Lessons from the Marshmallow Test

Stephanie M. Carlson, Ph.D.
Institute of Child Development
University of Minnesota
Reflection Sciences

Federal Reserve Bank of Minneapolis 2018

MEASURE WHAT MATTERS!

- *Disclaimer*: Stephanie Carlson and Philip Zelazo are entitled to royalties from the sale of the Minnesota Executive Function Scale by Reflection Sciences, LLC

- The University of Minnesota has a right to receive royalties from the sale of the Minnesota Executive Function Scale

Mischel and Ebbeson (1970)

Walter Mischel, PhD
1930-2018

“Marshmallow Test” is a Measure of Executive Function

- Working memory

I can remember the goal of this activity

- Inhibitory control

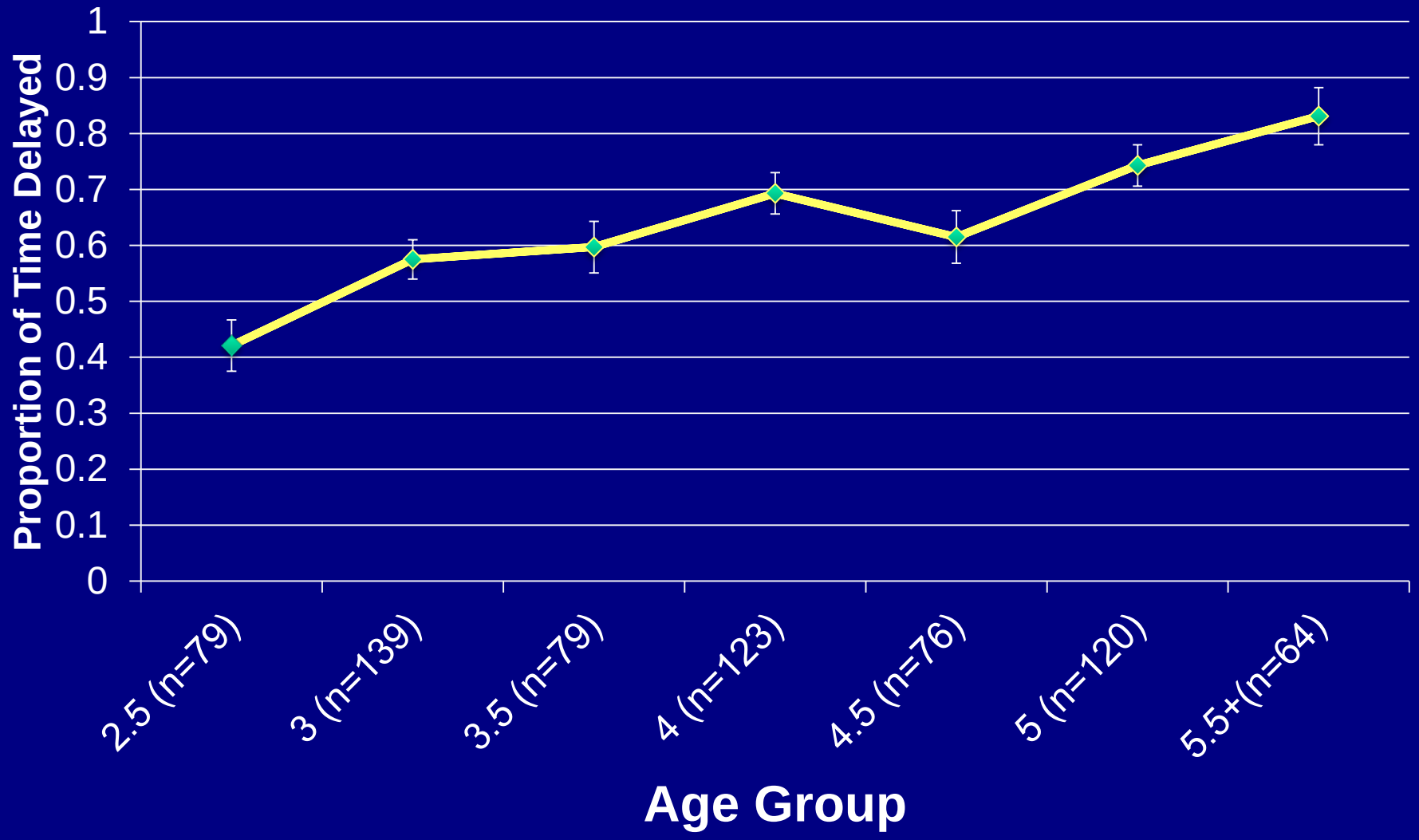
I can wait to respond and control my impulses

- Flexibility/shifting

I can think about this in a new way or distract myself

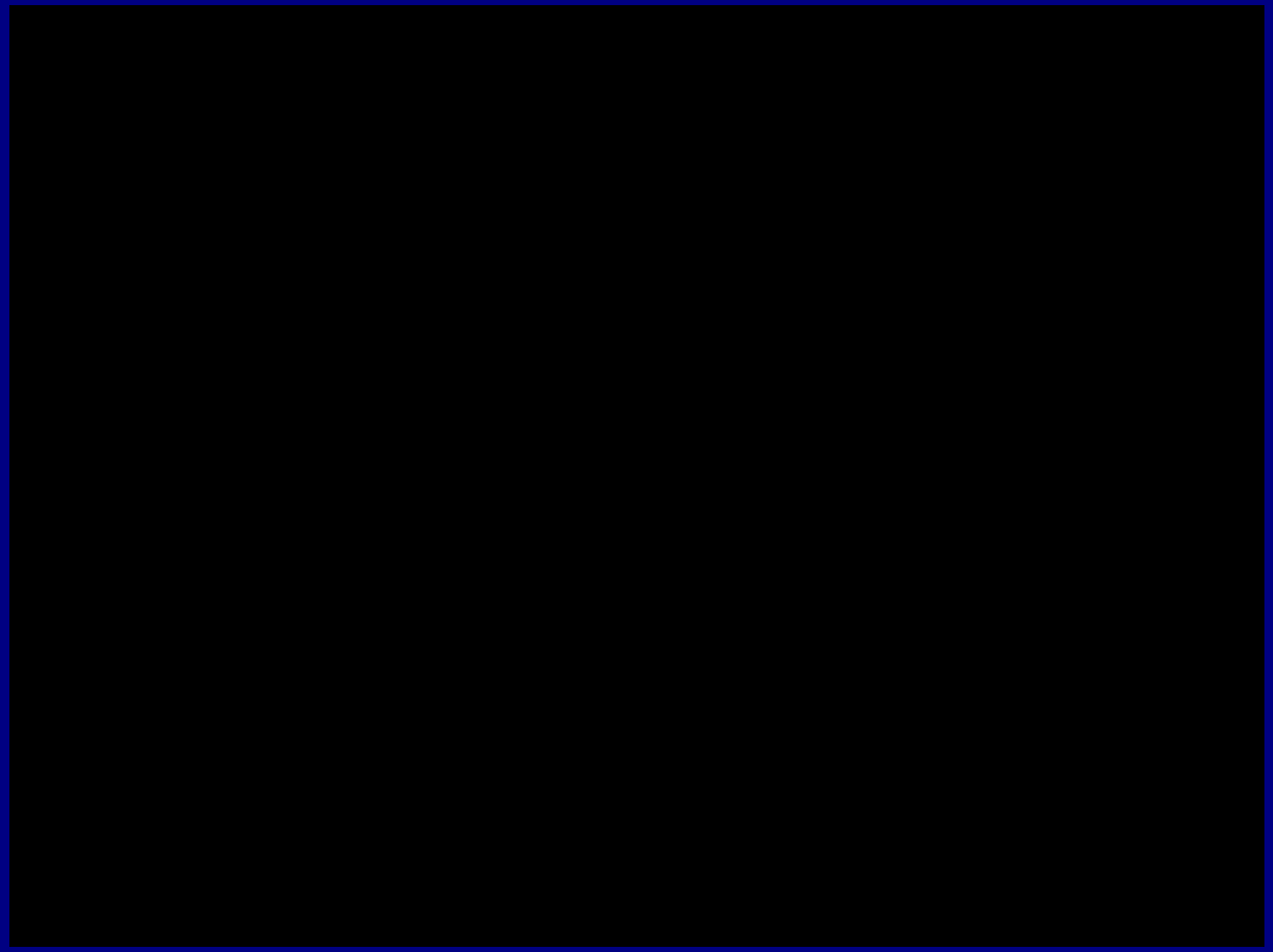
Mischel and colleagues

Proportion of Time Delayed as a Function of Age Group

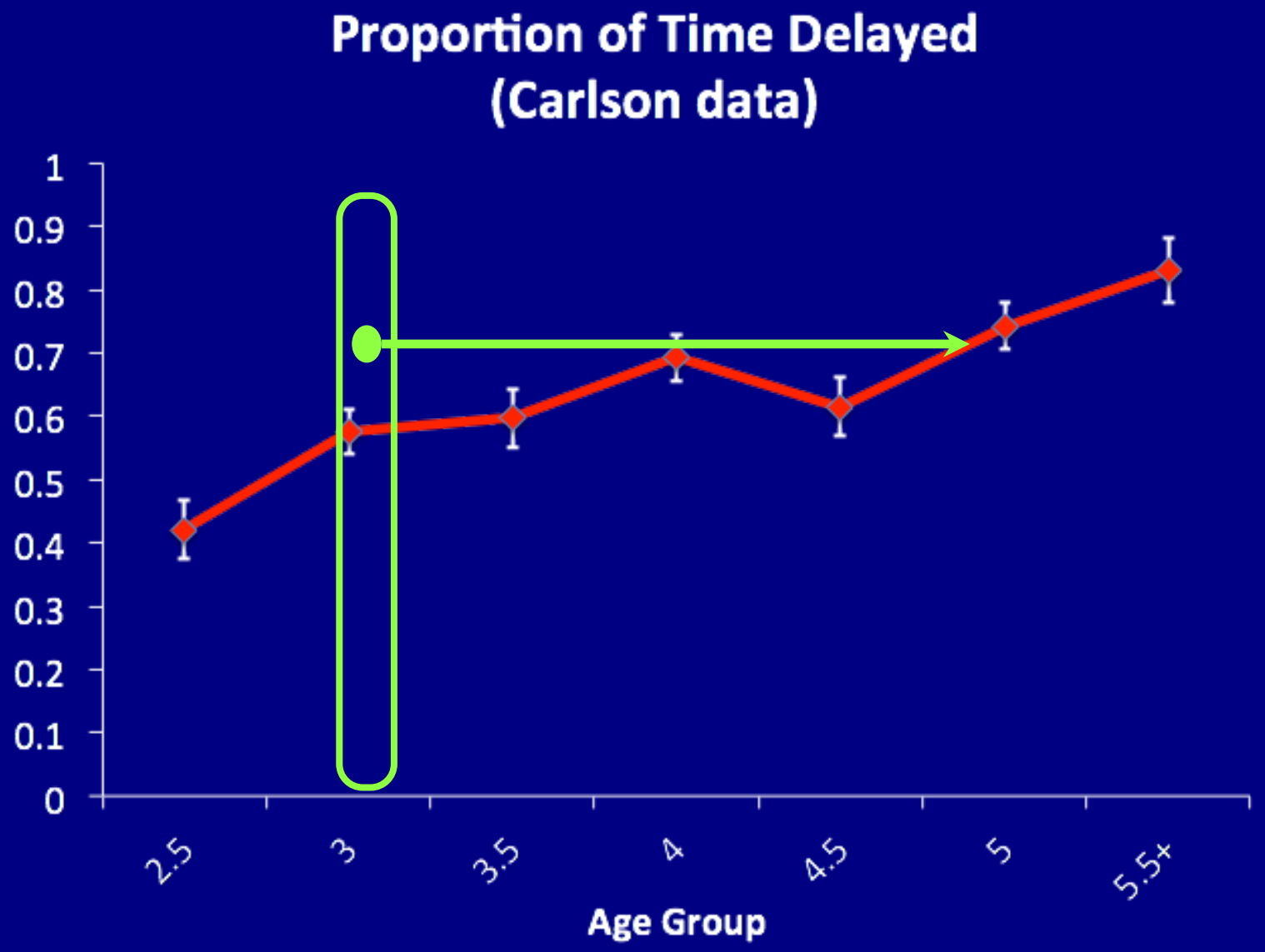


$F(6, 680) = 8.46$, $p < .001$; Carlson (2011)

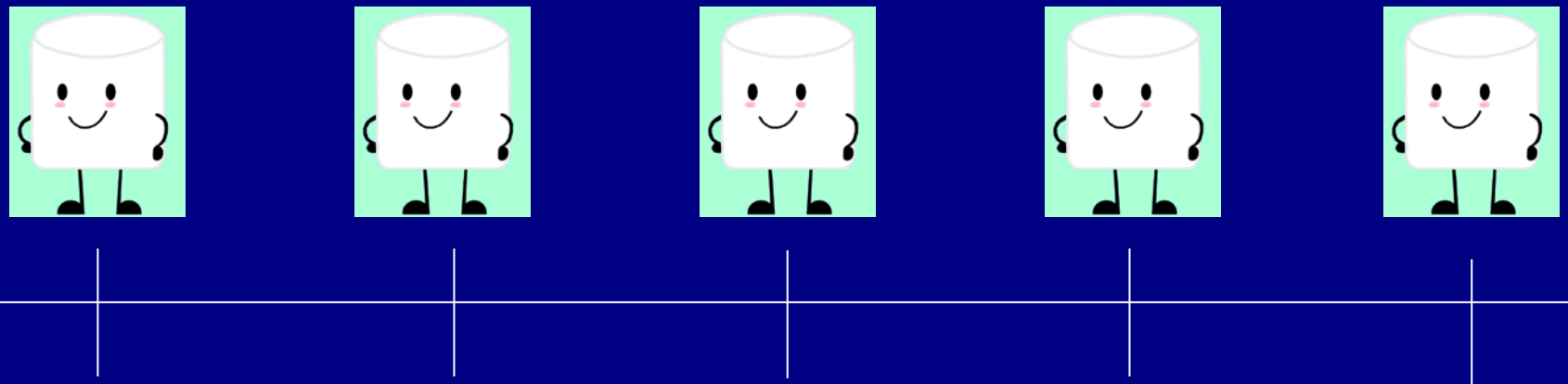
Reflection Self-talk



How did Jane do?



Long-term Outcomes Associated with Delay of Gratification at Age 4



4

Adolescence

- EF (Go-NoGo)
- Academic
- Social
- Coping with stress
- SAT scores

20s

- Fewer interpersonal and drug problems
- Higher ed level
- Better coping

30s

- Goal setting
- Self-regulation
- Lower BMI

40s

- Better regulated with reward stimuli, behaviorally and in the brain

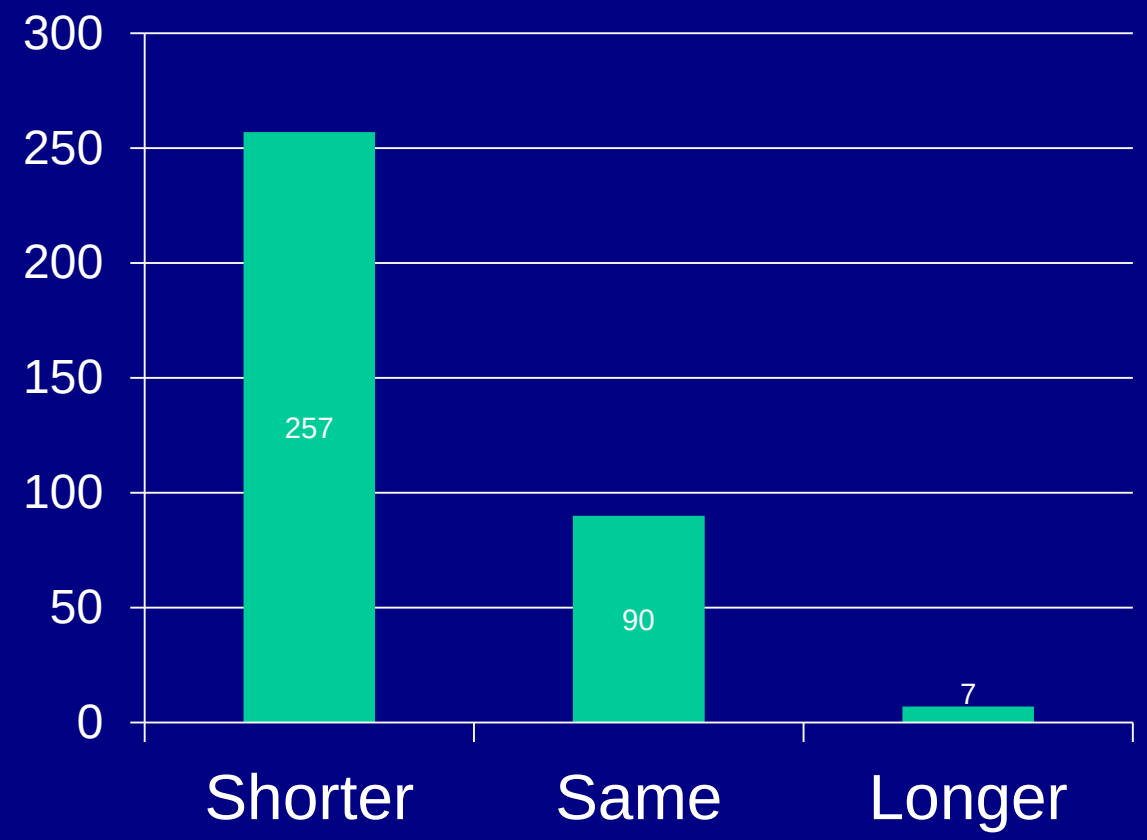
Has Children's Ability to Delay Gratification Changed Over Time?

"The children now love luxury; they have bad manners, contempt for authority; they show disrespect for elders and love chatter in place of exercise. Children are now tyrants, not the servants of their households. They no longer rise when elders enter the room. They contradict their parents, chatter before company, gobble up dainties at the table, cross their legs, and tyrannize their teachers."

U.S. National Survey Study

- Delivered via Amazon's Mechanical Turk (MTurk)
- 354 Adults (49% female)
- Age 20-69 years (*M* = 36)
- White (83%); African American (6.8%); Asian 6.8%); Hispanic (5.6%)
- 41 different states
- Household income $25k-200k (Median = $25k-50k)
- 54% were parents

When compared with children 50 years ago, do you think children today would wait…



Cohort Study

- 1960s: *N* = 165 (Stanford)
- 1980s: *N* = 135 (Barnard)
- 2000s: *N* = 540 (UWA and UMN)
- Ages 3-5 years; ~50% female
- Similar race/ethnicity and SES
- Followed the standard procedure w/ both rewards visible and a bell

What did the data show?

Why did adults get it wrong?

- Persistent complaints about “kids today”
  - Socrates
- Lack of perspective taking
  - Children *do* have less self-control than *you*
- Concerns about technology
- Rising standards

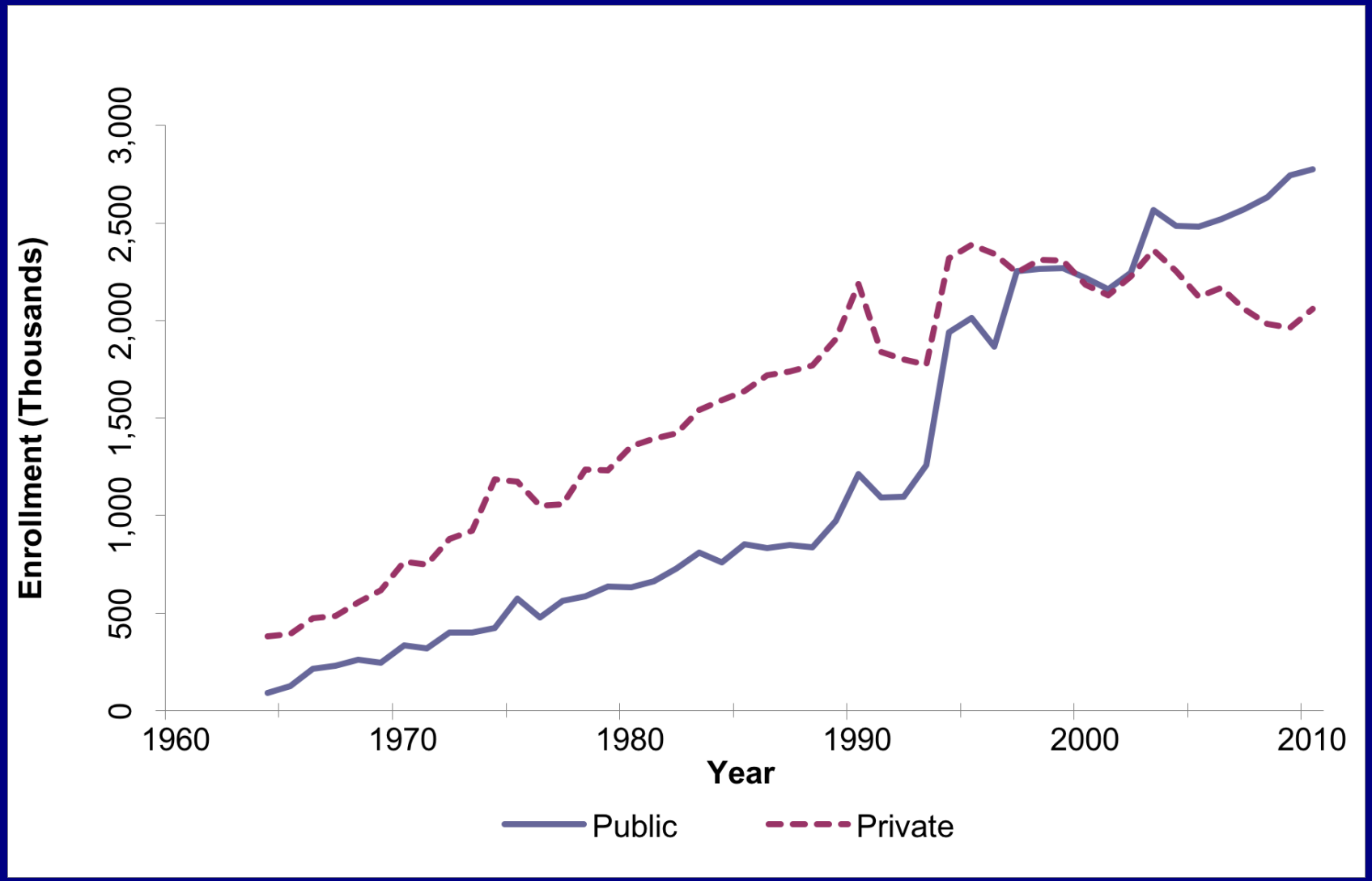
“As we care about more of humanity, we’re apt to mistake the harms around us for signs of how low the world has sunk rather than how high our standards have risen.”

--Steven Pinker (2017) *Enlightenment Now*

Why has delay of gratification improved?

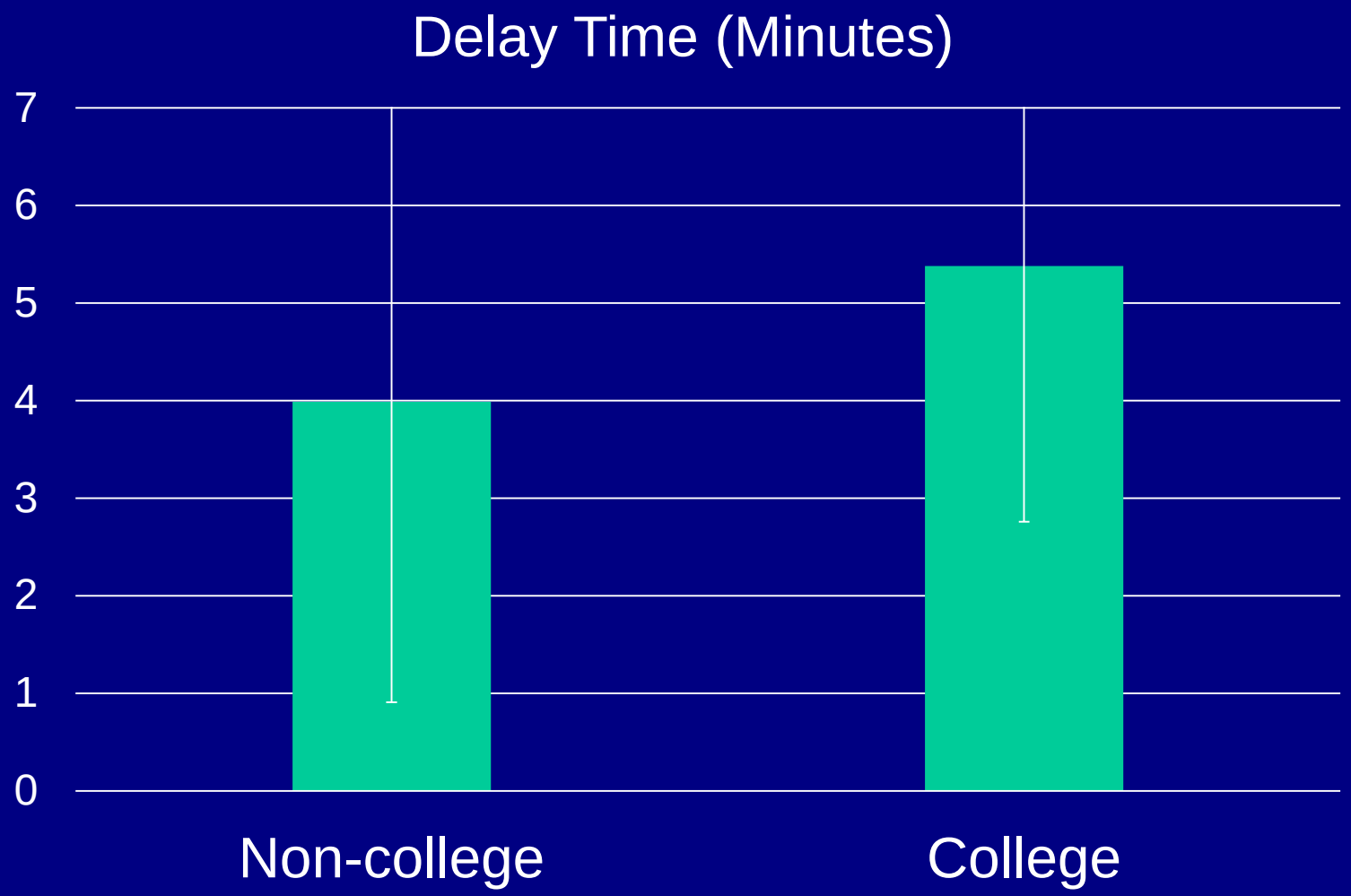
- The basics: GDP, nutrition, etc.
- Abstract thought -- a benefit of technology?
- Parenting -- more autonomy-supportive
- Preschool education

Preschool Enrollment



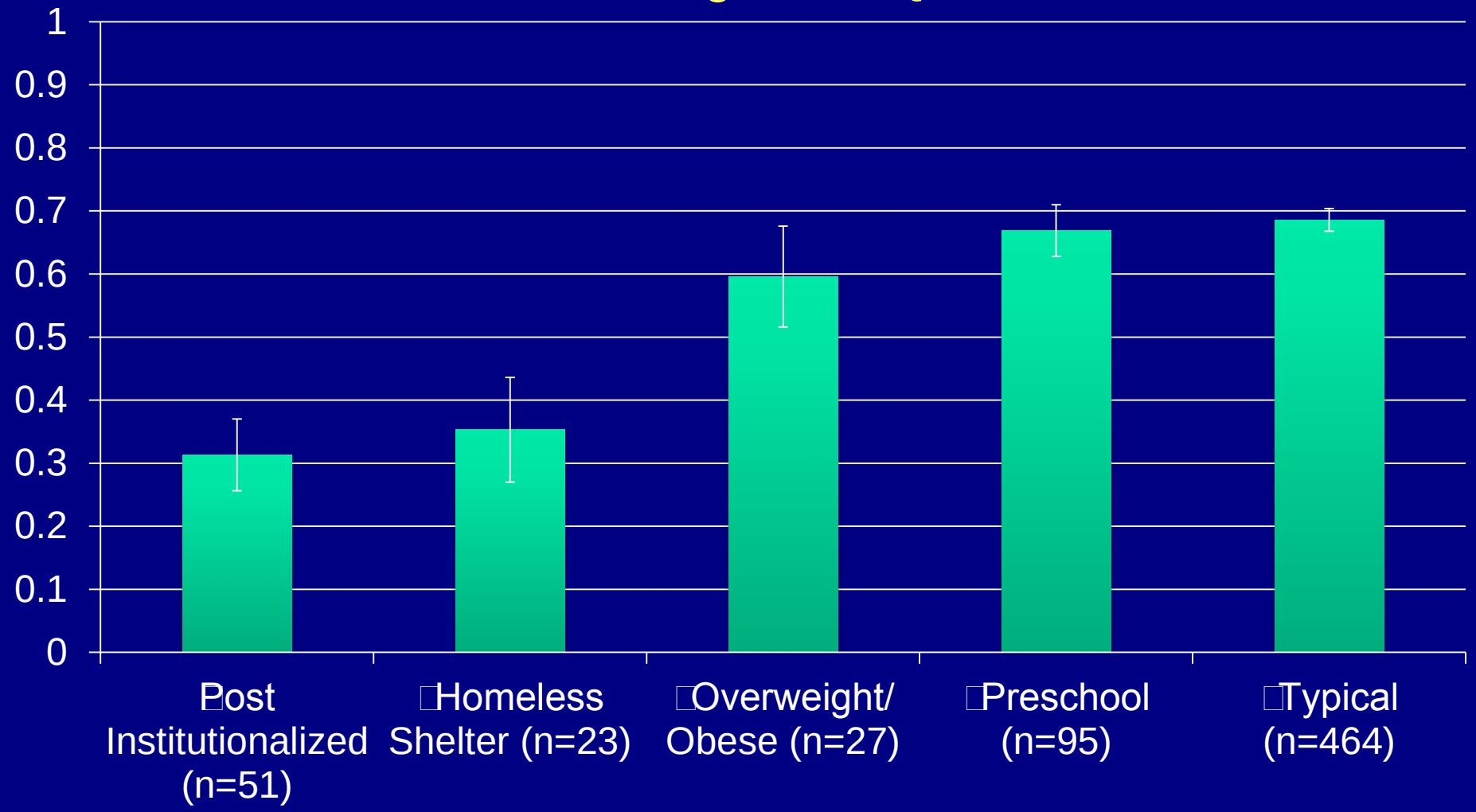
Karch (2013)

Our work is far from over…



Watts et al. (2018)

Proportion of Time Waited by Population Adjusting for Age and IQ



$F(6, 576) = 7.08$, $p < .001$; Carlson (2011)

Adaptation in Low-Trust, Low Resource Environments?

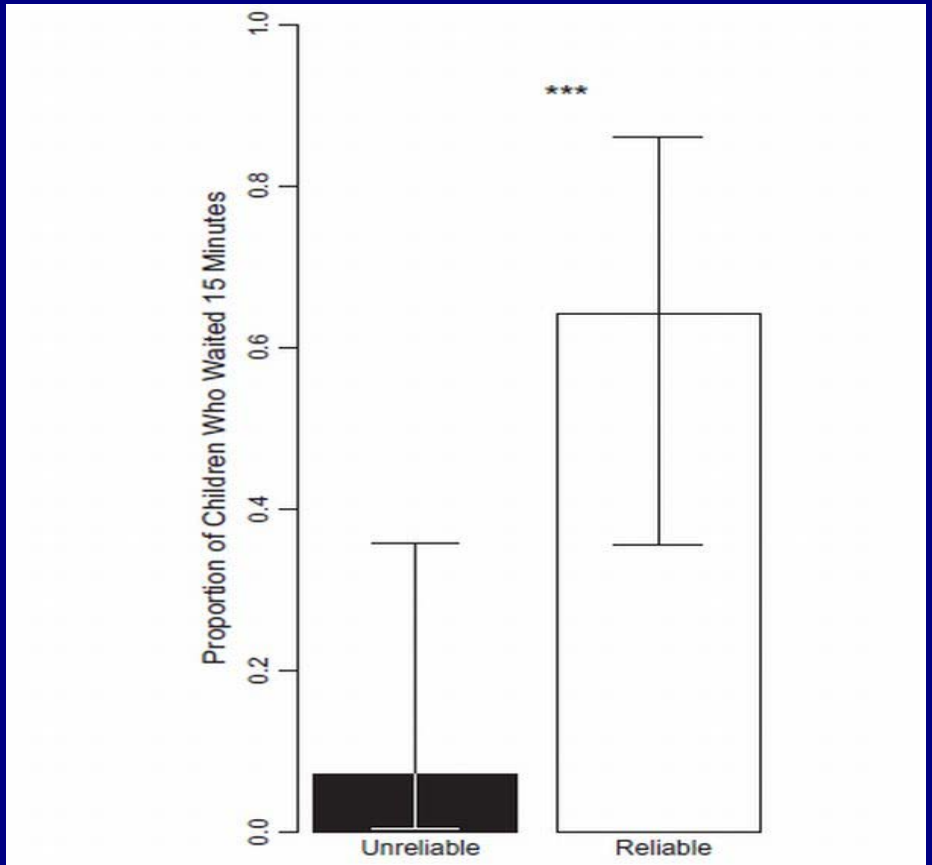


Fig. 2. Proportion of children who waited the full 15 min without eating the marshmallow by condition. Error bars show 95% confidence intervals. In the unreliable condition, only 1 out of the 14 children (7.1%) waited the full 15 min; in the reliable condition, however, 9 out of the 14 children (64.3%) waited. We tested the difference using a two-sample test for equality of proportions with continuity correction at $\alpha_{2\text{-tail}} = 0.05$. The test found it to be highly significant ($X^2 = 7.6222$, $df = 1$, $p < 0.006$).

Kidd et al., 2012

Limitations of the Marshmallow Test

- Food rewards
- Trust
- Not below age 3
- Not above age ~10
- Has a bimodal distribution (“fail” or “pass”)
- Not good for repeated assessment
- Not practical for school settings
- Not age-normed

Minnesota Executive Function Scale (MEFS™) App

Carlson & Zelazo, 2014

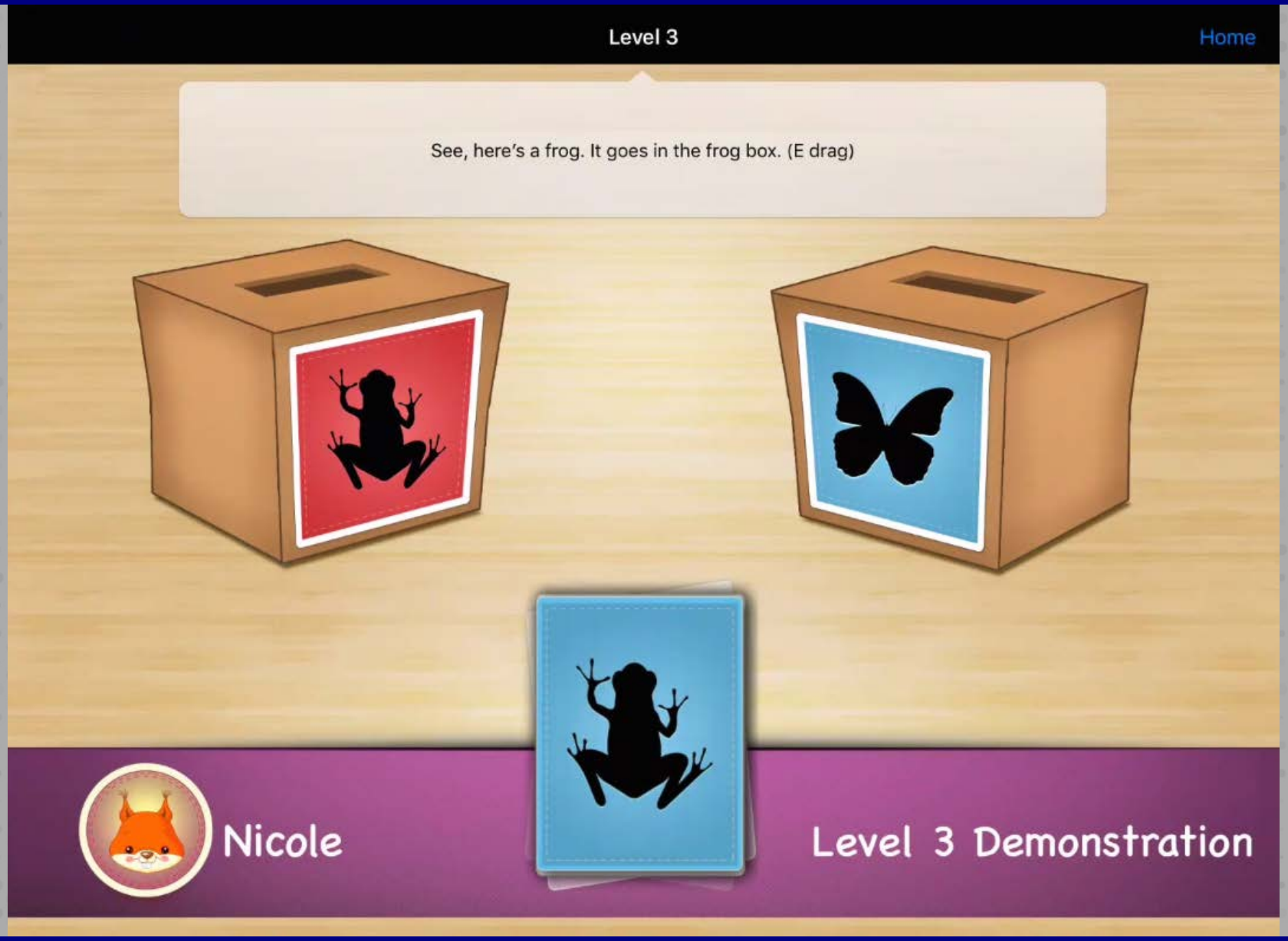
Level 3

Home

Look, I have these boxes here. This one has a frog on it and this one has a butterfly on it. This is the shape game. In the shape game, all the frogs go here and all the butterflies go here. (tap + button)

Nicole

Level 3 Demonstration


Level 3
Home
See, here's a frog. It goes in the frog box. (E drag)
Nicole
Level 3 Demonstration


Level 3
Home
And here's a butterfly. It goes in the butterfly box. (E drag)
Nicole
Level 3 Demonstration

The 3 Facets of EF

Cognitive Flexibility

Being able to switch between rules

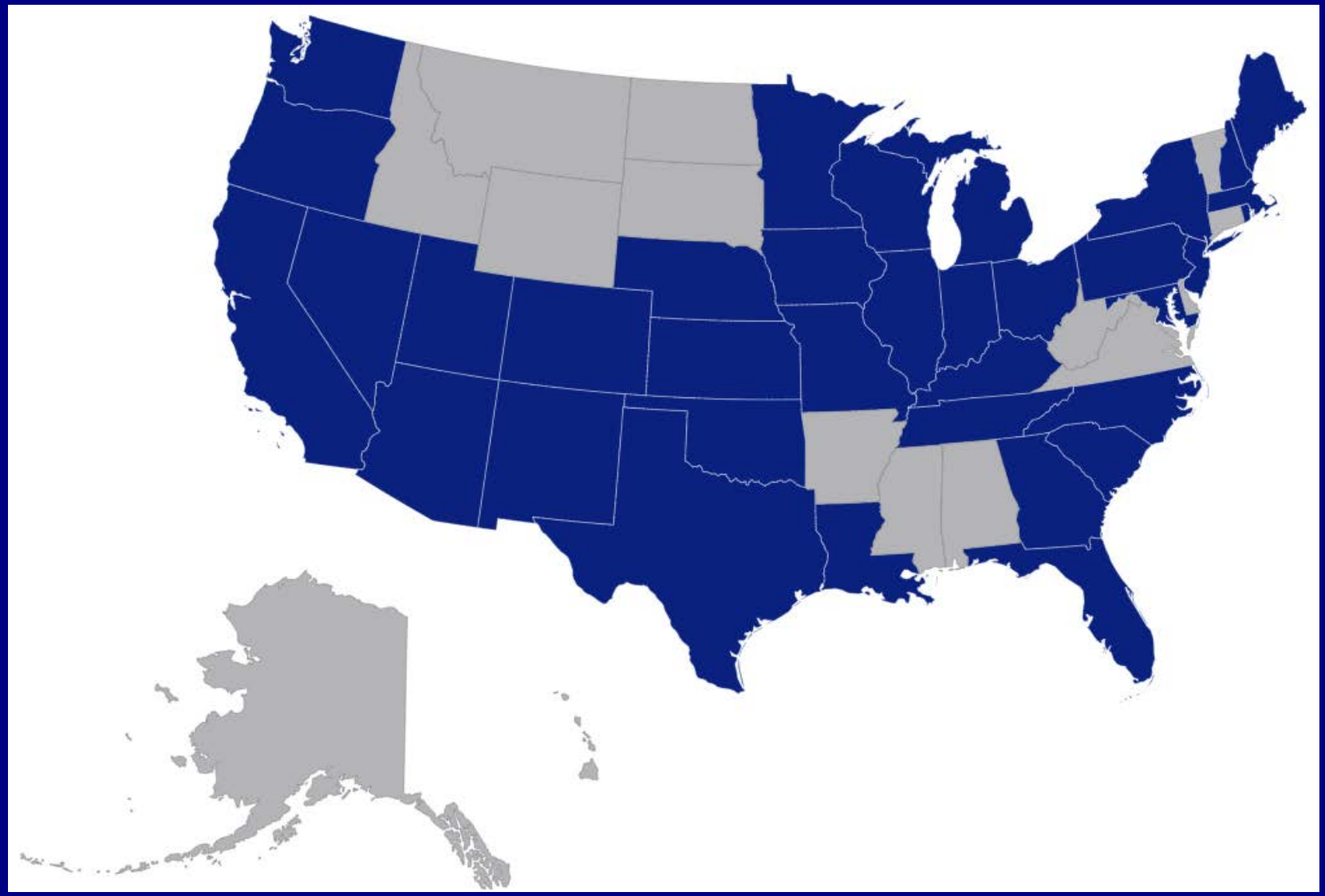
Working Memory

Remembering the rule and applying it in the correct way on a given trial

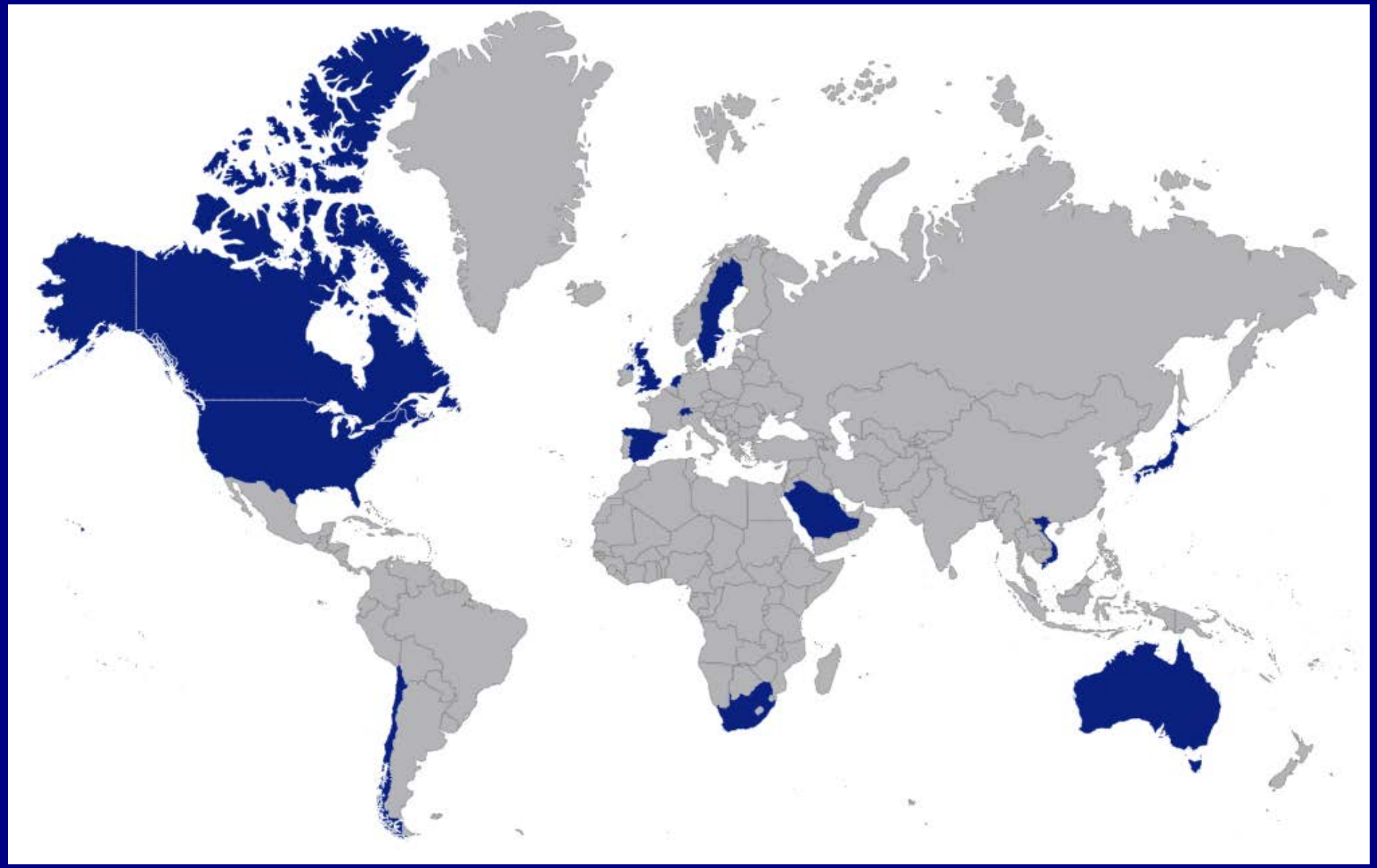
Inhibitory Control

Resisting reflexive response of sorting by the initial rule in order to sort by the correct but conflicting rule

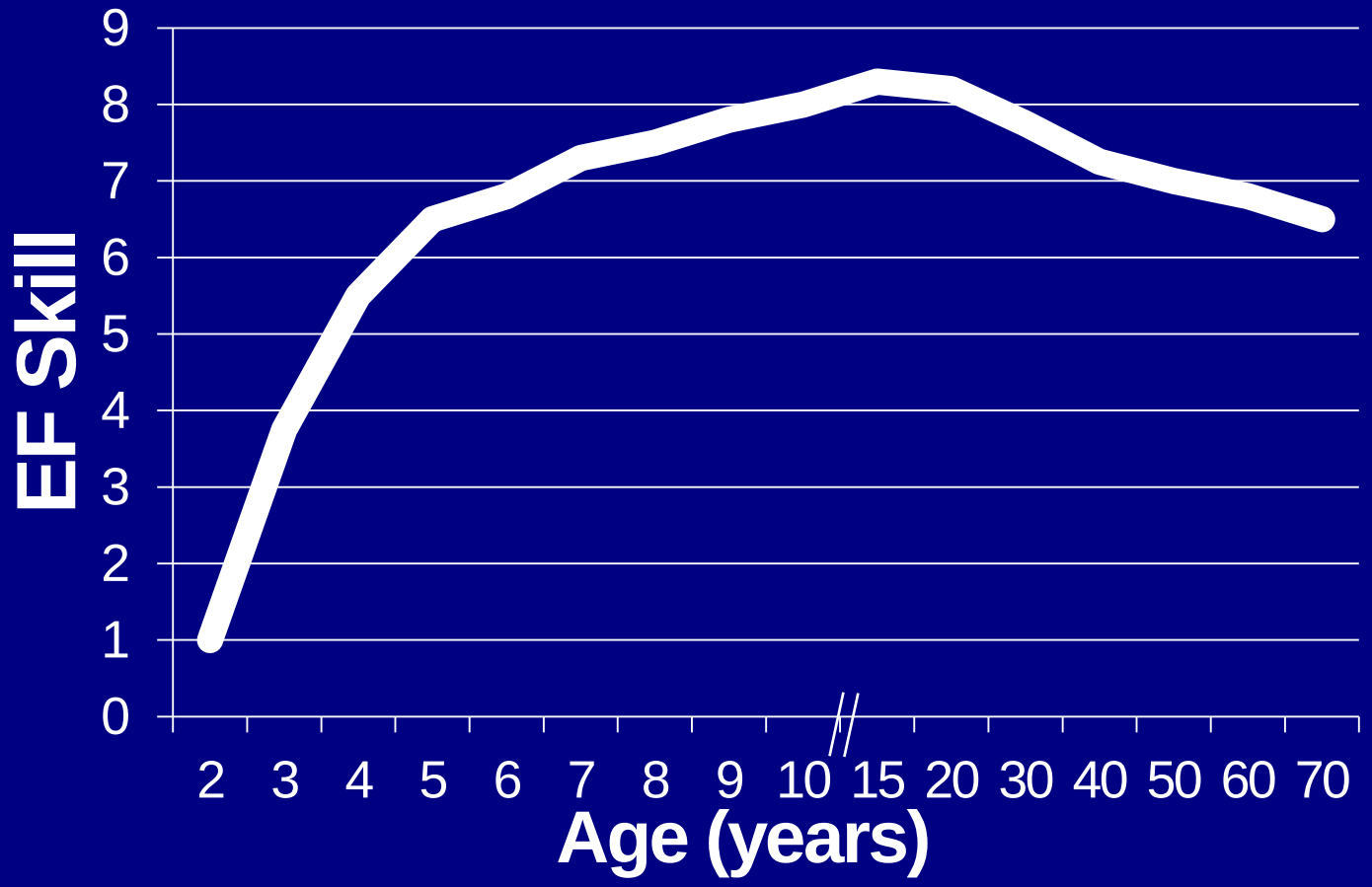
…170+ locations across 35 US States…



…and 13 Countries using 10 Languages

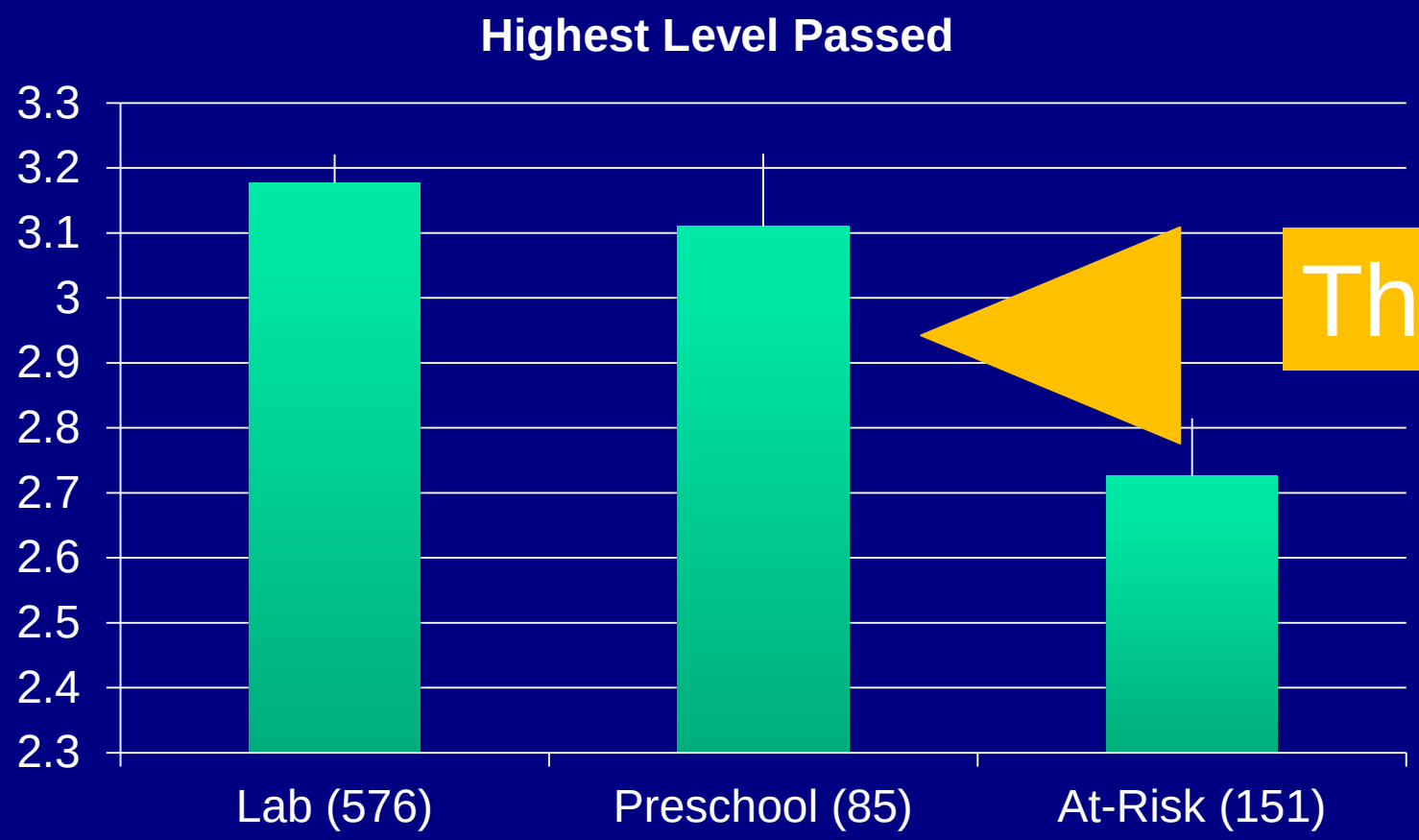


Executive Function from 2 to 70



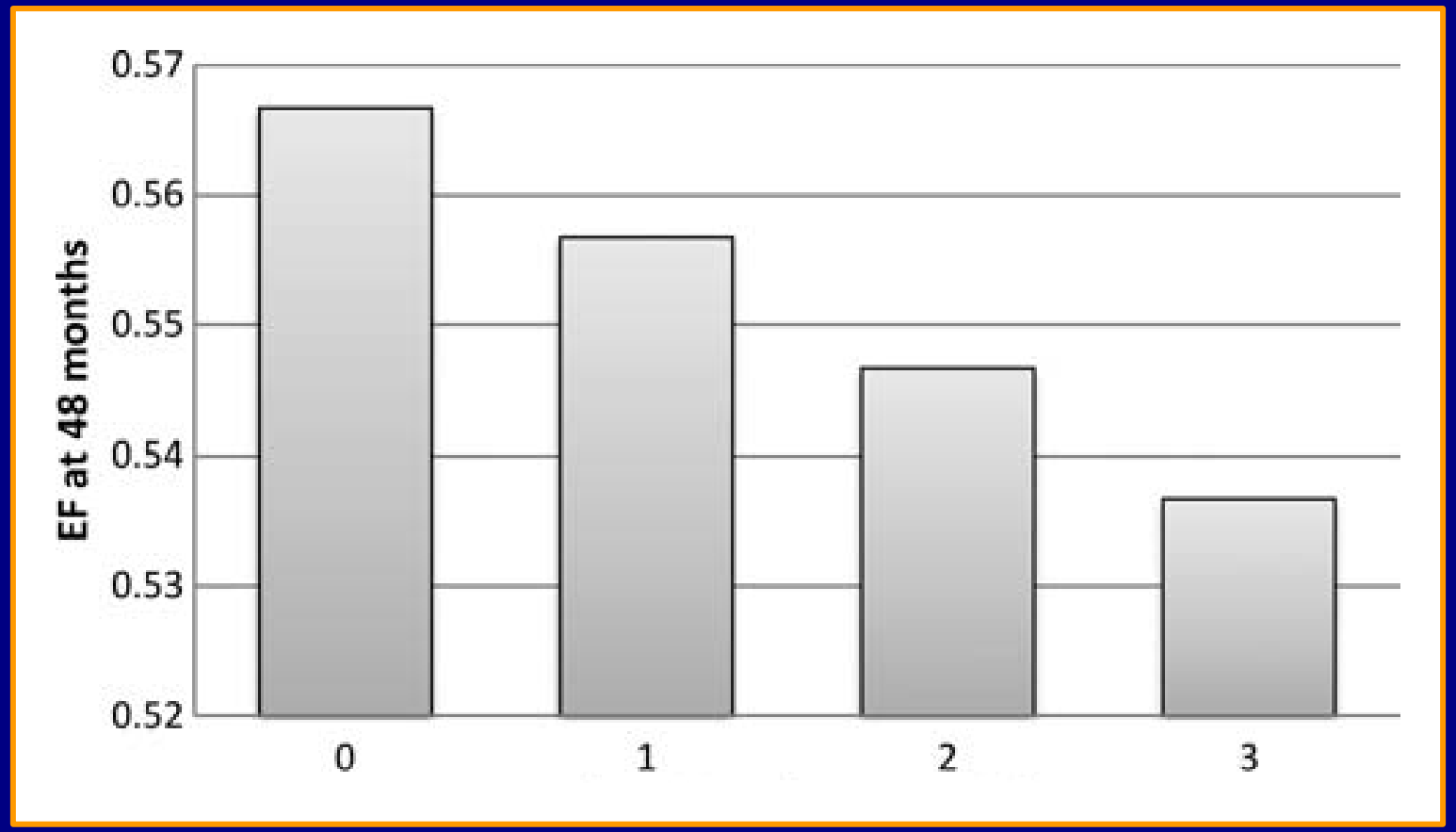
Zelazo et al., 2013; Carlson & Zelazo, 2015

MEFS Performance by Population Controlling for Age and IQ



N = 812. $F(2, 811) = 9.80$, $p < .001$, $\eta_p^2 = .024$, $R^2 = .65$. At-risk preschoolers scored significantly lower than lab and community preschool samples. Bars show *SE*.

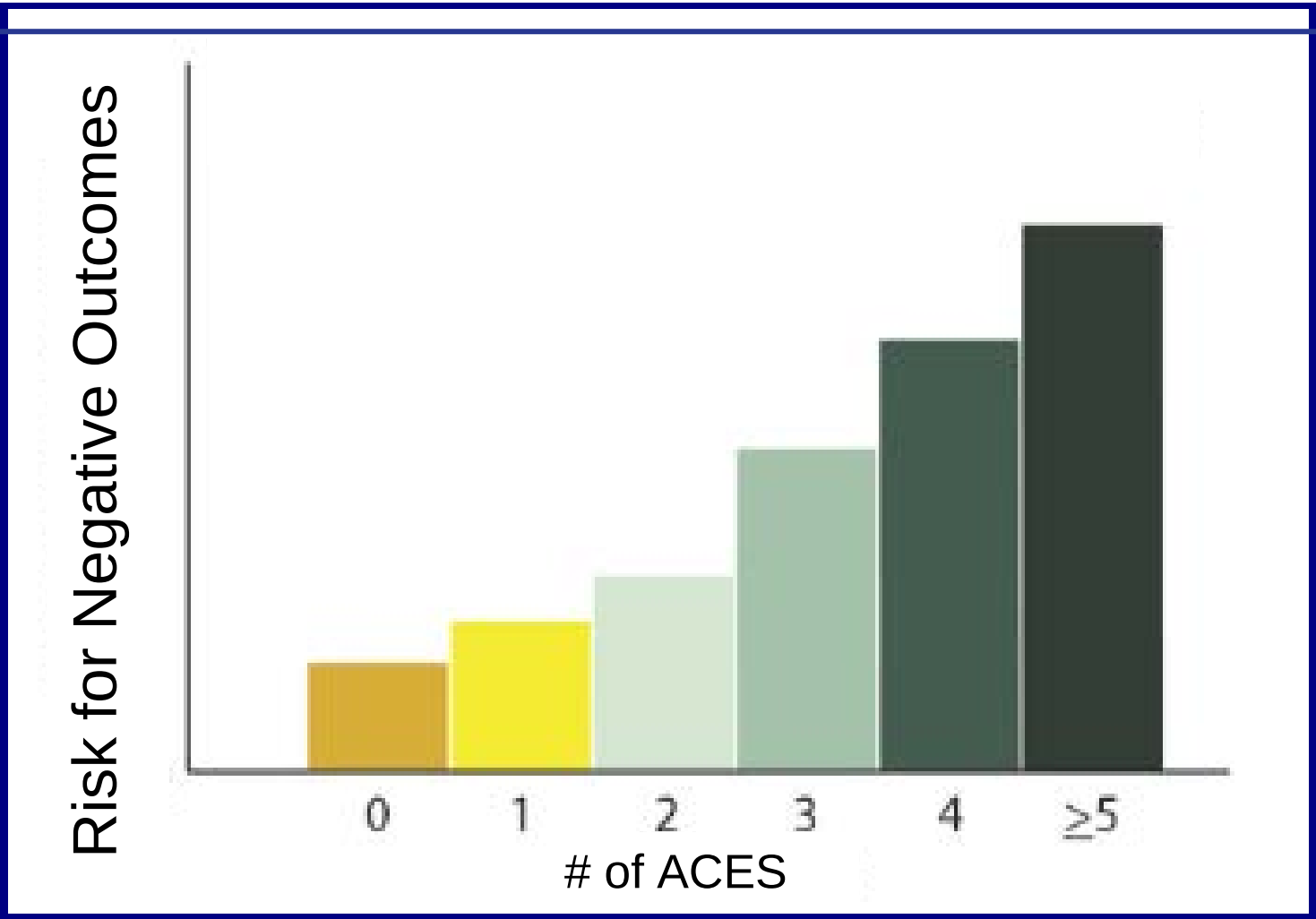
Lower SES is Associated with Lower EF



# of 12-month periods (after 15 months old) when family income was at or below U.S. poverty line

Adapted from Raver et al. (2013)

Trauma and EF

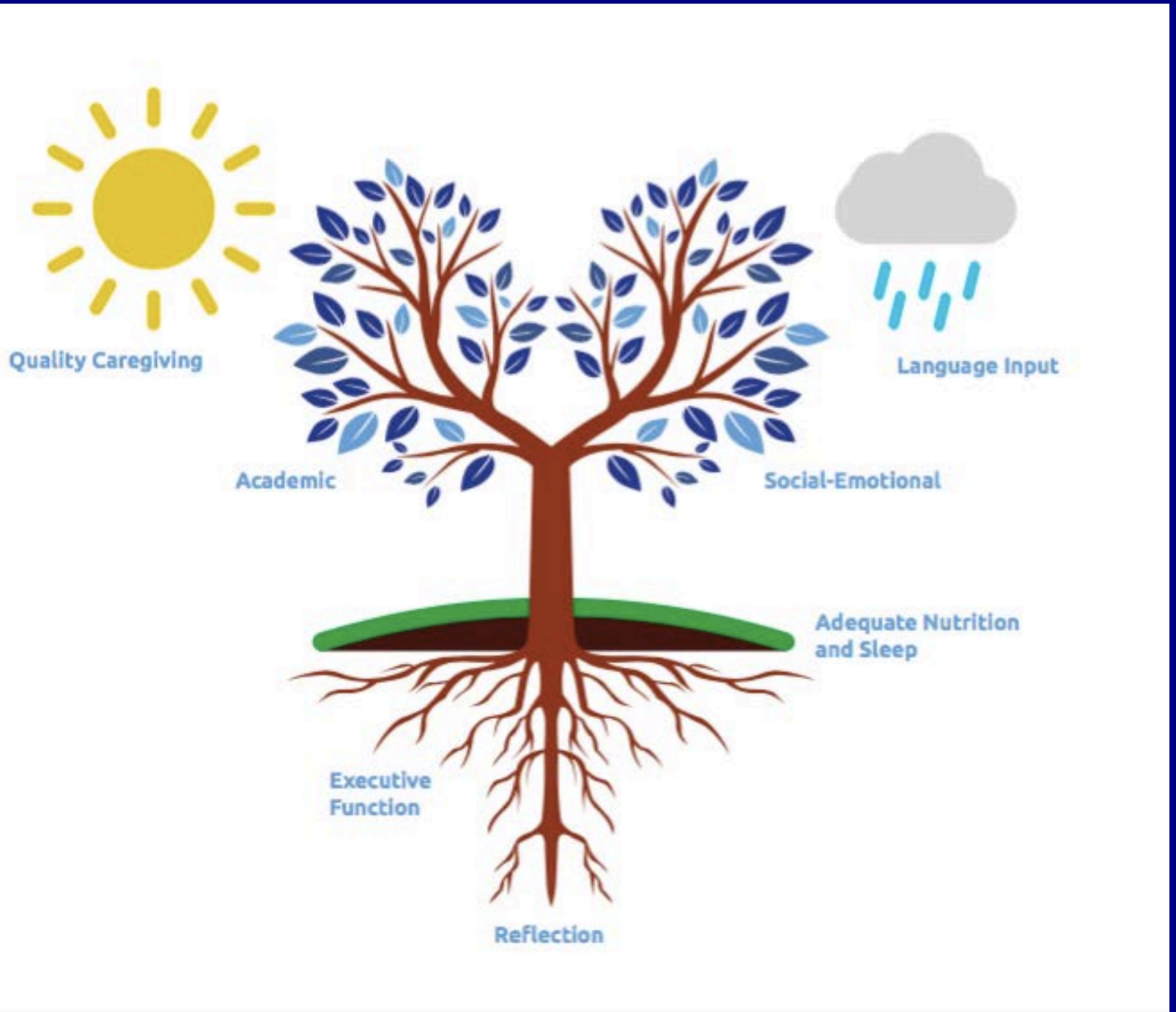


(ACEs Study Results – cdc.gov; Felitti, 2002; Felitti et al., 1998)

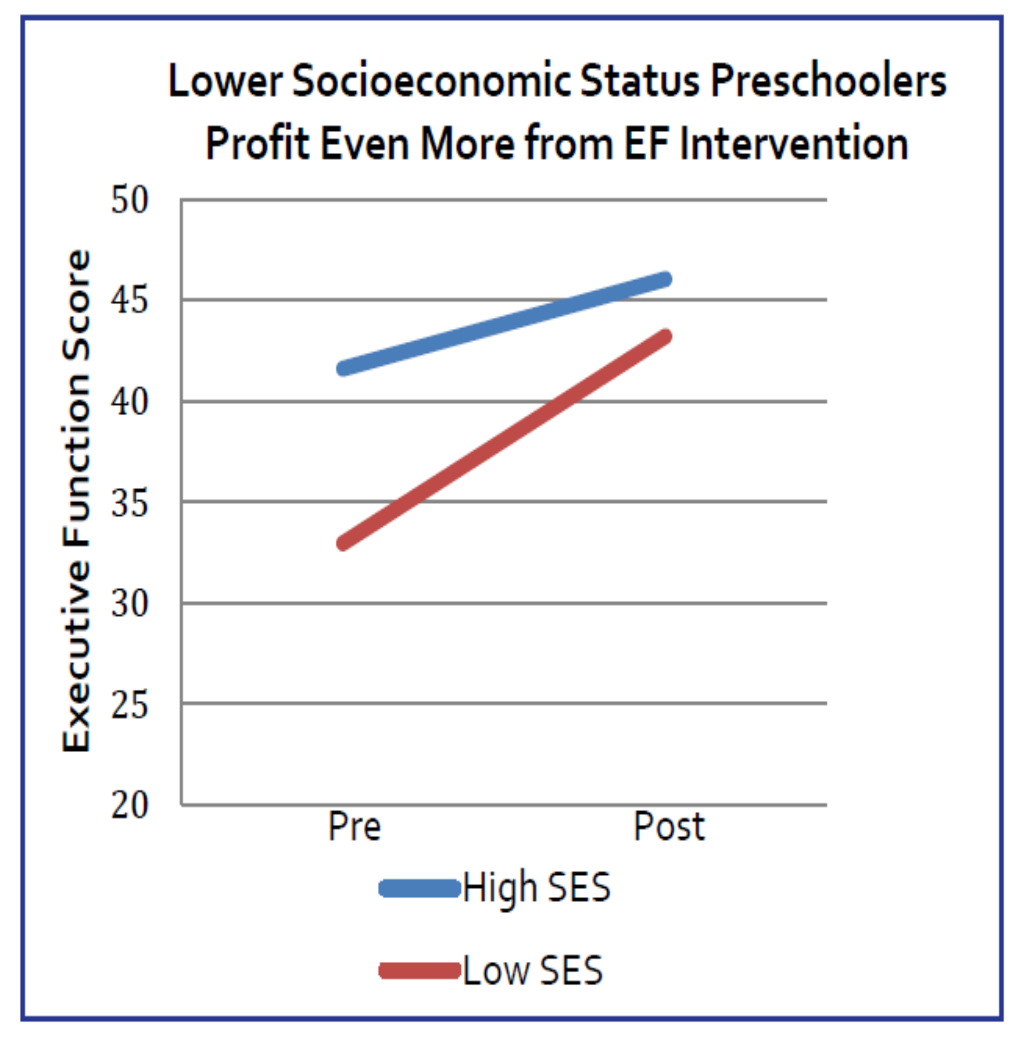
EF and School-Readiness

- Indirect role in creating optimal conditions:
  - Pay attention to teacher's directions
  - Sit still, keep hands to yourself
  - Persist when frustrated
  - Maintain positive peer relations
- Direct role in learning itself:
  - Hold new information in mind to work with it
  - Inhibit old information or biases that interfere
  - Think about problems in a new way

Early Intervention

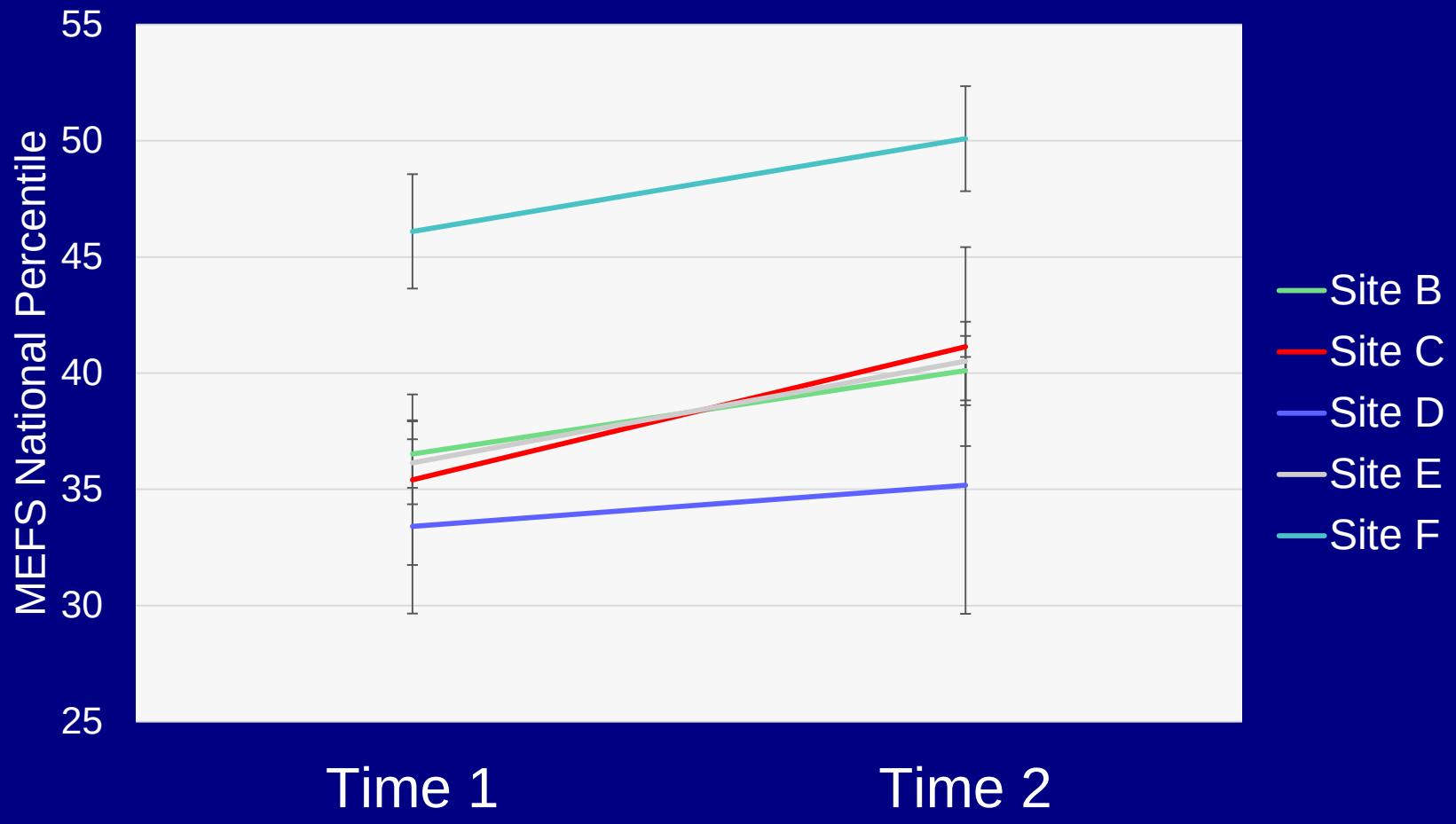


MEFS is Responsive to Interventions



Casey, 2016 dissertation

Closing the gap? Low-income Preschool Sites Varying in Quality

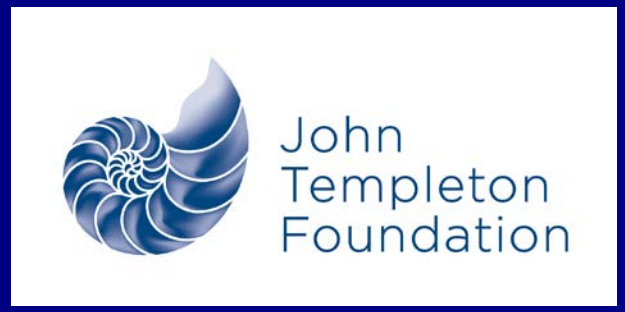


Take-home Messages

- Marshmallow Test is a classic measure of Executive Functioning
- Individual differences in delay are highly stable and predict later outcomes
- Despite adult predictions to the contrary, preschoolers' delay performance has increased since the 1960s
- Preschool education might be one reason why
- We do not know about lower SES children, who tend to perform less well
- Newer measures suggest an EF Gap, beyond the Marshmallow Test
- Minnesota Executive Function Scale (MEFS) is a standardized tool for measuring EF across the full spectrum, capturing growth sensitively, making data-informed decisions for ECE, comparing geographical regions, and tracking future trends

Thank you!

I am grateful to many colleagues, students, collaborators, as well as to those who provided funding for our research.



Questions?

Contact:
Stephanie M. Carlson
smc@umn.edu