A THREE-YEAR ACHIEVEMENT ANALYSIS
OF CHILDREN ENROLLED IN
THE FAMILY LITERACY INITIATIVE PROGRAM

December 2018

Prepared by:
Ronald E. Mertz, PhD
FOL Education Working Group

TABLE OF CONTENTS

Page

LIST OF TABLES

APPENDICIES

INTRODUCTION

This report summarizes test results of Liberian children with matched scores who were initially enrolled in the Family Literacy Initiative Program (FLI) in 2018, 2017, and 2016. The Family Literacy Initiative is a cooperative effort between The Friends of Liberia (FOL), HIPPY International (Home Instruction for Parents of Preschool Youngsters), and the WE-CARE Foundation, a Liberian literacy non-profit organization. WE-CARE provided local management and implementation, and HIPPY provided the three-year early childhood curriculum, instructional materials and staff training.

During the first year (2016), approximately 60 parents[1] were recruited in three communities (Caldwell, Duazon, and West Point) located in Montserrado and Margibi. For the second year (2017), the program recruited an additional group of approximately 20 parents from each of the three communities, while for the third year it added 10 new parents in each of the three original communities and 30 parents from a fourth community (Neezoe). Home visits were provided from January through August during each of the three years. In general, if families left the program during the instructional year, program staff recruited new families to replace them.

PROCEDURES

In order to obtain a standardized measure of children's school-related learning progress, the FOL Education Working Group (EWG) selected the *Bracken School Readiness Assessment, Third Edition.*[2] The 85-item test measures school readiness in five domains (colors, letters, numbers, sizes/comparisons, and shapes). Raw scores (the number of correct responses) are converted to scale scores and percentile ranks which can then be grouped into five descriptive categories - V*ery Delayed, Delayed, Average, Advanced, and Very Advanced.* These derived scores are based on a norming population in the United States. The age norms are in three-month intervals and range from 3 years to 6 years, 11 months. (See Appendix A for a more extended discussion regarding the test, data collection, scoring, and analysis that was provided in last year's report.)[3]

---

[1] The term "parents" sometimes includes grandparents or others if they had/shared primary responsibility for the child.

[2] The test, published by Pearson Corporation, was selected for several reasons. First, it does not require assessment professionals to administer; second, it provides age norms in three month intervals from three years to six years, eleven months; third, it has been used by a number of HIPPY programs in the US and was recommended by a staff member at the University of South Florida's HIPPY Training and Technical Assistance Center.

[3] Mertz, Ronald E. Achievement of Children Enrolled During the Second Year of the Family Literacy Initiative Program, October 2017.

Children in the analysis

Results in this report are based on matched scores. Children were first pretested shortly before or after the beginning of the instructional year and then shortly after the end of each year, typically in September. Therefore, the number of matched scores for each child used in the analysis ranged from two for children who were in FLI for the first time in 2018 to four for those who first entered in 2016. Children who left the program, entered late, or who were initially tested younger than 2 years, 10 months were not included in the analysis.

Three sets of matched score analyses were made, based on first-year enrollment - 2018, 2017, and 2016. In addition, first-year performance of children who initially enrolled in 2018 was compared to first-year performance of children who initially enrolled in 2017 and 2016.

One-year analysis
    Children initially enrolled in 2018 with matched pre/post scores: Total=53; 33 males, 20 females; average pretest age: 39.8 months

Two-year analysis
    Children initially enrolled in 2017 with pre/post, and 2018 scores: Total=48; 28 males, 20 females; average pretest age: 39.3 months

Three-year analysis
    Children initially enrolled in 2016 with pre/post, 2017, and 2018 scores. Total=35; 15 males, 20 females; average pretest age: 48 months

As indicated above, during the initial year of implementation in 2016, program staff recruited some children who were older than would typically be enrolled in the first-year curriculum resulting in an average pretest age of children in the three-year analysis of 48 months. During the second and third year, staff recruited children who were somewhat younger, resulting in an average pretest age of a little over 39 months.

FINDINGS

Finding 1. As in the previous two years, most children who entered the program in 2018 showed substantial progress at the end of their first program year.

Test results for children who had both pretest and posttest scores for their initial year in the Family Literacy Initiative Program over the last three years are provided in Table 1. As shown, most children experienced substantial growth during their first program year. (2018 results for all four communities are in Appendix B.)

Table 1. Comparison of first-year performance by children who entered in year one, two, and three.

Number of children in three descriptive classifications on the pretest	Percent of children scoring at three descriptive classification levels on the posttest		
	Very Delayed %	Delayed %	Average %
Year 1 Total N = 54			
Very Delayed (36)	20	47	33
Delayed (17)	0	41	59
Average (1)	-	-	100
Year 2 Total N = 57			
Very Delayed (32)	6	47	47
Delayed (24)	4	17	79
Average (1)	-	-	100
Year 3 Total N = 53			
Very Delayed (27)	7	52	41
Delayed (25)	4	44	52
Average (1)	-	-	100

 For example, 36 of 54 children who entered the program in year one (2016) had pretest scores that placed them in the *Very Delayed* descriptive classification category on the

pretest. However, on the posttest, only 20 percent of the 36 children had remained at this level, while 47 percent had progressed to *Delayed* and 33 percent to *Average.* During the same year, more than half of of the 17 children (50%) who had scored *Delayed* on the pretest reached the *Average* level on the posttest.

A similar pattern is evident for years two and three. In year two (2017), 32 of 57 children had scored *Very Delayed* on the pretest, but at the end of the year only two (6%) scored at that level while the others had reached *Delayed* (47%) or *Average* (47%). In year three (2018), about half of 53 children (27) had scored at the *Very Delayed* level on the pretest, but only two (about 7% of the 27 children) remained at this level on the posttest. in addition, among children in years two and three who scored *Delayed* on the pretest more than half reached the *Average* level on the posttest (79% and 52%).

Finding 2. While the majority of children who entered the program in 2016 showed substantial progress at the end of their initial year, many were not able to maintain their level of progress in relation to the norming sample in their second and third program year.

Although, as discussed under Finding 1, the majority of children showed substantial progress by the end of their first program year, analysis of matched scores for 35 children who entered in 2016 indicates that the majority of children have not been able to maintain their level of achievement in comparison to the norming population. For example, as shown in Table 2, while about two-thirds of the children had scored in the *Very Delayed* descriptive classification level on the the pretest, only 11 percent were at this level on the posttest. However, by the end of years two and three the percent of children scoring at this lowest level increased to 43 percent and 54 percent, respectively. Additionally, while 43 percent of the children had scored in the *Average* range on the 2016 posttest, the percent scoring *Average* declined to 20 percent in 2017 and 2018.

Table 2. Descriptive classification levels for 35 children with matched Bracken scores over a three-year period

Descriptive Classification	Test Dates			
	Pretest 2016 %	Posttest 2016 %	Summer 2017 %	Summer 2018 %
Average	3	43	20	20
Delayed	29	46	37	26
Very Delayed	69	11	43	54

Note: Total percents might not equal 100 due to rounding.

A decrease in performance relative to the norming sample does not mean that program children failed to make progress in learning those skills measured by the test. For example, as shown in Table 3, average raw scores among the 35 children with matched scores who entered in 2016 went from 9 to 38 after the first program year and then to 47 and 62 in years two and three, respectively. A similar pattern of raw score gains appears to be occurring for children who entered in 2017 and 2018.

Table 3. Average raw scores over the last three years for children with matched scores

Entry Year	Average pretest age (months)	Pretest	Posttest	Year 2	Year 3
2016 (35)	49	9	38	47	63
2017 (48)	39.3	6	30	45	_
2018 (53)	39.8	6	27	_	_

However, while most FLI children show progress (as measured by raw scores) from one year to the next, the level of progress does not keep up with that of the norming population. This can be illustrated in the following hypothetical example:

A child who is 3 years, 3 months old obtains a raw score of 3 on the pretest and is at the 1st percentile (Very Delayed) when tested before program entry. Then, the child, at age 4 years, one month is tested at the end of the first program year and obtains a raw score of 30 which is equivalent to the 21st percentile (Average). In order to remain at the 21st percentile (Average) the child must have a raw score of 49 at the end of the second year and 69 at the end of the third year.

Possible factors leading to this pattern are discussed below.

SUMMARY AND DISCUSSION

Over the last three years, the majority of FLI children have shown considerable progress at the end of the first program year as measured by the *Bracken School Readiness Assessment - Third Edition* in which performance of FLI children is compared to an American norming population using three-month interval age norms. During their second and third program year, however, many of the children do not maintain the same level of progress compared to the norming population. While most FLI children showed a pattern of growth in terms of raw scores (number correct) gains, the gains were less than those made by children in the norming population.

There are likely two major factors that are producing the patterns of growth indicated in this brief analysis. These are testing experience and the lack of comparable experiences by children in Liberia and children in the norming population.

As we have seen, FLI children typically score very low on the program entry "pretest" and then show considerable growth by the end of the first program year, only to regress in relation to the norming population in subsequent years.

There are a number of possible causes for very low pretest scores. One is that preschool age children in Liberia are less likely to have been exposed to variables such as colors, shapes, and letters than children in the norming population when they first take the test. This gap in exposure is likely substantially reduced the first year by their FLI experience, thereby increasing their performance relative to children in the norming sample.

Another possible factor is FLI children's lack of experience in a question/answer situation experienced during test administration. For example, American parents typically question their young children when they read to them or during other activities. It is also possible that some children will not be as open to interacting with a stranger, especially in a testing situation. This was most apparent during 2018 pretesting when six of eight children who scored zero on the pretest had made no attempt to answer any of the questions. All six of the children responded during the posttest and had raw scores ranging from 20 to 38 and descriptive classification levels of *Delayed* or *Average.*

A likely explanation for FLI children's test scores not increasing at a similar rate as that of the norming population is the lack of comparable quality formal learning opportunities. The HIPPY curriculum is designed to help compensate for the lack of a rich learning environment in the home and opportunities for formal learning experiences in preschool, kindergarten, and beyond. It is probable that these formal learning experiences were much greater for children in the American norming population, making it continuously more challenging for FLI children to meet or exceed performance by children in the norming population.

While monitoring FLI children's progress with the Bracken has been valuable in allowing the program to obtain a measure of children's progress in relation to a norming population, additional steps would contribute to providing evidence that the program is having its intended outcomes.

One possible step to be considered for the 2019 FLI program year would be to document and analyze parents' and children's progress throughout the year in completing and understanding HIPPY activities, and comparing that information to Bracken test results. It could be predicted, for example, that children who have completed a greater number of activities and have shown a higher level of understanding will show greater gains on the Bracken in comparison to their peers who have not completed as many activities and/or do not show the same level of understanding.

Another potential step would be to "pretest" a sample of three-year old children in the four communities or similar communities who are not served by the program and then test them at the end of the first and subsequent years. The expectation would be that the FLI children (treatment group) would outperform similar children who were tested, but not served (comparison group). While this study would not be as strong as one where children were tested and then randomly assigned to either a treatment (FLI) or a comparison group. It would, provide additional evidence for FLI's impact on learning outcomes that would go beyond the comparison of FLI children to the American norming population.

Now that the first group of FLI children has completed the three-year HIPPY curriculum, efforts should be made to follow FLI children's progress as they enter formal schooling and compare their performance to that of non-FLI children. This has been part of the original evaluation design prepared by the evaluator at the beginning of the program and updated in October 2015. As stated: *Student long-term (after curriculum year 5) outcomes will be assessed by obtaining information from schools of participating students and a comparison group, and will include the following:*

- *Student course grades*
- *Student attendance data*
- *Student deportment information*
- *Student promotion/retention information*[4]

Since the 2018-19 school year marks the first year that FLI children will have completed curriculum year 5, the FLI evaluation committee and program staff will pilot procedures for documenting these long-term outcomes. While it might not be possible to compare children in all four areas of data collection, efforts during this next year should provide a blueprint for 2020 and beyond.

---

[4] Mertz, Ron. Family Literacy Pilot Program-Liberia Evaluation Design, Updated 10-15-2015.

EXTENDED DISCUSSION OF PROCEDURES

The Test and Scores

The Bracken has a total of 85 items divided into five subtests. The subtests and number of items in each are: Colors (10), Letters (15), Numbers/Counting (18), Sizes/ Comparisons (22), and Shapes (20).

As discussed in the Examiner's Manual, several different scores can be generated from raw scores (percent correct).[5]   While the percent correct is useful, especially for classroom assessments, one purpose of a standardized test such as the Bracken is to see how children did on the test compared to others who took the test. Once a child enters formal schooling, the most common comparison is to other children in the same grade. However, for preschool children the most meaningful comparison is to children who are the same age. Therefore, the Bracken provides age norms in three-month intervals ranging from three years to six years, eleven months to convert raw scores into derived scores such as percentile ranks and standard scores.

Percentiles. The percentile indicates how children rank in comparison to children who took the test when it was given to a norming population. Since age norms are used, a percentile rank compares the child's performance on the test to children who are the same age (within a three-month interval). Percentile ranks typically range from 1 to 99. A score of, for example, 30 would indicate that the child scored higher than 30 percent of the children in the norming population.

Standard scores. Standard scores are derived from raw scores and range from 40 - 160 on the Bracken.

Descriptive classification categories. While percentiles are useful, the Bracken uses descriptive classification categories derived from standard scores to describe a child's rate of conceptual development. These categories, along with corresponding standard score and percentile ranges are shown below.

Very Delayed (standard scores 40-70; 2nd percentile or lower)
Delayed (standard scores 71-85; 3rd through16th percentile)
Average (standard scores 86-114; 18th through 82nd percentile)
Advanced (standard scores 115-129; 84th through 97th percentile)
Very Advanced (standard scores 130-160; 98th percentile or higher)
As indicated, most children in the norming population were in the *Average* category (18th through 82 percentile). Only the lowest two percent of children were in the *Very*

---

[5] Bracken, Bruce. *Bracken School Readiness Assessment, Third Edition*, Examiner's Manual. Pearson, San Antonio, Texas, 2007.

Delayed category and children who ranked between the 3rd and 16th percentile were in the *Delayed* category.

The importance of age norms

Since the Bracken is designed for testing children from three years to six years, 11 months and has different norms for every three month interval, it is important to understand that a child's performance on the test in terms of percentile rank and descriptive classification category is dependent on the child's age. For example, children three years to three years, two months who obtain a raw score of 23 on the test would have a percentile rank of 50 and would be classified as *Average*. However, children who are a year older with a raw score of 23 would rank at the eighth percentile and would be classified as *Delayed*, and those who were two years older with a raw score of 23 would rank in the first percentile and be classified as *Very Delayed*. Even over a shorter time period, such as the 30-week FLI instructional program, a child who made only small raw score gains could possibly show losses in terms of percentile rank or descriptive classification categories.

Test administration and scoring

The Bracken is individually administered. The child is shown a set of items in a *Stimulus Book* and its asked to point to the item designated by the examiner. Using the *Record Form*, the examiner records the child's answer as correct (1), incorrect (0), or "No response" (NR). After three consecutive incorrect responses in a subtest the examiner should move to the next subtest. When administration is completed, the examiner records the number of items answered correctly for each subtest.

For the FLI program, the record forms were scanned and then sent to the author as an email attachment. He reviewed each record form to check for observable recording and chronological age calculation errors. Using norms tables in the Examiner's Manual, he recorded the standard score, percentile rank, and descriptive classification for each child on the Record Form and then entered the information in a spreadsheet.

For each of the last three years, the program coordinator and assistant coordinator administered the test to children soon after they were recruited. Then, after the end of each program year, individuals recruited by WE-CARE administered the posttest. (In the first year, there were two examiners and in year two and three there were three).

Table 4. Bracken pre/post results for children who completed their first year in the Family Literacy Initiative Program in 2018.

Community (N)	Descriptive Classification Catetory	Number of children on pretest	Number of children scoring in three descriptive classification categories on posttest		
			Very Delayed	Delayed	Average
Caldwell (10)	Very Delayed	3	1	1	1
	Delayed	6	-	3	3
	Average	1	-	-	1
Duazon (10)	Very Delayed	5	1	3	1
	Delayed	5	1	2	2
	Average	0	-	-	-
Neezoe (27)	Very Delayed	14	0	8	6
	Delayed	13	0	5	8
	Average	0	-	-	-
Westpoint (6)	Very Delayed	5	-	2	3
	Delayed	1	-	1	-
	Average	0	-	-	-
Total (53)	Very Delayed	27	2	14	11
	Delayed	25	1	11	13
	Average	1	-	-	1