



**Save the Children**

## Sponsorship Mozambique: Nampula Midline Review 2016



# Background

## Research Design

- Treatment & Comparison groups not assigned randomly
- Originally hoped to be true baseline/endline study, but insufficient identifiers at baseline made it impossible to reliably interview the same children at midline
- Some issues with sampling and potential bias: children in comparison areas were selected at random from the community whereas children in treatment areas were selected at random from ECD centers
- While we attempt a number of analytical approaches to minimize bias, **strong conclusions about program impact cannot be made based on these analyses**



# Background

## Analyses & Limitations

---

### Analyses

- Three analyses presented in this report
  - Balance test of background characteristics at baseline and midline to assess comparability
  - Comparison of Comparison and Treatment adjusted mean scores at midline
  - Item-by-item midline scores regression analysis

### Limitations

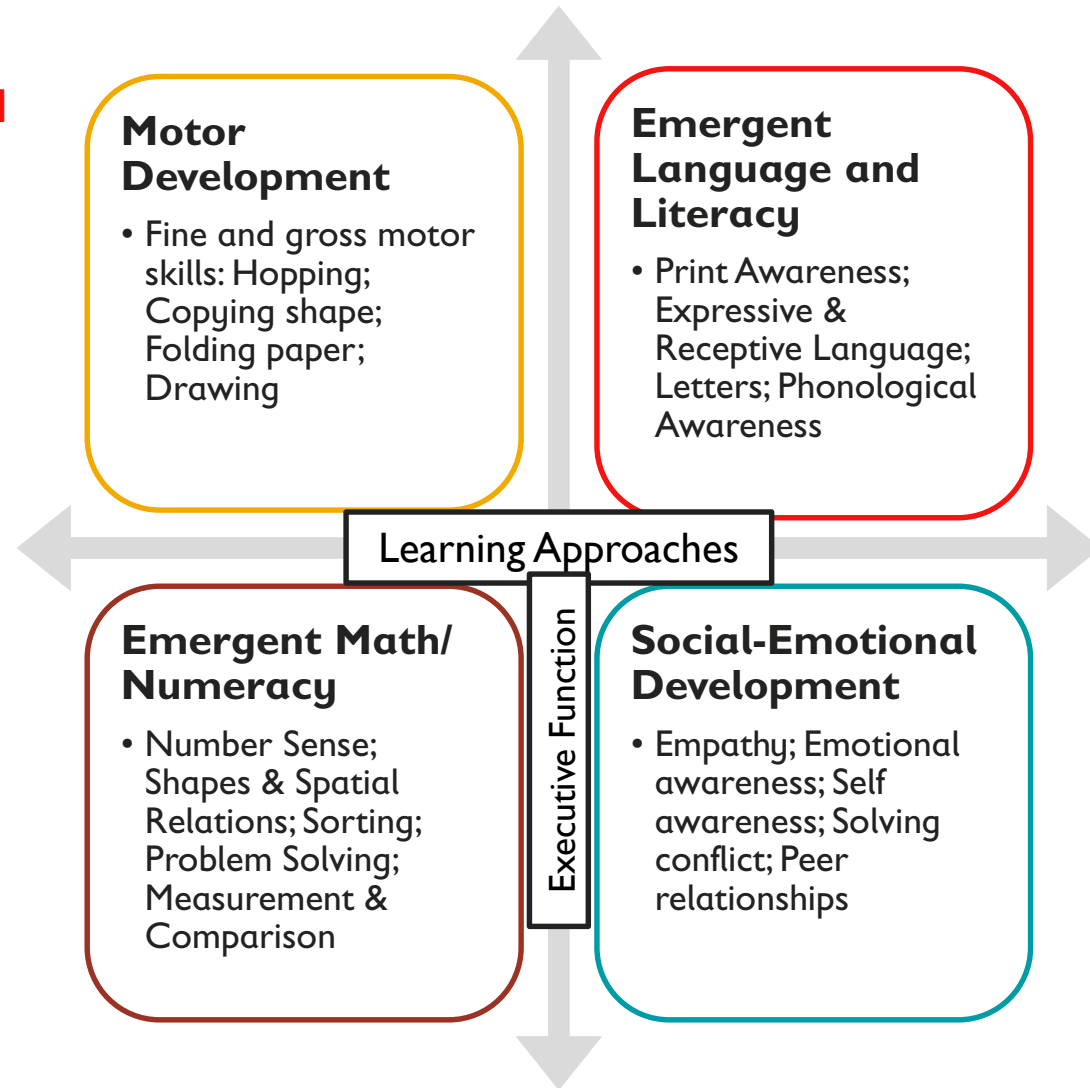
- Scoring problems on some IDELA items raises questions about fidelity of baseline data. Midline data appears to be of higher quality and is more reliable.
- Differences in sampling strategy in Treatment and Comparison areas limits comparability of full sample and may bias comparisons
- Systematic differences between children in Treatment and Comparison areas

# Background

## IDELA Instrument

### International Development and Early Learning Assessment

- Internationally tested and locally adapted measure of fundamental skills children require to be ready for school.



# Background

## Sample

---

### Full sample comprises 601 observations at baseline and midline

- More observations at midline than baseline
- Subsequent analyses use subsets of this data (data with missing variables is omitted for regression analysis)

	Comparison	Treatment	Total
Baseline	130	135	265
Midline	169	165	334
Total	<b>301</b>	<b>300</b>	<b>601</b>

# Balance test

## Explanation

---

### Purpose

Assess the comparability of treatment and comparison groups at baseline and midline.

### Analytical Approach

Compare mean values of background characteristics at baseline and midline, using t-tests with clustered standard errors to determine if differences are significant.

### Takeaway

If treatment and comparison groups are similar on background characteristics, it makes the argument that differences in outcomes between groups represent the impact of the program stronger. Significant differences between groups limit the strength of subsequent analyses.

# Balance test

## Baseline differences by treatment group (select variables)

Variable	Comparison Mean	Treatment Mean	Significant Difference?
Child's age	4.39	4.06	
Mother is Literate	36%	22%	
Father is Literate	59%	39%	*
Number of Children in Family	4.36	4.23	
Standardized SES Score	0.20	-0.14	
Number of Toy Types	3.95	3.26	
Family owns a shop-bought toy	45%	25%	*
Number of Reading Materials	1.46	1.10	
Sum of Positive Learning Activities	6.17	4.66	***
Parent reads to child	52%	31%	**
Parent teaches child letters	55%	28%	***
Sum of Negative Discipline	2.22	2.03	

Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$



# Balance test

## Midline differences by treatment group (select variables)

Variable	Comparison Mean	Treatment Mean	Significant Difference?
Child's age	4.12	4.68	***
Mother is Literate	9%	22%	
Father is Literate	29%	46%	
Number of Children in Family	4.71	5.15	
Standardized SES Score	-0.24	0.19	
Number of Toy Types	2.00	2.81	*
Family owns a shop-bought toy	15%	21%	
Number of Reading Materials	0.65	1.50	**
Family owns a story book	9%	29%	**
Sum of Positive Learning Activities	2.55	4.25	*
Parent reads to child	2%	23%	**
Parent teaches child letters	11%	40%	**
Sum of Negative Discipline	1.08	1.33	

Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$

# Balance test

## Summary

---

### Significant differences between Comparison & Treatment groups at both baseline and midline.

- At baseline, children in comparison areas appeared to come from families with more literate parents, slightly more learning materials, and parents who engaged them in more home learning activities.
- At midline, the reverse appeared to be true. Children from treatment areas appeared to have a better home environment, with more toys and reading materials, and parents who engaged them in more home learning activities.

### Questions about quality of baseline data

- There are large differences between baseline and midline results. Parents report significantly higher levels of engagement with children at baseline than midline in both treatment and comparison areas. This makes results questionable. For example, at baseline over 10% of respondents reported reading to their children despite neither parent being literate.
- **Comparing outcomes between these groups comes with substantial limitations.**

# Midline results

## Explanation

---

### Purpose

Examine item-by-item and domain averages and differences between treatment and comparison groups at midline. Assess strong and weak areas for children.

### Analytical Approach

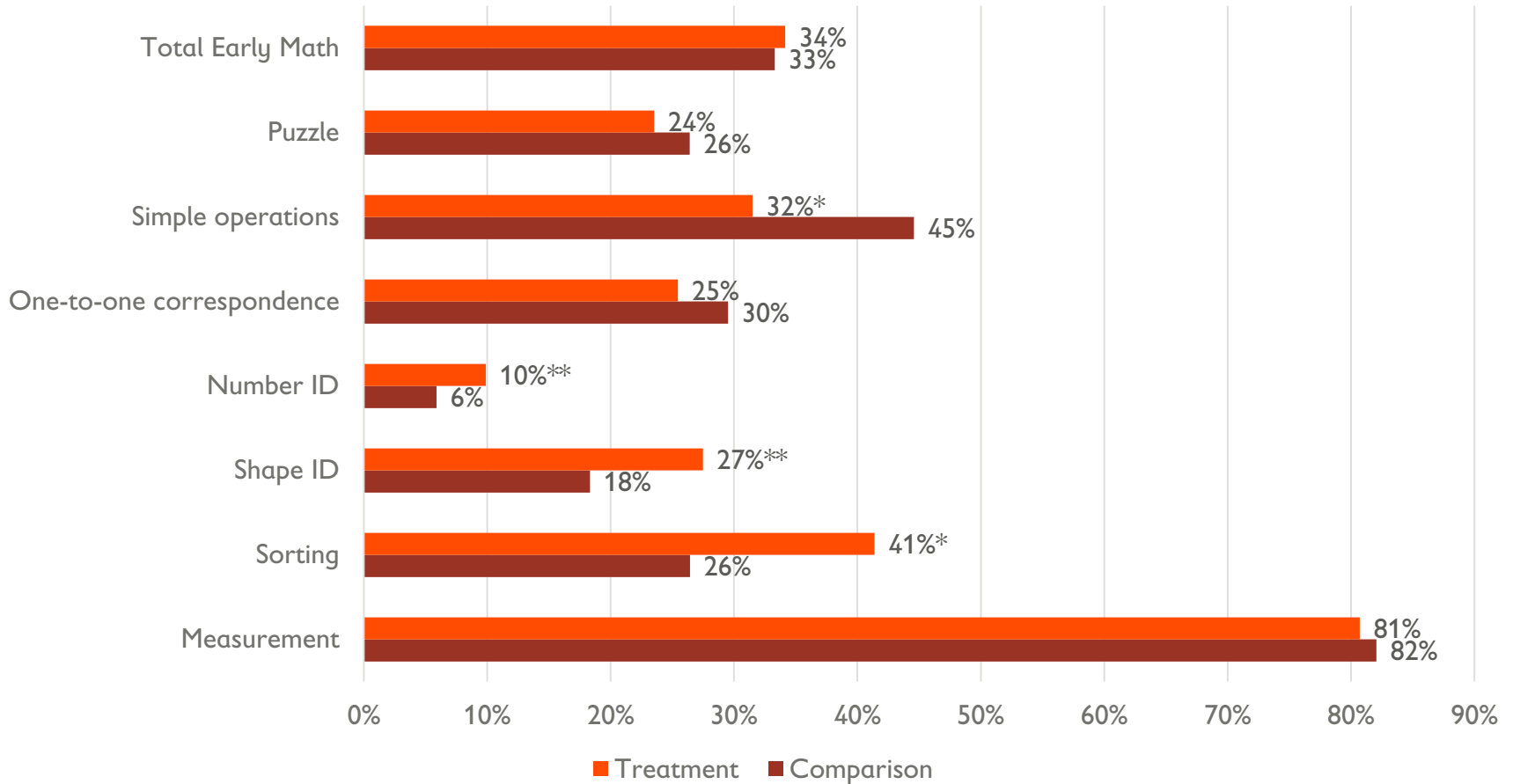
“Fitted mean” values for treatment and comparison children are presented after controlling for the child’s age, province, and caregiver-reported Socio-Economic Status, Home Learning Environment, Home Learning Activities, and accounting for village-by-village differences.

### Takeaway

**Midline differences between treatment and control should not be interpreted as program impact.** While we control for many factors, the results of our balance tests suggest there were large and significant differences between treatment and control children that may be expressed in unobserved variables. These differences may bias our findings.

# Midline results

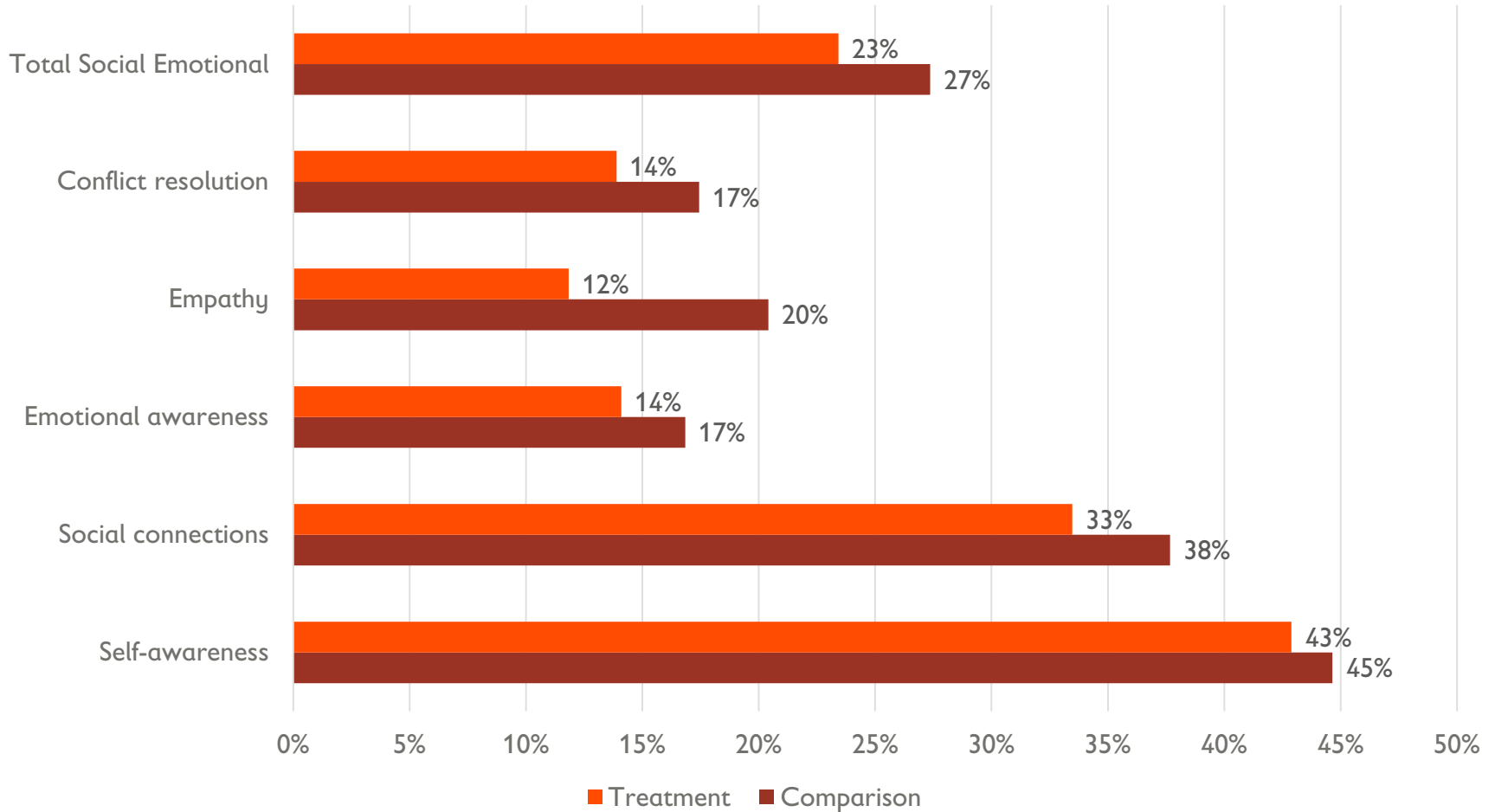
## Early Math



Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$

# Midline results

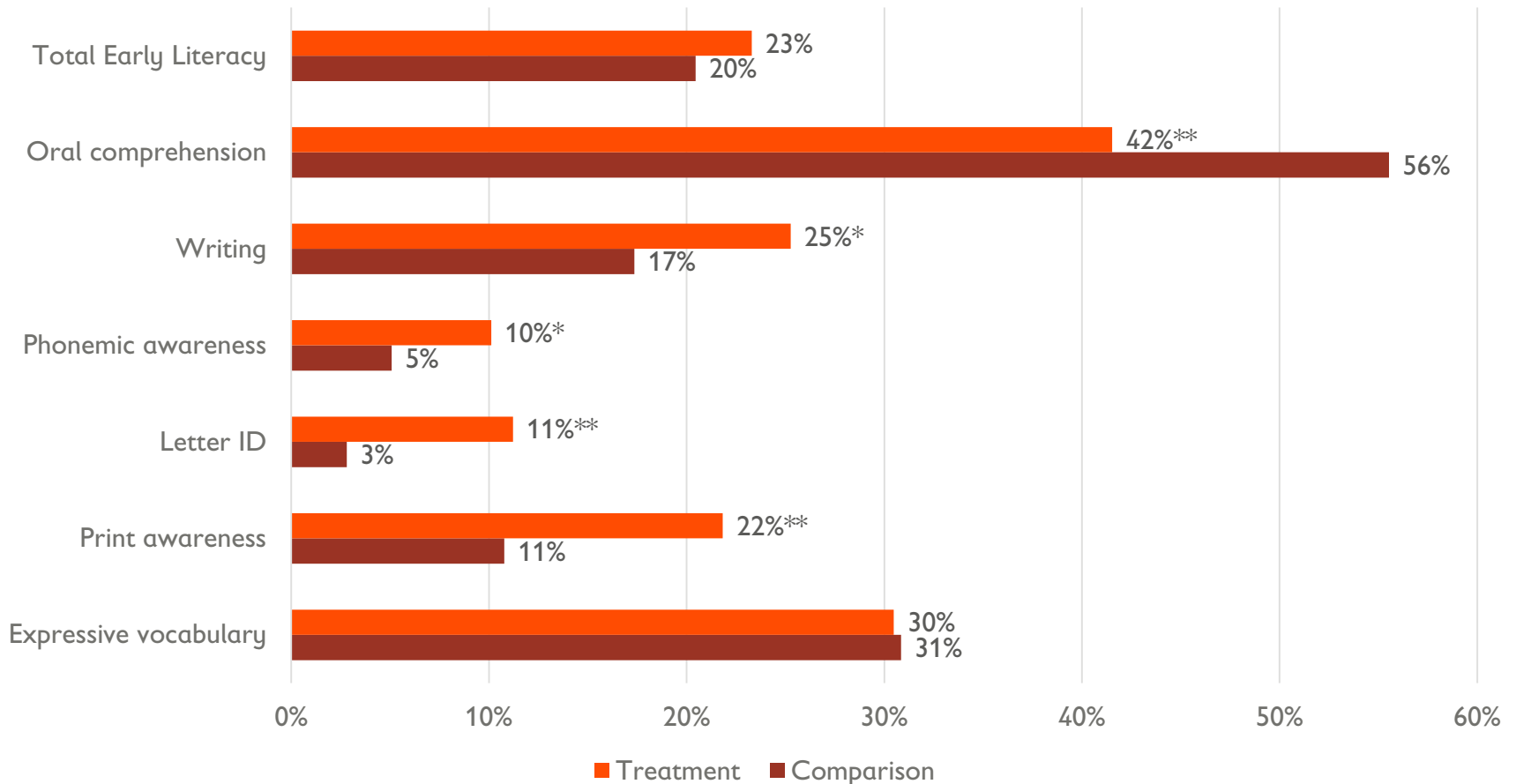
## Social Emotional Development



Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$

# Midline results

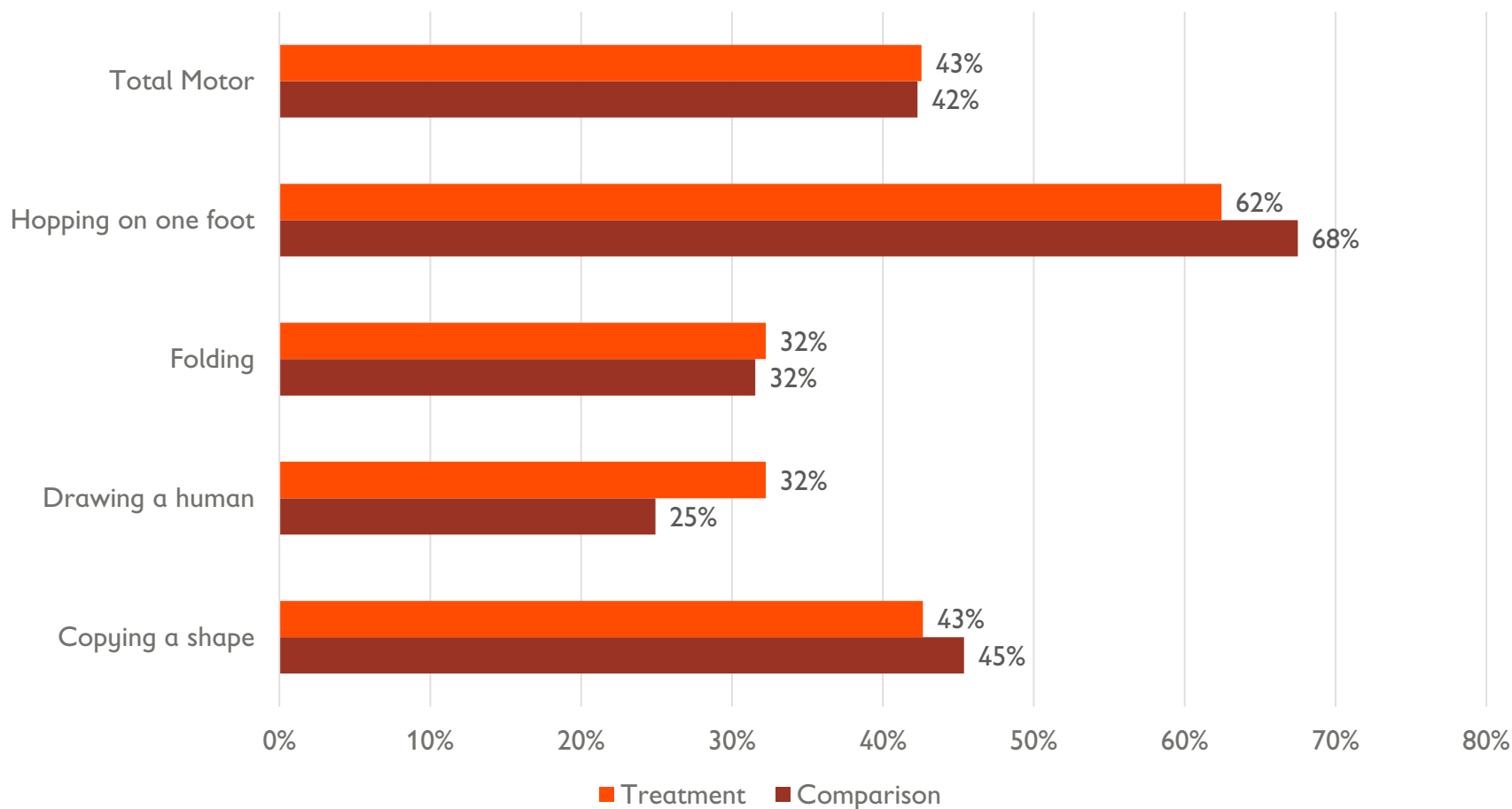
## Early Literacy



Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$

# Midline results

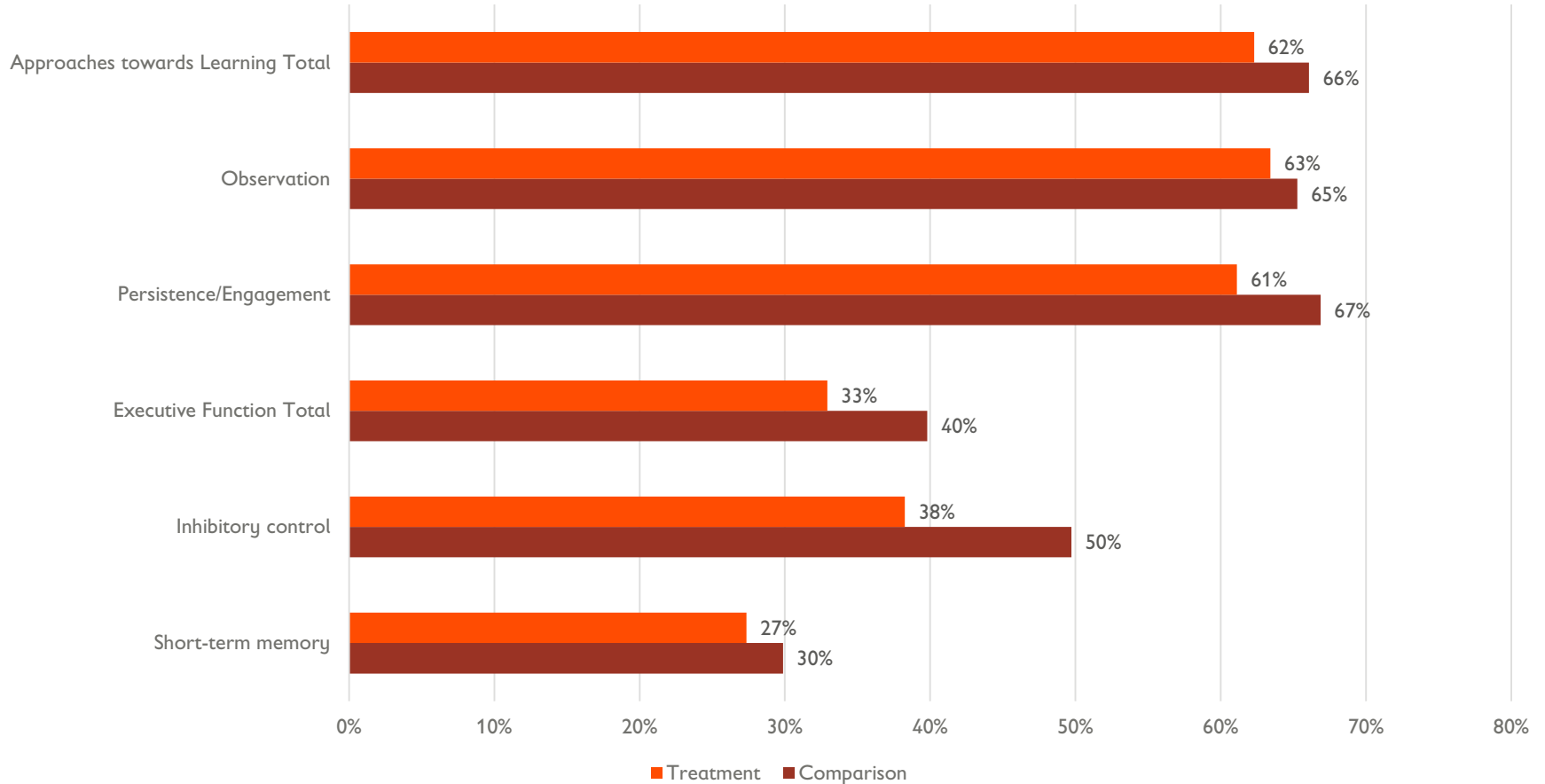
## Motor Development



Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$

# Midline results

## Executive Function & Approaches to Learning



Note: Significant differences are noted with \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$



# Midline results

## Summary

---

### **Differences in subtask performance in Early Literacy and Early Math Domains, but no overall difference in performance**

- In Early Literacy, children in treatment areas performed significantly better on Print Awareness, Letter ID, and Writing subtasks, but significantly worse on the Oral Comprehension subtask.
- In Early Math, children in treatment areas performed significantly better on Shape ID and Number ID, but significantly worse on Simple Operations.
- No significant differences in overall performance at the domain level.
- Inappropriate to interpret these differences as estimates of program impact

### **No differences in outcomes for Treatment and Comparison in Motor Development, Social-Emotional Development, Executive Function, or Attitudes towards Learning**

- Domain and subtasks results reveal no significant differences

# Comparison of means

## Explanation

---

### Purpose

Examine domain averages at baseline and midline and the differences between treatment and comparison groups and provinces. Estimate directionality of changes and make comparisons between provinces.

### Analytical Approach

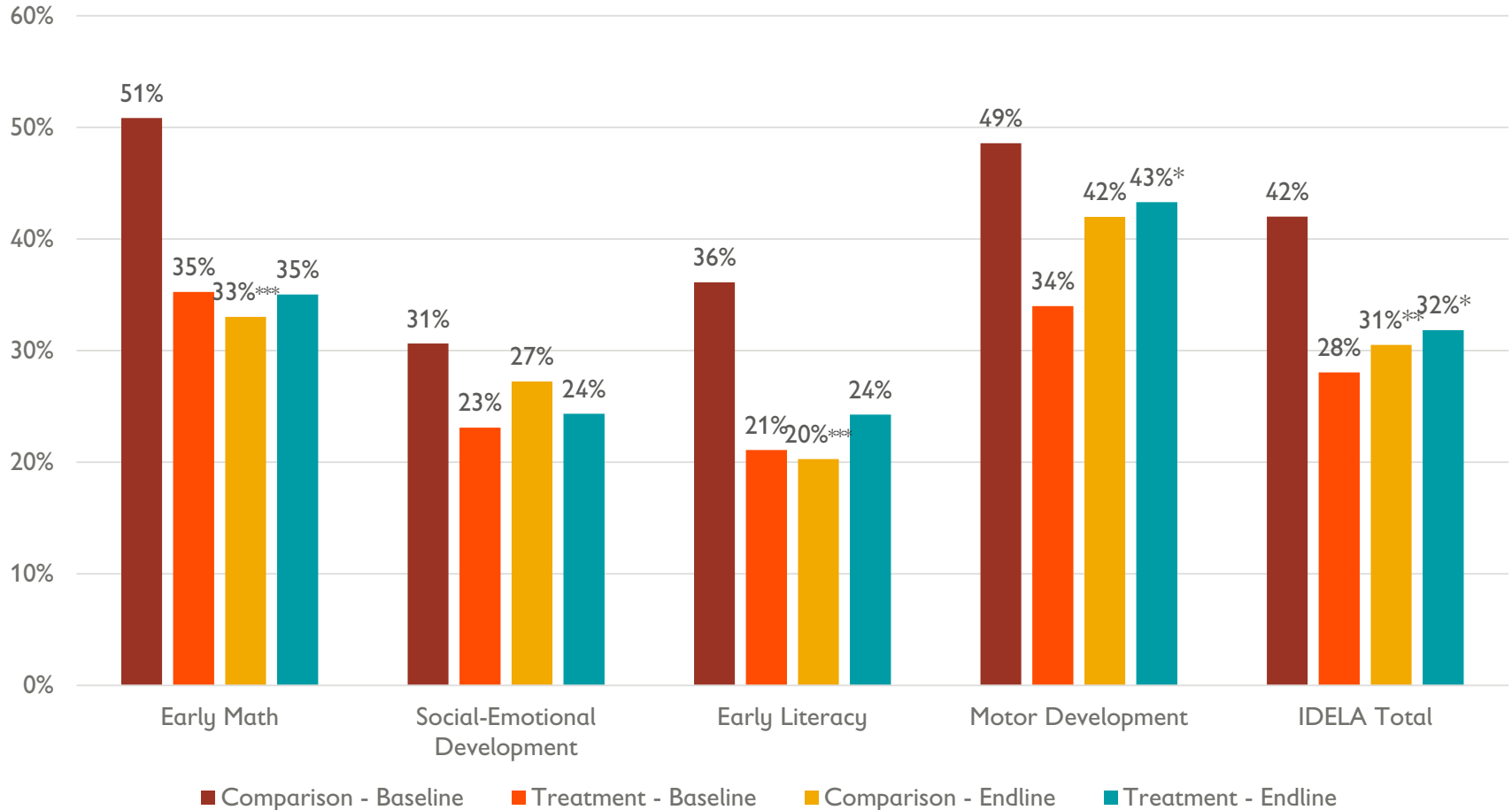
“Fitted mean” values are presented after controlling for the child’s age, province, and caregiver-reported Socio-Economic Status, Home Learning Environment, Home Learning Activities, and accounting for village-by-village differences. We compare baseline to midline, treatment to control, and the results from different provinces

### Takeaway

Differences between baseline and midline should not be interpreted as true change over time. The midline sample comprises mostly different children than the baseline sample. These should be interpreted as independent snapshots of the communities at baseline and midline.

# Comparison of means

## Domains by baseline/midline and treatment status



Note: Significant differences between baseline and midline mean scores are noted independently for treatment and comparison areas (\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ )

# Comparison of means

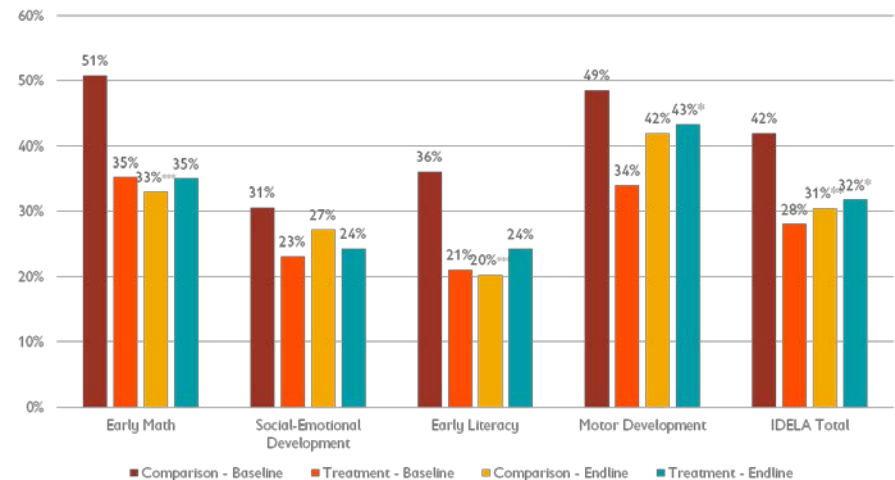
## Domains by baseline/midline and treatment status

### Baseline scores

- Large differences between treatment and comparison areas at baseline
- Comparison areas exhibit significantly better outcomes

### Midline scores

- Comparison areas had significantly lower midline scores (from baseline) in Early Math, Early Literacy, and the overall IDELA total
- Treatment areas had significantly higher (from baseline) Motor Development and overall IDELA total scores, but no difference in Early Math, Early Literacy, and Social-Emotional Development.
- The large negative difference from baseline to midline seen in comparison areas is not observed in treatment areas.



# Conclusion

## Summary of findings

---

### **Weak evidence of any impact of treatment**

- Evidence suggests positive impact on individual subtasks of Early Literacy and Math, but also negative impact on other subtasks and no overall effect

### **Populations in Control and Treatment areas exhibit substantial differences at both baseline and midline**

- Strange result of flipped directionality between baseline/midline (Comparison children were better off at baseline while Treatment were better off at midline)
- Background characteristics (especially Home Learning Environment and Home Learning Activities) may be affected by the program
- Bias in results may exist even after for controlling for different factors

### **Baseline data is of questionable quality**

- Paper-entry input allowed for logical a inconsistencies in scoring/entry

**THANK YOU**



**Save the Children**