

Special thanks to the following :

- a. Our team of data collectors:

Alinafe Chanunkha, Bridget Chalemera, Mary Ngwata, Isaac Mlagha, Edward Chipeta, Nomsa Chirwa, Sankhani Chilikhuma, Kumbukani Zgambo, Sylvia Kasongo, Richard Kam'mayani, Madalitso Kalepa, Bernard Chigaya, Moffat Jamu, Christopher Kamala, Joe Zonda, Grace Seyani, Jean Mapala Phiri, Chikondi Kuntho, Cynthia Akuzike Phiri, Eunice Palika, Scolastica Zamayere, Emmanuel Nyamwera, Memory Masina, Christopher Ndembo, Pauline Chisale, Doreen Square, Mary Ngwata and Anaph Phiri.

- b. Our team of Implementers for arranging logistics

Ruth Kamphanje and Enala Kamanga

Table of Contents

Executive summary 3

Introduction 3

 Research questions 4

 Context 4

Methods 5

 Sampling 5

 Measurement 5

 Data collection 6

 Analysis 6

Attrition 6

IDELA Results 8

 IDELA Overview 8

 Motor development 9

 Emergent Literacy 10

 Emergent Math 11

 Social-emotional Development 12

 Approaches to learning 14

SEARS – PRE Results 14

Equity analysis 16

 Gender 16

 Changes by baseline IDELA scores 17

Conclusion 18

 Limitations 19

 Next Steps 20

Appendix A 21

Executive summary

This report evaluates the effect of Save the Children programming on early learning and development among pre-school children in 40 early childhood development centers (CBCCs) in Zomba, Malawi. The evaluation examines the effects of two programs: the Early Literacy and Mathematics (ELM) package and the Healing and Education through the Arts (HEART) program. 10 CBCCs received ELM programming, 10 received HEART programming, 10 received both ELM and HEART programming, and 10 served as a comparison group.

To measure results, data collectors administered Save the Children's International Development and Early Learning (IDELA) assessment with 553 children. In addition, a new tool—the Social Emotional Assets and Resilience Scales for Preschool (SEARS)—was administered to CBCC teachers (caregivers).

The results from the IDELA assessment revealed that, without exception, students in Save the Children treatment groups (i.e. those who benefited from ELM, HEART, or both programs) gained significantly more than children in the comparison group. However, among the different types of interventions, there were no significant differences in early learning and development gains. ELM, HEART, and ELM + HEART treatment groups all exhibited similar gains. This held true even when looking at the gains in the subdomains specifically targeted by each of the programs. Children in the ELM group did not gain more in early mathematics and literacy than those in the HEART group, and children in the HEART groups did not gain more in the social emotional domain.

The results from the SEARS-PRE were more ambiguous. With an extremely low correlation between baseline and endline scores, there was no meaningful change in the results and few conclusions can be drawn.

The results of the evaluation suggest that Save the Children programming had a large and positive impact on early learning and development gains. However, there are a number limitations to these findings. Comparison schools were somewhat geographically separated from treatment areas to avoid contamination. However, as areas were not assigned randomly, there may have been unmeasured and unaccounted for differences between treatment and control CBCCs. Similarly, differential attrition among the treatment groups indicates that results may be biased in unpredictable ways.

In general, the results present a favorable view of Save the Children programming and may be used as an advocacy point. Additional research into why there were no differences among treatment areas (especially the content areas which ELM and HEART target) should be conducted to clarify the findings of this report.

Introduction

This report reviews the results from an endline assessment conducted to assess the impact of Save the Children programming in Zomba, Malawi. Save the Children programming consisted of the Early Literacy and Mathematics (ELM) package and the Healing and Education through the Arts (HEART) program.

Research questions

The key research questions explored in this study include:

1. How has the sample of children changed over time?
 - a. Looking at baseline characteristics, are children who could not be found different than those who could be?
 - b. Did attrition differ depending on type of treatment or control status?
 - c. Among children who could be found at endline, do they differ in baseline characteristics depending on type of treatment or control status?
2. Did average learning and socioemotional outcomes differ for children in the treatment and control groups?
 - a. Do learning outcomes/gains differ for children exposed to any Save the Children treatment package (ELM, HEART, or ELM + HEART) compared to the comparison group?
 - b. Do learning outcomes/gains differ for children depending on the type of Save the Children treatment programming they received?
 - c. How do different Save the Children treatment programs impact subgroups of skills (e.g. cognitive, pre-literacy, fine motor) differently?
3. How are learning gains distributed depending on baseline characteristics and treatment groups?
 - a. Are there any differences in learning and development gains by gender?
 - b. Did the gap between more highly developed and less highly developed children change over time?

Context

The study was conducted with a total of 553 five-year-old children enrolled at 40 early childhood development centers (CBCCs) in Zomba, Malawi. All the CBCCs are within the Sponsorship ECCD impact area separated by Education Zones.

For the purposes of the study, CBCCs were assigned to one of four groups:

1. ELM Only (10 centers and 139 children)
2. HEART Only (10 centers and 135 children)
3. ELM + HEART (10 centers and 140 children)
4. Comparison (10 centers and 139 children)

All groups (including Comparison) received basic Save the Children quality preschool intervention training. This involved a 13-day pre-service training package of capacity building for volunteers, community mobilization, material development, workshops, supplies of standard play and learning materials, mentorship and coaching for caregivers, parent education, and interactive radio instructions.

Among the intervention groups, CBCCs in the “ELM Only” group received Save the Children’s Emergent Literacy and Mathematics program (ELM) package, which provides teachers with pedagogical training focused on improving early literacy and math skills and also includes supplemental materials. CBCCs in the “HEART Only” group received Save the Children’s Healing through the Arts (HEART) program, designed to improve social-emotional skills by incorporating arts into the preschool curriculum. The

“ELM + HEART” Group received both ELM and Heart programs. Finally, the Comparison group did not receive any Save the Children programs beyond the basic quality preschool education package.

Methods

Sampling

CBCCs were assigned to treatment and control groups in a purposive manner. The 30 treatment CBCCs were assigned to intervention groups in a random and even manner. While all 40 study groups were within the Sponsorship ECCD impact area, the 10 comparison CBCCs were drawn from an area slightly further from the treatment area to prevent contamination of the treatment into control areas. Overall, the control and treatment CBCCs were very similar in composition and both benefitted from the SC Sponsorship initiatives.

At baseline, 14 children (preferably five-year-olds) were randomly sampled from the registers at each sampled CBCC. A few smaller CBCC centers had fewer than 14 5-year-old children available, and included 4-year-olds. The sample represented the gender distribution at CBCCs and was quite even: 47% boys 53% girls. This reflects total enrollment at all CBCCs – slightly more girls than boys.

Data collectors attempted to reach these same children for endline data collection. Overall, 79% of the original children were found and interviewed. The endline assessment encompasses a total of 436 children.

Measurement

In this study two main tools were used: the IDELA Child Assessment with children and the SEARS-PRE questionnaire with CBCC Caregivers (teachers). The IDELA Child Assessment was used to measure children’s early learning and development and was administered by data collectors with direct child observation. Items included in IDELA are listed in Table 1. The SEARS-PRE assessment was adapted from the original work of N. Ravitch of the University of Oregon and contains 25 items related to the social-emotional assets that children possess. The SEARS-PRE was administered in an interview where CBCC Caregivers were asked about the children in their class and reported the level of children’s behaviors on a 3-point scale. The same versions of both tools were used at baseline and endline.

Table 1. IDELA Child Assessment

Gross and Fine Motor Skills	Emergent Literacy	Emergent Math	Socio-emotional Development	Other items
Hopping	Print awareness	Size/length identification	Friends	Approaches to learning
Copying a shape	Expressive vocabulary	Sorting	Recognizing emotions in self	Inhibitory control
Drawing a human figure	Letter identification	Number identification	Recognizing emotions in others	Short term memory
Folding paper	Emergent writing	Shape identification	Conflict resolution	

Phonemic awareness	One-to-one correspondence	Personal information
Oral comprehension	Simple operations	
Puzzle completion		

Data collection

18 data collectors were involved in conducting the IDELA Child Assessment and SEARS-PRE teacher interviews at both baseline and endline, for a total of 28 data collectors (many participated in both baseline and endline). The data collectors were trained on IDELA and ELM parenting tool for three days. The training was facilitated by the technical team at Field Office where the data collectors were trained on the tool, practiced among each other, and finally conducted a live pretest in a selected school outside of the study. Data collectors used tablets with Tangerine software to collect and upload data to the cloud periodically. The data collection took three weeks.

Analysis

The primary purpose of this analysis is to investigate the children’s learning and development gains after one academic year (9 months of CBCC classes) and estimate the impact of treatments of ELM Only, HEART Only, and ELM + HEART treatments through quasi-experimental means.

First, an analysis of the comparability of learners in the treatment groups and comparison groups after attrition is presented. Then this report presents summary statistics on children’s overall and subtask scores on the IDELA and general results from the SEARS-PRE. To test whether or not groups exhibited significant differences in their changes over time, this report utilizes a difference-in-difference approach with regression analysis. It controls for differences in gender, age, and baseline scores and assumes unequal variance by clustering errors by CBCC center. Finally, this report will use additional multivariate regression models with interactions to explore the relationships between children’s developmental gains and baseline developmental levels.

Attrition

Of the 553 children surveyed at baseline, data collectors attempted to reach the same children for the endline assessment. 21% (117) of children could not be reached for a variety of reasons. Despite the high rate of attrition, there were few significant differences between the children who were found at the follow-up assessment and those who were not in terms of baseline characteristics. Children who were not found had similar IDELA and SEARS-PRE scores at baseline, were no more likely to be male or female, and were of a similar age. However, in the absence of a household survey, background data on children was relatively limited and therefore we are limited in our confidence of comparability of the children at baseline and endline.

In addition, there was a significant difference in attrition depending on treatment group. While the “ELM only”, and “HEART only”, and “Comparison” groups all had similar rates of attrition (20-26%), only 13% of children from the ELM & HEART treatment group were unable to be found at endline.

Table 3. Sample attrition

	Sample at baseline	# Children missing at endline	% Children missing at endline
Control	139	36	26%
HEART only	135	35	26%
ELM only	139	28	20%
ELM & HEART	140	18	13%
Total	553	117	21%

There were also differences in the reasons given for the child’s absence at endline by treatment group. In particular, a much larger percentage of children in the ELM only group were promoted to Standard 1. Conversely, a higher percentage of children in the HEART only group transferred out of their center.

Table 4. Sample attrition by reason

Reason	Control	HEART only	ELM only	ELM & HEART	TOTAL
Promoted to Standard 1	0	1	15	0	16
Sick	2	5	1	8	16
Transferred out	16	20	14	8	58
Other	18	2	5	2	27
Total	36	28	35	18	117

The following analyses look only at children for whom there is both baseline and endline data. The differences discussed above introduce some doubts about the comparability of students in the various treatment groups. Despite being similar in composition of age, sex, and baseline scores, the children who remain in each treatment groups might differ in important and unobserved ways. This differential attrition may bias findings in unpredictable ways. These differences do not necessarily invalidate later results, but are an important limitation the findings.

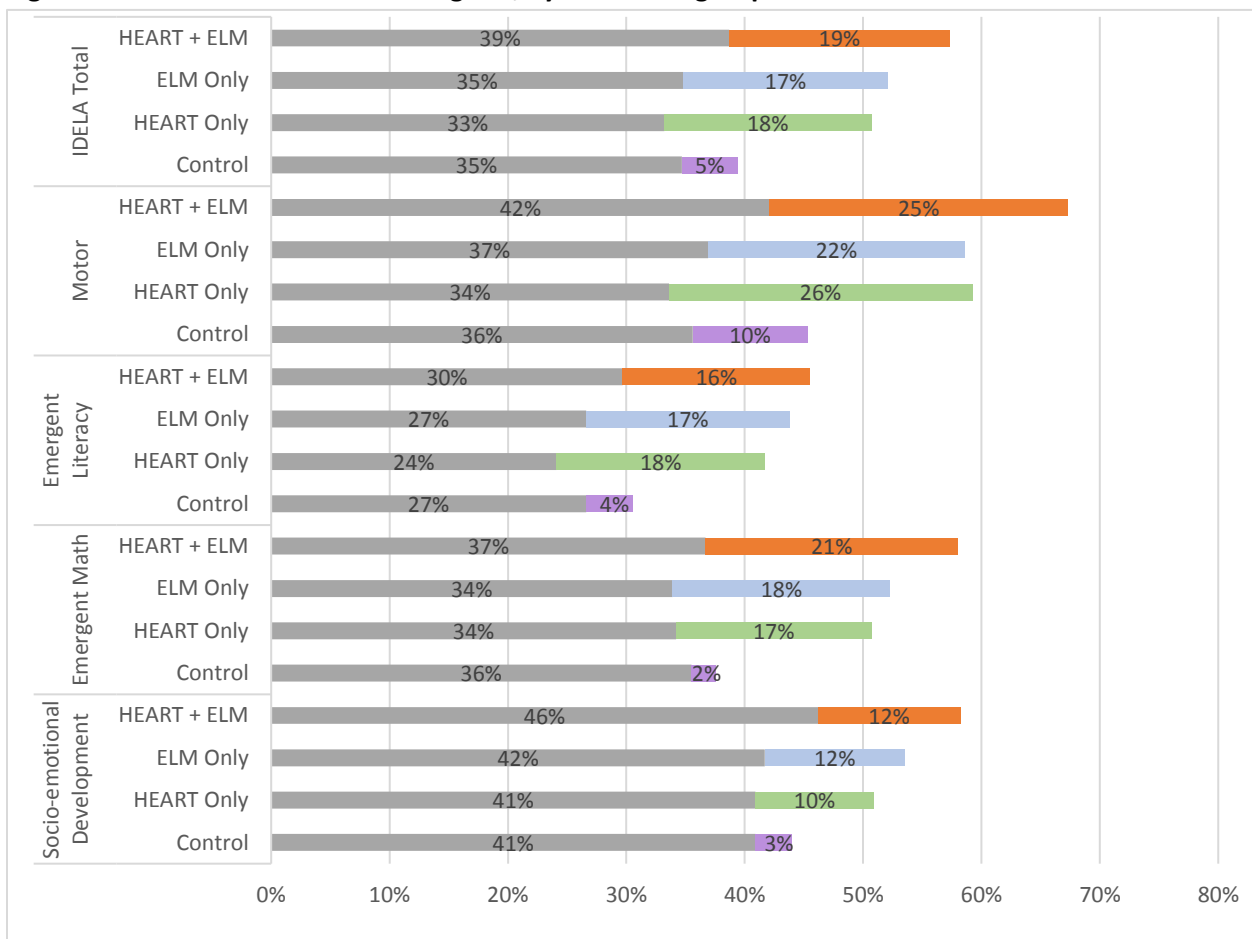
IDELA Results

This section details changes in children’s learning and development as measured by direct child assessment with the IDELA child assessment. IDELA items are organized into 4 categories: motor development, emergent literacy, emergent math, and social-emotional development. These domain items are then weighted evenly and added together to create the total IDELA score. Assessor observation items are not included in the total IDELA score because they are not a direct assessment of children’s skills, but help provide a more holistic picture of children’s early learning and development. After a brief overview of the general findings, we present a breakdown of the findings by domain.

IDELA Overview

The baseline assessment revealed that children in all treatment groups were similar in terms of learning and development; there were no significant differences between children on any of these domains. As can be seen in Figure 1, there were much larger gains in all domains among children in the treatment groups compared with the comparison group. These differences were highly significant. However, there were no significant differences in gains between children among the different treatment groups. The gains among children in ELM only, HEART only, and HEART + ELM were similar for all domains.

Figure 1. Overview of IDELA domain gains, by treatment group



Motor development

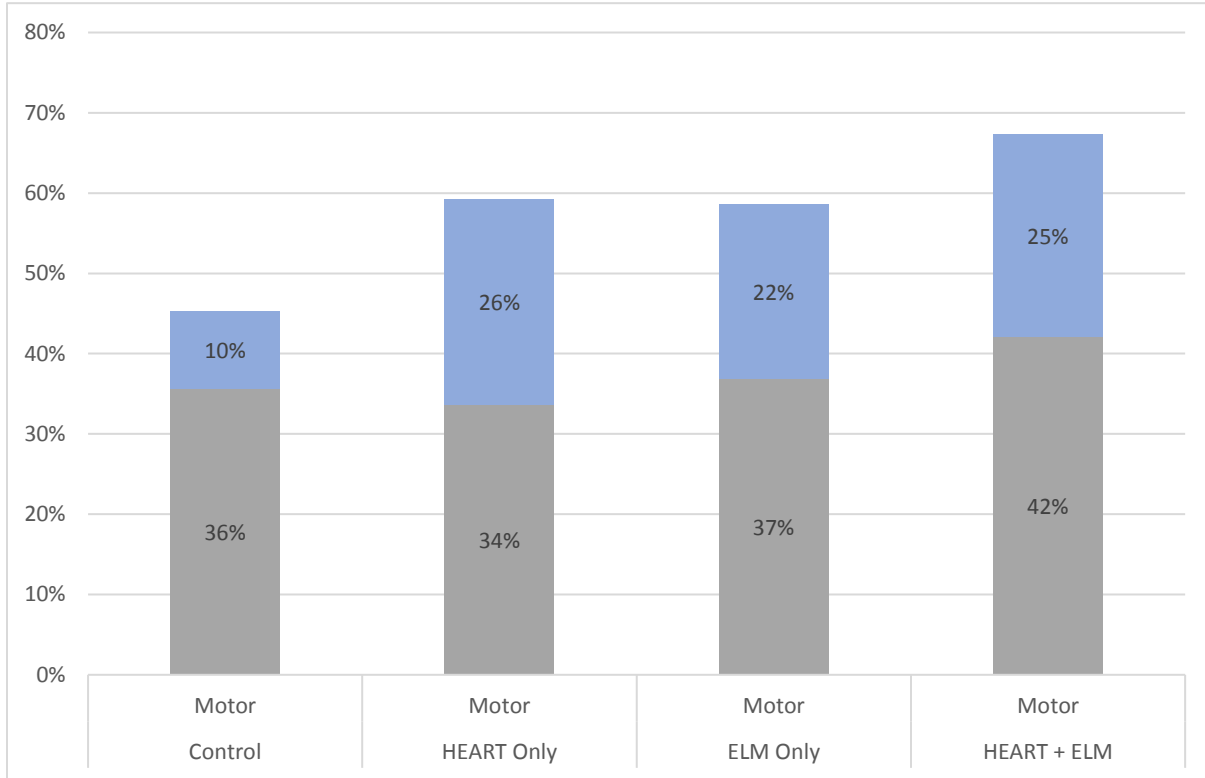
Table 5 displays children’s average motor development at baseline and endline. Figure 2 displays the gains in total motor development by treatment groups. **On the total scale of motor development, children in the ELM only, HEART only, and HEART + ELM treatment groups gained significantly more than children in the comparison group. Among the treatment groups, there were no significant differences between gains made by children in the ELM only, HEART only, and HEART + ELM treatment groups.**

Looking at individual items, children in the ELM only, HEART only, and HEART + ELM treatment groups scored significantly higher than children in the comparison group on the drawing a person and copying a shape subtask. However, only children in the HEART only and HEART + ELM treatment group scored significantly higher on the folding subtask. There were no significant differences between gains made by boys and girls.

Table 5. Motor development, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Drawing a person	35%	46%	26%	59%	33%	58%	40%	64%
Folding	40%	41%	41%	53%	41%	45%	44%	60%
Copying a shape	33%	49%	34%	66%	37%	74%	43%	79%
Total Motor Development	36%	45%	34%	59%	37%	59%	42%	67%

Figure 2. Motor development gains, by treatment group



Emergent Literacy

Table 6 displays children’s emergent literacy skills at baseline and endline. For all treatment groups, the letter ID and phonemic awareness subtasks proved most challenging. Children scored highest on the print awareness and oral comprehension subtask. Figure 3 displays the gains in emergent literacy by treatment group. **On the total scale of emergent literacy, children in the ELM only, HEART only, and HEART + ELM treatment groups all scored substantially higher and gained significantly more than children in the control group. There were no significant differences between gains made by children in the ELM only, HEART only, and HEART + ELM treatment groups.**

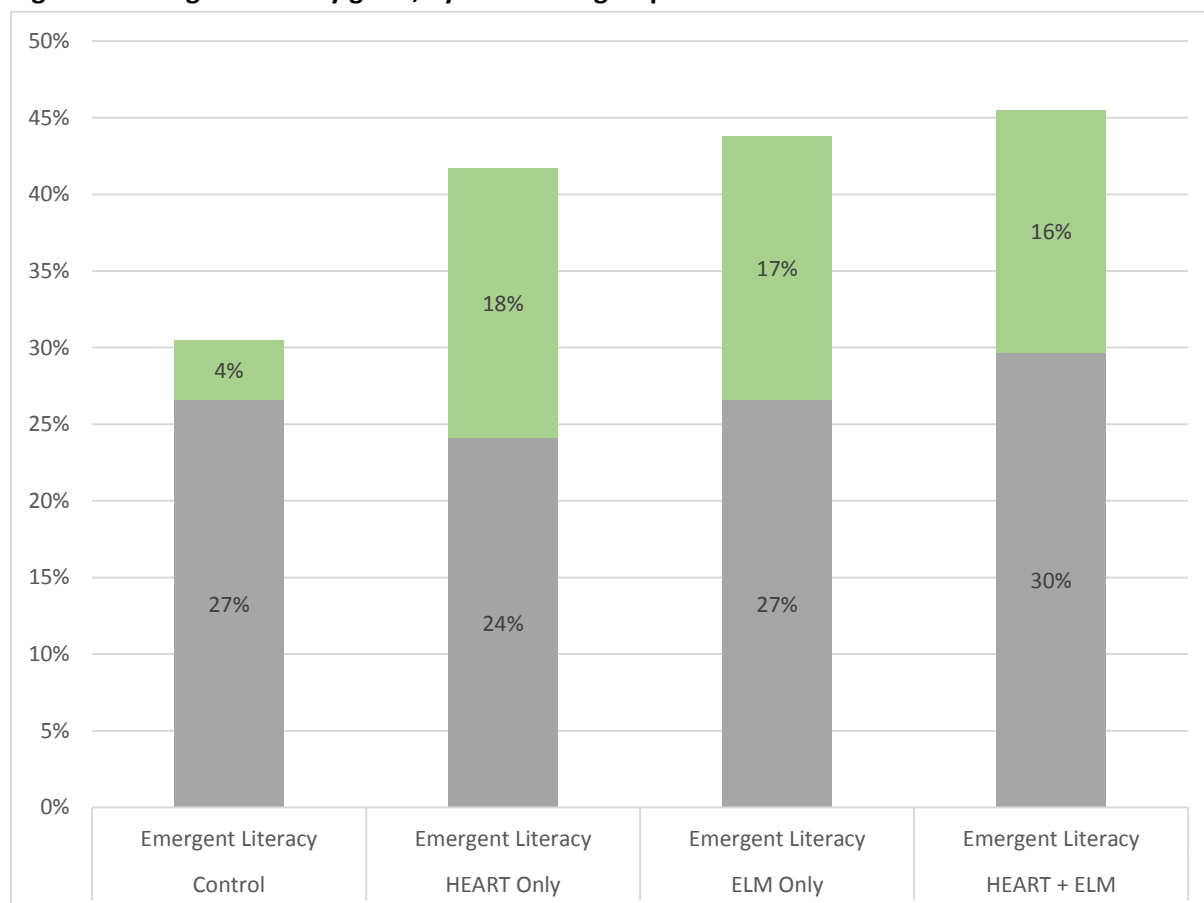
Looking at individual items, children in all of the treatment groups (ELM only, HEART only, and HEART + ELM) gained significantly more on the print awareness, letter ID, expressive vocabulary, oral comprehension and phonemic awareness subtasks than children in the comparison group. However, only children in the ELM + HEART treatment group gained significantly more than the comparison group on the writing subtask.

Table 6. Emergent Literacy, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Print awareness	49%	48%	47%	62%	50%	71%	56%	68%
Letter ID	1%	5%	2%	14%	1%	11%	2%	9%
Expressive vocabulary	28%	28%	24%	37%	28%	42%	32%	41%
Oral comprehension	46%	52%	41%	63%	45%	65%	48%	70%

Phonemic awareness	8%	11%	6%	30%	7%	28%	9%	27%
Writing	28%	39%	25%	44%	29%	47%	32%	58%
Total Emergent Literacy	27%	31%	24%	42%	27%	44%	30%	46%

Figure 3. Emergent Literacy gains, by treatment group



Emergent Math

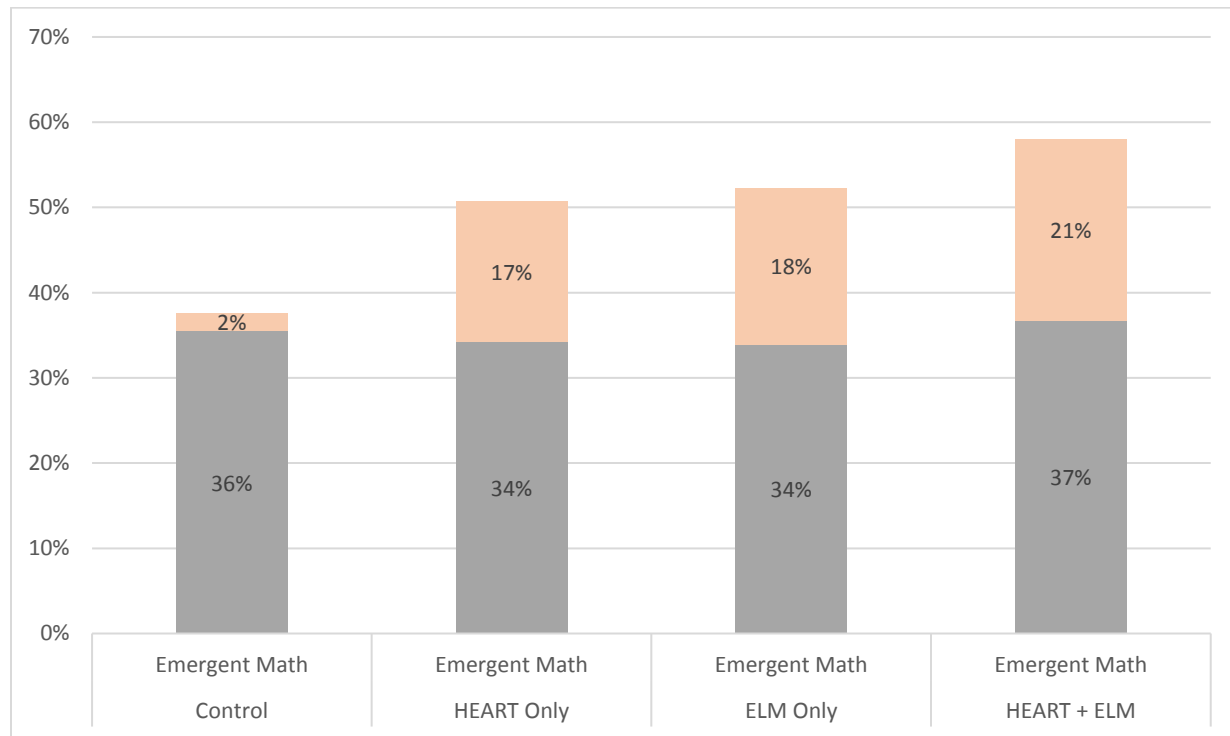
Table 7 displays children’s emergent math skills at baseline and endline. For all treatment groups, children scored lowest on the number ID subtask and highest on the size/length subtask. Figure 4 displays the gains in emergent math by treatment groups. **On the total scale of emergent math, children in the ELM only, HEART only, and HEART + ELM treatment groups all scored substantially higher and gained significantly more than children in the comparison group. There were no significant differences between gains made by children in the ELM only, HEART only, and HEART + ELM treatment groups.**

Looking at individual items, children in all of the treatment groups (ELM only, HEART only, and HEART + ELM) gained significantly more on all of the subtasks than children in the comparison group except for size/length. For size/length only children in the ELM treatment gained significantly more on average.

Table 7. Emergent Math, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Size/length	78%	80%	85%	90%	76%	87%	83%	90%
Number ID	2%	9%	4%	20%	3%	15%	3%	21%
One-to-one correspondence	28%	35%	20%	49%	24%	51%	25%	61%
Simple operations	48%	53%	40%	67%	45%	69%	55%	71%
Shape ID	26%	30%	27%	41%	29%	56%	27%	59%
Sorting	50%	42%	46%	59%	46%	58%	46%	64%
Puzzle	18%	15%	18%	29%	15%	31%	19%	42%
Total Emergent Math	36%	38%	34%	51%	34%	52%	37%	58%

Figure 4. Emergent Math gains, by treatment group



Social-emotional Development

Table 8 displays children’s social-emotional development at baseline and endline. Figure 5 displays the gains in social-emotional development by treatment group. **On the total scale of social-emotional development, children in the ELM only, HEART only, and HEART + ELM treatment groups all scored substantially higher and gained significantly more than children in the comparison group. There were no significant differences between gains made by children in the ELM only, HEART only, and HEART + ELM treatment groups.**

Looking at individual items, there was more variation than in the subtasks than in other domains. Children in the HEART only group gained significantly more on social connections and preferences recognition than children in the comparison group. Children in the ELM only group gained significantly more on the social connections and conflict resolution subtasks. Children in the ELM + HEART group gained significantly more than the comparison group only in the conflict resolution subtask, but had a large 24 percentage point gain.

Table 8. Socio-emotional development, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Social connections	48%	46%	41%	55%	47%	57%	51%	57%
Empathy	33%	51%	42%	59%	40%	61%	49%	72%
Conflict resolution	34%	33%	32%	42%	31%	45%	32%	56%
Personal information	49%	55%	48%	58%	49%	61%	53%	64%
Strengths recognition	19%	26%	20%	28%	20%	30%	21%	33%
Preferences recognition	54%	54%	47%	63%	58%	67%	63%	68%
Total Socio-emotional Development	41%	44%	41%	51%	42%	54%	46%	58%

Figure 5. Social-emotional Development gains, by treatment group



Approaches to learning

Finally, a number of assessor-rated items are included in the IDELA to measure the way in which children approach learning and problem solving. Table 9 reports the average baseline and endline measure on the changes between baseline and endline in persistence and approaches to learning.

At baseline, on average, children in the HEART treatment group had significantly lower approaches to learning than other children. At endline, this was reversed. Oddly, children in the comparison group appeared to regress on this domain. Among treatment groups, children in the HEART gained the most but these gains were not statistically different those in the ELM only and HEART + ELM groups.

Table 9. Approaches to learning, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline
Persistence during assessment	89%	76%	76%	87%	85%	87%	89%	91%
Observation after assessment	72%	61%	65%	71%	69%	69%	73%	74%
Total Approaches to Learning	81%	68%	71%	79%	77%	78%	81%	82%

SEARS – PRE Results

Teachers at each CBCC were also asked to respond to questions regarding their students' social-emotional wellbeing and assets through the SEARS-PRE. Parents reported whether their children rarely (coded as 1), sometimes (coded as 2) or often (coded as 3) displayed the target skill or behavior. On average, children were rated most highly on the items related to whether or not they were liked and accepted by other children and sharing their toys. They scored lowest on the ability to disagree with other without fighting. Table 10 displays the contents of all items and the baseline level and average gains by treatment group.

At baseline, there were no statistically significant differences between the SEARS-PRE results of children in different treatment groups. On average, children in the HEART and HEART + ELM treatment groups gained 5 percentage points from baseline to endline compared with 0 and 2 p.p. for comparison and ELM only groups, but this result was not statistically significant.

Table 10. Average SEARS-PRE scores, by treatment group

	Control		HEART only		ELM only		ELM & HEART	
	Base line	Gain	Baseline	Gain	Baseline	Gain	Baseline	Gain
Smiles and laughs when playing with other children (or with adults)	1.85	0.00	1.73	0.10	1.83	0.02	1.70	0.17
Says nice things to others or celebrates other's accomplishments	2.16	0.07	2.20	0.37	2.27	0.25	2.35	0.19
Disagrees with other people without fighting or arguing	2.02	0.11	1.93	0.16	2.14	-0.05	2.02	0.15
Responds appropriately to other people's facial expressions	2.32	-0.05	2.22	0.19	2.27	0.14	2.21	0.19
Is accepted and liked by other children	2.67	0.05	2.55	0.02	2.63	0.07	2.65	0.05
Adjusts well to new teachers or caregivers	2.50	0.04	2.40	0.20	2.63	0.08	2.55	0.11
Remains calm in disappointing situations	2.28	-0.15	1.99	0.15	2.16	0.05	2.14	0.11
Shows responsibility for things around her/him	2.36	0.02	2.42	0.14	2.50	0.09	2.34	0.25
Stands up for herself or himself	2.14	0.20	2.29	0.09	2.39	0.01	2.25	0.26
Uses words to solve problems	2.10	-0.16	2.10	-0.01	1.93	0.30	1.94	0.31
Uses clear communication when describing events	2.29	-0.04	1.98	0.48	2.32	0.09	2.24	0.27
Helps others solve problems	1.81	0.25	1.92	0.10	1.99	0.11	2.08	-0.07
Comforts other children who are upset	2.16	0.01	2.21	-0.02	2.21	-0.04	2.04	0.21
Calms down easily after being upset	2.27	0.13	2.28	0.15	2.37	0.13	2.35	0.12
Takes turns	2.45	0.03	2.34	0.11	2.38	0.18	2.45	0.16
Initiates play with others easily	2.52	0.02	2.34	0.28	2.50	0.16	2.60	0.08
Tells adults how he/she feels	2.37	-0.04	2.14	0.31	2.16	0.29	2.38	0.09
Is kind to others	2.45	0.04	2.45	-0.01	2.56	-0.02	2.51	0.02
Shares toys and other belongings	2.67	-0.32	2.36	0.18	2.54	0.11	2.61	0.07

Asks for help when necessary	2.28	-0.14	2.31	0.07	2.26	0.01	2.24	0.11
Tries to understand how others feel when they are angry, upset, or sad	2.30	-0.22	2.03	0.21	2.39	-0.24	2.12	0.11
Is interested and engaged in classroom activities	2.65	-0.07	2.52	0.15	2.52	0.11	2.45	0.27
Can express feelings/ emotions through words	2.26	0.09	2.20	0.20	2.41	0.05	2.43	0.18
Can express feelings/emotions/ experiences through arts activities	2.19	-0.06	2.24	0.08	2.19	-0.01	2.26	0.20
Responds appropriately to questions asked by other children	2.47	-0.08	2.26	0.16	2.44	-0.08	2.23	0.35
Total SEARS-PRE (out of 75)	57.5	-0.26	55.39	3.87	57.96	1.81	57.13	3.95
% SEARS-PRE	77%	0%	74%	5%	77%	2%	76%	5%

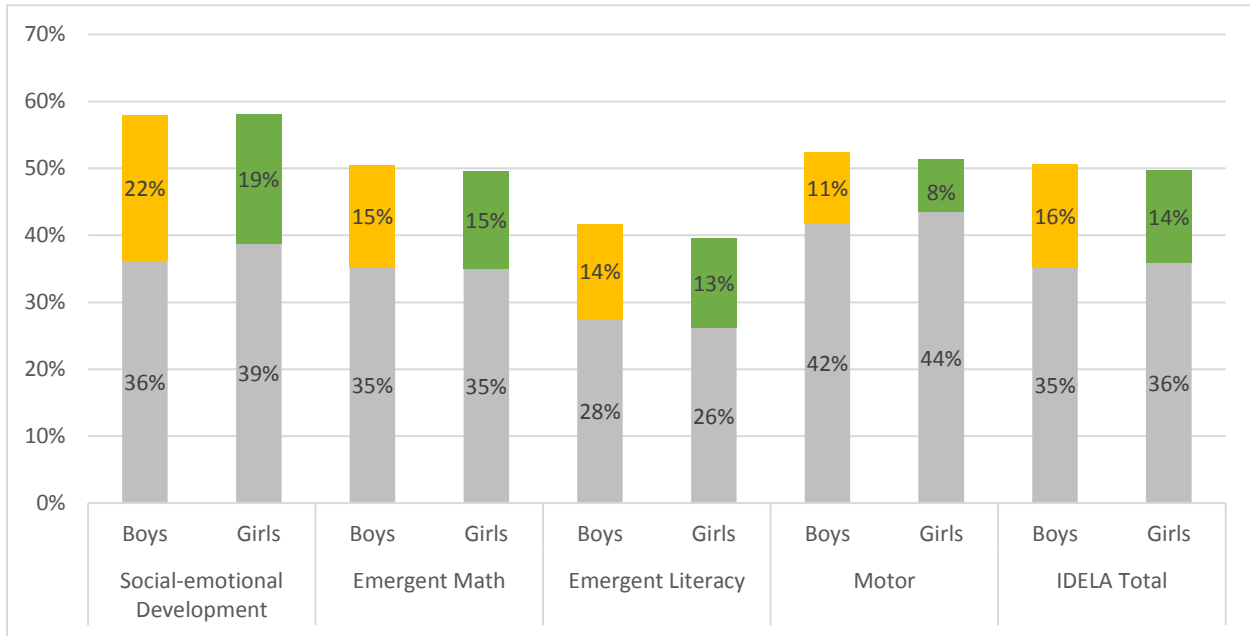
Equity analysis

Beyond simply examining how learning and developmental outcomes vary by treatment group, we also examine how children fared based on gender and different baseline scores. We use multivariate regressions to investigate whether or not gains in development between baseline and endline are evenly distributed among all children or if some children gained more than others.

Gender

There were no significant differences at either baseline or endline by student gender. In general, boys began slightly lower than girls and gained slightly more, but these differences were not significant. Figure 6 displays average baseline and gain scores by domain and gender.

Figure 6. Overview of IDELA domain gains, by gender



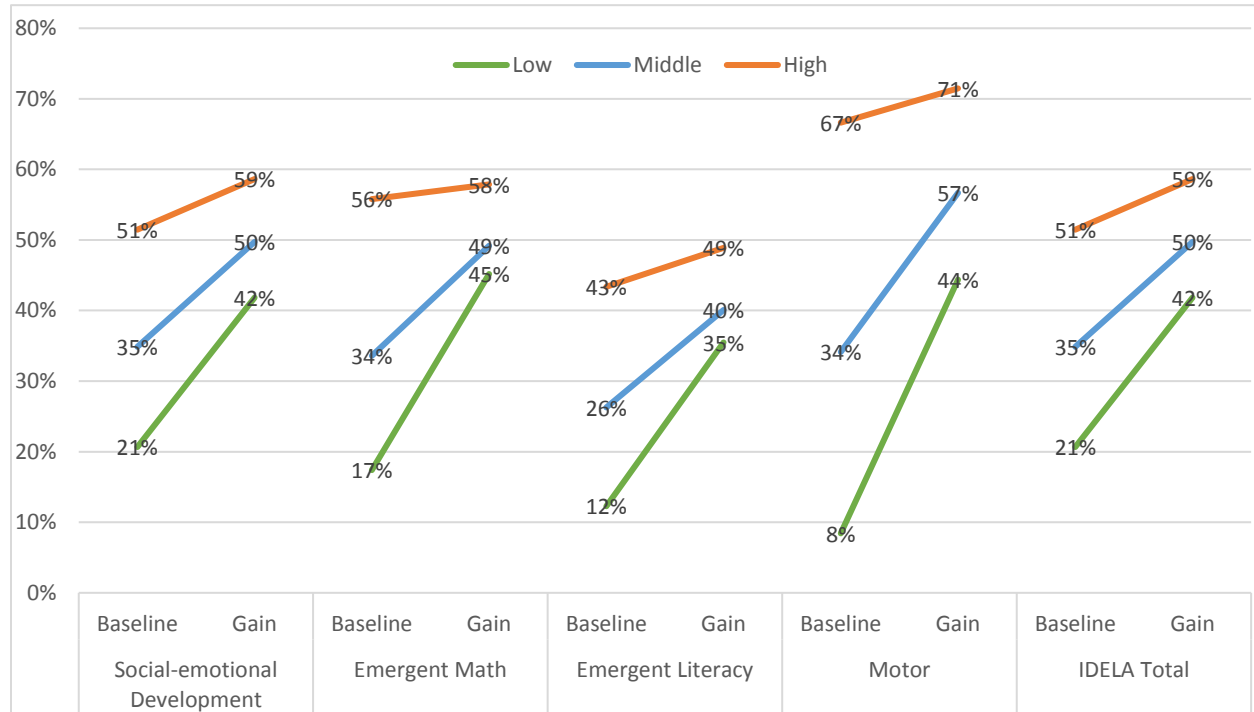
Changes by baseline IDELA scores

Typically in an equity analysis, gains are analyzed according to variables such as socioeconomic status, ethnicity, gender, and home learning environment. However, in this study, demographic data beyond gender and age was not collected. Instead of looking at results by demographic lines, the following analyses examine how gains varied by baseline measures. In all of the following, both gender and age are controlled for and children are clustered by CBCC to control for variation by center.

After controlling for age and gender, children are sorted into 3 categories by their score at baseline: low, middle and high. Students who scored +1 SD or above are considered “high” (equivalent to the 84th percentile and above), “low” consists of students who scored -1 SD or below at baseline (equivalent to 16th percentile and below), and “middle” comprises students within 1SD of the baseline mean (equivalent to the 17th to 83rd percentile).

Overall, children with low baseline measures are predicted to gain significantly more than children who began at a high or average baseline level. While all children are predicted to improve between baseline and endline, children who begin with low baseline scores are predicted to gain significantly more in all areas than children with middle and high scores. Children with high baseline scores gained significantly less than children with medium and low baseline scores. Figure 7 displays the predicted baseline and endline scores for children in all IDELA domains by low, middle, and high baseline score.

Figure 7. Predicted IDELA baseline and endline scores, by baseline score



In addition to looking at differences in gains by baseline levels, we also examine whether or not these differences in gains varied by study group. **Differences in gains associated with a “low”, “middle”, or “high” baseline level appear to be similar for all children, regardless of treatment.** The differences in gains by baseline status (i.e. that children with lower baseline scores gained more, on average) were not any larger for children who were in the comparison, HEART, ELM, or HEART + ELM group.

These results are encouraging. There is no comparison group that received no ECD, but they are suggestive that ECD programming can help narrow the gaps that children face in development. However, further analysis is necessary to determine whether these results actually are a product of differential gains or are an analytical artefact of regression to the mean.

Conclusion

Children in all of Save the Children’s ECD treatments (ELM only, HEART only, and HEART+ELM) had significantly larger gains in developmental outcomes (IDELA) than children in the comparison CBCCs. This is an encouraging result that suggests Save the Children’s ECD programs are of a higher quality than the status quo. We found no significant findings about the SEARS-PRE.

Looking more deeply, **we fail to find meaningful differences in outcomes between Save the Children’s different treatment groups.** The results of this study do not suggest that HEART programming was associated with greater gains in social-emotional skills or ELM led to larger gains in emergent literacy and math than HEART. Given the large differences in programming, this is a surprising result.

A more encouraging finding was about the result of ECD as a whole. **Children with lower baseline scores saw greater gains between baseline and endline than children who started out with middle or high**

baseline scores. This suggests the ECD programs can benefit the neediest children in Zomba's communities, and can be effective in reducing learning and developing inequities that exist between children before then enter primary school.

Overall, the results show striking gains for children who received Save the Children programming. Further study on the fidelity of implementation, spillover between treatment groups, and further background information would be useful at better understanding the results and generating additional hypotheses for testing. Ultimately, a more rigorous impact evaluation would be required to establish a concrete casual connection between Save the Children programming and gains in developmental outcomes.

Limitations

While the study was undertaken with a fairly strong research design, there are limitations to attributing a causal connection between Save the Children programming and the impressive effects.

The fact that CBCCs all came from the same geographic area and both comparison and treatment schools received other Save the Children programming is a substantial strength of the study. However, with little demographic and background data available, it is difficult to empirically assess the comparability of treatment and control CBCCs. The fact that comparison schools were in a separate area from treatment CBCCs helped prevent contamination of the comparison group, but may have also biased the sample. However, on the available characteristics (baseline development levels and gender), there were no significant differences between children in treatment and comparison CBCCs.

At endline, the differential attrition, and in particular the differences in reasons for attrition, generate some concern about the comparability of treatment and control groups. Even if they were relatively comparable at baseline, if children left treatment and control groups at different rates and for different reasons, the results may be biased in unpredictable ways.

At endline, we are limited to a regression analysis which means that there may be unobserved differences between children in different treatment groups that bias our findings. We are also unable to control for additional factors, such as SES or Home Learning Environment. However, the limited available data shows that, at least in terms of age, gender, and developmental status at baseline, that all treatment groups were similar, a strength of the study. The fact that CBCCs were randomized into treatment and control groups is a substantial strength of the study.

Considering the findings about differential changes by baseline score (i.e. that children who started with a low baseline score gained more), a limitation is that we do not know if IDELA is a "flat" measure. It may be more sensitive to changes at the lower end of the scale than at the higher end.

The results study are suggestive of a substantial impact of Save the Children programming on early development outcomes. However, as a results of the limitations discussed above, the results should be considered correlational and do not prove an impact of the programs.

Next Steps

The current study raised a number of issues and indicates directions for next steps.

1. Results for all Save the Children intervention groups are suggestive of a substantial impact. Dissemination of these findings can be used to promote Save the Children programming and to lobby for additional buy in from the government, donors, and other partners implementing ECCD programs.
2. A national scale up of HEART and ELM interventions could lead to large learning and development gains for children and also provide additional opportunities to evaluate the impact of these programs more precisely.
3. From a research perspective, one of the most interesting results from this study was the similarity of results in the three treatment groups, despite treatments that, at least on paper, varied substantially.
 - Project staff have hypothesized that, due to previous interventions and cultural settings, that the difference in programming may have been larger in theory than in practice. For example, ELM CBCCs may have also been incorporating arts into the curriculum and that HEART CBCCs may have used ELM methods in the past. Investigating the reasons for this more deeply would require qualitative research on the actual practices in each CBCC.
 - As a goal of HEART is to create conducive play and learning environments, an assessment of CBCC environments would help explain differences in outcomes.
4. A large number of children advanced to primary school during the study. Ensuring that CBCCs provide facilitate a smooth (and not premature) transition into primary school should be a goal of future programming.

Appendix A

Table A1. Multivariate equity analysis results clustered by village, all children

VARIABLES	(1) Motor Gain	(2) Literacy Gain	(3) Math Gain	(4) Social- emotional Gain	(5) IDELA Gain
Child's age (years)	0.035	0.077***	0.072***	0.027	0.044**
Child was female	0.010	0.007	0.007	0.007	0.011
Child received ELM	0.076**	0.116***	0.074***	0.078***	0.079***
Child received HEART	0.089**	0.077**	0.045*	0.039*	0.067***
Child baseline score was low	0.134***	0.122***	0.094***	0.137***	0.063**
Child baseline score was high	-0.177***	-0.135***	-0.083***	-0.191***	-0.077***
Constant	-0.034	-0.319**	-0.274**	-0.099	-0.145~
R-squared	0.203	0.280	0.178	0.307	0.226
Observations	436	436	436	436	436

*** p<0.001, ** p<0.01, * p<0.05