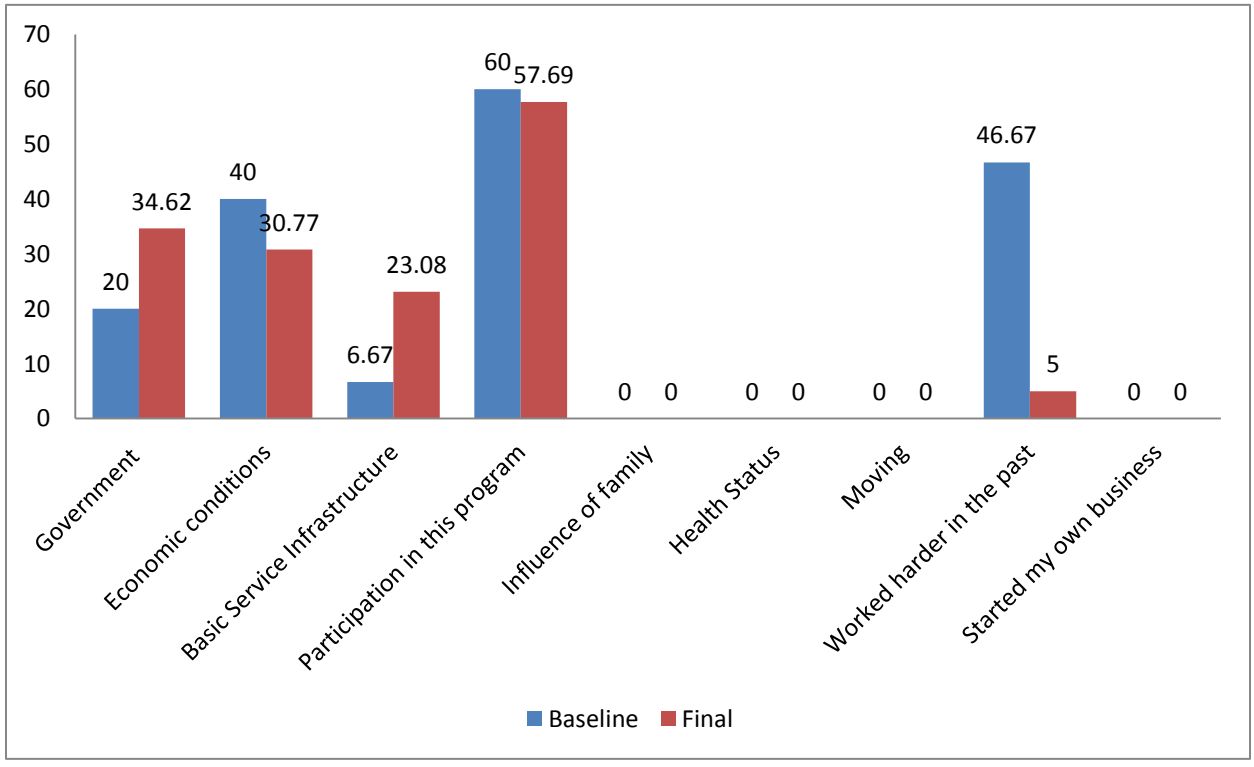


Figure 34: ALP Teachers' Views on What Might Have Caused Any Changes in Their Lives

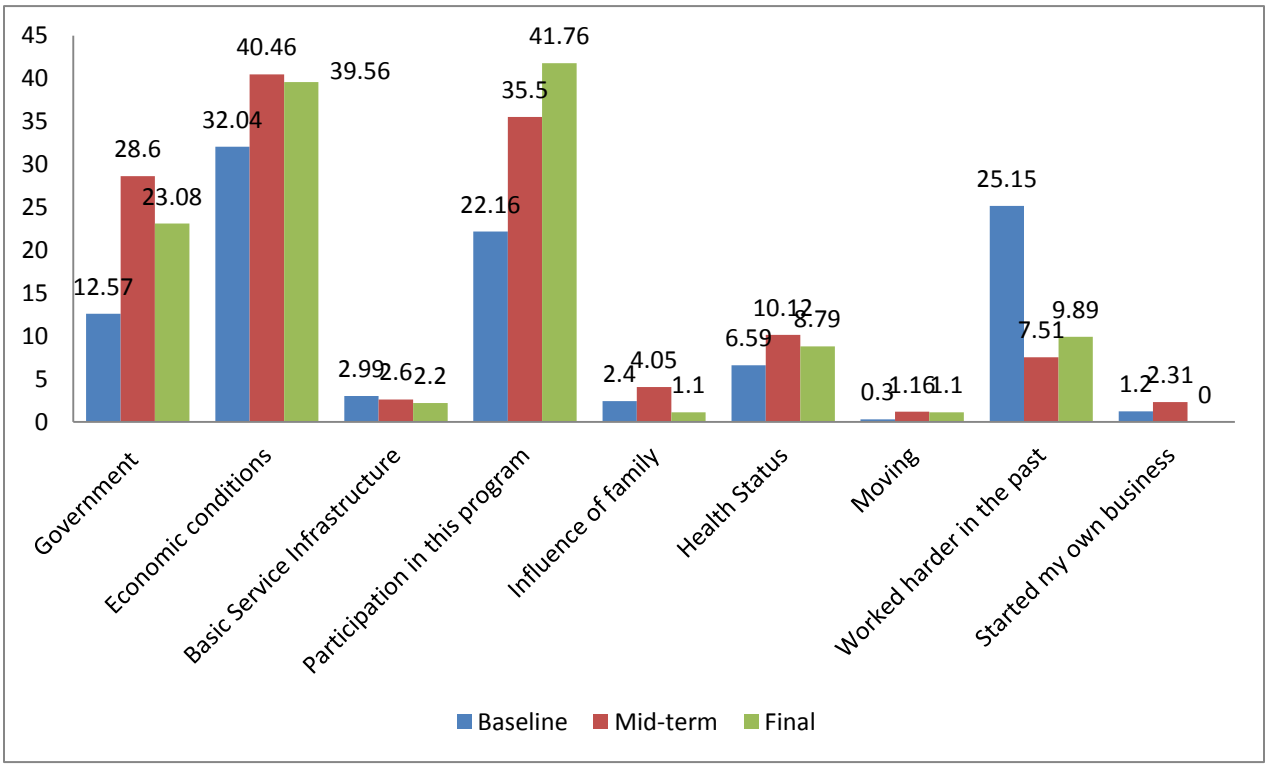


Figure 35: NFE Teachers' Views on What Might Have Caused Any Changes in Their Lives

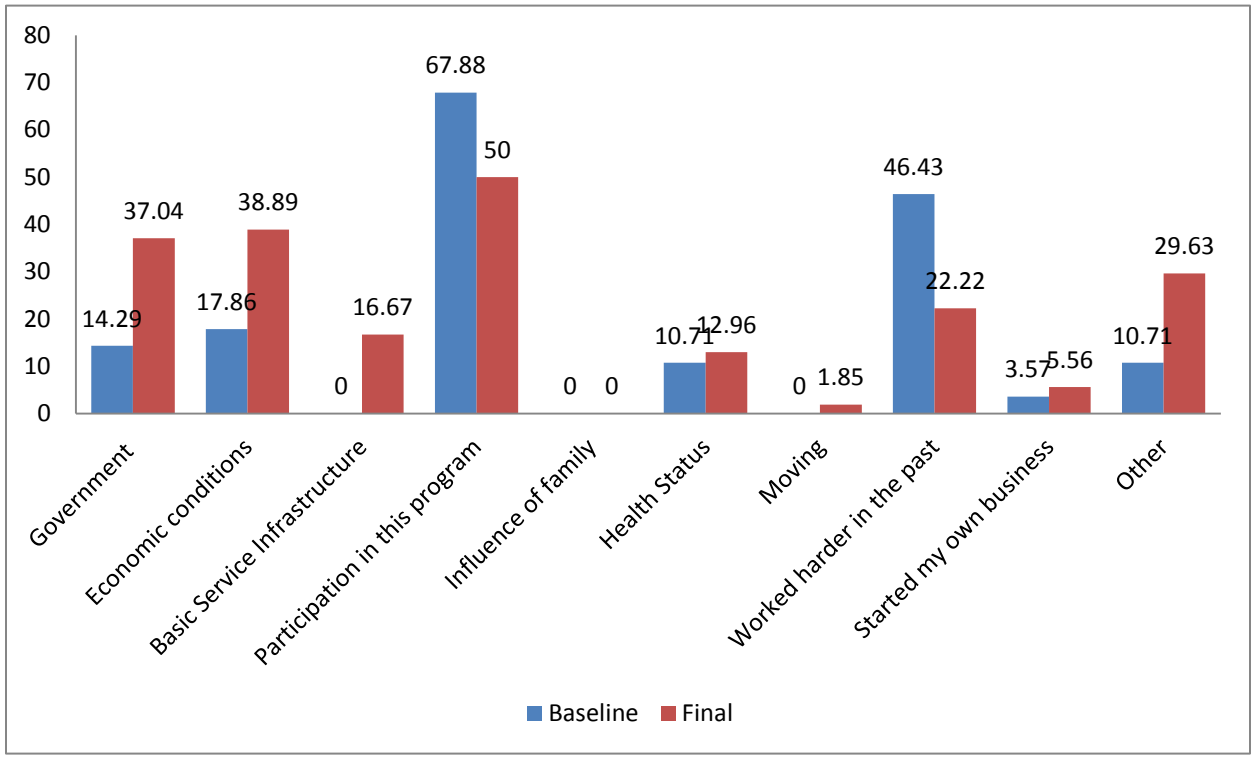


Figure 36: ALP Students' Views on What Might Have Caused Any Changes in Their Lives

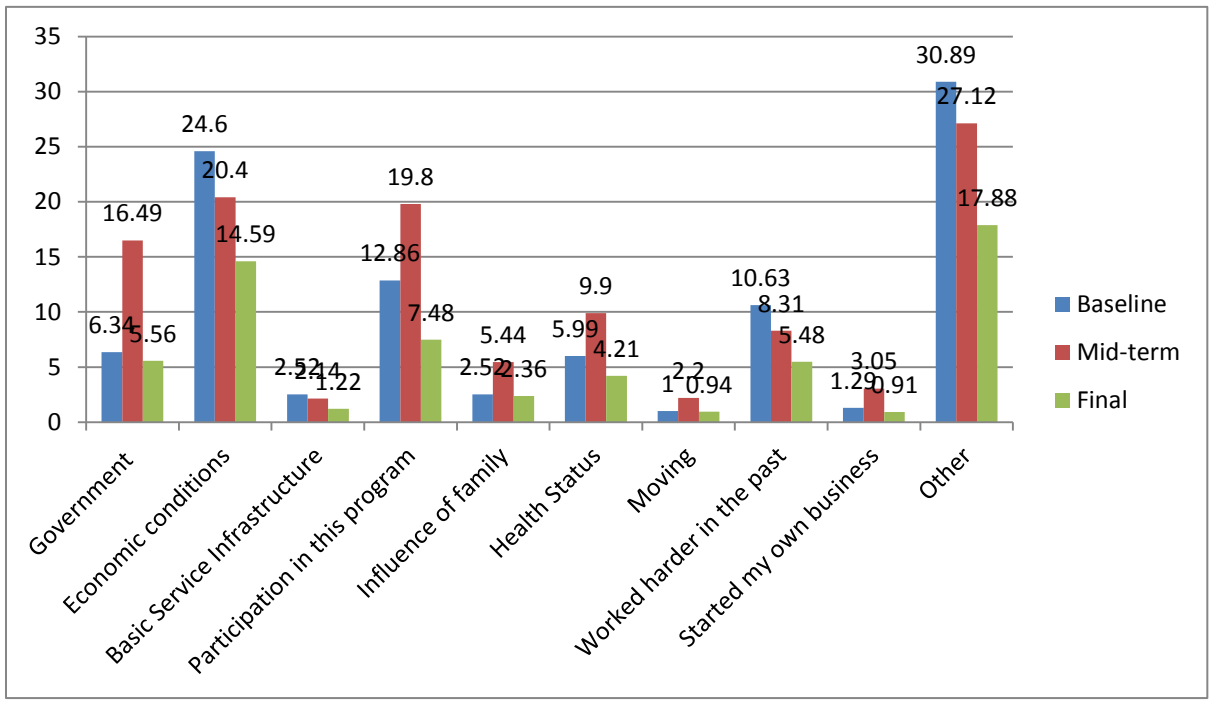
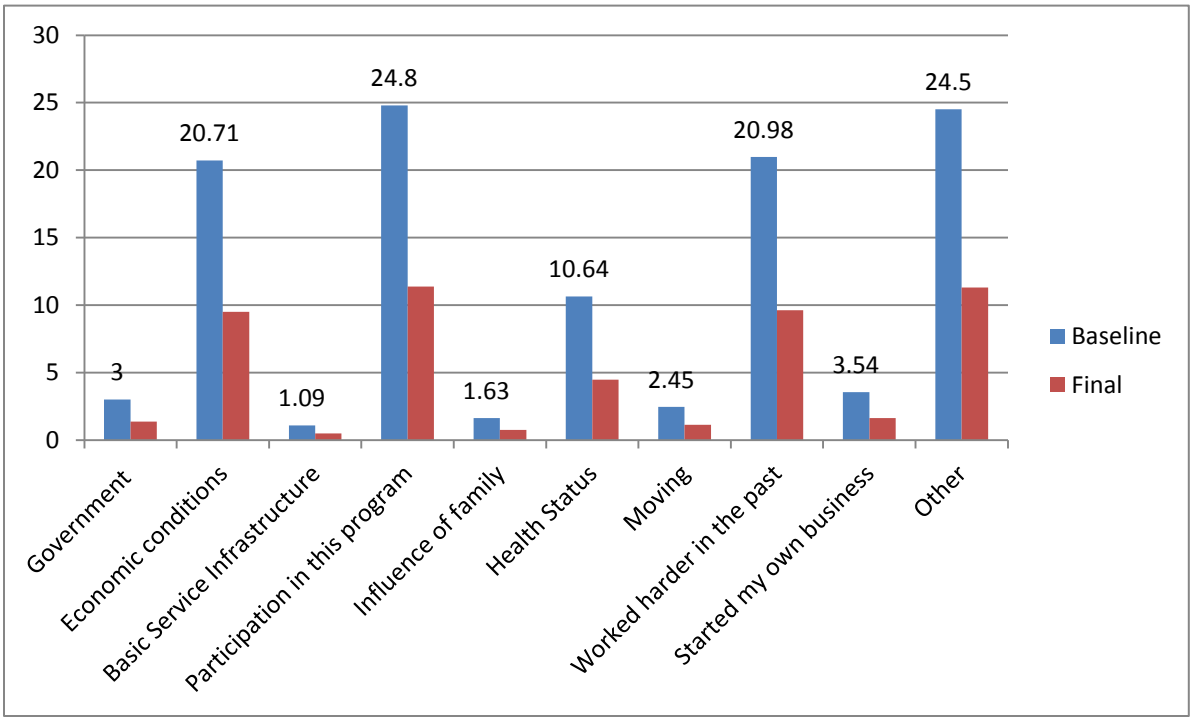


Figure 37: NFE Students' Views on What Might Have Caused Any Changes in Their Lives



14. Perceptions of Safety and Violence

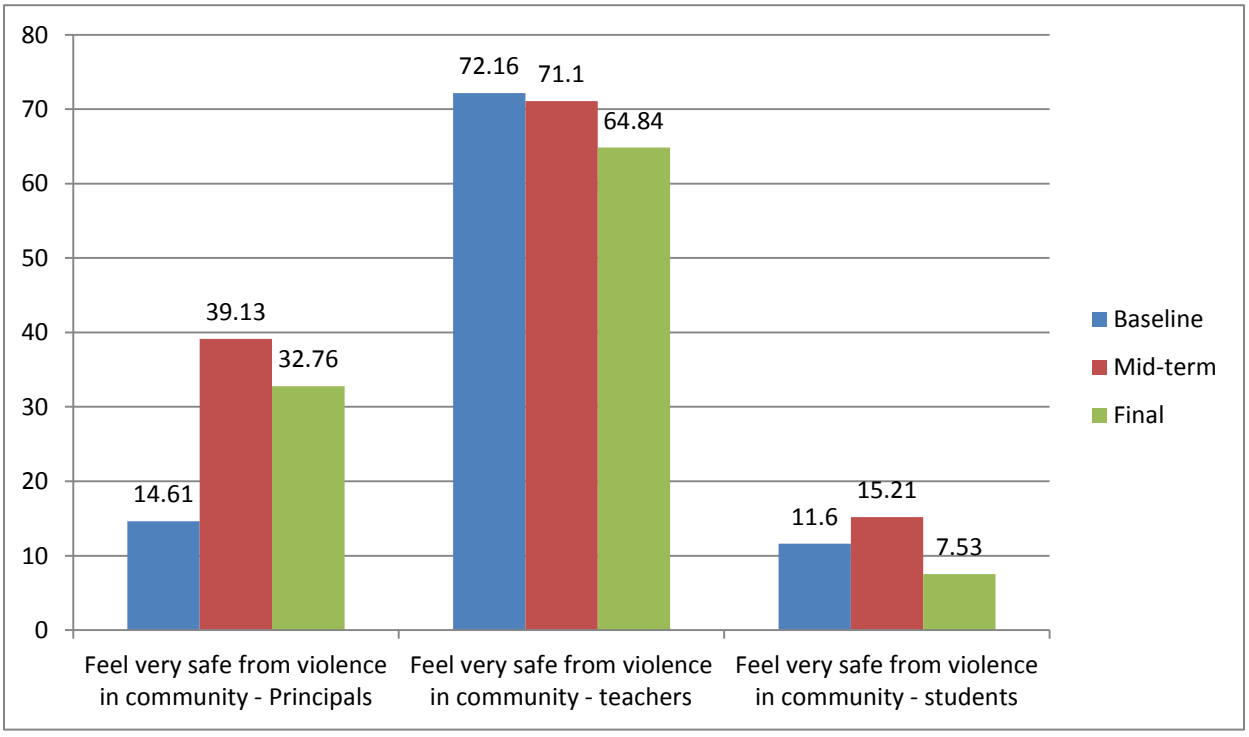
Respondents were asked to answer three questions related to safety and violence in their communities and schools:

1. "How safe do you feel with respect to violence in your community?" Although we reported only "very safe" answers, note that a significant number of respondents also said "safe," which is important.
2. "Do you think your students are safe from violence in your school?" We reported on the number of respondents who said "yes."
3. "Do you know of anyone who has been approached in a harassing way?" Again, we reported on the number of "yes" answers.

14.1 ALP Respondents

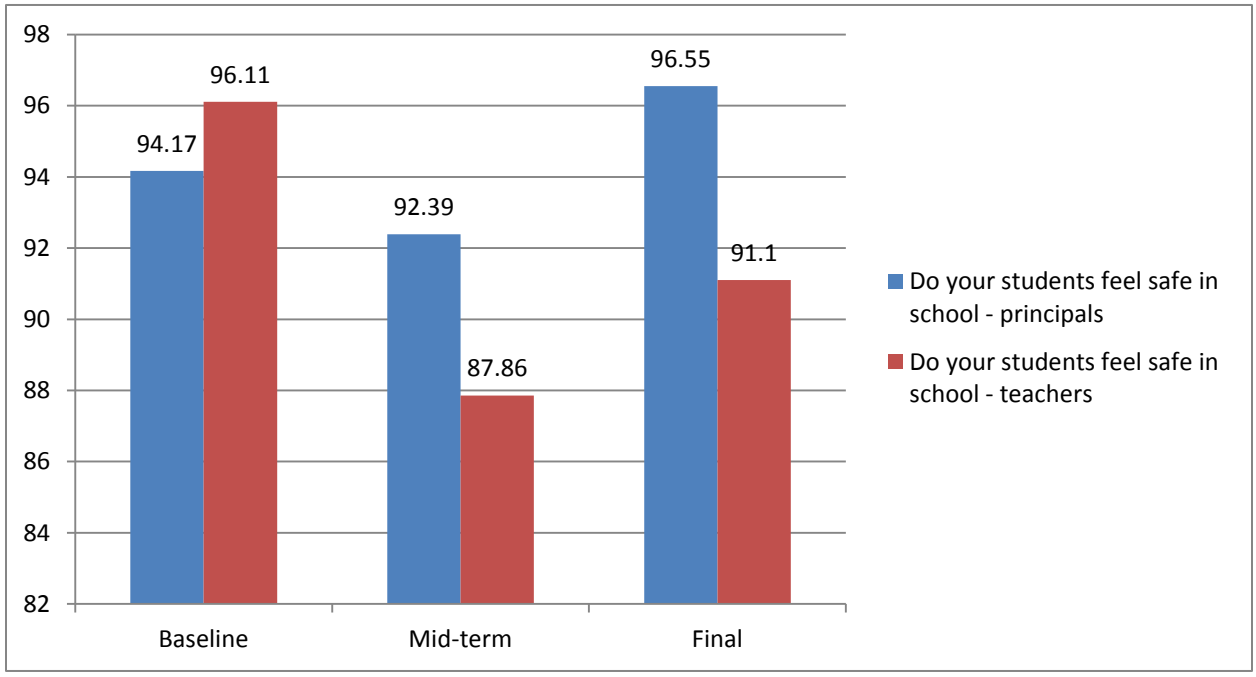
With respect to the first set of questions, ALP principals indicated that they felt safer at the final assessment than at the time of the baseline, but less safe than at the time of the midterm assessment (see ***Figure 38***). Teachers' responses were somewhat more consistent, but indicated a gradual decline, in their view, in safety in their community. Students' responses were also consistent, but they felt the least safe in their community, with rather low numbers responding that they felt "very safe." We also noted a decrease in this opinion between baseline and midterm and between midterm and final assessment.

Figure 38: ALP Principal, Teacher, and Student Views on Safety in their Community (% "yes")



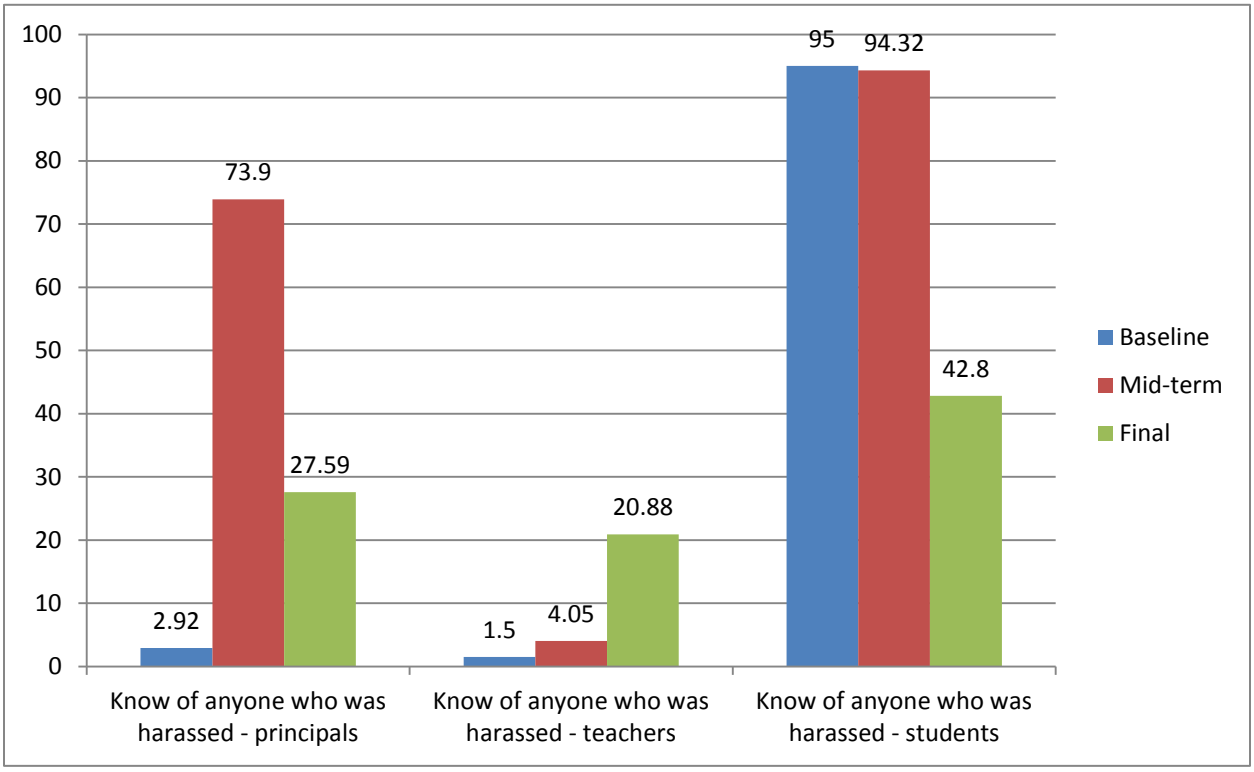
When asked about their opinion on the safety of learners in their schools, the principals' responses remained positive across all assessments and improved by a few percentage points over time. As for teachers, their responses also remained rather high, with slight but not significant decreases from baseline to midterm to final assessment (see ***Figure 39***).

Figure 39: ALP Principals' and Teachers' Views on Safety in School (% "yes")



ALP student data regarding their views on harassment incidents indicate a significant improvement across assessments—that is, a significant decrease in the number of students who knew of someone who had been approached in a harassing way. The numbers went down from about 90% to slightly more than 40%. As for the principal views on this point, the numbers reported went up significantly from baseline to midterm, but then dropped significantly at the final assessment. As for the teachers, unfortunately, we noted an increasing trend across assessments in the number of teachers who knew of someone being harassed (see ***Figure 40***).

Figure 40: ALP Principal, Teacher, and Student Views on Harassment (% "yes")



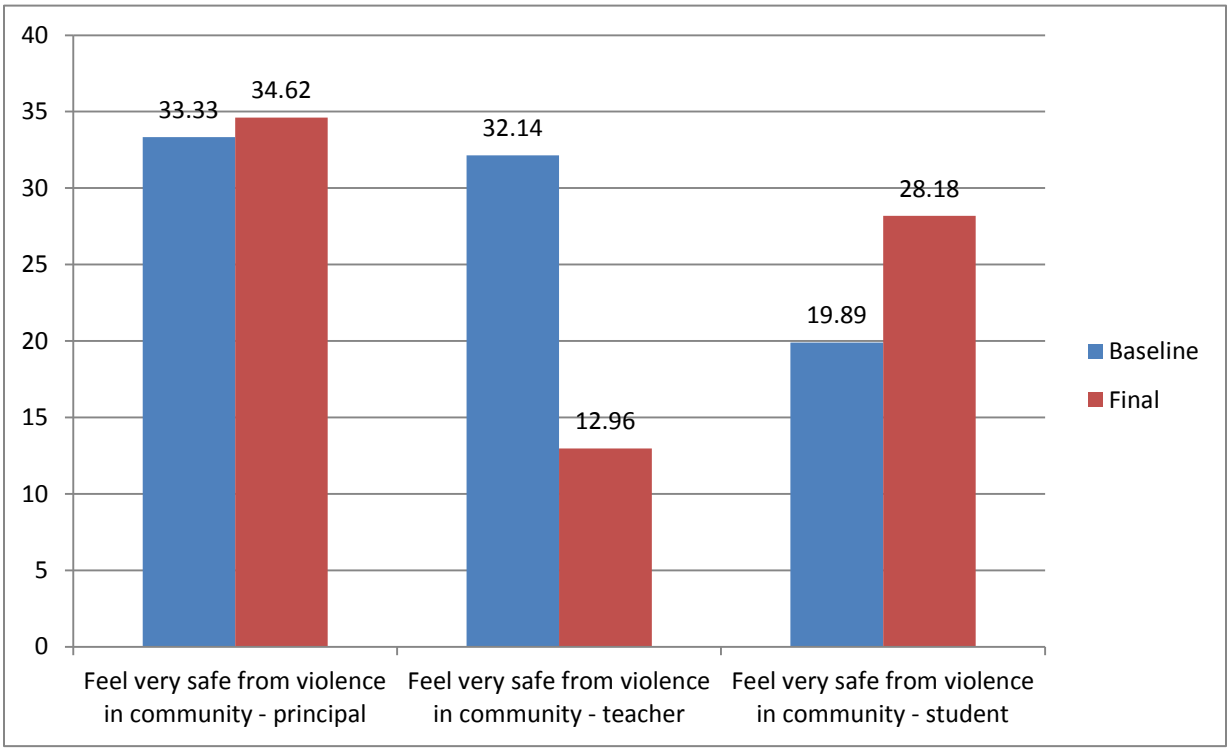
14.2 NFE Respondents

With respect to the NFE program, we noted more consistent views across the different respondents—principals, teachers, and students—on all three questions asked. For easier reference, we repeat the questions from the beginning of Section 14:

1. "How safe do you feel with respect to violence in your community?" Although we reported only "very safe" answers, note that a significant number of respondents also said "safe," which is important.
2. "Do you think your students are safe from violence in your school?" We reported on the number of respondents who said "yes."
3. "Do you know of anyone who has been approached in a harassing way?" Again, we reported on the number of "yes" answers.

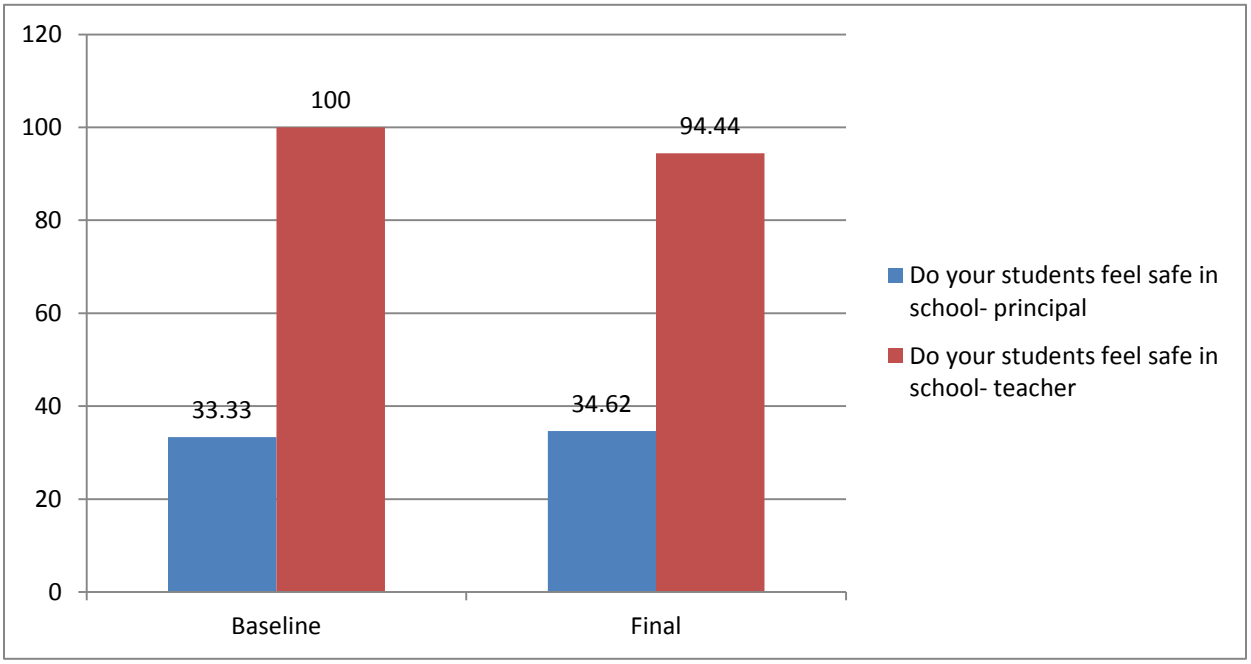
With respect to the first set of questions, a significant number of respondents said that they felt "safe" in their communities, but we chose to report on the "very safe" response since that indicates an absolute positive answer (see ***Figure 41***). There was no significant change between baseline and final assessment for principals. However, there was a decrease in the number of teachers who said they felt very safe in their communities. On the other hand, the number of students who said they felt very safe in their communities *increased.*

Figure 41: NFE Principal, Teacher, and Student Views on Safety in Their Community (% "yes")



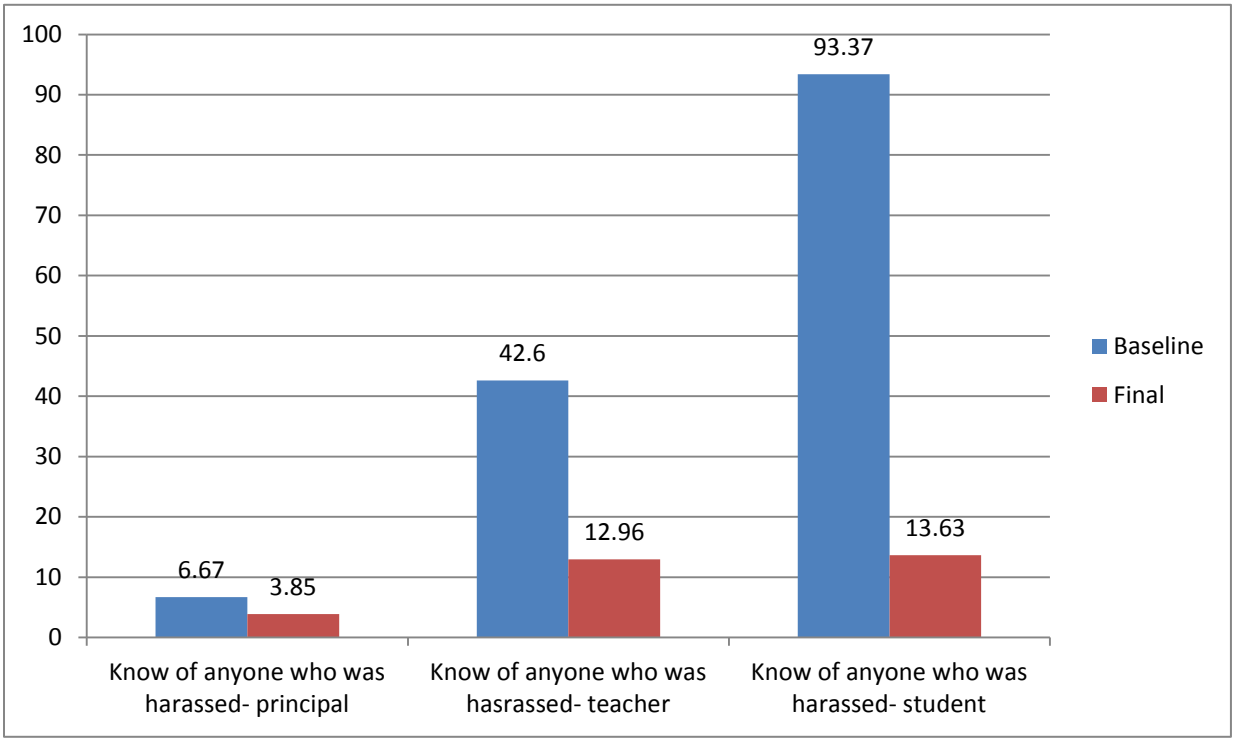
With respect to the safety of learners in their schools, as expressed by teachers and principals, we found no significant changes between baseline and final assessment (see ***Figure 42***). All of the teachers at the time of baseline felt that their students were safe, as opposed to 94.44% of them feeling this way at the time of the final assessment. The responses from principals registered a slight increase.

Figure 42: NFE Principal and Teacher Views on Safety in School (% "yes")



With respect to the views on harassment, this is a question where we saw significant decreases in the number of respondents who had heard of such incidents (see ***Figure 43***). The largest improvements were noted in the proportion of students who had heard of harassment incidents, which decreased from about 94% to about 14%. A similar trend occurred in the case of NFE teachers, as well as in the case of NFE principals.

Figure 43: NFE Principal, Teacher, and Student Views on Harassment (% "yes")



15. Recommendations

The CESLY program showed a positive impact on reading achievement. At the same time, it is not possible to determine the impact of the program as currently conceived in comparison to other types of reading interventions or standard instruction. Further, the NFE program was assessed after a short period of implementation, with three and a half months between the two testing periods, as well as final assessment occurring after only eight and a half months of class operation. Thus, NFE is not directly comparable to the ALP program, which represents a cumulative investment of three years in reading or language instruction. With these limitations in mind, the following recommendations are proposed:

- **Consider sustaining and scaling up the CESLY program.** In May 2011, the ALP program has been phased out and replaced with the NFE program. However, both ALP and NFE programs showed student learning increases in reading. Given the brief implementation period of NFE in particular, it stands to reason that a longer intervention would result in more reading gains. It is strongly recommended that the CESLY program focus its support toward teachers and schools, so that NFE learners acquire all of the basic skills, as tested by EGRA and EGMA, before the end of Level 1—that is, to learn how to read and perform basic mathematical functions before learners transition to the next level.
- **Improve female reading achievement.** The findings suggest that males outperformed females at baseline, and the gaps in achievement by sex increased over the course of the intervention. For both programs, additional emphasis should be placed on raising female achievement. During this project period, CESLY has already hired a Gender Specialist, launched retention campaigns for females, conducted female club strengthening, produced reading materials specifically oriented to females, and reoriented the curriculum to meet females' needs (e.g., by placing prevention of pregnancy and gender violence at the start of the lessons before females drop out). However, these and other efforts to strengthen female achievement should be continued and strengthened.
- **Focus on reading comprehension.** Students appeared to make marked improvements in several early literacy skills, such as letter identification. However, the percentage of students who successfully answered 80% of reading comprehension questions remained low. In the CESLY program, basic components of reading were emphasized in Year 1, while Year 2 focused on reading comprehension and practice of writing. In future programming, reading comprehension should be a focus of instruction from program inception, and it should be increasingly emphasized as basic reading skills are acquired. It is recommended that the CESLY program review the EGRA Plus: Liberia reading intervention, which provides clear scope and sequence as to how to approach reading. This reading intervention has proved to be successful in ensuring that students learn how to read in less than one year[9] and has been adjusted for Grades

[9] The effectiveness of the reading interventions developed by RTI are not only limited to Liberia, but also have proven successful in other countries like Kenya and South Africa.

1–3. It is currently being implemented through the Liberia Teacher Training Program: Phase 2.

- **Focus on mathematics.** Some improvements have been made on the numeracy front, but these are rather low, and more emphasis and focus needs to be given to the mathematics area. The NFE curriculum and support provided to teachers need to be more intense so that all learners acquire basic numeracy skills, as assessed by the EGMA tool, before the end of Level 1. The assessment results indicate that students can perform some tasks (number identification, simple additions and subtractions), but not all basic mathematical functions (multiplication, quantity discrimination, divisions) that are key foundational skills for complex mathematical functions that will come in subsequent years and curriculum.
- **Calibrate the curriculum.** Even though at the end of Level 1, NFE learners are expected to have rudimentary literacy knowledge according to the CESLY program, the assessment after only three months of intervention showed that important gains are possible. Reading and mathematics skills that were assessed using EGRA and EGMA are the most basic skills that anyone can learn, and learning them quickly is critical for future learning. It is recommended that the NFE curriculum and teacher training material be calibrated so that these skills are acquired before NFE learners transition to Level 2.
- **Conduct a quasi-experimental study of CESLY.** To fully gauge the impact of the ALP and NFE programs, a quasi-experimental study would be an ideal scenario. The lack of a control group means that it is not possible to accurately and entirely attribute the learning gains seen in CESLY participants to the program itself. Creation of a control group was considered at the beginning of the project; however, it was determined that it would not be possible to create one given the sampling population. For instance, all of the ALP schools had been receiving support for a few years before the CESLY program started, and they were to continue receiving support through the CESLY program. In the context of randomized control trials, this type of scenario leads to the "contamination" of the control group in case one would have been selected. In other words, no true control group would exist because there would be no schools that would not be receiving the intervention. The same issue was noted for NFE schools. All of the NFE schools were to receive the intervention, resulting in the inability to create control schools. After reviewing different possibilities, it was agreed in the end that measuring progress over baseline would be sufficient to gauge the impact of the project.
- **Consider further investigating the implementation of the mathematics intervention.** On more than half of all mathematics tasks, the percentage of zero scores (in other words, the number of students who could not complete the tasks) increased from baseline to final assessment, and those findings were statistically significant. On the others, zero scores decreased, and this, too, was statistically significant in nearly all cases. This finding suggests a certain amount of disequilibrium in the quality of the math intervention by subtask; explanations for this variation should be investigated, so that appropriate steps can be taken to improve mathematics achievement across the board. At the very least, the CESLY

program could investigate the following two possibilities. First, for the NFE program, it may be the case that the curriculum is calibrated in such a way that some of the EGMA-assessed tasks are taught in later levels, e.g., Level 2. This would not apply to the ALP students, because they should have already learned this material. Second, it would be very important to determine if teachers' content knowledge of some of these skills are at the level desired and needed.