

Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures

Rebecca Thornton

University of Illinois at Urbana-Champaign

Jason Kerwin

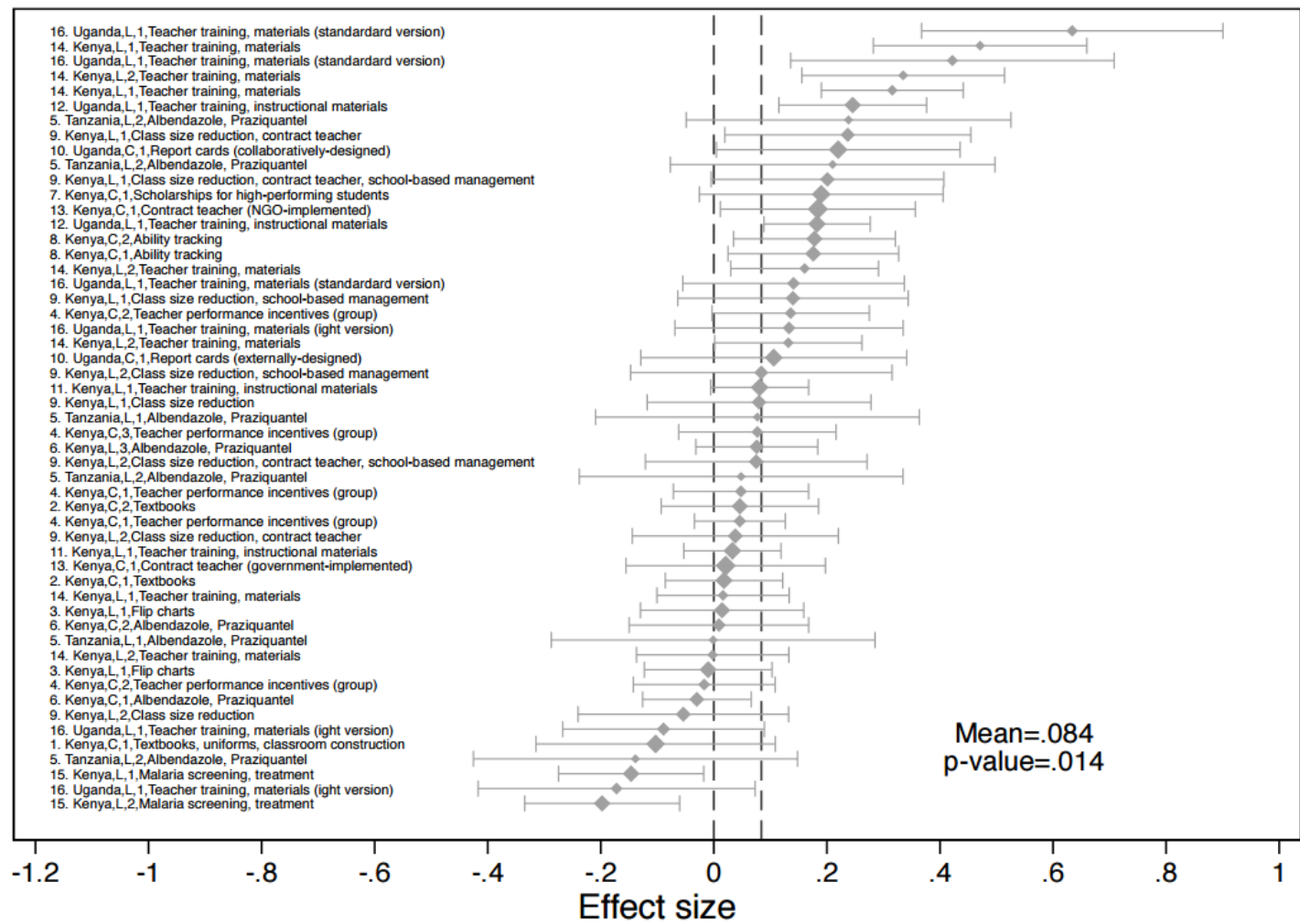
University of Minnesota

How to improve learning?

- Hundreds of program effectiveness studies
- Meta-analyses, systematic reviews, meta-meta-analyses

“What works”

Figure 2: Effects of treatments on language or composite test scores



How to (use research to) improve learning?

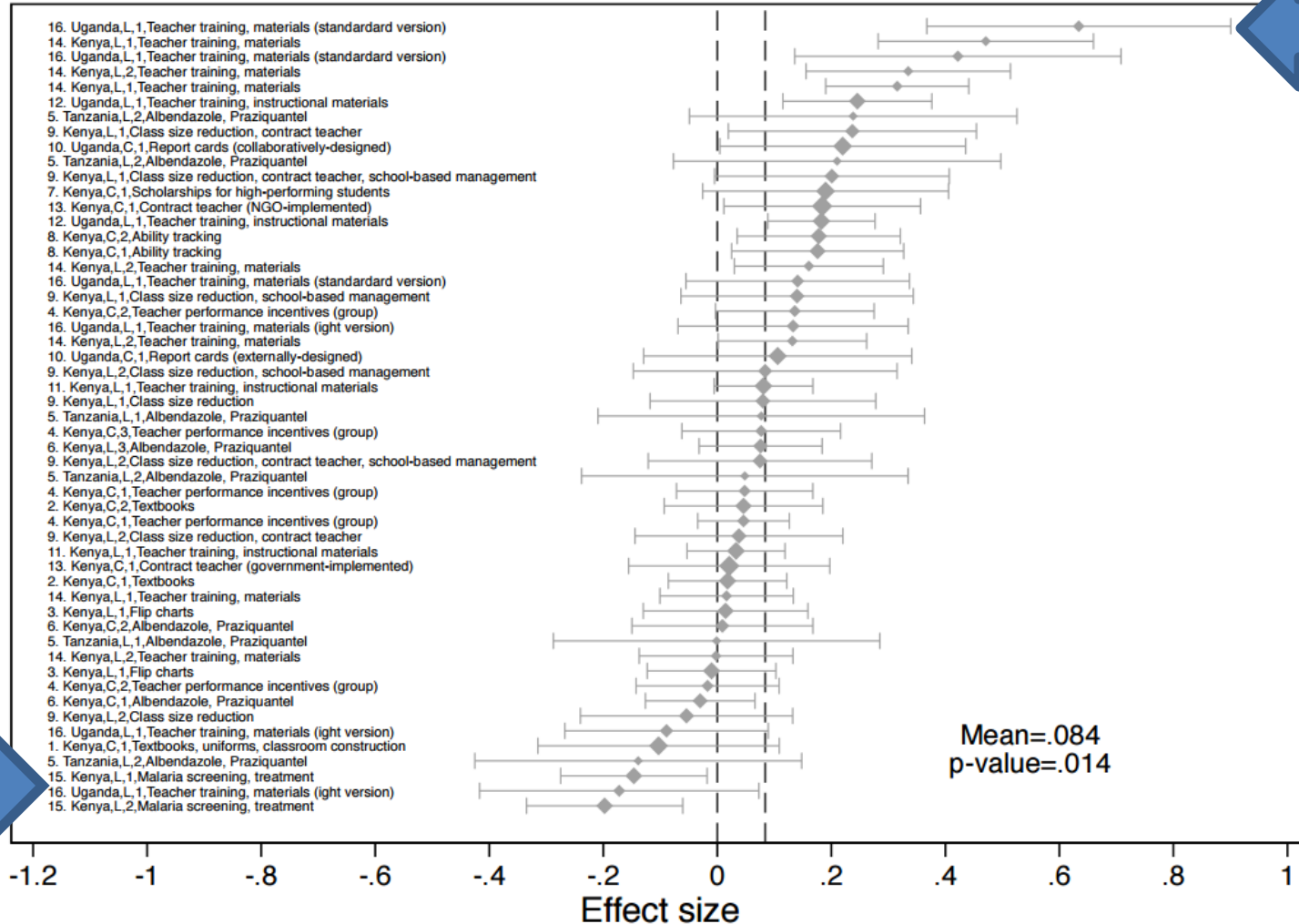
- Never able to “replicate” implementation of a program
 - Context
 - Budget constraints
 - Logistical constraints
 - Are enough details even provided?

What happens as we modify a program?

- Large variation in effectiveness across programs
 - Across setting & intervention type
 - Within setting & intervention type (Evans and Popova 2016b, Vivaldi 2017)
- Most evidence across studies, not within (McEwan 2015)
- This paper: examines variation within a single study, holding context and intervention type constant

Outcomes from a single study

Figure 2: Effects of treatments on language or composite test scores



Mango Tree Literacy Program

- Community engagement
- Pedagogy
 - Mother tongue, slower pace, phonics, scripted lessons
- Materials



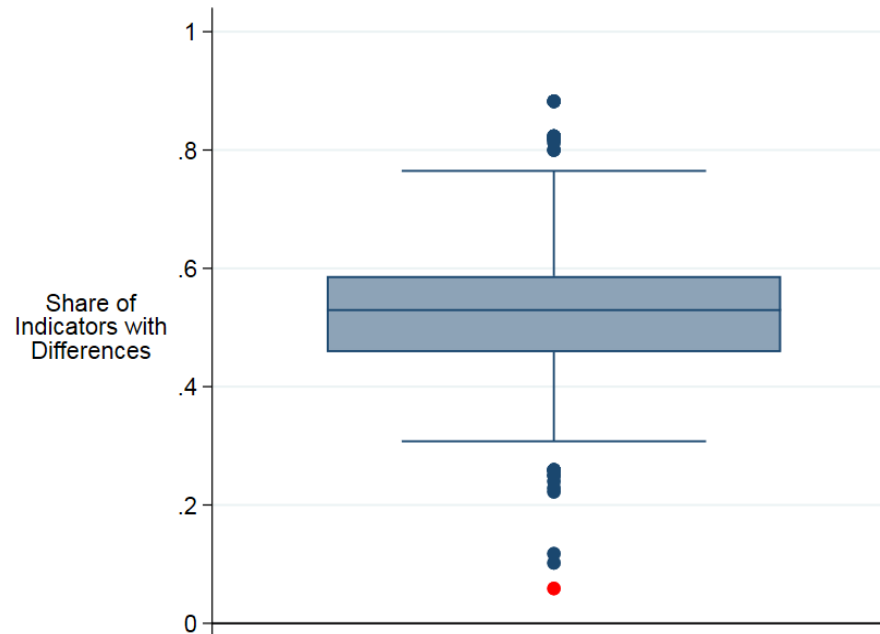
- Teacher training and support
 - Each term: 1 residential + 3 non-res workshops, 3 class visits

Reduced-cost version

- Modified to resemble implementation at scale
 1. Cascade model of training and support (non-res)
 2. Fewer classroom visits (5 vs. 2)
 3. No slates or wall clocks

Differences?

- Teacher training indicators (Arancibia, Popova, and Evans 2016)
 - Codes 26 teacher training programs, including NULP
 - Out of 51 indicators, three (5.9 percent) differ
- 325 pairwise combinations, compute % indicators different



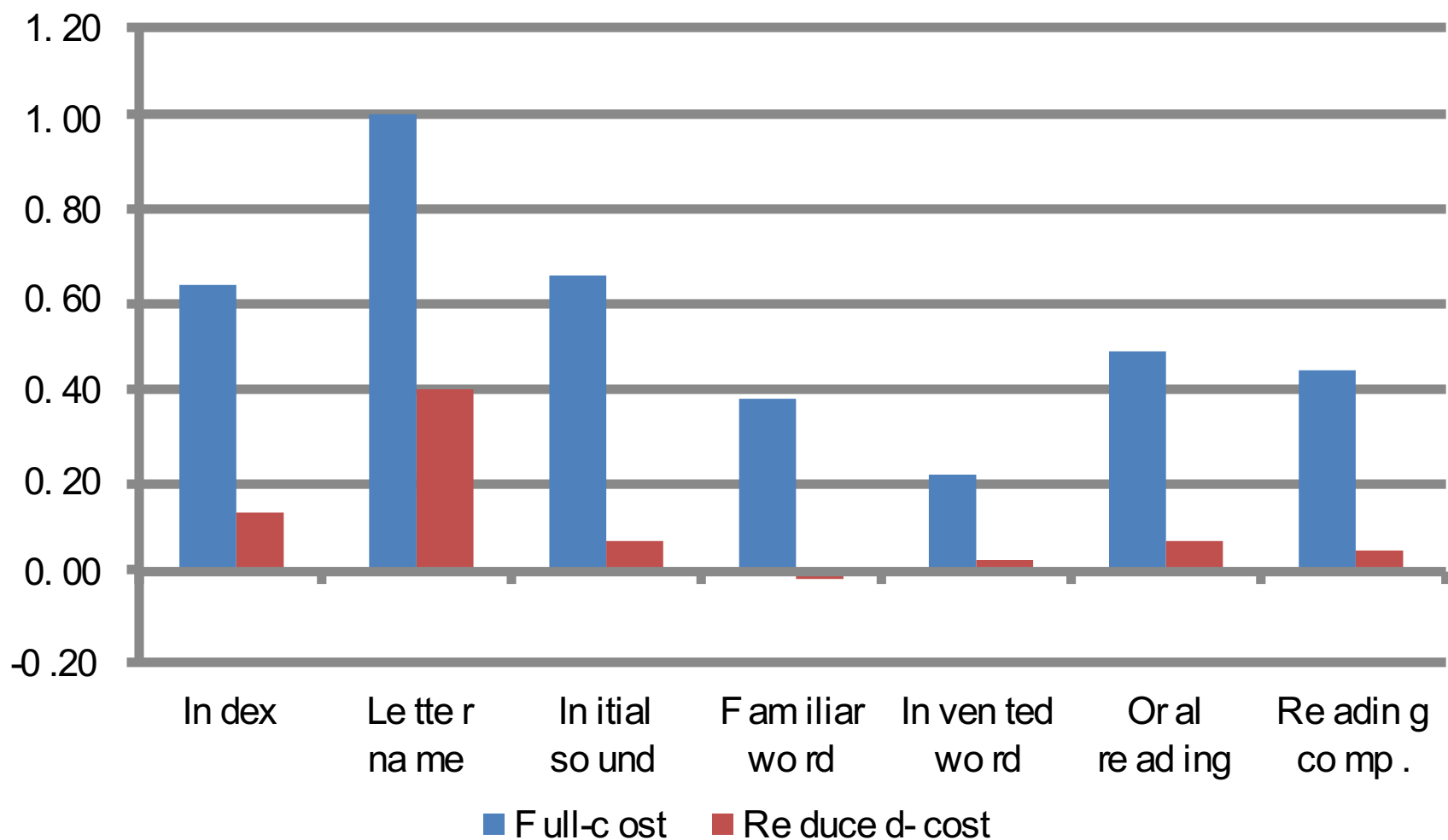
Research design

- 38 primary schools
- 50 grade one students/school (N=1,900)
- Public randomization
 - Control (Government status quo)
 - Full-cost program (Mango Tree)
 - Reduced-cost program (Cascade model)

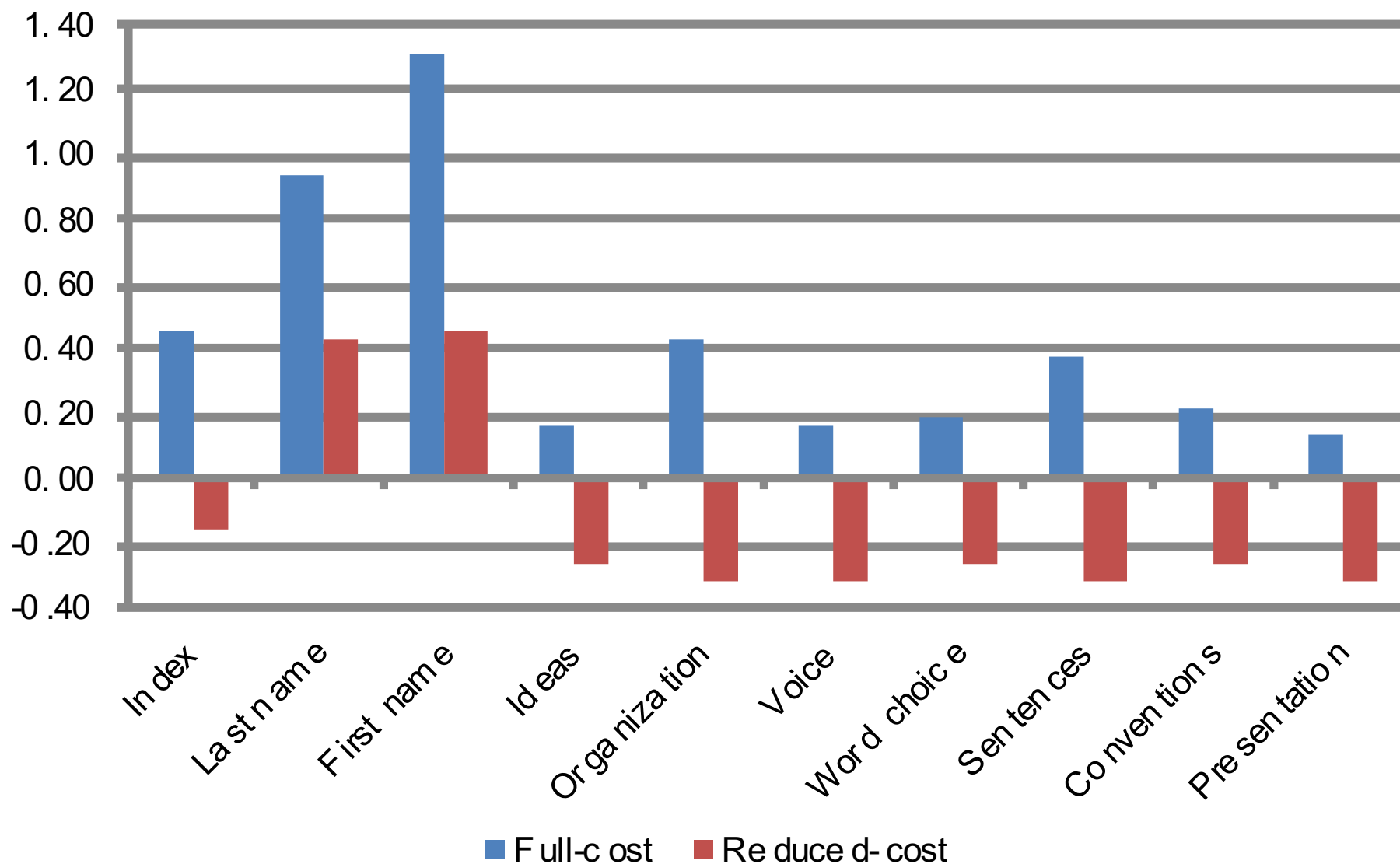
Research design

- Exams
 - Baseline
 - Endline (78% of baseline, N=1481)
 - Outside examiners blinded to study arm
- Learning
 - Reading Leblango (EGRA)
 - Writing Leblango (EGWA)
- Results
 - Each module + PCA index
 - Normalize
 - Randomization inference p-values

Program effects on reading



Program effects on writing



Mechanisms

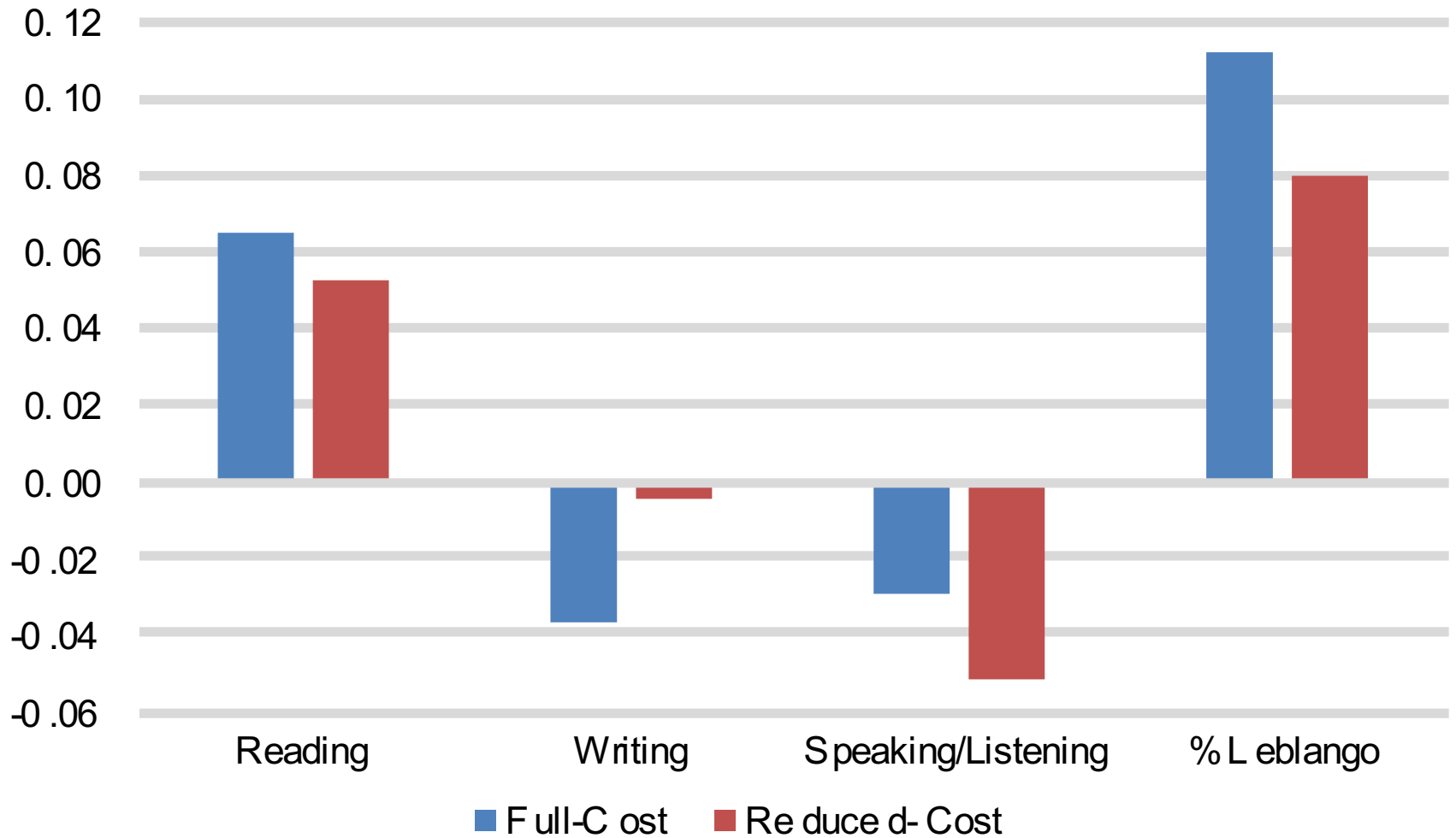
- What led to the success of the full-cost program?
- What led to the failure of the reduced-cost program?
- Use classroom observations data to explore mechanisms
 1. Time on task
 2. Productivity
 3. Complementarities

Classroom observations

- 3 visits: two 30-min literacy lessons/classroom
- Factors

Specific Lesson Actions			
Time	Teacher actions	Pupil actions	
FIRST		Reading	
10 minutes:	<u>Positive actions:</u> <ul style="list-style-type: none"> <input type="checkbox"/> Refers to TG or lesson plan while teaching <input type="checkbox"/> Moves freely around the classroom <input type="checkbox"/> Calls on individual pupils by name <input type="checkbox"/> Encourages pupil participation and keeps their attention <input type="checkbox"/> Brings pupils back on task when needed <input type="checkbox"/> Observes and records pupils' performance 	<input type="checkbox"/> Sounds <input type="checkbox"/> Letters <input type="checkbox"/> Words <input type="checkbox"/> Sentences	<input type="checkbox"/> Whole class <input type="checkbox"/> Smaller group <input type="checkbox"/> Individual at seat <input type="checkbox"/> Individual at board <input type="checkbox"/> On board <input type="checkbox"/> In primer <input type="checkbox"/> In reader <input type="checkbox"/> Other: _____ <input type="checkbox"/> English <input type="checkbox"/> LL
(start time)		Minutes on pupil reading tasks _____ min.	
(end time)		% of pupils participating in reading task _____%	
	<u>Negative actions:</u> <ul style="list-style-type: none"> <input type="checkbox"/> Lesson does not appear planned <input type="checkbox"/> Remains at the front of the class <input type="checkbox"/> Does not call on individual pupils by name <input type="checkbox"/> Very little pupil participation and attention <input type="checkbox"/> Ignores or does not address pupils who are off task <input type="checkbox"/> Does not record pupil performance 	Writing	
	<u>Other:</u> <ul style="list-style-type: none"> % time speaking English _____% % time speaking LL _____% Minutes out of class _____ min. Minutes in class but not teaching _____ min. Minutes teaching _____ min. 	<input type="checkbox"/> Pictures <input type="checkbox"/> Letters <input type="checkbox"/> Words <input type="checkbox"/> Sentences <input type="checkbox"/> Name	<input type="checkbox"/> Air writing <input type="checkbox"/> Handwriting practice <input type="checkbox"/> Copying teacher text from the board <input type="checkbox"/> Writing own text <input type="checkbox"/> On slate <input type="checkbox"/> On paper <input type="checkbox"/> On board <input type="checkbox"/> English <input type="checkbox"/> LL
		Minutes on pupil writing tasks _____ min.	
		% of pupils participating in writing task _____%	
		Speaking/Listening	
		<input type="checkbox"/> To a partner <input type="checkbox"/> To a small group <input type="checkbox"/> To the whole class <input type="checkbox"/> To the teacher	<input type="checkbox"/> English <input type="checkbox"/> LL
		Minutes on pupil speaking/listening tasks _____ min.	
		% of pupils participating in speaking/listening task _____%	

Classroom time (%)



Productivity?

Returns to time on task (SDs)

	Full-cost program	Reduced-cost program	Control
Hour reading	0.011	0.004	0.011
Hour writing	0.024	0.008	0.002

- Reading: more time on sounds (not sig between full & reduced)
- Writing: more focus on names
- Teachers more engaged in both reading and writing

Complementarities?

- Did not randomize each input
- Across inputs (materials, human capital)
 - Reading: more use of materials (& in full-cost)
 - Writing: large differences
- Across skills (reading and writing, advance and basic)

Complementarities?

- Mediation analysis with linear regression (Sequential g -estimator Acharya, Blackwell, and Sen 2016)

	<u>EGRA</u>	<u>Writing</u>
% explained by mediators	0.020	0.037

- Machine-learning allowing for complementary inputs and non-linearities

	<u>EGRA</u>	<u>Writing</u>
R-squared	0.80	0.99

Summing up

- Massive gains in learning possible
 - Even in most resource-deprived schools and using existing teachers
- Small changes in inputs may dramatically change program effectiveness
- Cutting costs may leave some students worse off
- What we measure is crucial for seeing the entire picture

Implications

- Researchers
 - Focus on isolating individual inputs
 - Systematically underestimate possible effects
 - Attention to types of learning metrics
- Policy makers
 - Limited resources – infeasible to provide most-effective programs
 - Almost always need to modify or eliminate some inputs
- Take advantage of complementarities, rather than focus on individual inputs
- What is the most effective research for knowing what/how to implement what works?