



# The Long Road to High-Quality Universal PreK in Boston

Catherine E. Snow

(with particular thanks to Christina Weiland  
and Meghan McCormick)

ProLEER, 2 October 2019

# Acknowledgements

## MDRC

JoAnn Hsueh  
Rebecca Unterman  
Michelle Maier  
Meghan McCormick  
Marissa Strassberger  
Sara Stazak  
Nicole Leacock  
Rama Hagos  
Sharon Huang

## Boston Public Schools

Jason Sachs  
Brian Gold  
Abby Morales  
Marina Boni  
Melissa Luc  
David Ramsey

## Harvard GSE

Catherine Snow  
Nonie Lesaux  
Sibyl Holland  
Maia Gokhale  
Students in Three Successive  
H233/234 Cohorts

## University of Michigan

Christina Weiland  
Deborah Ball  
Amanda Ketner  
Lillie Moffett  
Paola Guerrero-Rosada

## Cradle to Kindergarten

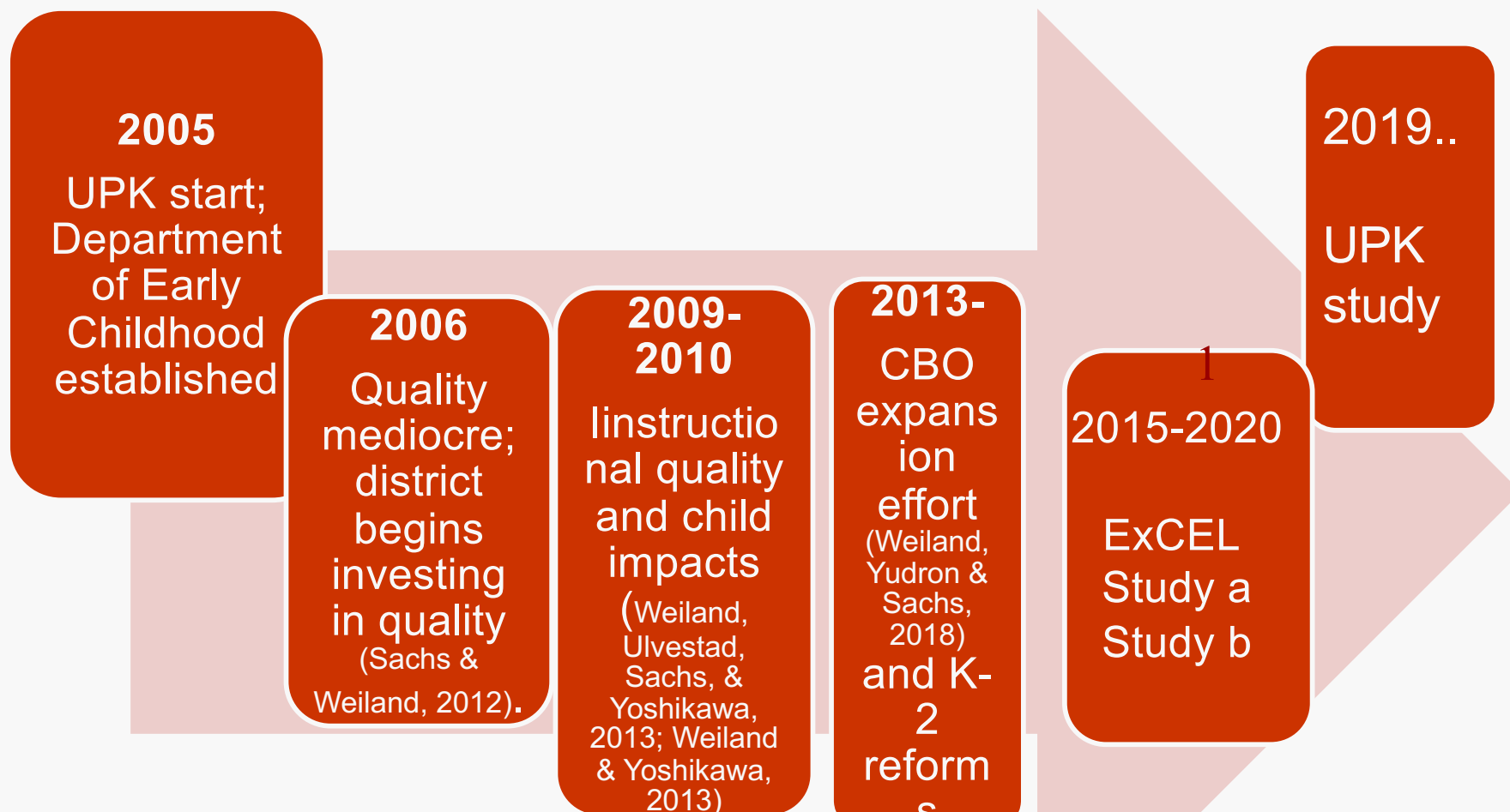
Ajay Chaudry  
Taryn Morrissey  
Hirokazu Yoshikawa

## Primary

## Funder

IES

# Boston Prekindergarten History



# Impacts of BPS K1 on Children's Early Numeracy, Language, Literacy, Executive Functioning, and Emotional Development

Christina Weiland & Hirokazu Yoshikawa  
Harvard Graduate School of Education

Funded by the U.S. Department of Education, Institute of Education Sciences

# Study Motivation: BPS

- Significant investment of city resources in K1 and in K1 quality (curricula, coaches, training)
- Are these investments paying off in terms of better child development?
- Helps us understand how K1 is contributing to closing achievement gaps and promoting the success of all children
- Opportunity to get rich data on two cohorts of K1 students for use in studying longitudinal impacts of K1

# Research Questions

- RQ1: What is the causal impact of the Boston Public Schools prekindergarten program on child early mathematics, language, literacy, executive functioning, and emotional development outcomes?
- RQ2: Do some student subgroups benefit more from the program than others?

# K1 basics

- Pre-K: About 28-35% of city 4-year-olds enrolled; enrollment open to any 4-year-old in the city
- Teachers paid on same scale and subject to same educational requirements as K-12 teachers
- Uniform curricula - *OWL* (Schickedanz & Dickinson, 2007) and *Building Blocks* (Clements & Sarama, 2007)
- Early childhood coaching system – one set of coaches supporting two curricula

# More info about OWL

[https://www.youtube.com/watch?v=IC\\_W9jmgchY](https://www.youtube.com/watch?v=IC_W9jmgchY)

**NOTE THAT THE OWL WAS BEING USED IN  
2008-2009, HAS NOW BEEN SUPERCEDED BY  
A BPS-ADAPTED VERSION**

<https://sites.google.com/bostonpublicschools.org/earlychildhood/focus-on-k1/unit-1-family>



# Fidelity of Implementation

- Observations conducted in 74 prekindergarten classrooms during treatment year
- Curricula were moderately to very fully implemented

# Sample

2,018 children  
(in 67 schools)

969  
before cutoff  
(Pre-K 2008-2009)

1,049  
after cutoff  
(Pre-K 2009-2010)

*Final sample represents 85% of schools & 70% of eligible children in those schools*

## Race/ethnicity

11% Asian, 27% Black, 41% Hispanic, 3% Other, 18% White

## Home language

50% English, 27% Spanish, 22% Other

## Gender, Free/Reduced Lunch, and Students with Disabilities

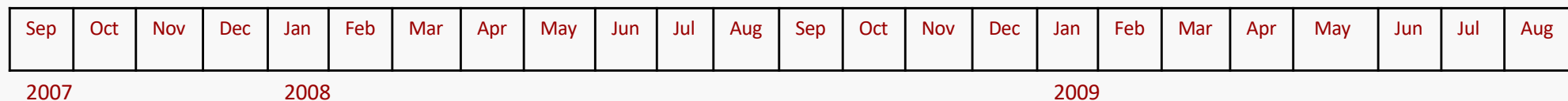
51% male, 69% receive free/reduced lunch, 9% students with disabilities

# Study design for child-level impacts: Regression discontinuity

**SEPTEMBER 1  
BIRTHDAY CUTOFF**

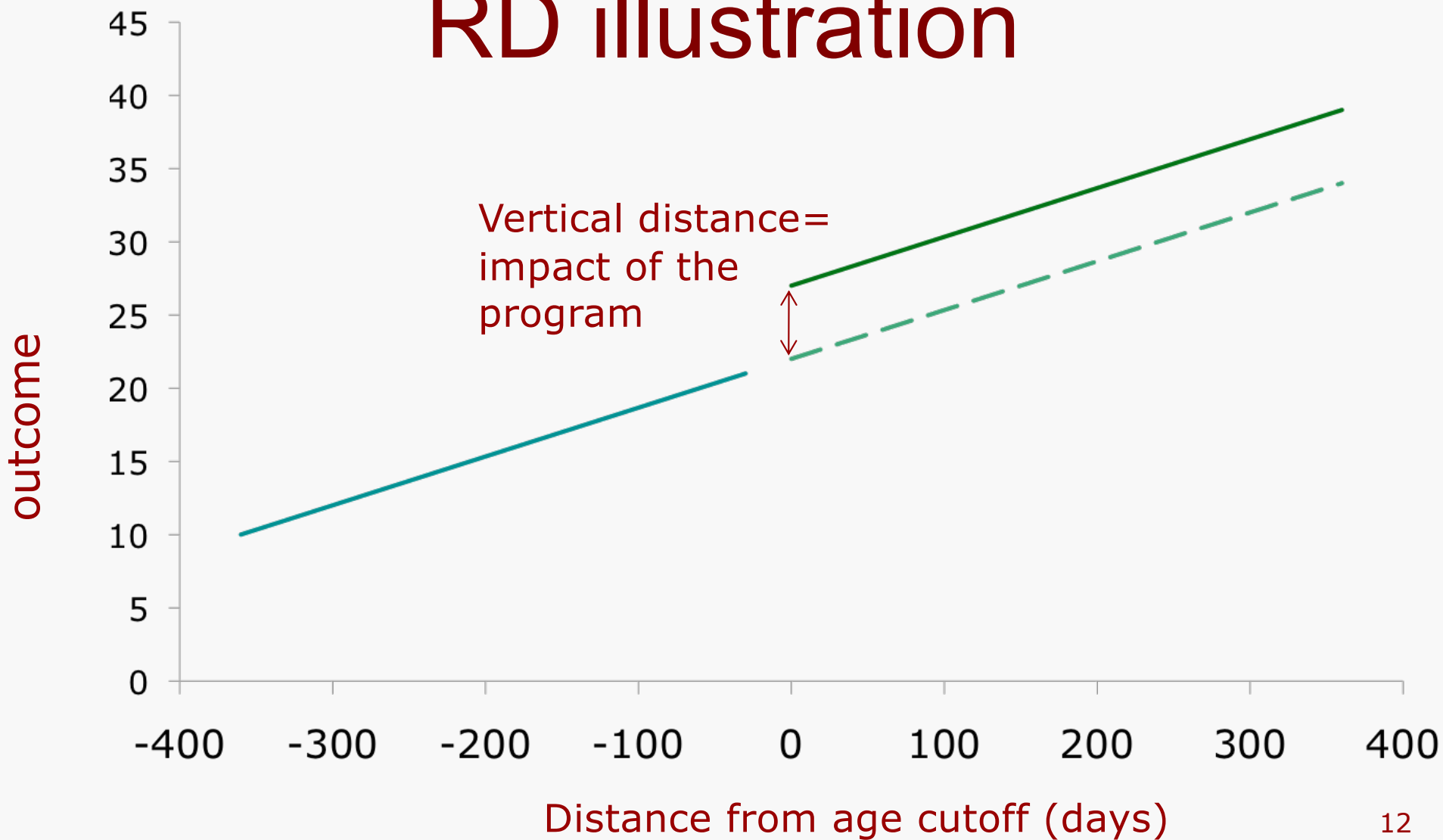
**“Treatment” Group  
(attend prek in 2008-2009)**

**“Control” Group  
(attend prek in 2009-2010)**



Weiland and Yoshikawa, 2013

# RD illustration



# Outcome Measures: Math, Language and Literacy Skills

- A trained assessor tested children one-on-one on a battery of tests, including:
  - Early math: *Woodcock-Johnson Applied Problems subscale* (Woodcock, McGrew & Mather, 2001) and *Research-based Early Math Assessment Short Form* (Weiland et. al, 2010)
  - Language: *Peabody Picture Vocabulary Test-III* (Dunn & Dunn, 1997)
  - Literacy: *Woodcock-Johnson Letter-Word Identification subscale* (Woodcock, McGrew & Mather, 2001)

# Outcome Measures: Executive Function Skills

- Executive Function:
  - Working memory: *Forward and Backward Digit Span* (Gathercole & Pickering, 2000; Wechsler, 1986 )
  - Inhibitory control: *Dimension Change Card Sort* (Frye, Zelazo & Palfai, 1995), *Pencil Tap* (Diamond & Taylor, 1996)
  - Attention shifting: *TOQ Attention* (Smith-Donald, et al., 2007)

# Measures: Emotional Development

- Emotional Development:
  - Emotion labeling: *Emotion Recognition Questionnaire* (Ribord, Camras, Stafani, & Spacarelli, 1988)
  - Positive emotion: *TOQ Positive Emotion*, (Smith-Donald, et al., 2007)
  - Impulse control: *TOQ Impulse Control* (Smith-Donald, et al., 2007)

# What sorts of questions?

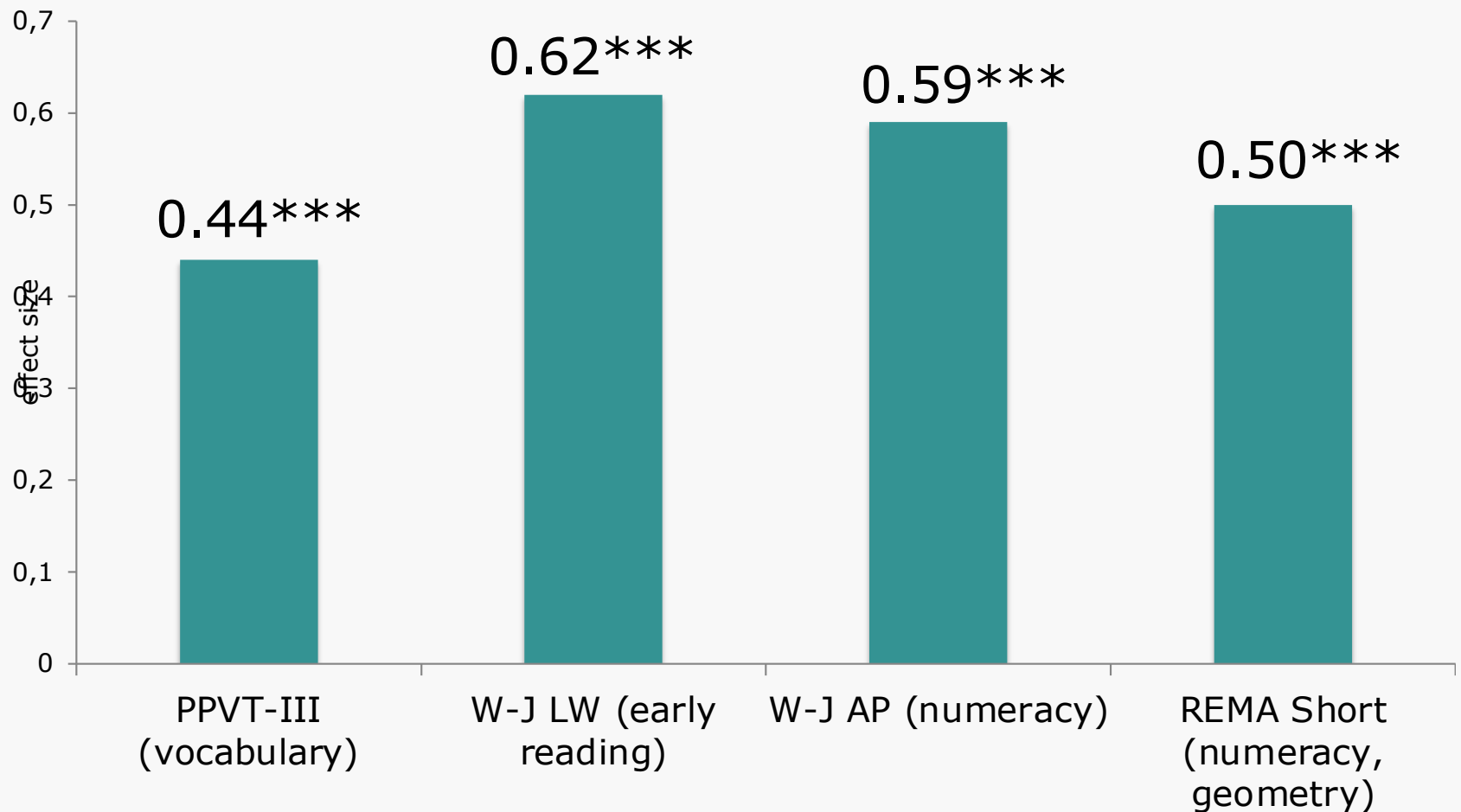
- Program effects
- Subgroup differences in effects
  - Which subgroups?
- Mediation by implementation features
  - What aspects of implementation?



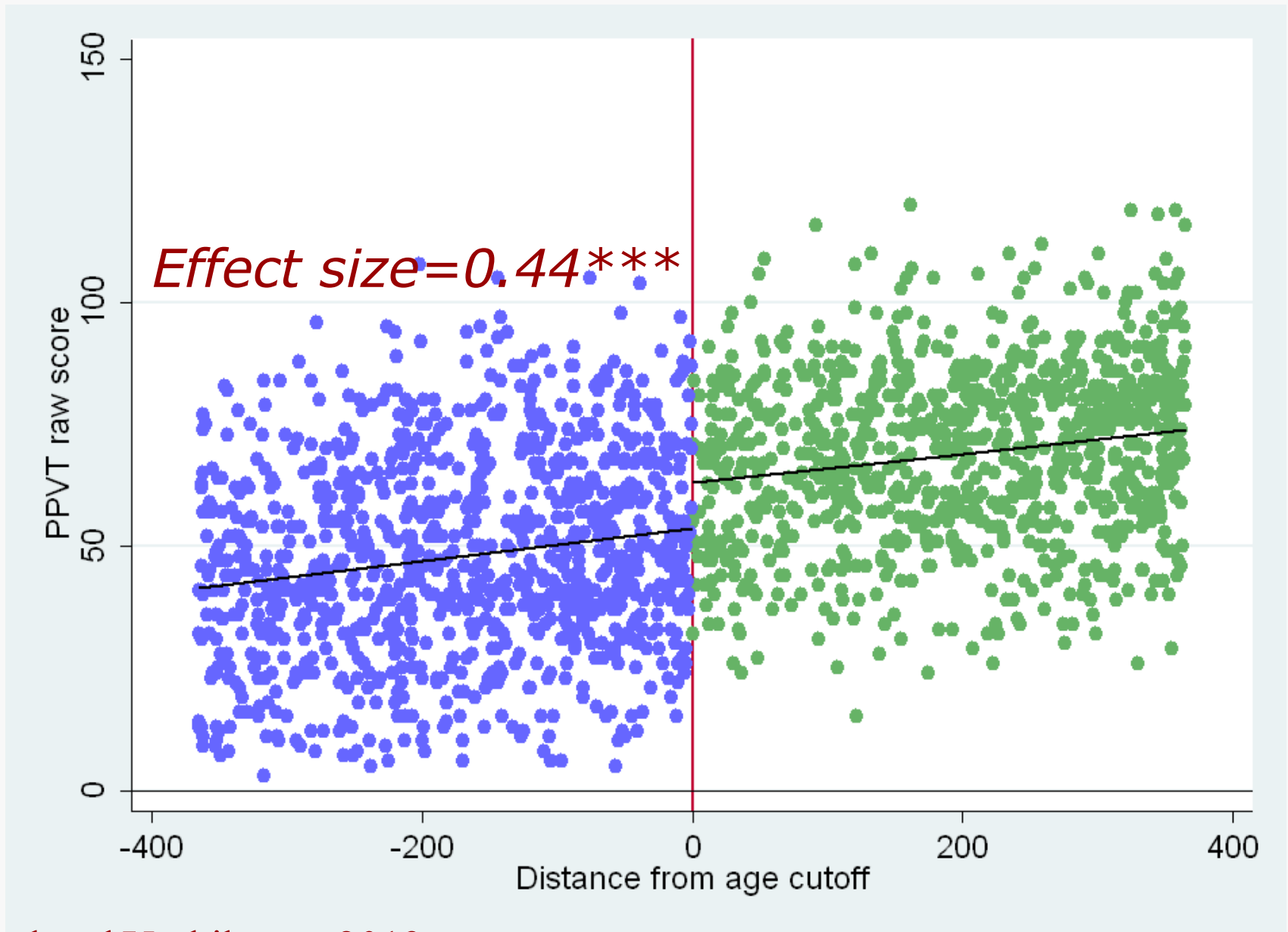
# Results: Format of child impacts

- Translated into effect sizes – a standardized measure that allows to compare results across studies
- Typical effect size in an educational intervention is around 0.20
- Small effect:  $<0.30$
- Moderate effect: 0.40-0.60
- Larger effect:  $>0.60$

## Results: Language, Literacy, and Mathematics

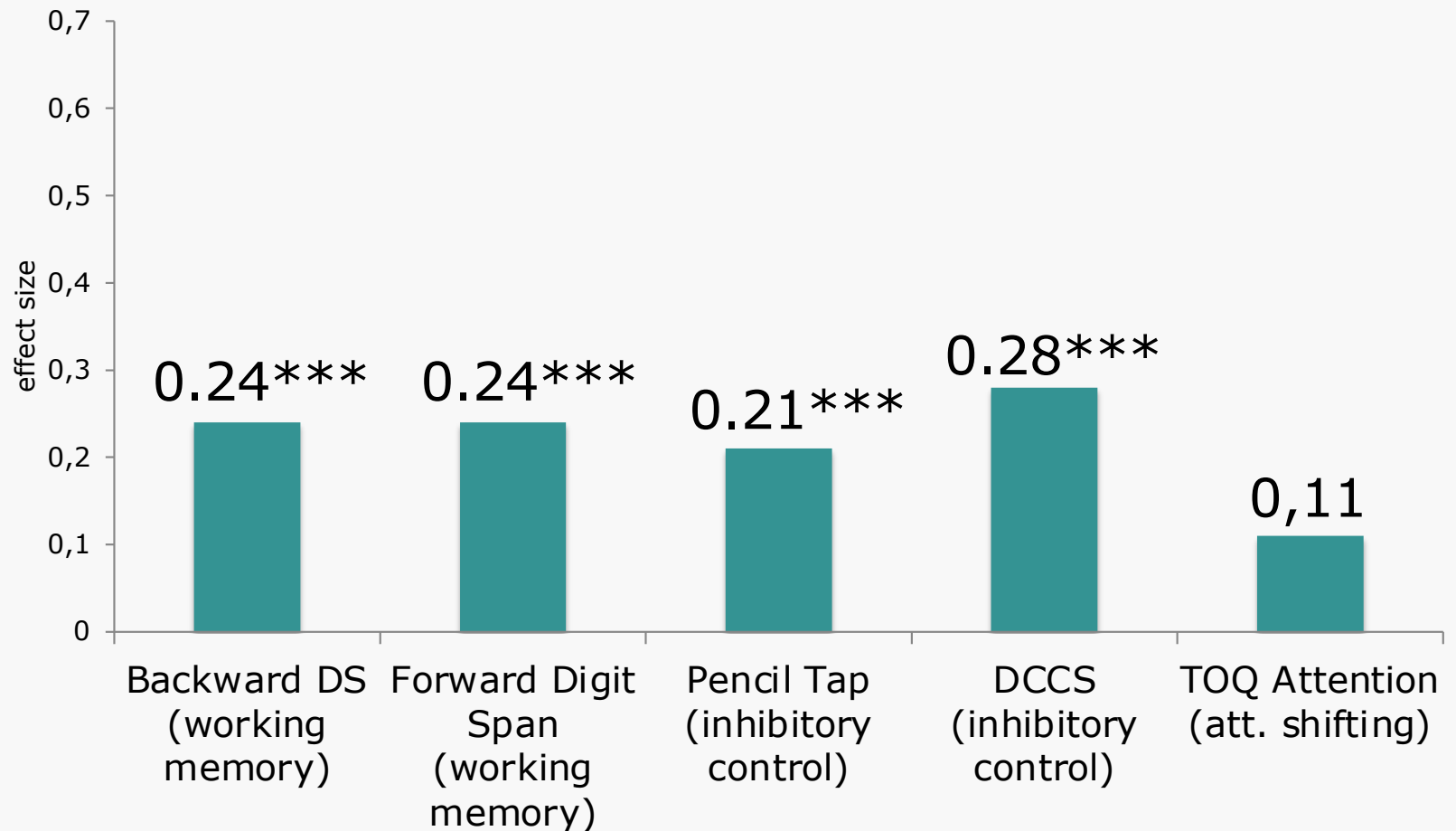


# Plot of the PPVT Effect

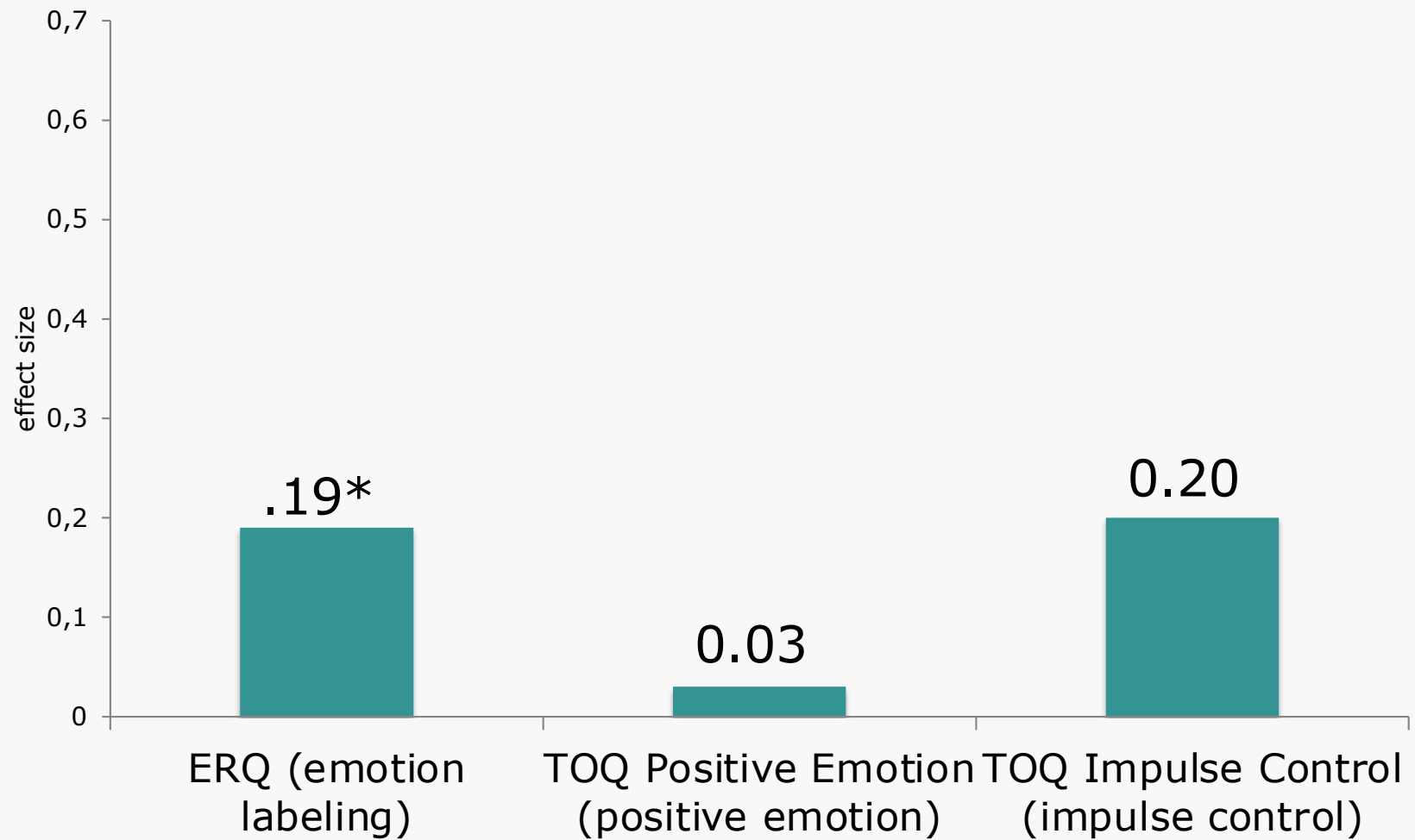


Weiland and Yoshikawa, 2013

# Results: Executive Function



## Results: Emotional Development/Regulation



## Comparison of Boston effects to other recent public preK evaluations

	PPVT-III	Letter Word Identification	Applied Problems	REMA Short
<b>Boston</b>	<b>0.44***</b>	<b>0.62***</b>	<b>0.59***</b>	<b>0.50***</b>
Tulsa 2005	--	0.80***	0.38*	--
Tulsa 2008		0.99***	0.36***	
Michigan	-0.16	--	0.47*	--
New Jersey	0.36*	--	0.23*	--
South Carolina	0.05	--	--	--
West Virginia	0.14	--	0.11	--
Oklahoma	0.29*	--	0.35	--
New Mexico, Y1	0.35+	--	0.38+	--
New Mexico, Y2	0.25+	--	0.50+	--
New Mexico, Y3	0.17+	--	0.43+	--

\*\*\*p<0.001; \*\*p<0.01; \*p<0.05

+ results statistically significant but level of significance not reported.

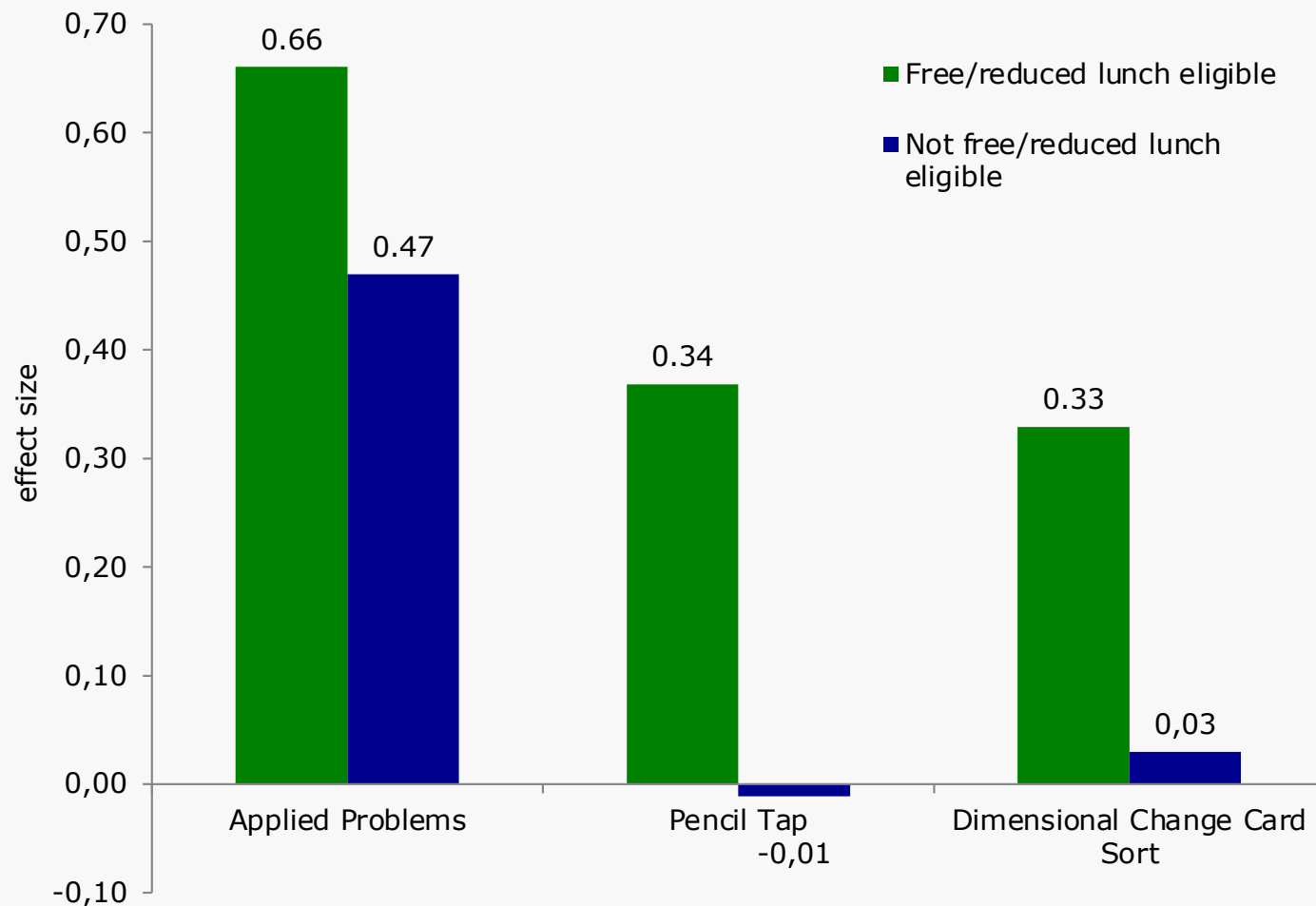
Citations: Tulsa (Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, & Gayers, 2008); MI, NJ, SC, WV, OK (Wong et al., 2007); NM (Hustedt, Barnett, Jung & Goetze, 2009).

Note: All cited studies use the standard deviation of the control group as the denominator in calculating effect sizes. Boston models all use a bandwidth of 365 days and linear functional form between the outcome and age.

## RQ2: Subgroup effects

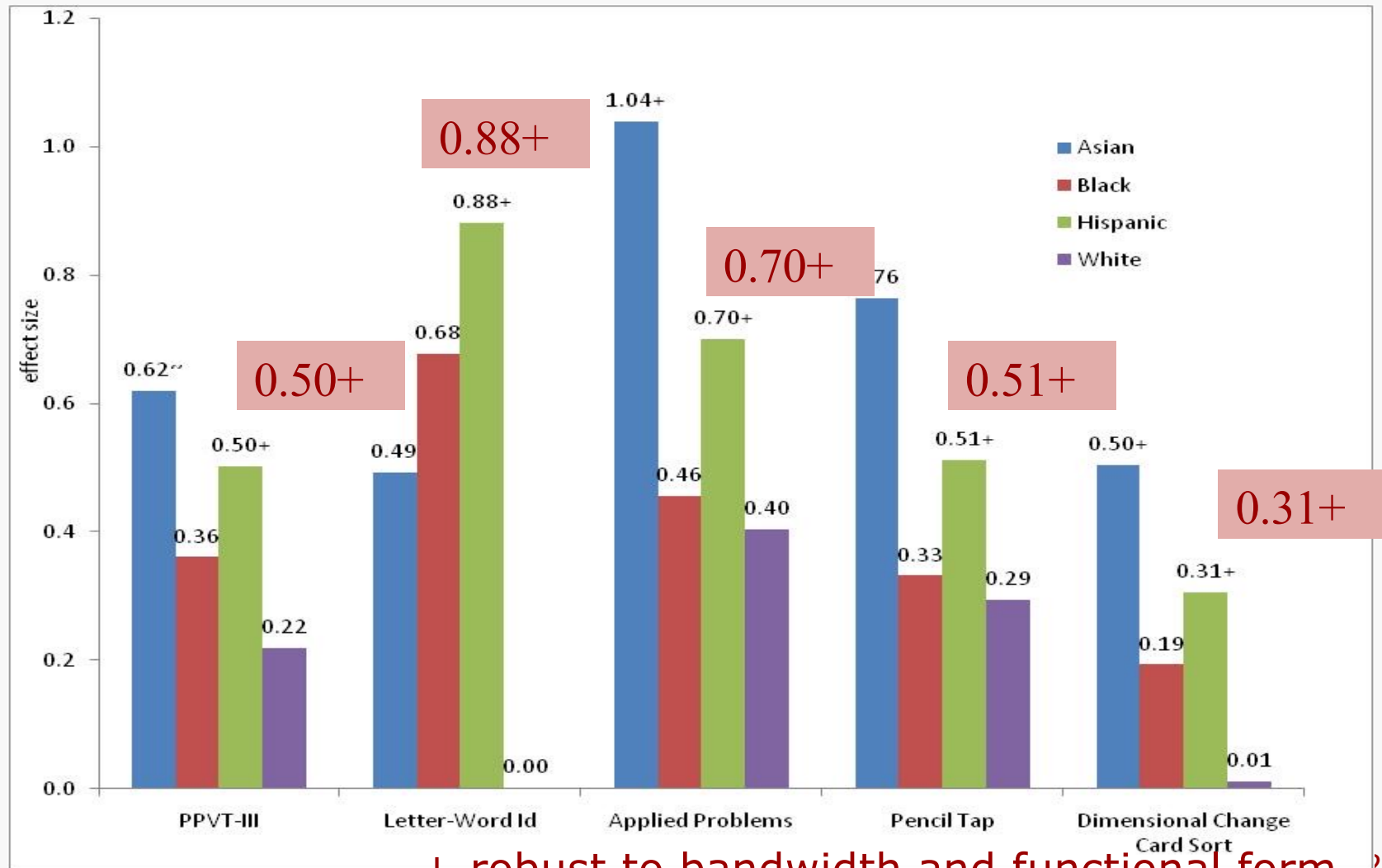
- Subgroups of interest:  
Free/reduced lunch,  
race/ethnicity, language, and  
gender
- Strategy: Same  
analytical/modeling approach  
but included interaction terms  
for subgroups of interest

## Results: Free/reduced lunch subgroup effects





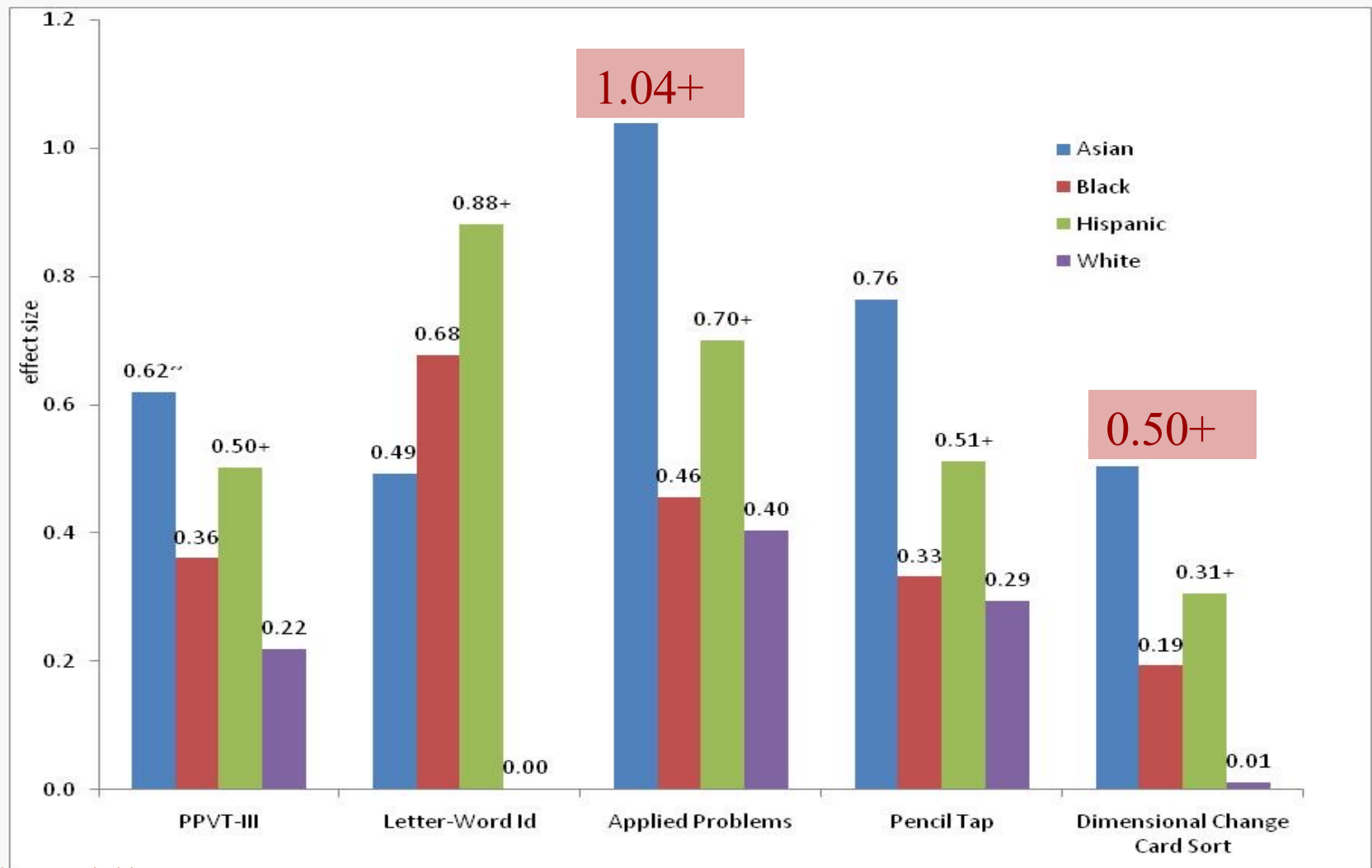
# Race/ethnicity subgroup effects



+ robust to bandwidth and functional form 25

Weiland and Yoshikawa, 2013 ~ not robust to bandwidth and/or functional form

# Race/ethnicity subgroup effects

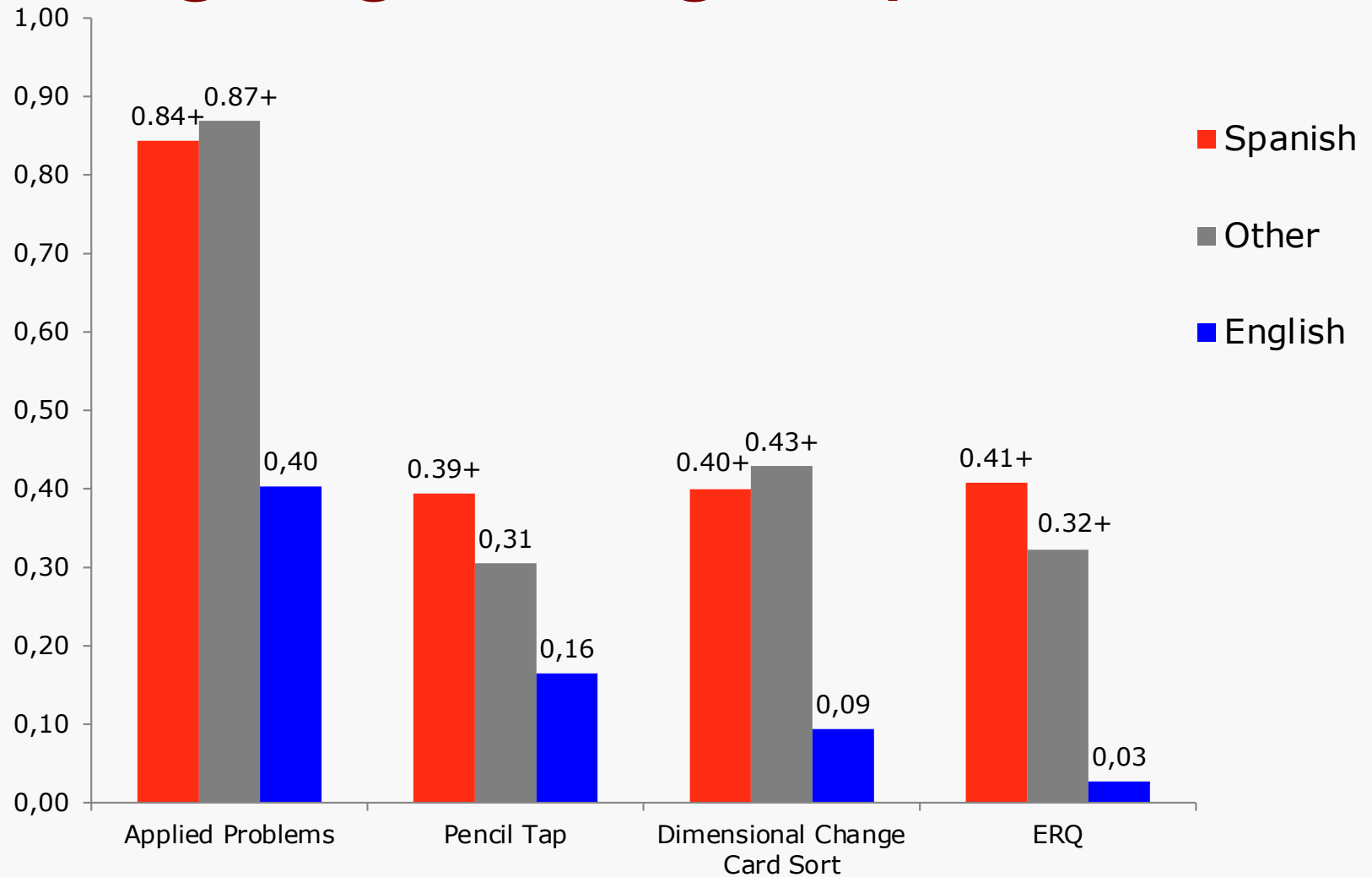


Weiland and Yoshikawa, 2013

+ robust to bandwidth and functional form

~ not robust to bandwidth and/or functional form

# Language subgroup effects



# Summary: Mathematics, Language, and Literacy

- Largest increases to date on vocabulary and mathematics in evaluations of public prekindergarten at scale
- Investment in curricula specific to these domains produced *substantial and meaningful* gains
  - Fidelity-to-curricula data suggest curricula implemented reasonably well
  - Consistent with theory and some empirical work (Clements, Sarama, Spitler, Lange & Wolfe, in press; Harrison, McLeod, Berthelsen, & Walker, 2009; NAEYC & NAECS/SDE, 2003)

# Summary: Executive Function

- Increases in executive function skills from targeting language and mathematics skills most likely due to the curricula
- Critical planning, attentional and self-regulation skills for later school success
  - Mechanism unclear but possibly due to spillover from cognitively focused curricula
  - Some parts of curricula align with EF, particularly math

# Summary: Emotional Development

- Increase in emotion recognition
  - Directly targeted by the OWL
- No impact on emotional outcomes that were not so strongly targeted by the curricula

# Limitations

- Results only generalize to students at the cutoff
- Results only generalize to children whose parents agreed to let them participate
- Cannot definitively identify the causal mechanisms behind detected effects

# ExCEL Study, 2015-2020: Issues Addressed

- Constrained vs unconstrained skills
- Maintenance of implementation quality
- Fadeout of child effects
- Network-related goals



## Research work: Part of IES Early Learning Network

- Network's aim:
  - Identify the malleable home, classroom, school, and system factors that promote children's gains from P-3



**CURRICULAR  
FEATURE**

**PREVIOUS PRACTICE**

**FOCUS CURRICULUM**

**Content of instruction**

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>■ Substantial repetition of preschool content in elementary school</li> <li>■ Lessons are focused on basic skill development, not integrated into thematic lessons directed at content knowledge</li> <li>■ Subjects (literacy, language, math, science, social studies) taught separately</li> <li>■ Shallow content instruction, spread across many content areas (e.g., 16 topics for language/literacy in kindergarten)</li> </ul> | <ul style="list-style-type: none"> <li>■ Content builds from preschool to second grade with little repetition</li> <li>■ Lessons are theme-based and focus on building critical thinking and content knowledge</li> <li>■ Connections are made across subject areas</li> <li>■ Deep content instruction (e.g., 4 themes for language/literacy in kindergarten, 6 in first grade)</li> </ul> |
|---|---|

**Format of instruction**

- |  |   |
|--|---|
| <ul style="list-style-type: none"> <li>■ Kindergarten/elementary school structures and formats not aligned with preschool</li> <li>■ Primarily whole-group</li> <li>■ Teacher-directed, with mostly passive listening and individual seatwork</li> </ul> | <ul style="list-style-type: none"> <li>■ Structures and formats mirror preschool</li> <li>■ Primarily small-group</li> <li>■ Student-directed, with teacher support</li> <li>■ Promotes active engagement with materials and tasks that relate to broader themes</li> <li>■ Project-based, including collaborative work with peers</li> </ul> |
|--|---|

## ExCEL P-3 Data Collection Overview

Data Collection Activity	Fall 2016	Winter/Spring 2017	Fall 2017	Winter/Spring 2018	Fall 2018	Winter/Spring 2019	Fall 2019	Winter/Spring 2020	Fall 2020	Winter/Spring 2021
Student Grade	PreK	PreK	K	K	1 <sup>st</sup> Grade	1 <sup>st</sup> Grade	2 <sup>nd</sup> Grade	2 <sup>nd</sup> Grade	3 <sup>rd</sup> Grade	3 <sup>rd</sup> Grade
Direct Child Assessment	X	X	X	X		X		X		X
Parent Survey	X		X		X		X		X	
Teacher Reports on Children	X	X	X	X		X		X		X
Videotape Observation (CLASS Scores) Live Fidelity Observation (Fidelity Scores)		X		X		X		X		X
Teacher Reports and Survey		X		X		X		X		X

Data Collection Activity	Fall 2016	Winter/Spring 2017	Fall 2017	Winter/Spring 2018	Fall 2018	Winter/Spring 2019	Fall 2019	Winter/Spring 2020	Fall 2020	Winter/Spring 2021
Student Grade	Pre K	PreK	K	K	1 <sup>st</sup> Grade	1 <sup>st</sup> Grade	2 <sup>nd</sup> Grade	2 <sup>nd</sup> Grade	3 <sup>rd</sup> Grade	3 <sup>rd</sup> Grade
<b>Direct Child Assessment</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>		X		X		X
Parent Survey	X		X		X		X		X	
Teacher Reports on Children	X	X	X	X		X		X		X
Videotape Observation (CLASS Scores)		X		X		X		X		X
Live Fidelity Observation (Fidelity Scores)										
Teacher Reports and Survey		X		X		X		X		X

## ExCEL P-3 Assessment Overview

ASSESSMENT	ALL KIDS?	Assessment Purpose/Notes
PreLAS	No	<ul style="list-style-type: none"> <li>• Y1, given to all students</li> <li>• Y2, only given to students who have Spanish listed as home language</li> <li>• Assesses student's language ability to take the assessment</li> </ul>
REMA	Yes	<ul style="list-style-type: none"> <li>• Not administered in Fall 2016</li> <li>• Math Assessment that was developed by the same people</li> </ul>
PPVT	Yes	<ul style="list-style-type: none"> <li>• Assesses language/vocabulary</li> </ul>
Renfrew Bus Story	No	<ul style="list-style-type: none"> <li>• Assesses student's autonomy in repeating/telling story</li> <li>• Only administered to students who were in the descriptive study</li> </ul>
Woodcock Johnson Applied Problems	Yes	<ul style="list-style-type: none"> <li>• Assesses math skills</li> </ul>
Hearts and Flowers	Yes	<ul style="list-style-type: none"> <li>• Assesses attention and inhibitory control/ executive functioning</li> </ul>
Digit Span	Yes	<ul style="list-style-type: none"> <li>• Assesses working memory and executive functioning</li> </ul>
Woodcock Johnson/Munoz Picture Vocabulary	No	<ul style="list-style-type: none"> <li>• Administered to small sample of students (&lt;100 for English/Spanish)</li> <li>• Assesses language</li> </ul>
PSRA	Yes	<ul style="list-style-type: none"> <li>• Assesses student's demeanor and behavior during the assessment</li> </ul>

# Fadeout Hypotheses, BPS Partnership

**Measurement:**  
Constrained vs.  
Unconstrained Skills

**Alignment:**  
Curriculum to Align Instruction  
across PreK & Kindergarten

**BPS RPP**

**Quality of Elementary  
School:**  
Sustained Effects Dependent  
on Quality of Kindergarten

**Peer Effects:**  
Sustained Effects Dependent  
on Kindergarten Peers

# Research questions

(McCormick, Maier, Weiland, Hsueh, Sachs, & Snow, 2018)

1. What does fidelity look like across prekindergarten public school classrooms in BPS?
  - Does fidelity vary systematically by classroom composition?
  
2. Is fidelity to the BPS PreK model associated with children's language and math scores in the Spring of PreK?
  - For which groups of students does fidelity appear most predictive of Spring outcomes (e.g., dual language learners, racial/ethnic minority students)?

School-level characteristic	% for study schools	% for school district
School structure: PreK – 5 <sup>th</sup> grade	30%	50%
School structure: PreK – 1 <sup>st</sup> grade	5%	8%
School structure: PreK – 8 <sup>th</sup> grade	55%	32%
% Students economically disadvantaged	48%	51%
% Students Black	26%	31%
% Students White	16%	16%
% Students Hispanic	46%	42%
% Students Asian	9%	6%
% Students whose first language is not English	49%	42%
% Met or exceeded expectations on 2015 – 2016 ELA exam	40%	36%
% Met or exceeded expectations on 2015 – 2016 math exam	44%	42%

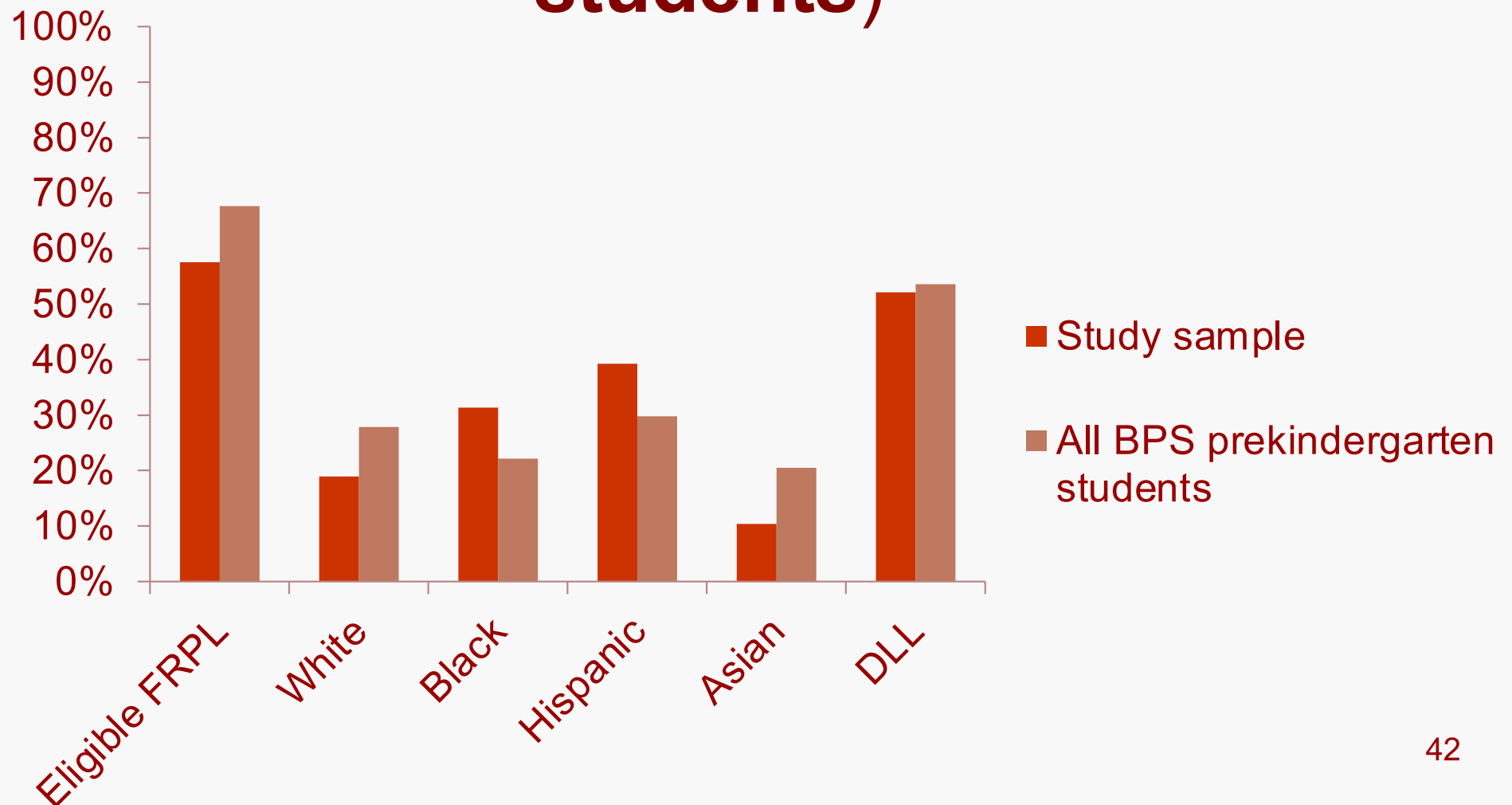


# Classroom & teacher participants

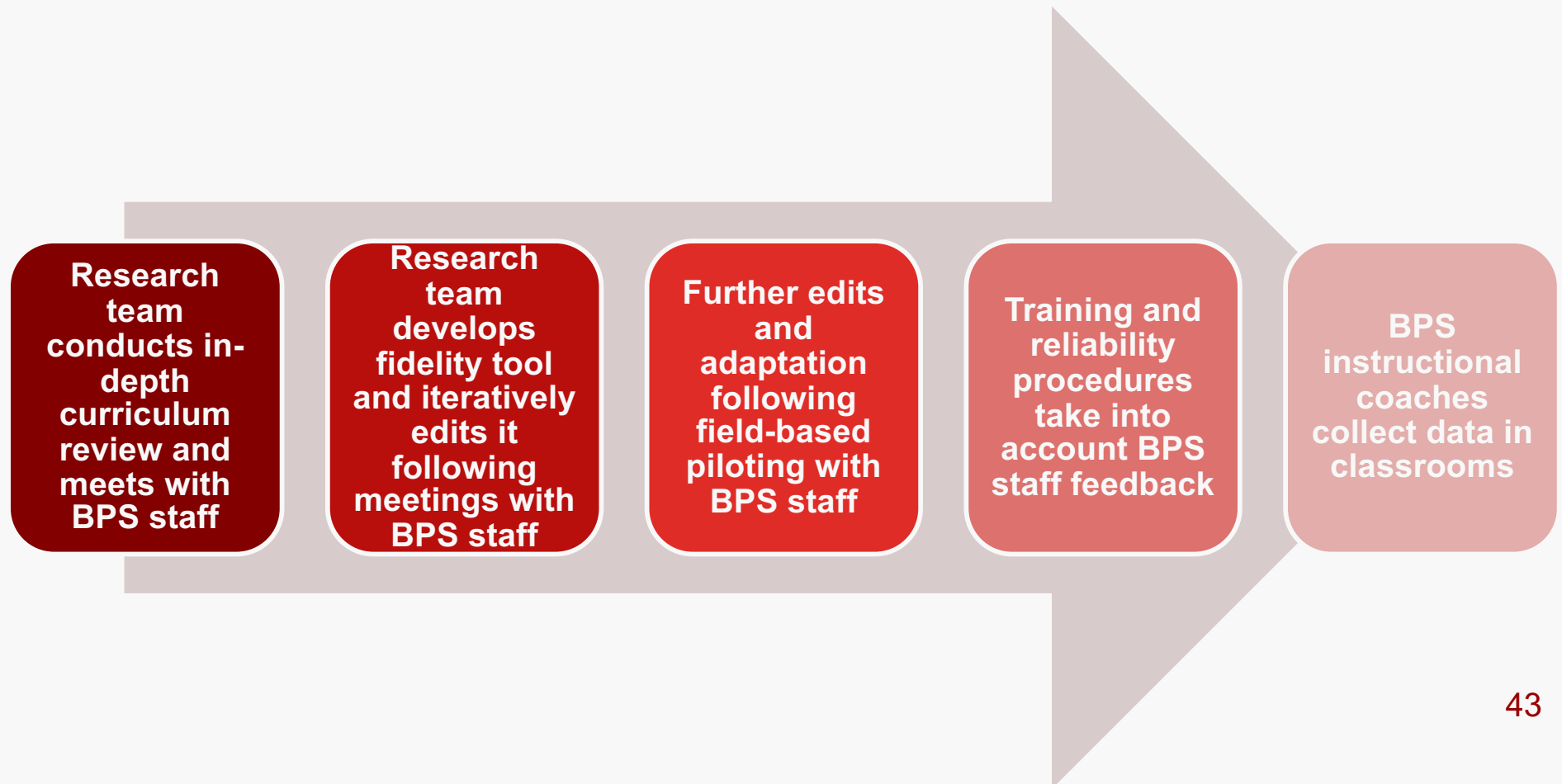
(*N* = 41 public school classrooms in 20 schools)

Teacher characteristic	%age/Mean
Teacher age	44.0 (SD = 9.4)
Years teaching	14.8 (SD = 9.3)
Years teaching prekindergarten	8.6 (SD = 7.4)
Years teaching at current school	7.8 (SD = 8.0)
Teacher has master's degree	90%
Teacher female	100%
Teacher Black	22%
Teacher White	49%
Teacher Hispanic	13%
Teacher Asian or other race	16%
Classrooms per school	1.35 (SD = .42)

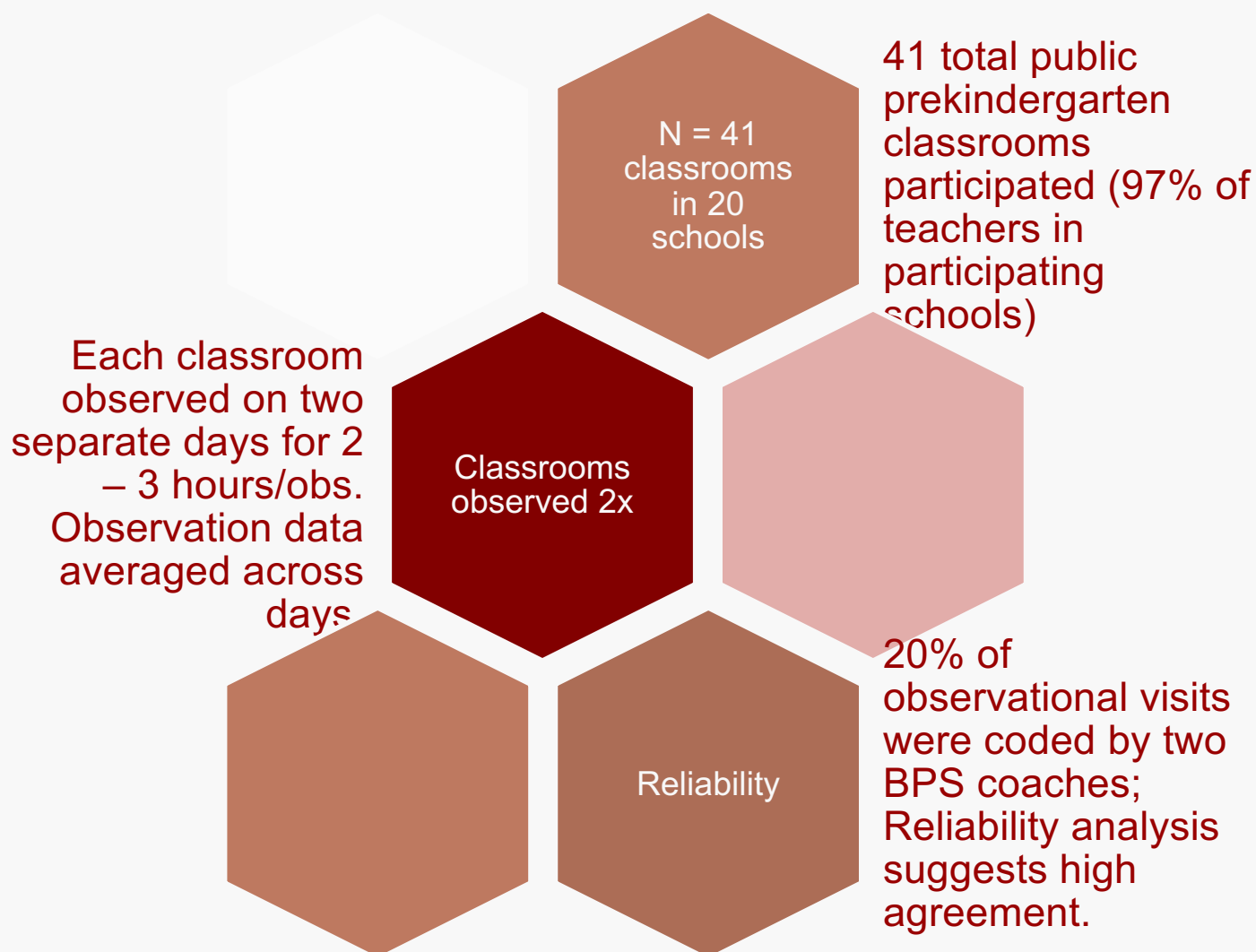
# Student sample (*N* = 299 BPS prekindergarten students)



# Research & BPS teams Co-construct Tool to Measure Fidelity of Implementation



# Fidelity Data in Public School Classrooms



# Example fidelity items

## Q17. Teachers and children have sustained, substantive discussions around unit content throughout Centers.

(Talking turn = teacher speaks, child returns; 2 turns = teacher speaks, child returns, teacher returns)

- (5) Most teacher-child interactions (90% or more) can be characterized by **sustained and substantive interactions (5+ turns) around unit content** where teacher prompts for extended talk, asks follow-up questions and children provide contingent responses
- (4)
- (3) Interactions are a mix of shallow, brief interactions and a few sustained interactions **(5+ turns)**
- (2)
- (1) Most interactions between teacher and children are **brief (1-2 turns)**

## Q12. The teacher encouraged mathematical reflection.

- (5) Highly intentional in stating “big idea” (e.g., these are different ways to count, to tell us how many) and relating activity to previous ones or real world (e.g., we count to keep track, like when taking attendance...what else do we track?)
- (4)
- (3) Provides cursory reflection on the activity; may not be at the level of “big idea”
- (2)
- (1) No mention of big math ideas or relation to previous activities or real world

## Analyzing fidelity data

1. Examine dosage, adherence, and quality of implementation
2. Examine fidelity scores within curriculum components
3. Consider variation within and across components
4. In order to make fidelity more relevant to district - create measures that **cut across components** and **operationalize core practices** that are central to curriculum

# Cross-component fidelity measures

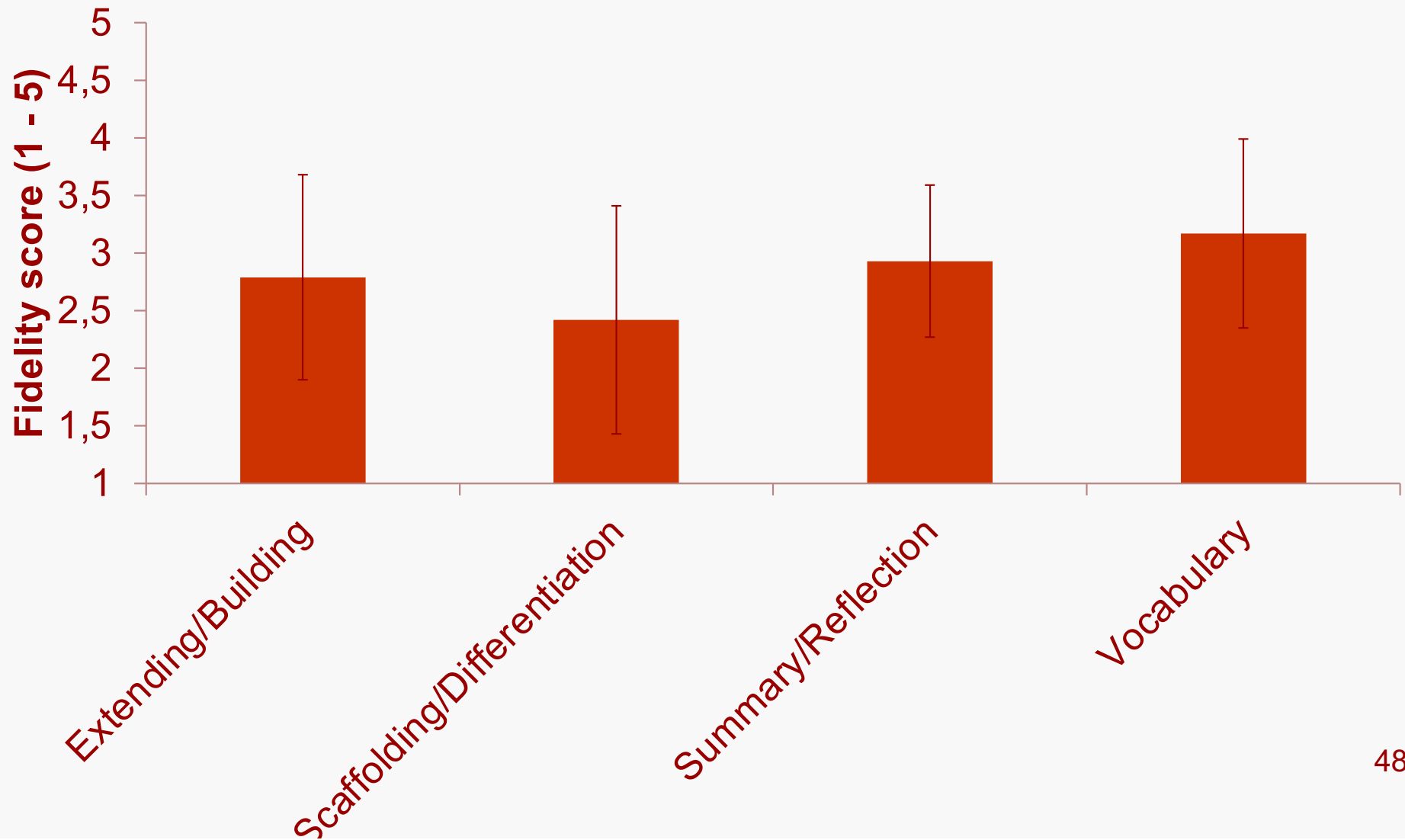
Vocabulary  
( $\alpha = .91$ )

Extending/Building  
( $\alpha = .91$ )

Summary/  
Reflection/Making  
Connections  
( $\alpha = .79$ )

Scaffolding/  
Differentiation  
( $\alpha = .82$ )

# What does fidelity look like overall in BPS public school prekindergarten classrooms?





## How do fidelity measures relate with CLASS?

	Instructional support	Emotional support	Classroom org.
Instructional support	1.0		
Emotional support	.67	1.0	
Classroom org.	.69	.85	1.0
Extending/Building	.18	.16	.10
Summary/Reflection	.22	.10	.14
Vocabulary	.01	.01	-.07
Scaffolding/ Differentiation	.35	.21	.22 <sub>49</sub>

# A couple of conclusions

Curricula and professional development are key to the process of change.

Quality improvement is iterative.

Quality improvement is iterative!

# Implications: Policy and Practice

- Adds to evidence base for publicly funded Pre-K
  - First evidence of causal effect on EF and emotion recognition
- High-quality coaching system can be implemented to support two curricula
- Math results particularly compelling
- Some evidence of larger effects for some subgroups on some assessments (particularly Latino students), but benefits largely accruing to everyone
- Contributes to discussion around the choice between increasing access and improving quality

# UPK Expansion

- Same measures as in the ExCEL study
- +/- 30 classrooms
- Lots of qualitative data about coaching in particular
- Hoping for funding!