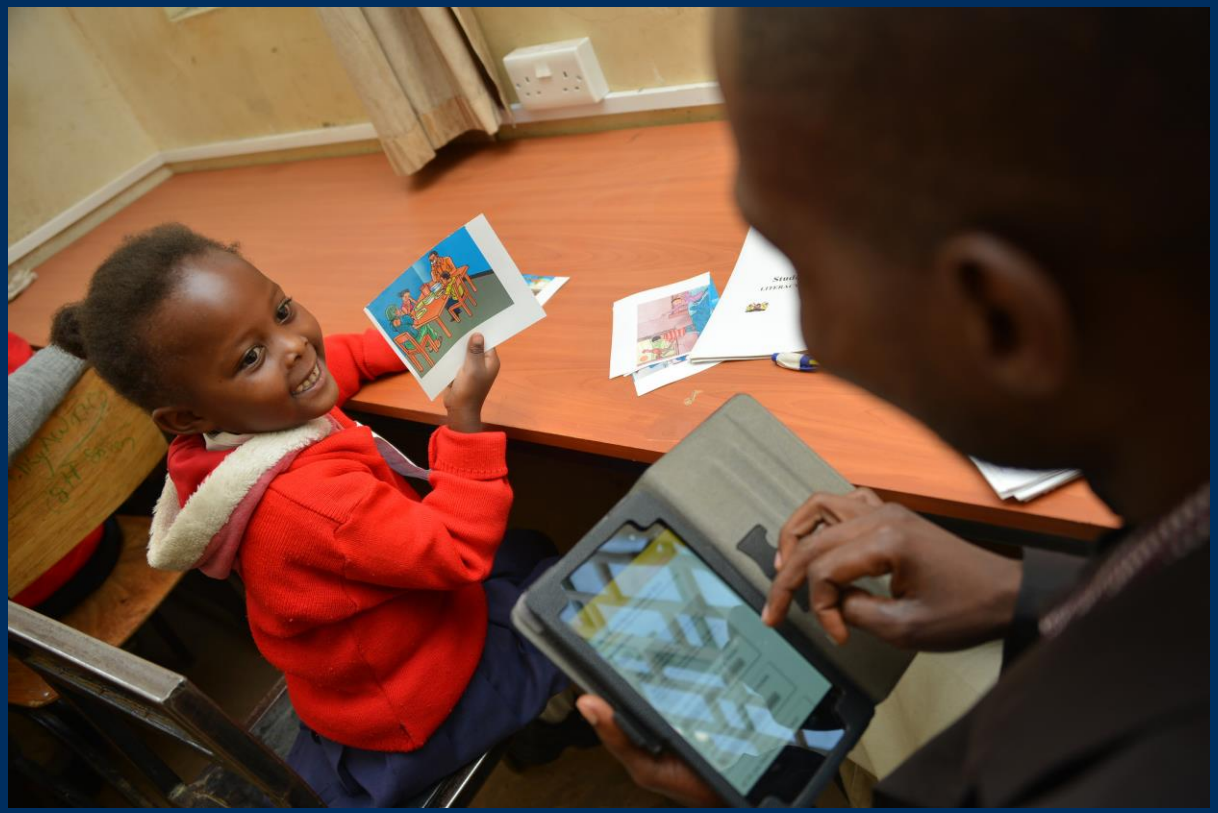
# Innovations in Early Childhood Development Assessment



Wednesday, October 10, 2018
1:00 p.m. – 5:00 p.m.

@RTI_INTL_DEV | #RTILearns | #EarlyYears

RTI INTERNATIONAL

# Agenda

| | |
|---|---|
| 1:00 PM—1:10 PM | Welcome Remarks |
| 1:10 PM—2:00 PM | Infant/Toddler (age 0-3) Assessment |
| 2:05 PM—2:55 PM | Language and Literacy Assessment |
| 2:55 PM—3:10 PM | Break |
| 3:10 PM—4:00 PM | Executive Function Assessment |
| 4:05 PM—4:55PM | Social Emotional Learning Assessment |
| 4:55 PM | Concluding Remarks |

# Cross-Cutting Themes

- Cultural Transport of Assessments
  - Content
    - Are we aiming to measure the same domains in a new setting?
    - Do those domains manifest themselves differently in a new setting?
  - Methods
    - Can we assess domains with the same methods?
- Agenda for Research and Practice
  - What are the knowns, unknowns and pressing issues in international assessments in each domain?

# Infant/Toddler (age 0-3) Assessment

**Dana Charles McCoy**
Assistant Professor
*Harvard Graduate School of Education*

**Maureen Black**
Distinguished Fellow, Early Childhood Development, *RTI International*
Clinical Professor, *University of Maryland School of Medicine*

# Measuring Population-Level Development for Children Under Three:

## Evidence from the Caregiver-Reported Early Development Instruments

Dana Charles McCoy
Marcus Waldman
Günther Fink

October 10, 2018
Innovations in ECD Assessment, RTI

CREDI
Caregiver-Reported Early Development Instruments

Harvard
Graduate School of Education

# BACKGROUND

# APPROACHES TO MEASURING ECD

## Screeners

- Used to identify children with early signs of developmental delay who are in need of additional services

## Programmatic Assessments

- Used in research to measure impacts of programs or policies

## Population Assessments

- Used to monitor the overall ECD status of a community, country, or region

Precise
Resource intensive
Culturally specific

Imprecise
Quick/cheap
Cross-culturally valid

# WHY DO POPULATION ESTIMATES MATTER?

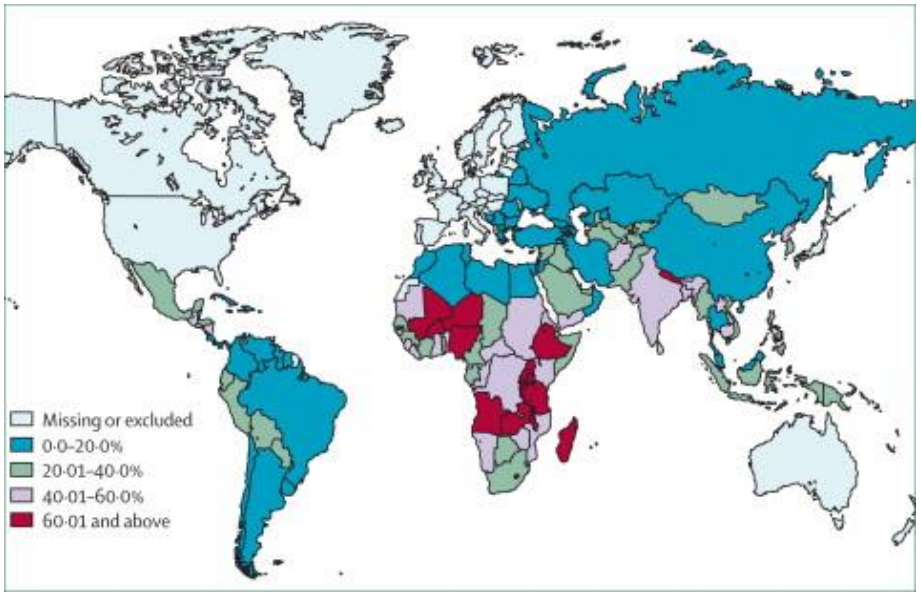
- Population estimates of ECD allow us to:
  - raise awareness of developmental inequities
  - make informed decisions re: policies & resource allocation
  - determine effectiveness of large-scale intervention efforts
  - monitor progress in achieving goals (e.g., MDGs, SDGs)


TARGET

**4.2 by 2030 ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education**

# HISTORY OF POPULATION ASSESSMENT

- Historical reliance on **risk factor proxies** like mortality, stunting, poverty
  - e.g., "over 200 million children under 5 are not fulfilling their developmental potential" (Grantham-McGregor et al., 2007)




- **Problem**: These are increasingly insufficient
  - Rates of risk factors are rapidly decreasing, but little is known about persistence of developmental challenges
  - Stunting and cognitive development are correlated at <.3 (Sudfeld et al., 2015)

# MODERN POPULATION ASSESSMENT

| Age | Assessment | Format |
|---|---|---|
| 3.5 - 7 | EDI | Teacher Report |
| 4 - 6 | MELQO | Direct Assessment<br>+ Teacher/Parent Report |
| 3 - 4 | ECDI (UNICEF) | Parent Report |
| 2 - 4 | PRIDI (Inter-American Development Bank) | Direct Assessment<br>+ Parent Report |
| 0 - <3 | CREDI | Parent Report |
| 0 - <3 | IYCD (WHO) | Parent Report |
| 0 - <3 | GCDG | Direct Assessment<br>+ Parent Report |

# THE CREDI

# THE CREDI

- Aim: To develop a **population-level measure of ECD** for children under three
- All items must:
  - capture core developmental domains for 0-36mo
  - be clear/simple enough to be easily reported by caregivers and implemented quickly with minimal training
  - be "culturally neutral" for global use
  - be psychometrically valid/reliable
- All materials (forms, translations, guides) are **freely available** via project website

CREDI

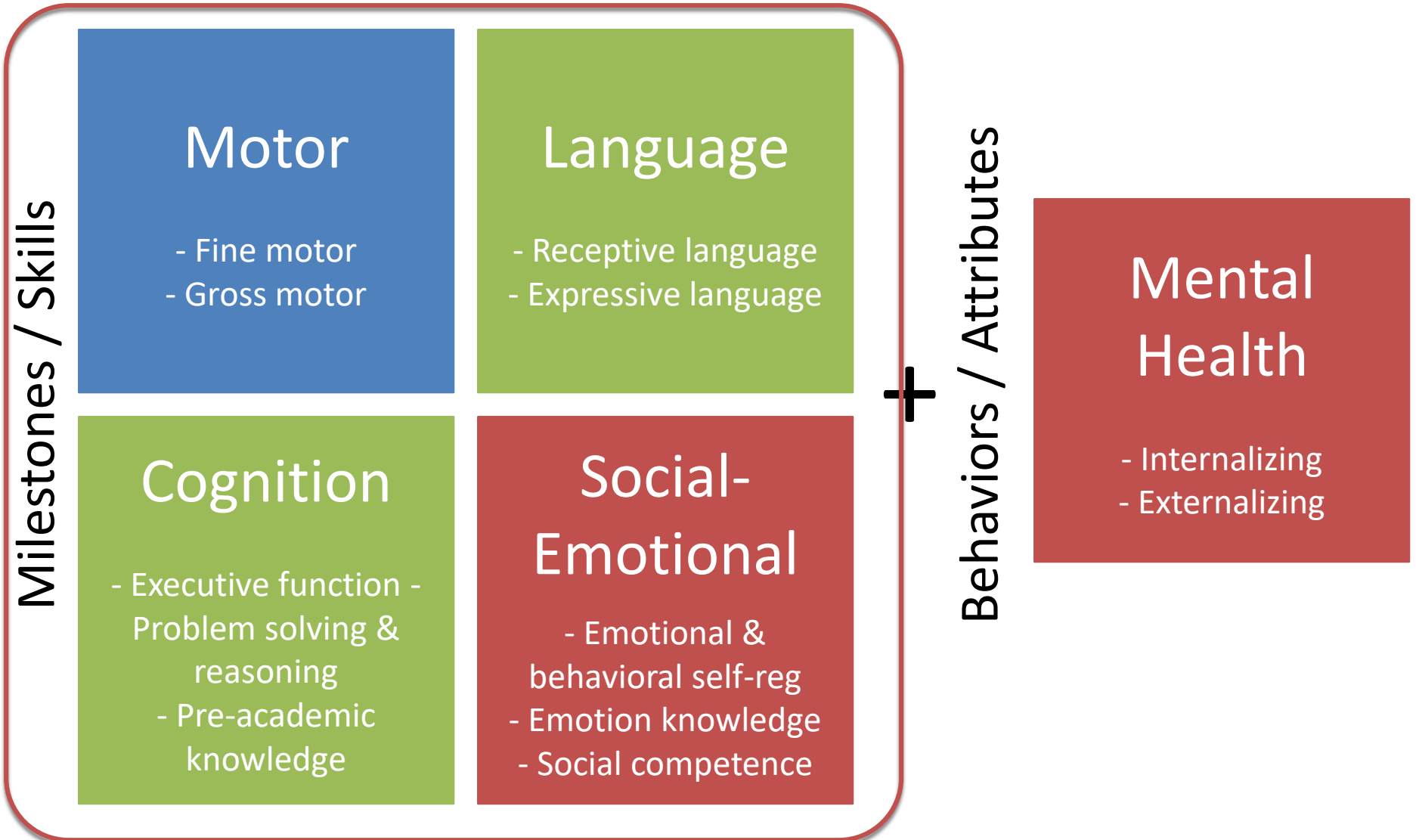
Caregiver-Reported Early Development Instruments

# VERSIONS OF THE CREDI

| | **Short Form** |
|---|---|
| *Purpose* | Population-level monitoring |
| *Length* | 20 items |
| *Format* | 6-mo age brackets |
| *Administration time* | ~5 mins |
| *Domain inclusion* | One score representing all domains (“overall ECD”) |

- Both versions allow for oral or written administration

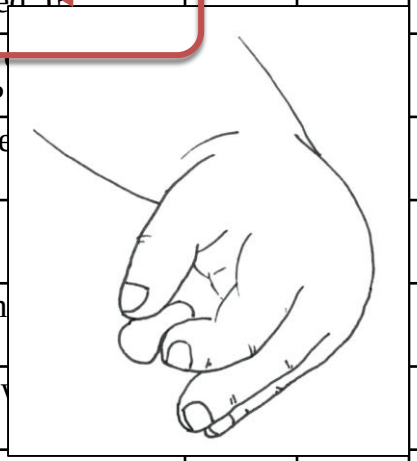
# CREDI DOMAINS



# SAMPLE ITEMS

| | | | Yes | No | Don't Know |
|---|---|---|---|---|---|
| **Motor** | Gross motor | Can the child climb onto an object such as a chair or bench? | 1 | 0 | 8 |
| | Fine motor | Can the child pick up a small object (e.g., a small toy or small stone) with just his/her thumb and a finger? | 1 | 0 | 8 |
| **Language** | Expressive language | Can the child say one or more words (e.g., names like Mama or "ba" for "ball")? | 1 | 0 | 8 |
| | Receptive language | Can the child follow orders or instructions that have more than one part (e.g., "Go get water and go to bed")? | 1 | 0 | 8 |
| **Cognitive** | Problem solving & reasoning | Can the child figure out how to turn a spoon or you give it to him/her the wrong way around? | | | 8 |
| | Literacy/numeracy | Can the child count up to five objects (e.g., finge people)? | | | 8 |
| **Social-Emotional** | Emotion Regulation | he/she calm down | | | 8 |
| | Social Competence | getting along with children? | | | 8 |
| | Executive Function | Is the child often impatient or unwilling to wait ask him/her to? | | | 8 |
| **Mental Health** | Internalizing | Is the child frequently sad, worried, or anxious? | 1 | 0 | 8 |
| | Externalizing | Does the child often kick, bite, or hit other children or adults? | 1 | 0 | 8 |

Examples



# TRAINING & TRANSLATION

- **Training**
  - Half day CREDI-specific training
  - No specific educational requirements beyond basic literacy, numeracy, research skills
- **Translation** & **adaptation**
  - Must be carefully done!
  - Strong emphasis on meaningful translation facilitated by Item Guide
  - Back-translation a must
  - Minimal cultural/linguistic adaptation needed
    - *Exception*: examples

# VALIDATION

# PILOT SITES


*N* = **149** items, **21** sites, **17** countries, **14** languages, **16,029** caregivers

# ITEM SELECTION

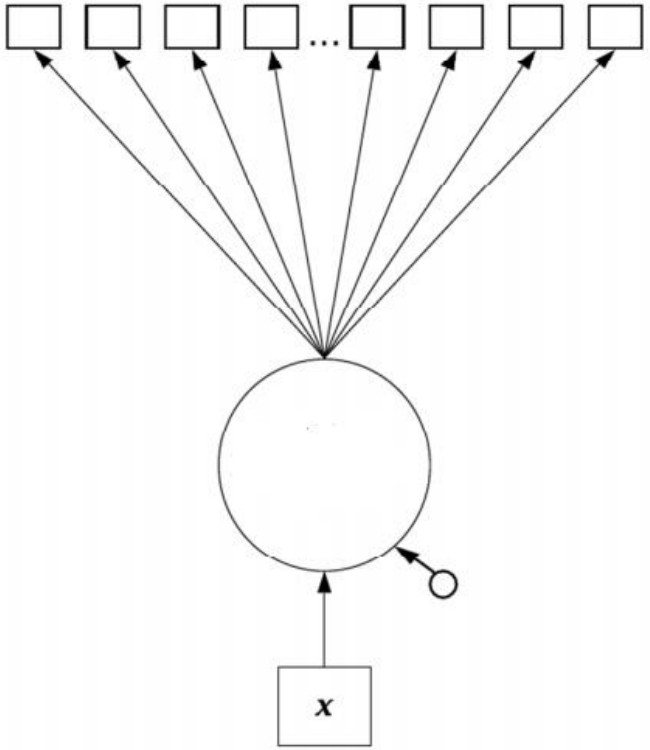
- Beginning with 149 items…

1. **Weeded** out "bad" items
   - **Unclear** items (cognitive interviews, "don't know")
   - **Unreliable** items (test-retest reliability)
   - **Non-invariant** items (difficulty, discrimination not similar across countries)
2. **Selected** items for final forms
   - Used **two-parameter (2PL) item response theory (IRT)** to minimize SE of measurement (maximize total information)
   - **Short Form (*N* = 62 items)**
   - **Long Form (*N* = 108 items)**
   - Used IRT results for continuous scoring across age

# IRT APPROACHES

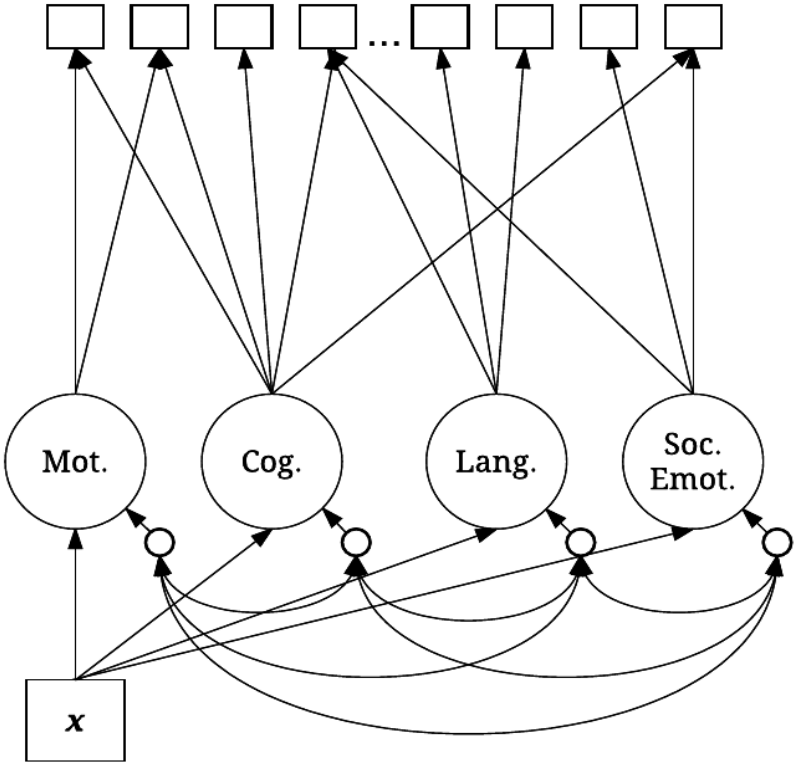
**Short Form**

Unidimensional 2PL IRT




**Long Form**

Multidimensional 2PL IRT



# SHORT FORM RELIABILITY & CRITERION VALIDITY

- **Internal consistency**
  - Range = .80-.89
- **Test-retest**
  - *M* Kappa = .62 (*SD* = .13, range = .41-.86)

| Measure | Site |
|---|---|
| **ASQ:SE** | Chile |
| **BSID cognition** | Pakistan |
| **BSID cognition** | Tanzania (Ifakara) |
| **INTERNDA** | Zambia (Chipata) |
| **MacArthur Bates CDI** | Chile |
| **PRIDI** | Brazil (Sao Paulo) |

# DISCUSSION

# OPEN QUESTIONS & NEXT STEPS

- Mental health scale
- Predictive validity
- Development of norms and standards
  - “on track” and “off track” status
- Adaptive testing
- Cohesion with other tools (IYCD, GCDG)
  - Stay tuned for Maureen…!

# THANK YOU!

dana_mccoy@gse.harvard.edu

sites.sph.harvard.edu/credi

Project supported by:
Grand Challenges Canada®
Grands Défis Canada

**Collaborators for data collection**: Elisa Altafim, Alexandra Brentani, Andreana Castellanos, Alexandra Chen, Anne Marie Chomat, Wafaie Fawzi, Cristina Gutierrez de Piñeres, Jena Hamadani, Natalia Henao, Pamela Jervis, Codie Kane, Jeffrey Measelle, Patricia Medrano, Lauren Pisani, Muneera Rasheed, Peter C. Rockers, Jonathan Seiden, Christopher R. Sudfeld, Fahmida Tofail, Christine Wong, Dorianne Wright, Aisha Yousafzai

**Advisory panel**: Frances Aboud, Amina Ali, Jere Behrman, Maureen Black, Kim Boller, David Bravo, Pia Britto, Claudia Cappa, Amanda Devercelli, Tarun Dua, Melissa Gladstone, Jena Hamadani, Magdalena Janus, Pamela Jervis, Patricia Kariger, Joan Lombardi, Sally Grantham-McGregor, Jeff Measelle, Lauren Pisani, Beth Prado, Abbie Raikes, Nirmala Rao, Jack Shonkoff, Fahmida Tofail, Aimee Verdisco, Susan Walker, Hiro Yoshikawa, Aisha Yousafzai, Stephanie Zuilkowski

# Development & validation of the D-Score for measurement of Early Childhood Development

Maureen Black, PhD

## Global Child Development Group

# Word Bank Toolkit for Measuring Early Childhood Development, 2017


**147 measures of early childhood development**

# Project Objectives

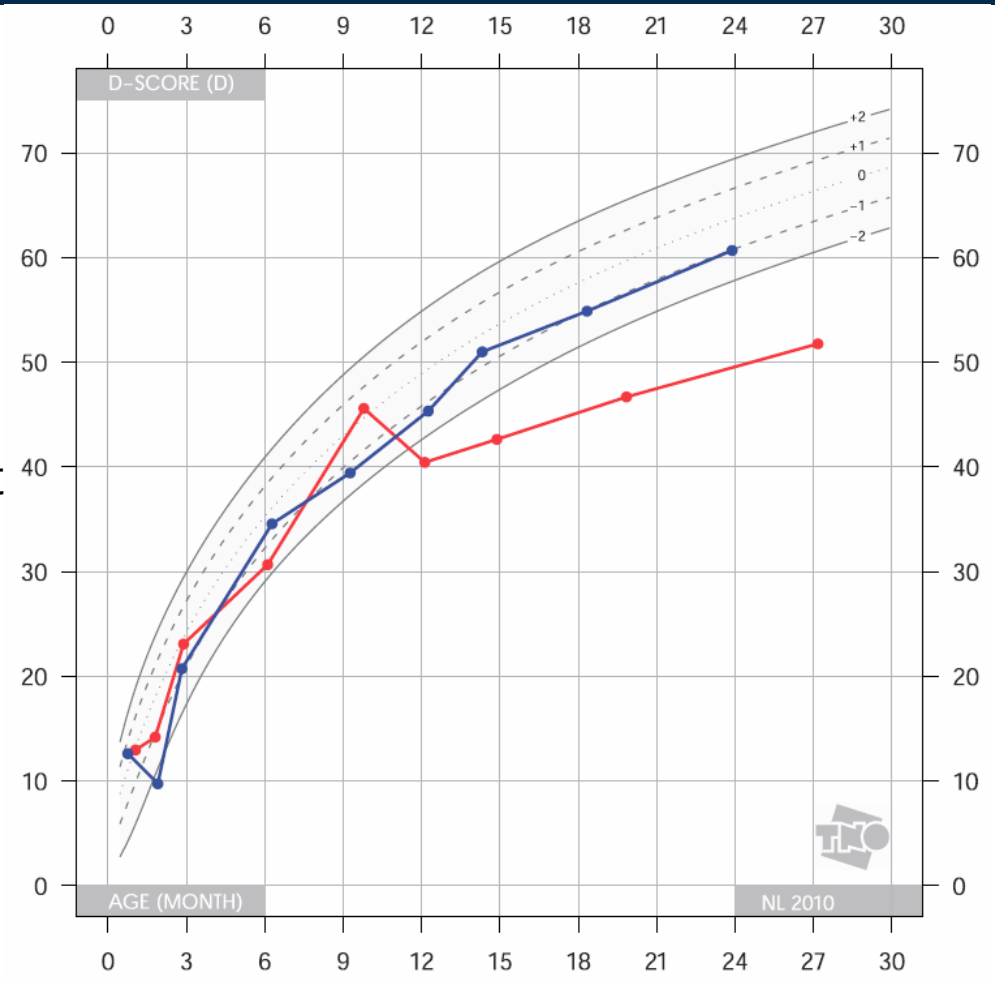
- Develop a global measure of early childhood development to assess **0-3 year-olds** that:
  - Is administered by direct assessment
  - Is reliable & valid across different contexts (e.g. culture, income)
  - Relatively easy to administer with few materials
  - Items are culturally neutral or easily modifiable for cultural differences
  - Is feasible for use in population-level surveys (short form)
  - Can be used for program evaluation
  - Is predictive of later school-age outcomes

# Development score

- Underlying assumption that development can be described by a continuous latent variable that represents multiple domains of early development
- Possible to derive an interval scale with common numerical unit or Development score **(D-score)**
- Would allow quantitative comparisons across different ages and contexts
- Age standardized D-score, D-score for age z-score (DAZ), would enable comparison of development  similar to HAZ for growth

# Example of a D-score reference chart

D-score unit




Age in months

Source: S. Van Buuren (2014) Growth charts of human Development. Statistical Methods in Medical Research

# Collaboration: Existing Data

- Advisory Board: Longitudinal data sets
  - Child development measured by standard instrument
  - Time 1 measure < age 36 months
  - Time 2 measure > age 5
- Data sharing agreement
- Build structure for data mapping
  - Similar items across existing instruments

# Assembled data from existing longitudinal studies

**16 cohorts in 11 countries**

**> 36,000 children**

- Africa
  - Ethiopia
  - Madagascar
  - South Africa
- Americas
  - Brazil (2)
  - Chile (2)
  - Colombia (2)
  - Ecuador
  - Jamaica (2)
- Asia
  - Bangladesh
  - China
- Europe:
  - The Netherlands (2)

**Time 1 Measures for ages 0-48 months**

- Bayley I, II, III
- Griffiths
- Denver
- Dutch Scale
- Battelle
- Barrera Moncada
- Others

Item level coded (pass =1, no =0)

**Time 2 Later measures age 5-18 years**

- WPPSI/WAIS
- Ravens
- PPVT
- Others

Birth cohorts, instrument validation studies, intervention evaluations

# Different instruments use similar items to assess the "same" developmental skill

Mapped equivalent items across different instruments to the Bayley-III

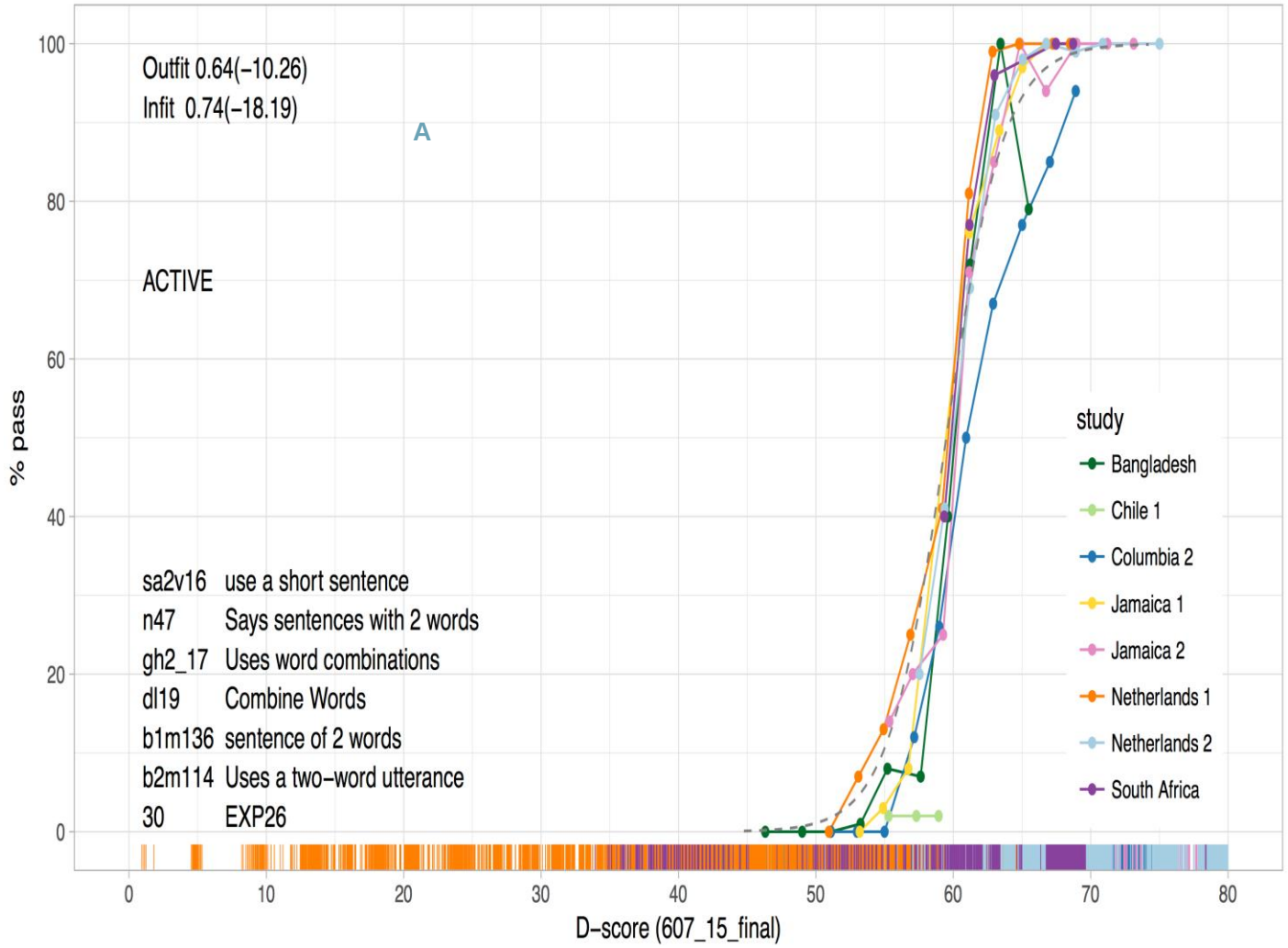
Example: Language items  across 3 instruments & expert opinion of mapping quality

| Bayley 3 Item Description | Griffiths Item Description | Mapping Score | Denver Item Description | Mapping Score |
|---|---|---|---|---|
| Child imitates at least four different repetitive consonant-vowel combinations | Babbled phrases: 4 + syllables | excellent | Child repeats the same syllable 3 or more times, eg. "Dadada" "Gagaga"? | moderate |
| Child uses at least two different words appropriately | Says 2 clear words | very good | Says 2 words | very good |
| Child correctly names at least four colors | Knows 6+ colors | moderate | Child names color of 4 blocks | excellent |

# Data organization

- Some datasets multiple waves (e.g.  12  and 24 months)
- Data organized as matrix:
  - child-wave rows and items as columns
- Items – pass, fail, missing
- 1339 items (after removing those with <10 responses in either pass or fail)
- Mapping led to 95 possible ‘equate’ groups containing at least 2 same-skill items from different instruments

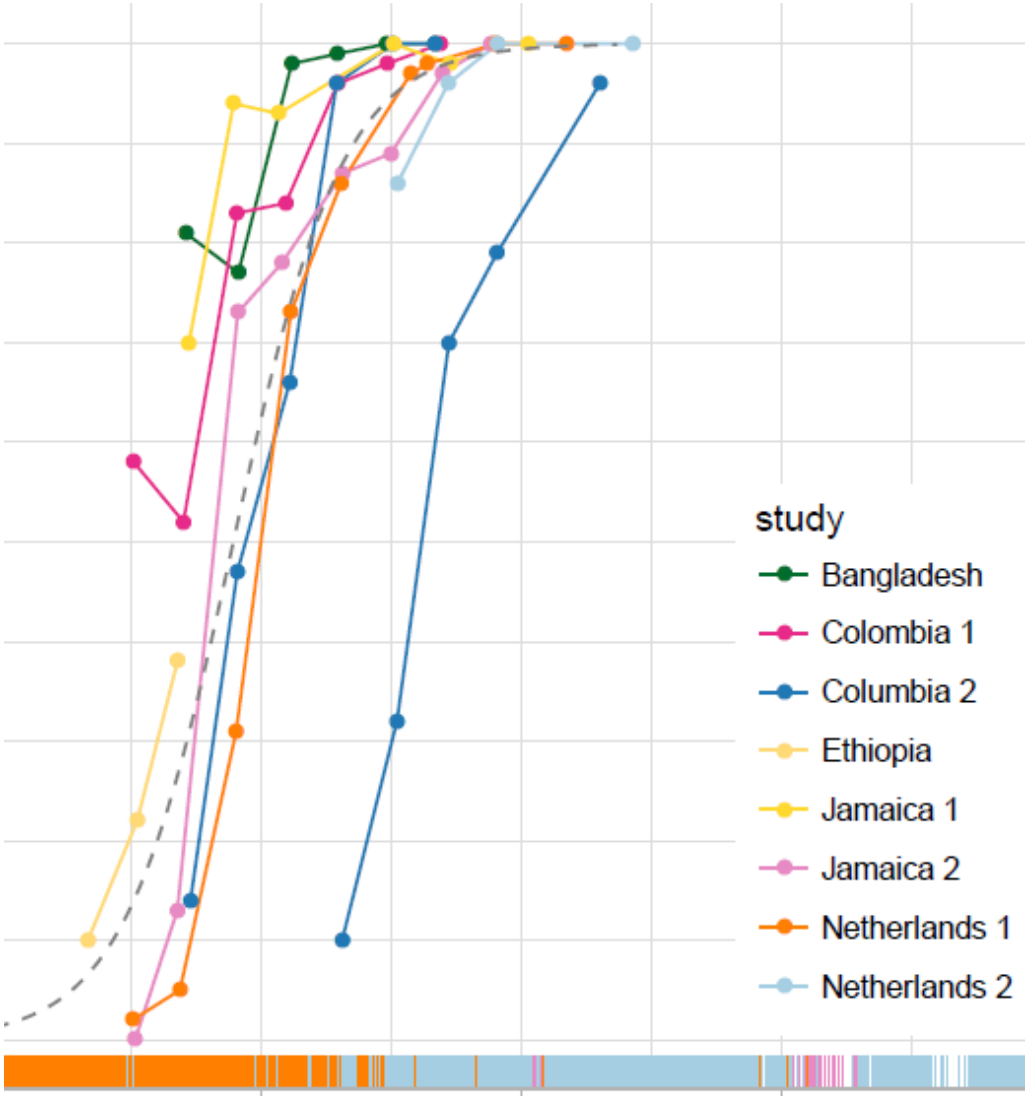
# Example of  a successful equate group



# Equate group – variable item difficulty

| | |
|---|---|
| n43 | Throws ball without falling down |
| ge2_14 | (Eye and hand) Can throw a ball |
| dg23 | Throw Ball Overhand |
| b3g44 | Throws ball |
| b2p64 | Throws ball |
| 229 | GM44 |

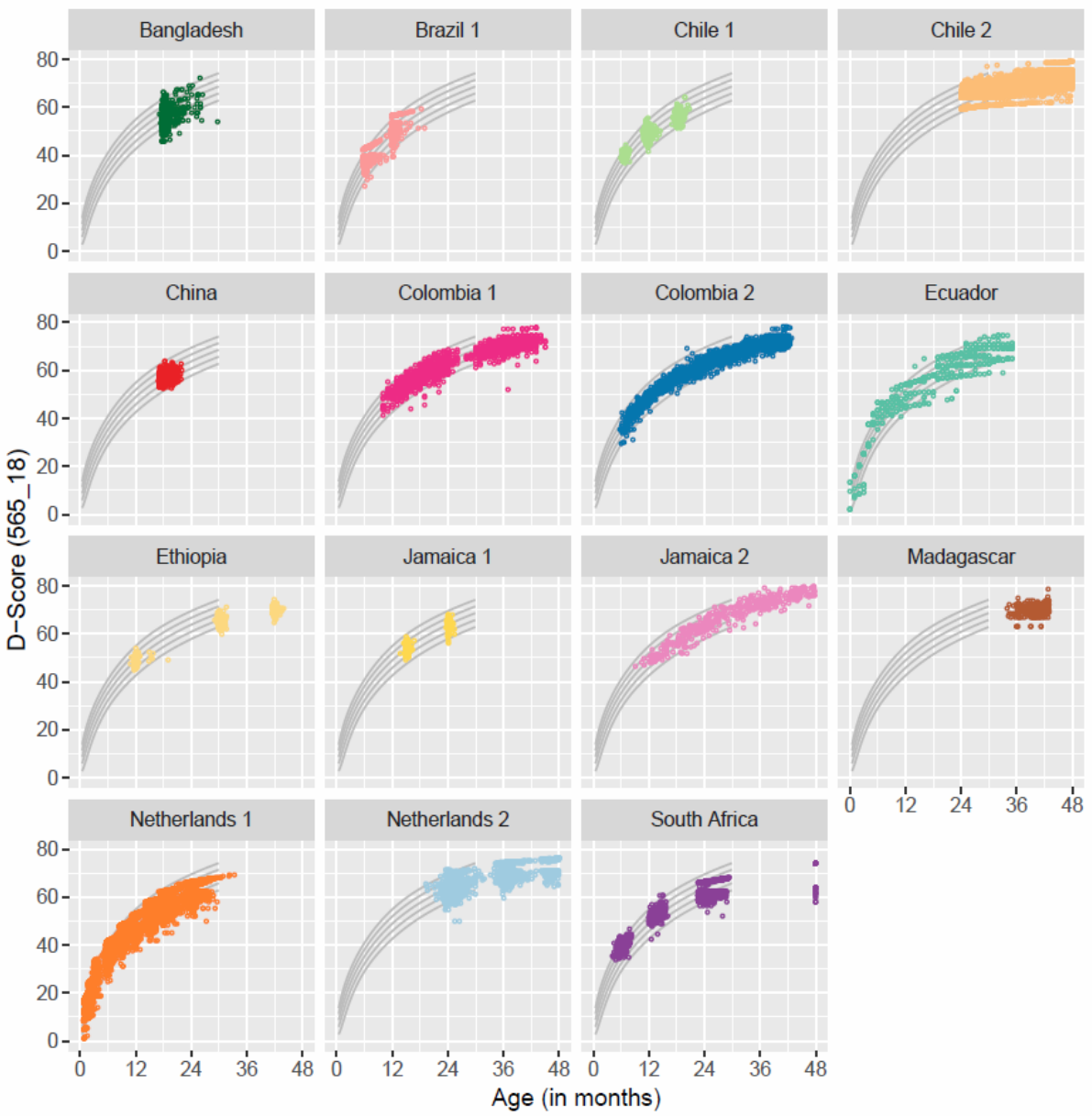
Outfit 1.45(6.31)
Infit 0.94(−3.37)



# Model estimation

- Rasch Model – Probability of passing item is function of difference between child ability and item difficulty
- Active equate groups:  same-skill items from different instruments constrained to have same difficulty level. Used to connect instruments to common scale
- Built using iterative approach varying i) active equate groups  and ii) cut points for acceptable fit to the model
- Final model retained items in active equate groups and individual items with infit and outfit statistics < 1
- 565 items (from 11 instruments) & 18 active equate groups

# Distribution of the D-score by age and cohort



# Validation

- Age conditional distribution of D-scores across cohorts - calculate D-score for age (DAZ) using LMS method.

- Concurrent validity correlation of DAZ with age standardized score from original instrument for each cohort/wave
- Discriminant validity comparing DAZ by birth weight, stunting and maternal education
- Predictive validity correlation of DAZ at Time 1 with school age outcomes and compared with correlation of original instrument with late outcome

# DAZ in children < 48 months correlated with original concurrent developmental measures

| Cohort | Age range (months) | Bayley--I, II, III[b] | | | Other Measures | |
|---|---|---|---|---|---|---|
| | | Cognition | Language | Motor | | |
| | | MDI | | PDI | Total Score | Measure |
| Bangladesh | 18 | 0.797 | | 0.503 | | |
| Brazil 1 | 5-11 | | | | 0.859 | Denver-II |
| | 11-19 | | | | 0.926 | |
| Chile 1 | 6 | 0.861 | | 0.438 | | |
| | 12 | 0.880 | | 0.361 | | |
| | 18 | 0.835 | | 0.249 | | |
| Chile 2[d] | 24-35 | | | | 0.768 | Tepsi |
| | 36-47 | | | | 0.855 | |
| China | 18 | 0.541 | - | 0.458 | | |
| Colombia 1 | 10-26 | 0.710 | 0.809 | 0.775 | | |
| | 28-45 | 0.742 | 0.840 | 0.672 | | |
| Colombia 2 | 6-17 | 0.386 | 0.333 | 0.675 | 0.758 | |
| | 18-29 | 0.671 | 0.837 | 0.651 | 0.642 | Denver-II |
| | 30-42 | 0.649 | 0.811 | 0.620 | 0.795 | |

Weber et al., under review

# DAZ in children < 48 months correlated with original concurrent developmental measures

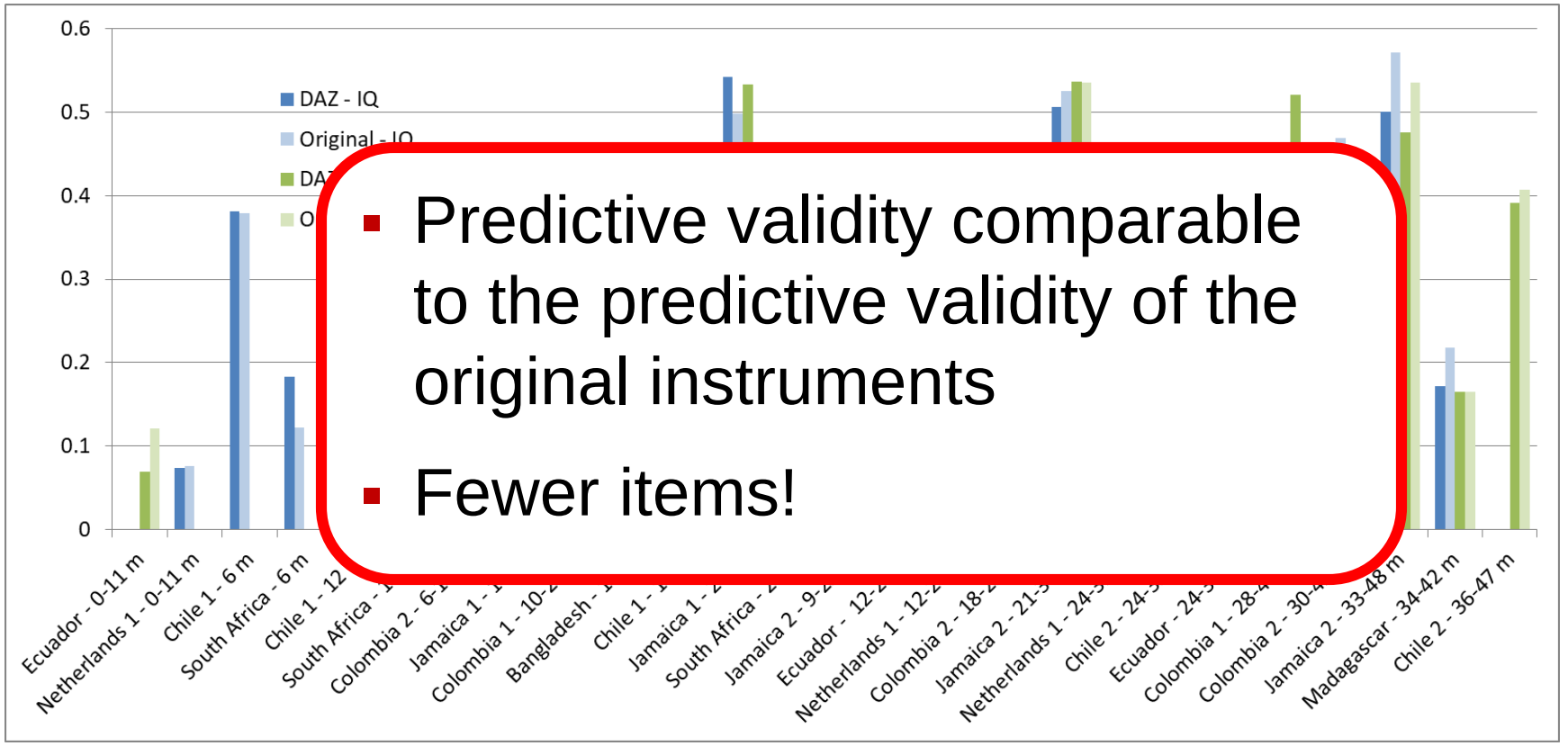
| Cohort | Age range (months) | Bayley--I, II, III[b] | | | Other Measures | |
|---|---|---|---|---|---|---|
| | | Cognition | Language | Motor | | |
| | | MDI | | PDI | Total Score | Measure |
| Ecuador | | | | | | |
| Ethiopia | | | | | | |
| Jamaica | | | | | | |
| Jamaica | | | | | | |
| Madagas | | | | | | |
| Netherla | | | | | | |
| | 24-34 | | | | 0.486 | |
| South Africa | 6 | 0.791 | | 0.775 | 0.868 | Griffiths DQ[c] |
| | 12 | 0.763 | | 0.659 | 0.725 | |
| | 24 | | | | 0.729 | Vineland |

- Moderate to strong concurrent validity.
- Expected as D-score is calculated from subsets of items from the original instruments

# Discriminant validity of DAZ with birthweight, nutritional status, and maternal education

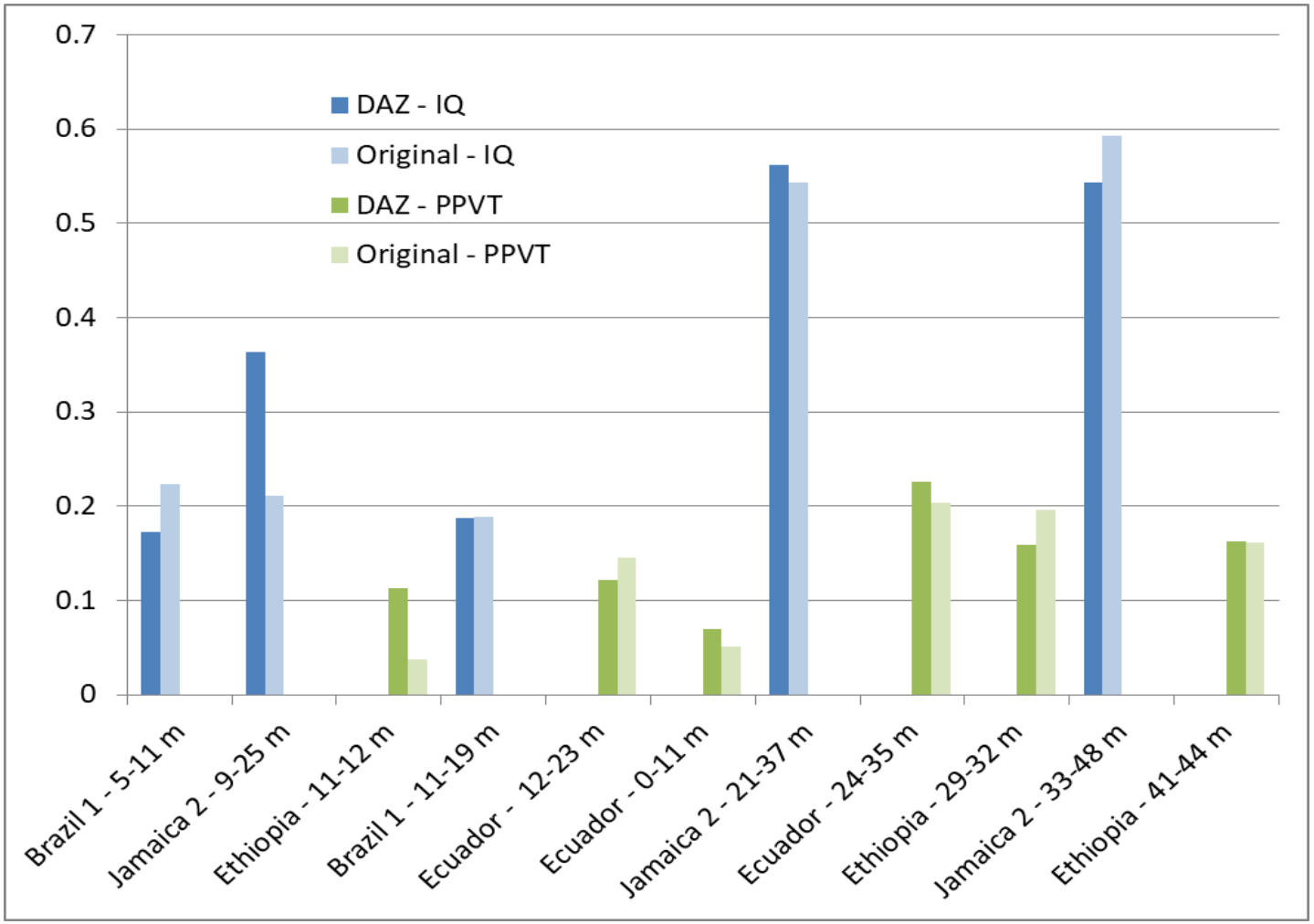
- Children with NBW higher DAZ than LBW (< 2.5kg). Significant in 18/26 cohort/waves
- Non-stunted children had higher DAZ than stunted.  Significant in 21/28 cohort/waves
- DAZ scores increased with maternal education in most cohorts.

# Correlation of DAZ and age-adjusted original measures in children under 48 months with IQ and receptive language measures at 4.5 to 9 years




- Predictive validity comparable to the predictive validity of the original instruments
- Fewer items!

# Correlation of DAZ and age-adjusted original measures in children under 48 months with IQ and receptive language measures at 9.5 to 18 years



# Summary

- D-score estimated with far fewer items than used in the original assessment, suggesting the feasibility of a relatively **short instrument** to assess ECD while maintaining validity.

- **Interval-scale property** of the D-score enabled quantitative comparisons across ages, as with trajectories of height and weight.

- Model enables estimation of **D-scores for pre-existing studies** from other countries, thus enabling external validation.

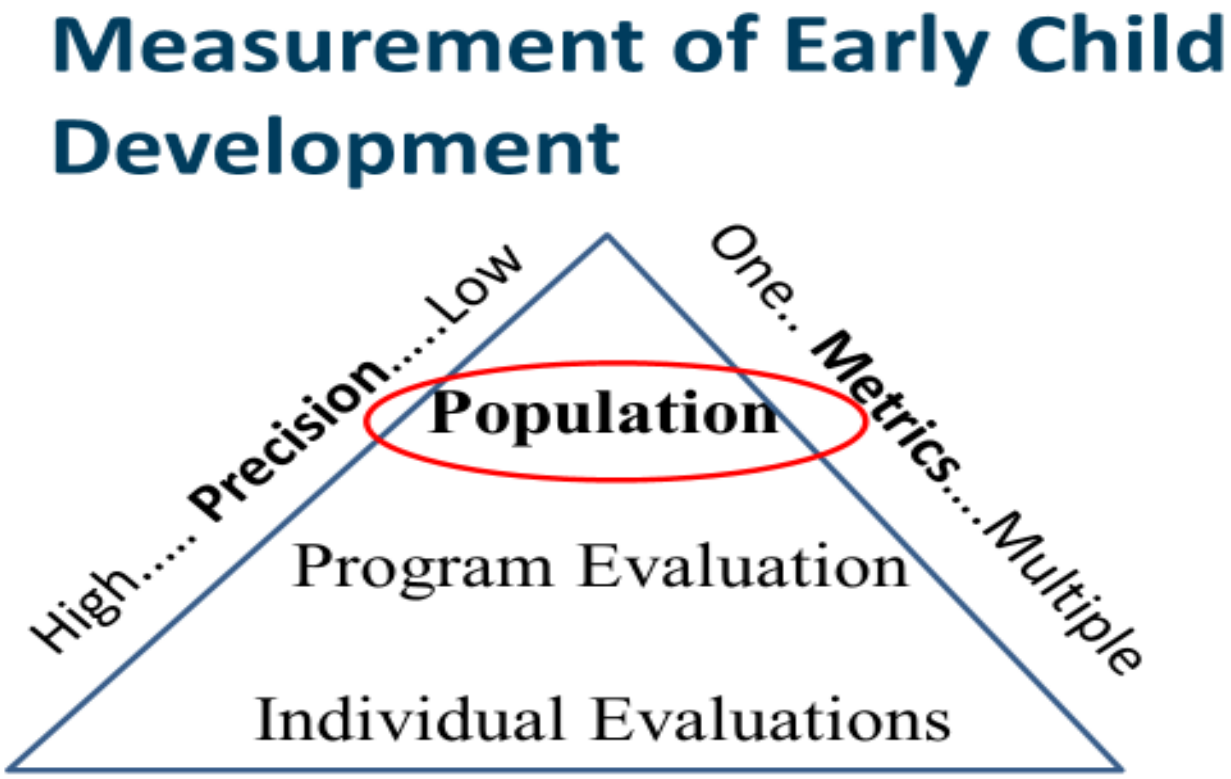
# Current status

- Good model fit and acceptable validity
- D-score has demonstrated potential as **simple low cost instrument valid** for global use for assessment of development in children 0-3 years .
- 165 items (approx. 20-25 per 6 month age group): non-duplicative, easy to train and administer, feasible in the field with minimal set of materials (blocks, pegboard, ball).
- Simulation showed high correlation with full model. Set of candidate items for global measure

# Current Progress

- External validation of the D-score with new data sets
- Development of training guidelines and administration procedure, materials
- Pre-testing items for feasibility
- Field testing
- Adaptive testing – order of administration contingent on pattern of pas/fail.

# Population and Program Measures



# Global Scale of Early Development (GSED)

**2 instruments:**

Population measure

Program Evaluation

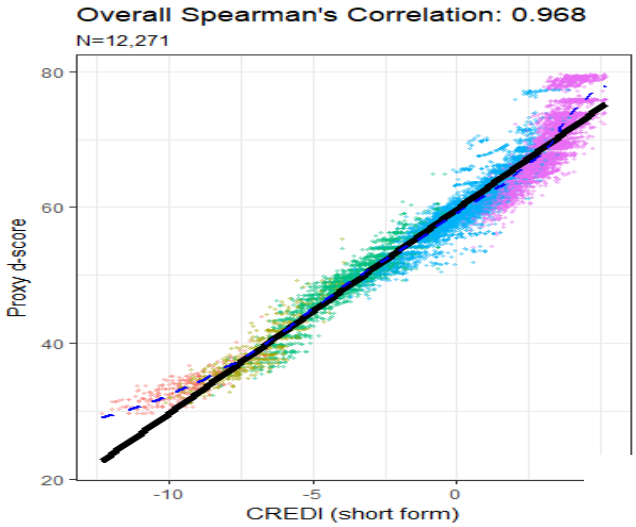
**CREDI**: Caregiver Reported Early Development Instrument

**IYCD**: Infant & Young Child Development

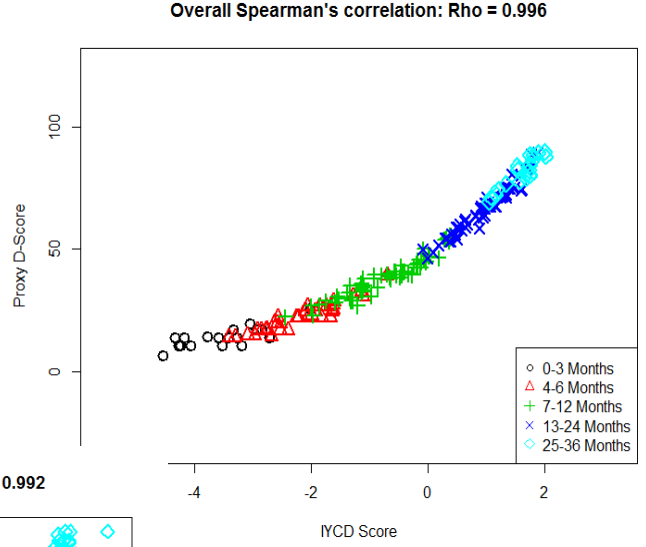
**D-Score**: Developmental Score Global Child Development Group

# Condordance among D-score, CREDI, and IYCD

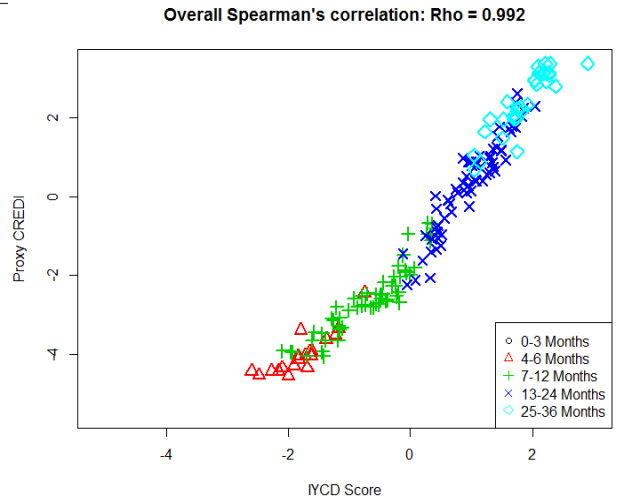
## D-Score – CREDI, Rho=0.968




## D-Score – IYCD, Rho=0.996




## CREDI – IYCD, Rho=0.992



# Next Steps

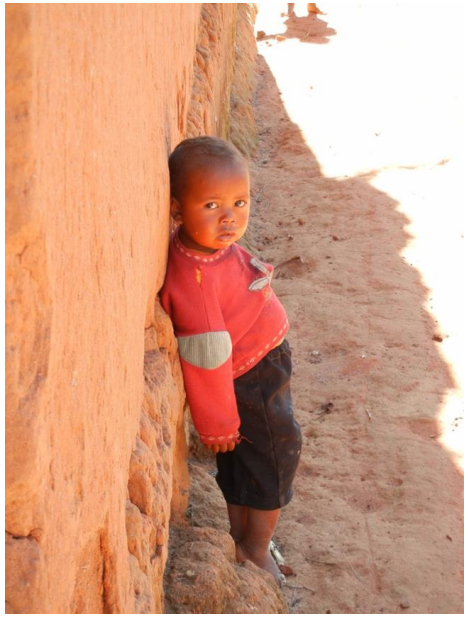
- Harmonize across 3 sets of items
  - Subject Matter Experts (SMEs)
  - Statistical concordance
- Develop short and long instruments for field testing
- Standard Operating Procedures (instructions)
- Training Materials
- Tablet administration and recording
- Site selection

# Next Steps

- Field test
  - Healthy children
  - Wide representation of urban/rural, socioeconomic status, etc.
  - Preliminary norms
- Validity
  - Concurrent and discriminant validity
  - Short term (6 month) predictive validity
- Open access
  - Feedback with local updates
- Additional field tests

# Acknowledgements

**Funders**
Bill & Melinda Gates Foundation
Bernard van Leer Foundation



**Global Child Development Group –Advisory Board**

- Ann Weber
- Marta Rubio-Codina
- Stef van Buuren
- Iris Eekhout
- Susan Walker
- Maureen Black
- Sally Grantham-McGregor
- Caridad Araujo
- Susan Chang
- Lia Fernald
- Jena Hamadani
- Charlotte Hanlon
- Simone Karam
- Betsy Lozoff
- Linda Richter

# Global Scale for Early Development

**D-Score**

Susan Walker
Stef van Buuren
Maureen Black
Ann Weber
Sally McGregor
Marta Codina Rubio
Susan Chang
Iris Eckhout
Lia Fernald
Betsy Lozoff
Linda Richter
Jena Hamadani
Simone Karam
Lisy Zarasoa

**IYCD**

Tarun Dua
Magdalena Janus
Melissa Gladstone
Gillian Lancaster
Gareth McCray
Patricia Kariger
Vanessa Cavallera

**CREDI**

Dana Charles McCoy
Gunther Fink
Marcus Waldman

**Donors**

Bill & Melinda Gates Foundation
Grand Challenges Canada
Bernard van Leer Foundation
Hilton Foundation

# Infant/Toddler (age 0-3) Assessment

**Dana Charles McCoy**
Assistant Professor
*Harvard Graduate School of Education*

**Maureen Black**
Distinguished Fellow, Early Childhood Development, *RTI International*
Clinical Professor, *University of Maryland School of Medicine*

# Language and Literacy Assessment

**Margaret "Peggy" Dubeck**
Senior Literacy Researcher
*RTI International*

**Karen Ford**
Assistant Professor, Research Faculty
*University of Virginia Curry School of Education*

# Multi-Language Assessment (MLA) for Young Children: An Instrument in Development

# Background

- Many children come to school with language skills that differ from the language of instruction (Walter & Benson 2012)
- 387 million children in primary schools not reaching minimum proficiency levels in reading (UNESCO, 2017)
- Oral language skills have been shown to have a small role in non-native word reading for non-native speakers (Geva & Yaghoub Zadeh, 2006; Quiroga et al., 2002). Threshold not known.
- Minimal information about oral language skills limits understanding why some children do not respond to literacy instruction

# Domains of Knowledge

- Multilingual children have domains of knowledge in various languages (e.g., home, playground, classroom) (Bedore et al., 2005)

- Most language assessments administered in one language describe results from a deficit approach (Core et al., 2013)

# Recent Efforts to Capture Assets

- Recognize the need for adjusted measurement (De Villiers 2015; Gatt et al., 2015)
- Responses in multiple languages (Kan & Kohnert, 2005)
- Common underlying proficiency (Leseman, 2000)
- Conceptual scoring (Gross et al., 2014; Mancilla-Martinez & Vagh, 2013)

# Language assessment needs

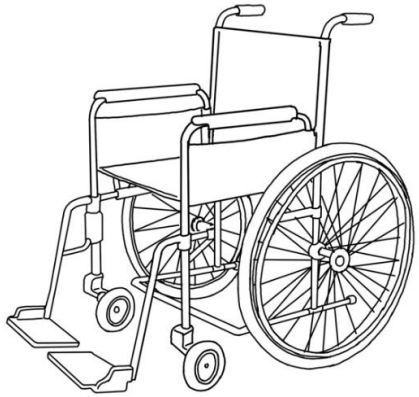
1. that is valid for measuring expressive language
2. that avoids subject effects
3. that can be administered reliably
4. that can be scored reliably by providing a structured protocol for raters to reach agreement
5. that can be produced inexpensively

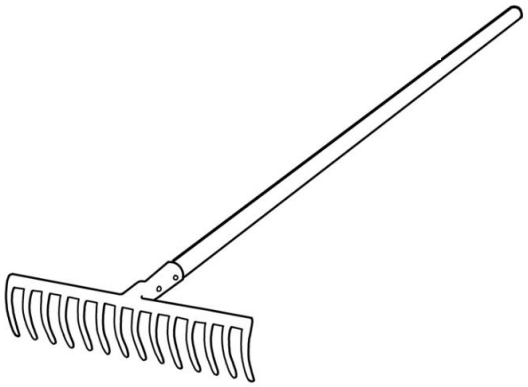
# Items (Images)

- Culturally relevant
- Exposure via the community, family, media, prior to schooling
- Intended to yield variable distribution
- Identifiable in isolation with a common prompt: “What is this?”
- Produced quickly (line drawings)

# Domains

- animals near, animals distant, architecture, clothing, community, food, furniture, household objects, human body, nature, personal objects, tools, transportation





# Administration Practice

- Practice item to encourage responses in any language.

**Let's look at this picture together. I know this. I call it a dog. Some people call it mbwa. What do you call it?**

- Practice item to encourage describing use.

**Let's look at another picture. I do not know its exact word. But I know you can drink from it. Can you tell me what you know about it?**

- Practice item to encourage domain.

**Let's look at another picture. I do not know its exact word. But I know that it is a vegetable. Can you tell me what you know about it?**

# Administration Support

- Prompt: **What is this? You can tell me in any language.**
- Follow-up: **Do you know anything about it?**
- 36 black and white drawings, four per A4 page
- Engaging pace, 5-6 minutes
- Scoring: Utterance written. Categorized after the child is dismissed. Confirmed by a second rater.

# Child Responses - Categories

- **English, Kiswahili, Kikamba:** The child names the item.
- **Appearance**: The child describes something that has a strong resemblance to the image.
- **Connection**: The child shares a personal connection to the item that shows knowledge of the item.
- **Domain**: The child provides a synonym or a word that is directly related to the concept.
- **Feature**: The child accurately describes the size, shape, color, or the texture of the item.

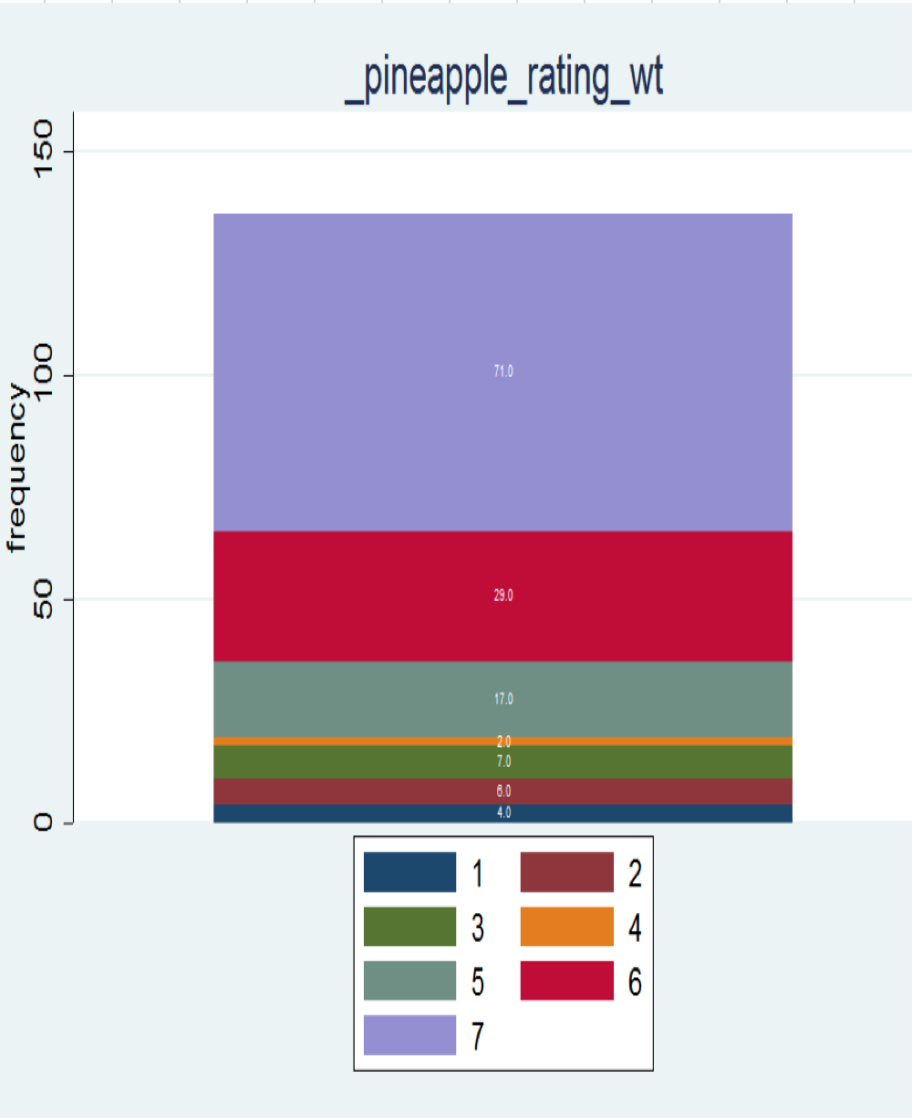
# Child Responses - Categories

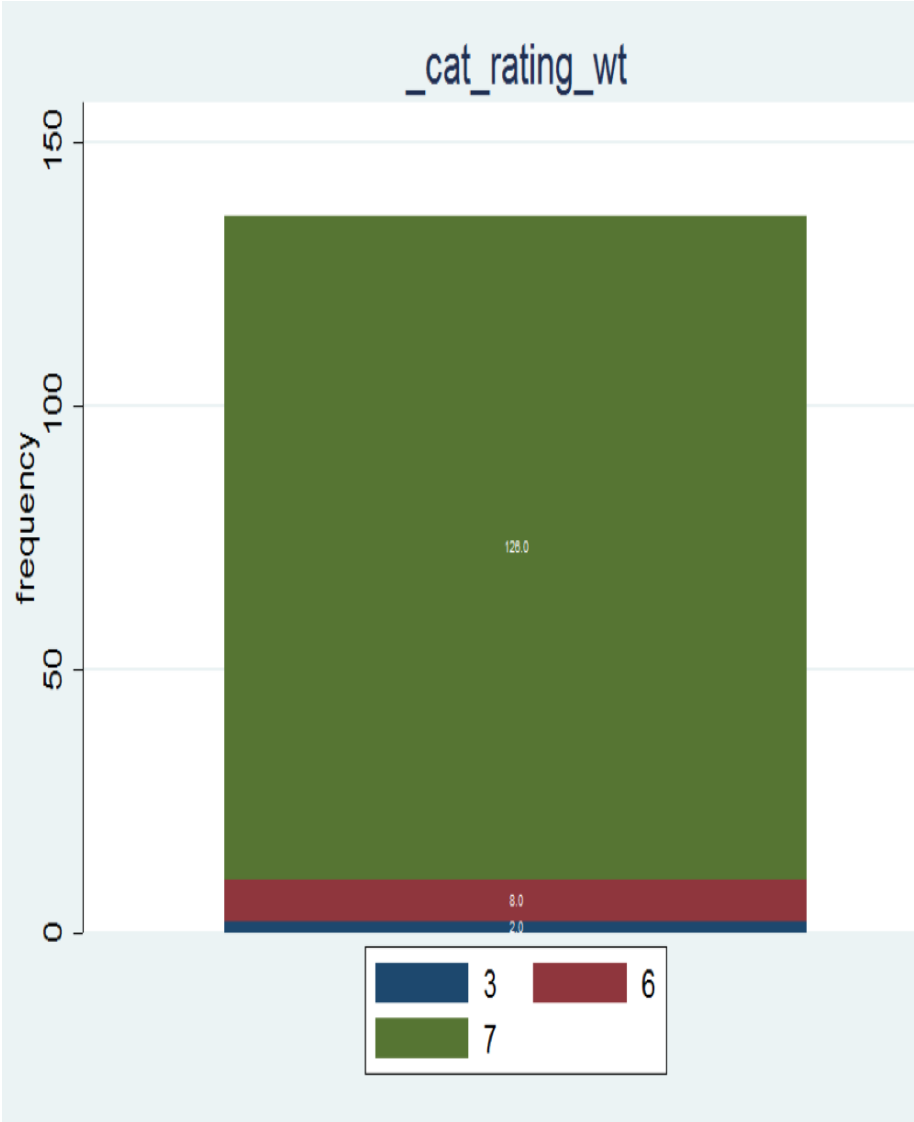
- **Non-Responsive:** The child is silent.
- **Nonsense**: The child says something that is not a word in any language.
- **Phonology**: The child says something that shares phonological elements of the target word.
- **Use**: The child accurately describes how the item is used.
- **Wrong:** The child's response is wrong and does fit in any category.

# Assessment Development

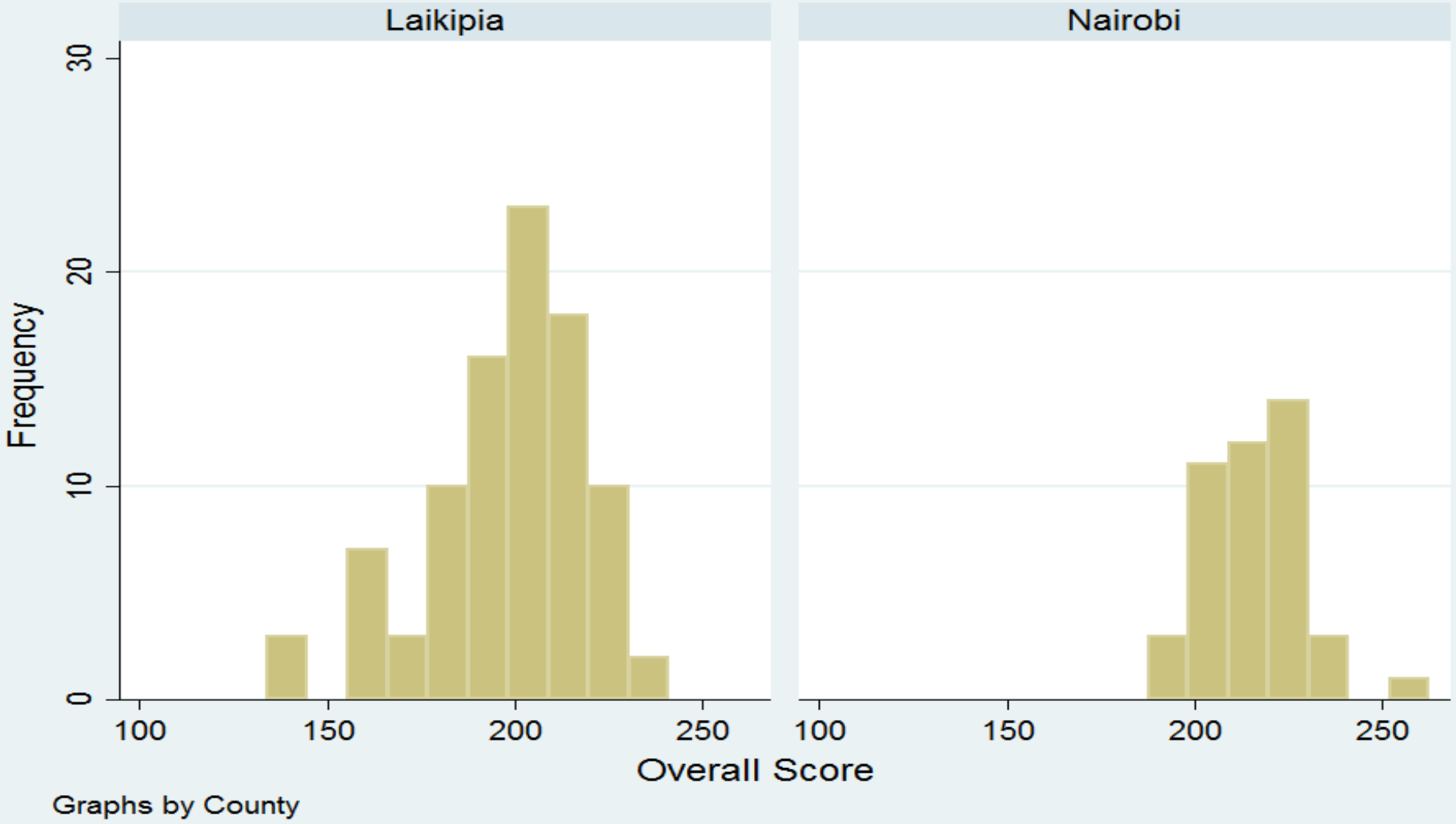
1. Bank of images reviewed & prioritized
2. Items requested from local illustrator
3. Illustrations shown to children using a receptive procedure "show me" to capture the validity of the items
4. Language translated and back translated
5. Field tests: Scoring categories expanded
6. Test - retest to explore consistency
7. Field & pilot test in 3 counties (Laikipia, Machakos, Nairobi)
8. Semantic fluency for construct validity =.68
9. Internal frequencies across items
10. SEM to explore category weights

# Comparison of Two Items

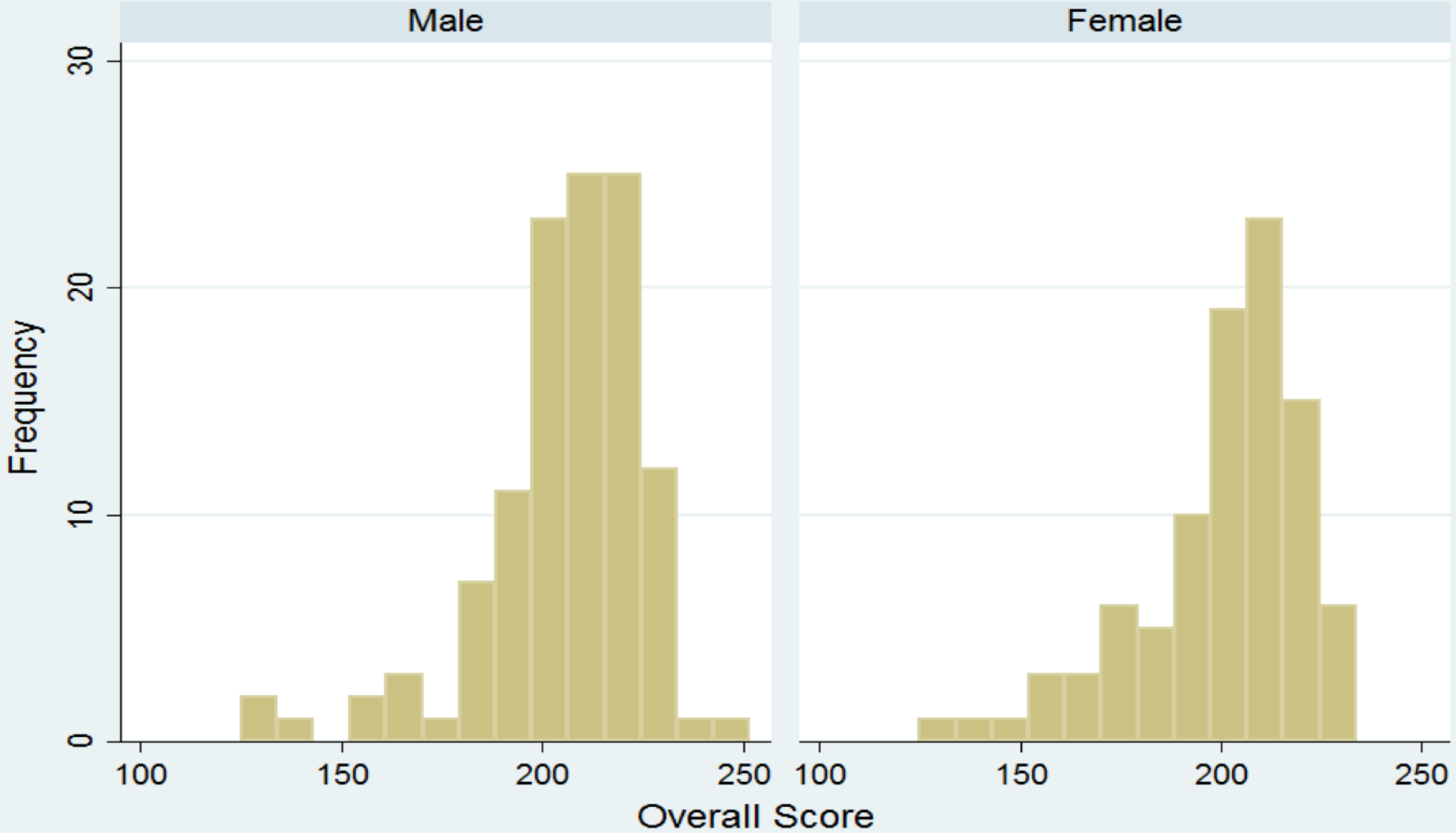


# Results – Rural - Urban



# Results – Gender



# Next Steps

- Primary purpose is to understand if young children's language skills across languages is predictive
- KEMRI SERU (Scientific Ethics and Review Unit) approval longitudinal study, two time points
- Used alongside other literacy measures
- RQ: Do children's expressive language scores across multiple languages help to explain their literacy outcomes in the early years of formal schooling?
- Other examinations of technical adequacy

# Conclusion

- Initial pilot suggest that this tool can capture language skills across languages at a single time point.
- Results from a longitudinal study will help to prioritize the value of developing young children's language skills

# References


Bedore, L. M., Peña, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools*, 36, 188–200. doi:10.1044/0161-1461(2005/020)

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: implications for assessment. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1637-1649.

de Villiers, J. (2015, May). Taking account of both languages in the assessment of dual language learners. In *Seminars in speech and language* (Vol. 36, No. 02, pp. 120-132). Thieme Medical Publishers.

Geva, E., & Yaghoub Zadeh, Z. (2006). Reading efficiency in native English-speaking and English-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading*, *10*(1), 31-57.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574-586.

Leseman, P. P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. Journal of Multilingual and Multicultural Development, 21, 93–112. doi:10.1080/ 01434630008666396

Mancilla-Martinez, J., & Vagh, S. B. (2013). Growth in toddlers' Spanish, English, and conceptual vocabulary knowledge. *Early Childhood Research Quarterly*, 28, 555–567. Retrieved from http://dx.doi.org/10.1016/j.ecresq.2013.03.004

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology*, *40*(1), 85-111.

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders,* 4, 1–21. doi:10.1080/14769670500215647

Umbel VM, Pearson BZ, Fernández MC, Oller DK. Measuring bilingual children's receptive vocabularies. Child Dev 1992; 63 (4) 1012-1020

Walter, S. & Benson, C. (2012) Language policy and medium of instruction in formal education. In Spolsky, Bernard (ed) The Cambridge Handbook of Language Policy, 278-300. Cambridge: Cambridge University Press

# Thank you!  Asante! Muvea!

**Margaret “Peggy” Dubeck, PhD**
Senior Literacy Researcher
RTI International
@pegdubeck
dubeck@rti.org

# Developing Early Literacy Assessments in Spanish

Karen Ford

October 10, 2018

Innovations in ECD Assessment, RTI

University of Virginia
Curry School of Education

pals™ español
Phonological Awareness Literacy Screening

# Background

## Phonological Awareness Literacy Screening (PALS)

- Marcia Invernizzi, P.I.
- Statewide in VA since 1997
- Identifies children in need of literacy intervention
- VA partially subsidizes intervention

# Initial Considerations

- Does it make sense to do this?
- What is known about how literacy develops in Spanish?
- Is literacy development different for bilingual vs. monolingual children?

# Considerations, cont.

- Translation can be threat to validity
- Need for linguistic and cultural neutrality
- Need for stability across groups (gender, instructional program, etc.)

# PALS español: 2004 - 2009

**Preliminary development efforts**

- Research, task/item development
- Informal piloting
  ($n$ = 2,000 in K-3)
- Exploring funding sources

# PALS español: 2009 - 2014

**IES grant to develop K and 1-3**

- Item development, piloting, parallel forms

  (4x/2 yrs./*n* = 400)

- Field testing

  (4x/2 yrs./*n* = 5,000)

# PALS español K Tasks and Domains

| Tasks | Domains | | |
|---|---|---|---|
| | Phonological Awareness | Alphabet Knowledge | Orthographic Knowledge |
| **Rhyme Awareness** | ✓ | | |
| **Beginning Sound Awareness** | ✓ | | |
| **Alphabet Recognition** | | ✓ | |
| **Letter Sound Knowledge** | | ✓ | ✓ |
| **Spelling** | ✓ | ✓ | ✓ |
| **Concept of Word** | ✓ | ✓ | ✓ |

# PALS español 1-3 Tasks and Domains

| Tasks | Domains | | | |
|---|---|---|---|---|
| | Orthographic Knowledge | Word Recognition in Context | Alphabet Knowledge | Phonemic Awareness |
| **Word Recognition in Isolation** | ✓ | | ✓ | ✓ |
| **Spelling** | ✓ | | ✓ | ✓ |
| **Oral Reading in Context** | ✓ | ✓ | ✓ | ✓ |
| **Alphabet Recognition** | | | ✓ | |
| **Letter Sound Knowledge** | | | ✓ | |
| **Concept of Word in Text** | ✓ | | ✓ | ✓ |
| **Blending** | | | | ✓ |
| **Segmenting** | | | ✓ | ✓ |

# PALS español: 2013 - 2018

**IES grant to develop PreK**

- Item development, piloting
  (2x/1 yr./*n* = 677)
- Field testing
  (5x/3 yrs./*n* = 3,900)

# PALS español PreK Tasks

| Tasks | Domains | | | |
|---|---|---|---|---|
| | Phonological Awareness | Alphabet Knowledge | Print Knowledge | Oral Language |
| **Syllable Clapping** | ✓ | | | |
| **Rhyme Awareness** | ✓ | | | |
| **Beginning Sound Awareness** | ✓ | | | |
| **Alphabet Recognition** | | ✓ | | |
| **Letter Sound Knowledge** | | ✓ | | |
| **Print and Word Awareness** | | | ✓ | |
| **Name Writing** | | ✓ | ✓ | |
| **Language and Listening Comprehension** | | | | ✓ |

# Vertical Scale PreK-K

- Need to track literacy development across PreK and K
- Created a vertical scale from fall PreK through spring K using a common item design with concurrent calibration

## Table 4 PreK Spring Summed Score to Scaled Score Conversion Table

| Summed Score | Scaled Score | Percentile Rank | Summed Score | Scaled Score | Percentile Rank | Summed Score | Scaled Score | Percentile Rank | Summed Score | Scaled Score | Percentile Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 100 | 0.0 | 38 | 254 | 30.1 | 76 | 350 | 63.8 | 114 | 452 | 84.9 |
| 1 | 102 | 0.1 | 39 | 256 | 31.0 | 77 | 353 | 64.4 | 115 | 455 | 85.2 |
| 2 | 105 | 0.2 | 40 | 259 | 32.4 | 78 | 355 | 64.9 | 116 | 458 | 85.7 |
| 3 | 108 | 0.3 | 41 | 262 | 34.0 | 79 | 358 | 65.6 | 117 | 461 | 86.1 |
| 4 | 111 | 0.4 | 42 | 264 | 35.0 | 80 | 360 | 66.0 | 118 | 465 | 86.7 |
| 5 | 116 | 0.6 | 43 | 267 | 36.6 | 81 | 363 | 66.7 | 119 | 468 | 87.2 |
| 6 | 121 | 0.9 | 44 | 269 | 37.5 | 82 | 365 | 67.1 | 120 | 471 | 87.5 |
| 7 | 127 | 1.2 | 45 | 272 | 38.9 | 83 | 368 | 67.9 | 121 | 475 | 88.2 |
| 8 | 133 | 1.5 | 46 | 275 | 40.3 | 84 | 370 | 68.5 | 122 | 478 | 88.5 |
| 9 | 139 | 1.6 | 47 | 277 | 40.9 | 85 | 373 | 69.3 | 123 | 482 | 88.9 |
| 10 | 146 | 1.9 | 48 | 280 | 41.9 | 86 | 375 | 69.8 | 124 | 485 | 89.2 |
| 11 | 151 | 2.2 | 49 | 282 | 42.8 | 87 | 378 | 70.5 | 125 | 489 | 89.6 |
| 12 | 157 | 2.8 | 50 | 285 | 44.4 | 88 | 380 | 71.0 | 126 | 493 | 90.1 |
| 13 | 162 | 3.3 | 51 | 287 | 45.0 | 89 | 383 | 71.6 | 127 | 497 | 90.6 |
| 14 | 168 | 3.8 | 52 | 290 | 46.1 | 90 | 386 | 72.0 | 128 | 501 | 90.9 |
| 15 | 172 | 4.4 | 53 | 292 | 47.0 | 91 | 388 | 72.5 | 129 | 505 | 91.2 |
| 16 | 177 | 5.2 | 54 | 295 | 48.2 | 92 | 391 | 73.2 | 130 | 509 | 91.4 |
| 17 | 182 | 6.1 | 55 | 297 | 48.9 | 93 | 393 | 73.6 | 131 | 514 | 91.8 |
| 18 | 186 | 6.7 | 56 | 300 | 50.2 | 94 | 396 | 74.2 | 132 | 518 | 92.1 |
| 19 | 190 | 7.6 | 57 | 302 | 50.9 | 95 | 399 | 74.7 | 133 | 523 | 92.4 |
| 20 | 194 | 8.8 | 58 | 305 | 51.9 | 96 | 401 | 75.2 | 134 | 528 | 92.8 |
| 21 | 198 | 9.6 | 59 | 307 | 52.6 | 97 | 404 | 75.7 | 135 | 533 | 93.1 |
| 22 | 202 | 10.5 | 60 | 310 | 53.5 | 98 | 406 | 76.2 | 136 | 539 | 93.5 |
| 23 | 206 | 11.5 | 61 | 312 | 54.0 | 99 | 409 | 76.9 | 137 | 544 | 93.8 |
| 24 | 210 | 12.3 | 62 | 315 | 54.8 | 100 | 412 | 77.7 | 138 | 550 | 94.2 |
| 25 | 213 | 13.4 | 63 | 317 | 55.5 | 101 | 414 | 78.2 | 139 | 556 | 94.6 |
| 26 | 217 | 14.8 | 64 | 320 | 56.5 | 102 | 417 | 78.8 | 140 | 563 | 95.2 |
| 27 | 220 | 15.9 | 65 | 322 | 57.0 | 103 | 420 | 79.3 | 141 | 569 | 95.5 |
| 28 | 224 | 17.4 | 66 | 325 | 57.7 | 104 | 423 | 79.9 | 142 | 575 | 95.9 |
| 29 | 227 | 18.2 | 67 | 327 | 58.1 | 105 | 425 | 80.3 | 143 | 580 | 96.4 |
| 30 | 230 | 19.1 | 68 | 330 | 58.8 | 106 | 428 | 81.1 | 144 | 585 | 96.8 |
| 31 | 233 | 20.3 | 69 | 332 | 59.3 | 107 | 431 | 81.7 | 145 | 589 | 97.4 |
| 32 | 236 | 21.9 | 70 | 335 | 60.0 | 108 | 434 | 82.2 | 146 | 592 | 97.7 |
| 33 | 239 | 23.0 | 71 | 337 | 60.6 | 109 | 437 | 82.6 | 147 | 595 | 98.3 |
| 34 | 242 | 24.5 | 72 | 340 | 61.4 | 110 | 440 | 83.0 | 148 | 597 | 98.7 |
| 35 | 245 | 25.8 | 73 | 342 | 61.9 | 111 | 443 | 83.5 | | | |
| 36 | 248 | 27.2 | 74 | 345 | 62.7 | 112 | 446 | 84.0 | | | |
| 37 | 251 | 28.7 | 75 | 348 | 63.2 | 113 | 449 | 84.4 | | | |

# Our goals - for PALS español to be:

- psychometrically sound
- broad-based
- developmentally appropriate
- culturally and linguistically sensitive
- instructionally useful

## Contact information:

Karen Ford
Curry School of Education, University of Virginia
kford@virginia.edu
434-924-7569
https://pals.virginia.edu

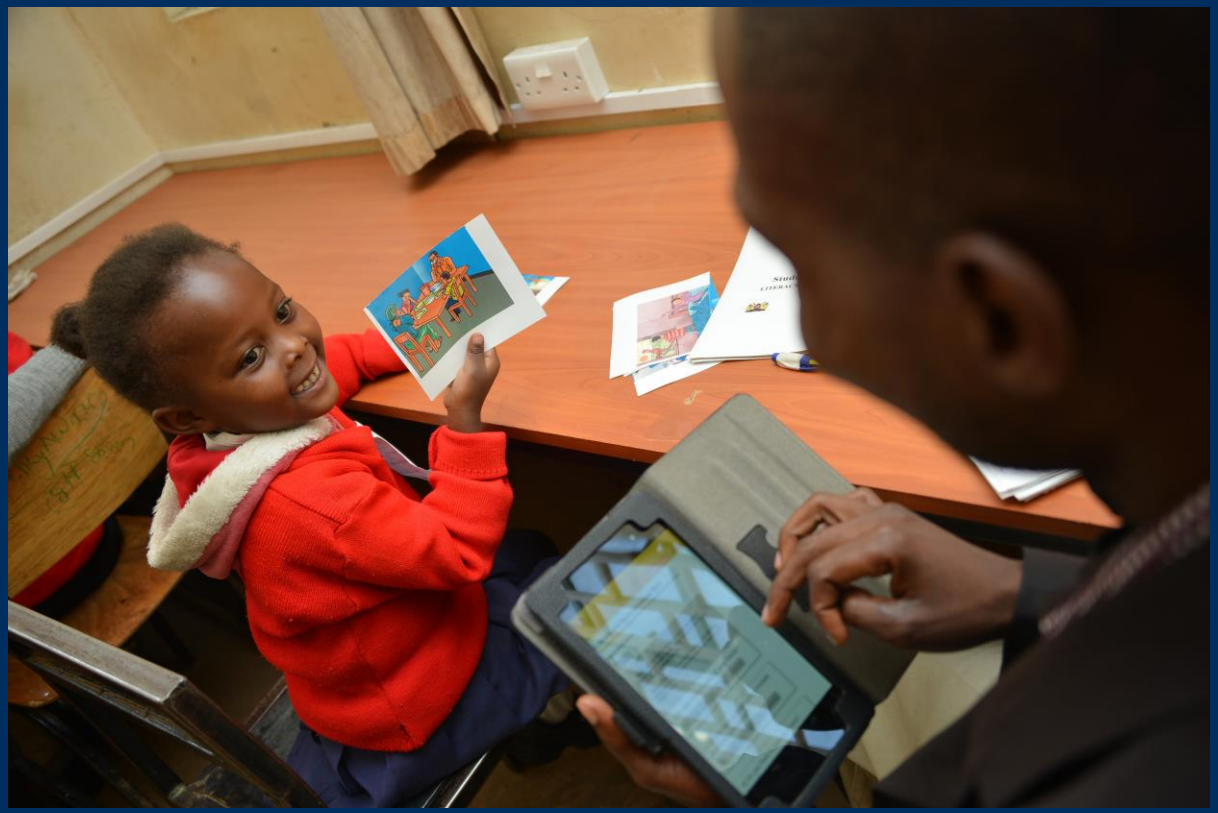
The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A130469 and R305A090015 to The University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

# Language and Literacy Assessment

**Margaret "Peggy" Dubeck**
Senior Literacy Researcher
*RTI International*

**Karen Ford**
Assistant Professor, Research Faculty
*University of Virginia Curry School of Education*

# Innovations in Early Childhood Development Assessment



Wednesday, October 10, 2018
1:00 p.m. – 5:00 p.m.

@RTI_INTL_DEV | #RTILearns | #EarlyYears

RTI INTERNATIONAL

# Executive Function Assessment

**Jelena Obradović**
Associate Professor, Developmental and Psychological Sciences Program
*Stanford Graduate School of Education*

**Michael Willoughby**
Fellow and Senior Research Public Analyst
*RTI International*

# STUDYING CHILDREN'S EXECUTIVE FUNCTIONS IN A GLOBAL CONTEXT

**Dr. Jelena Obradović**
**Stanford University**

Stanford | GRADUATE SCHOOL OF EDUCATION


SELF-REGULATION
attention, behavior, emotions
1. SELECT & PURSUE GOALS
3. MONITOR & EVALUATE PROGRESS
2. PLAN & ORGANIZE BEHAVIOR
4. DELAY GRATIFICATION

# EXECUTIVE FUNCTIONS (EFs)

- **INHIBITORY CONTROL**
  - Suppress impulsive thoughts or behaviors
  - Resist distractions and temptations

- **WORKING MEMORY**
  - Hold, update, and manipulate information in the mind

- **COGNITIVE FLEXIBILITY**
  - Shift attention or responses between competing rules or mental states

# WHY STUDY CHILDREN'S EXECUTIVE FUNCTIONS IN LMIC SETTINGS?

# EFs: AN INDEX OF EARLY DEVELOPMENT

1. Basic building blocks of various competences
   - cognitive, social, emotional
2. Set important developmental cascades in motion
3. Can promote resilience in at-risk children
4. Malleable; can be improved via interventions
5. A marker of early experience
   - a proxy for early neuro development

# EFs: A CULTURALLY UNIVERSAL MEASURE

1. Assessment can be adapted to be both developmentally appropriate and culturally sensitive. It depends less on educational/schooling experiences than IQ/academic tests.
2. Can help us identify children who are doing well or have a capacity to do well in culturally and ethnically diverse settings.

# HOW DO WE MEASURE EXECUTIVE FUNCTIONS?

# EF ASSESSMENT

- **PARENT & TEACHER REPORT**
  - Feasibility (pre-K, scale)
  - Reliability (literacy)
  - Bias (demographic groups)
  - Validity (general conduct)
- **STANDARDIZED TASKS**
  - Table-top or computerized tasks in a lab-like setting

# EF ASSESSMENT IN LMIC

- Build in time for adaptation (and get/give funding for it)
- Work with local experts & leaders
- Build local capacity by advancing conceptual, methodological, and pragmatic expertise
- Ongoing quality control checks:
  - Training, certification, and ongoing supervision of child assessors
  - Daily briefings, weekly progress reports, monthly observations, double-coding, conference calls

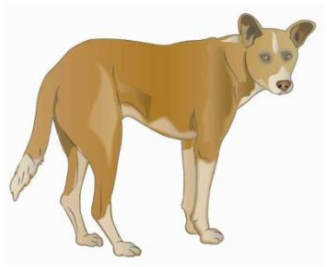
# COMPUTER-BASED ASSESSMENT OF EFs

+ can be administered quickly (also in group settings)
+ with minimal training
+ in an ecologically valid setting
+ yielding data for all students
= a pragmatic, cost-effective, scalable direct assessment

# QUALITY ASSURANCE: COMPUTER TASKS

- Easy to collect bad data!
- Contextualize task instructions!
- Data checks: instruction, practices trials, perseverative responding
- Convergent validity: Assessor reports
- Divergent validity: IQ data

Example: GO/NO-GO TASK

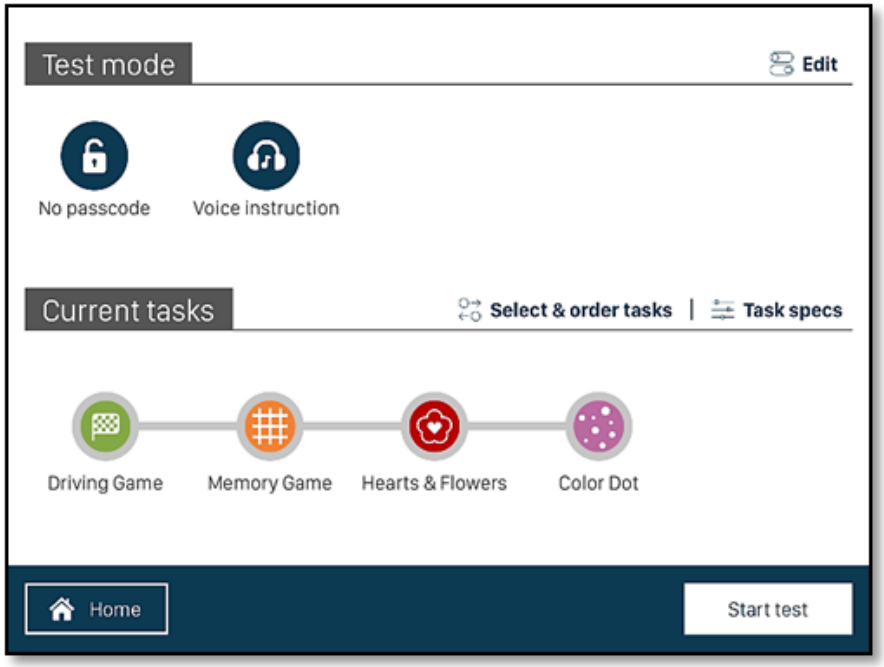



# BEYOND EFs . . .

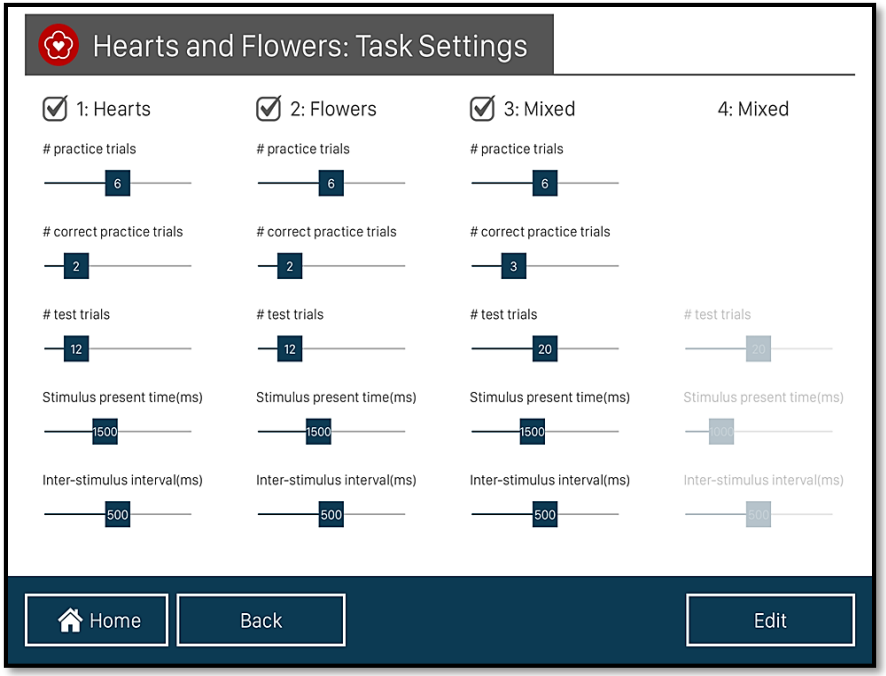

- We need to study children's EF skills together with their motivation (persistence, challenge preference) and emotion regulation (frustration tolerance) to fully understand how self-regulation promotes learning and adaptation.

# ASSESSMENT OF MOTIVATION, EFFORT & SELF-REGULATION

## Customizable administration




## Customizable task settings



# HOW DO WE PROMOTE EXECUTIVE FUNCTIONS?

# INFANCY

- **Hiding games**: following sequences, managing suspense

- **Singing games**: predictable rhymes, anticipatory reactions, hand gestures, repetitions

# PRESCHOOL

- **Role play: holding and updating complex play rules and scripts, staying in character, controlling impulses, taking turns, flexibly adjusting to your partner's ideas, repurposing objects**

# ELEMENTARY SCHOOL

- <u>Games, puzzles</u>: turn taking, strategy, taking other person's perspective, collaborating

- <u>Physical activities, dancing, cooking, playing music</u>: following rules, quick decision making, constant monitoring, respond flexibly

# ALL AGES

- Conversations: actively focusing and sustaining attention, ignoring distractions
- Storytelling: holding, updating, manipulating, and organizing information in their heads
- Routines: planning, goal setting, self-monitoring
- Bilingual experience

# HOME CONTEXT & PARENTING PRACTICES

- **Cognitive stimulation**: reading, high quality child care, educational toys and experiences

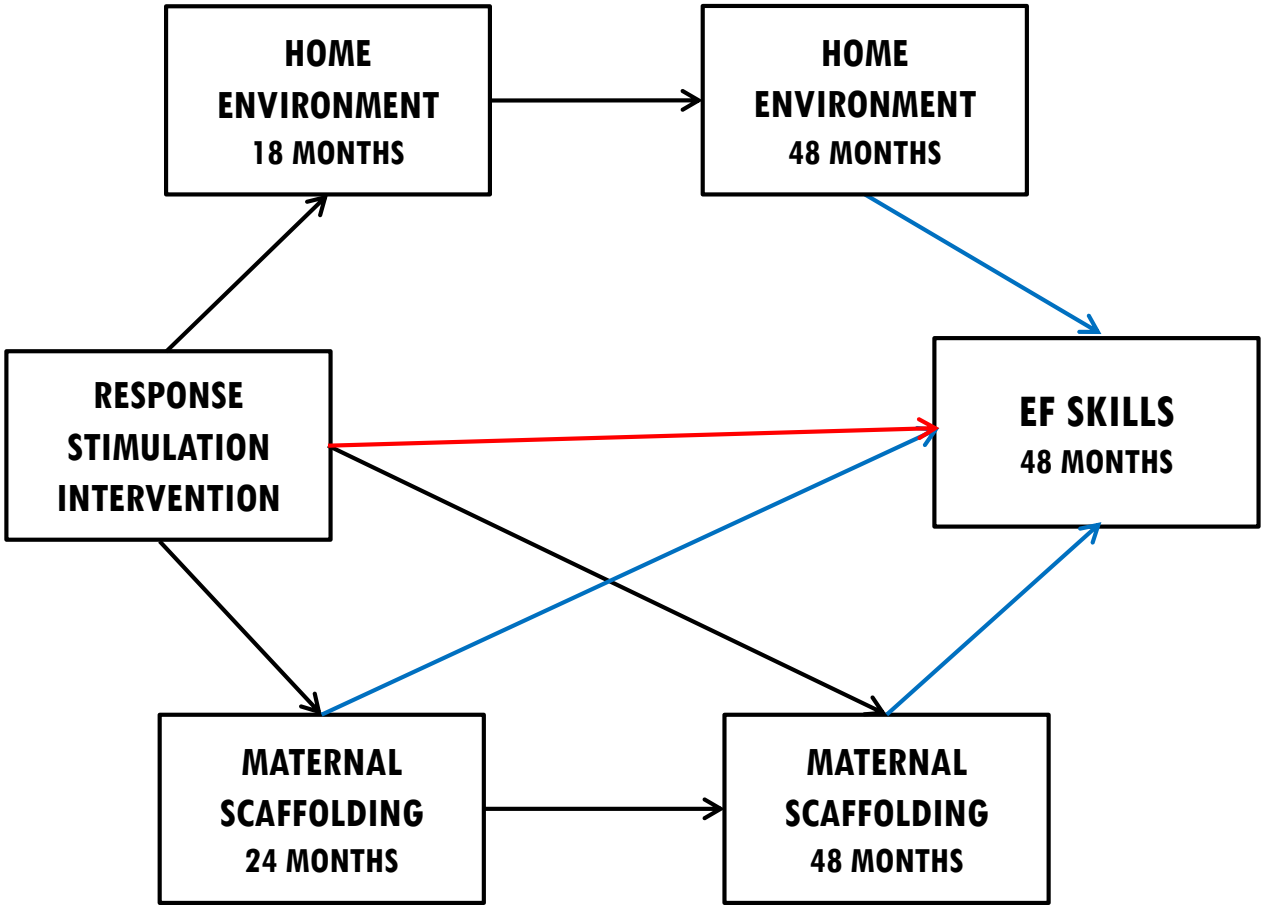
- **Parental scaffolding**: autonomy support, verbal and physical promoting, praise & elaboration, maintaining & redirecting attention

# ECD INTERVENTION

**Conducted by Dr. Aisha Yousafzai**

- **RCT in high risk Rural Pakistan**
- **Observation, coaching & feedback via monthly home visits and community groups**
- **Activities focus on**
  - **sensitive & responsive caregiving**
  - **mother/child play; communication**
  - **cognitive stimulations**
- **Conducted follow-up of 1144 very disadvantaged preschoolers**

**Yousafzai et al., 2014**




Obradović et al., 2016

**ECD intervention effect remained significant after controlling for physical growth, age 2 & age 4 cognitive skills, and maternal cognitive skills.**

# PARENTAL EFs → PARENTING BEHAVIORS

- Greater parental EF skills predict:
  - less harsh and reactive parenting (Deater-Deckard et al., 2010)
  - higher levels of supportive, responsive, and sensitive parenting (Shaffer & Obradović, 2017)
  - higher levels of maternal cognitive scaffolding (Obradović, Portilla, et al., 2017)

# TWO-GENERATIONAL APPROACHES

- There is a need to design interventions that target two generations, promoting self-regulation in both children and caregivers, especially those who parent or teach in stressful and chaotic contexts.
- We need to understand how older siblings and peers promote EFs in young children from LMIC settings.

# THANK YOU!

e-mail: jelena.obradovic@stanford.edu
website: http://www.stanford.edu/group/sparklab/

# ACKNOWLEDGMENTS


- My lab: Ximena Portilla, Nicole Tirado-Strayer, Parissa Ballard, Sarah Bardack, Jenna Finch, Elisa Garcia, Liam Aielo, Victoria Rodriguez, Michael Sulik.
- My collaborators: Aisha Yousafzai, Muneera Rasheed and AKU STAR team, Amanda Tarullo, Chuck Nelson, Anne Shaffer, Janette Herbers, Nicole Bush, Tom Boyce, Keith Burt, Ann Masten, Frosso Motti-Stefanidi
- Many Stanford undergraduate students
- School district administrators, principals, teachers
- Children and their families


UBC HUMAN EARLY LEARNING PARTNERSHIP

NIMH National Institute of Mental Health

Les Fiducies Killam Trusts

CIFAR CANADIAN INSTITUTE for ADVANCED RESEARCH

JACOBS FOUNDATION

WILLIAM T. GRANT FOUNDATION

Grand Challenges Canada™ Grands Défis Canada™

Stanford | GRADUATE SCHOOL OF EDUCATION

# Innovations in ECD Assessment

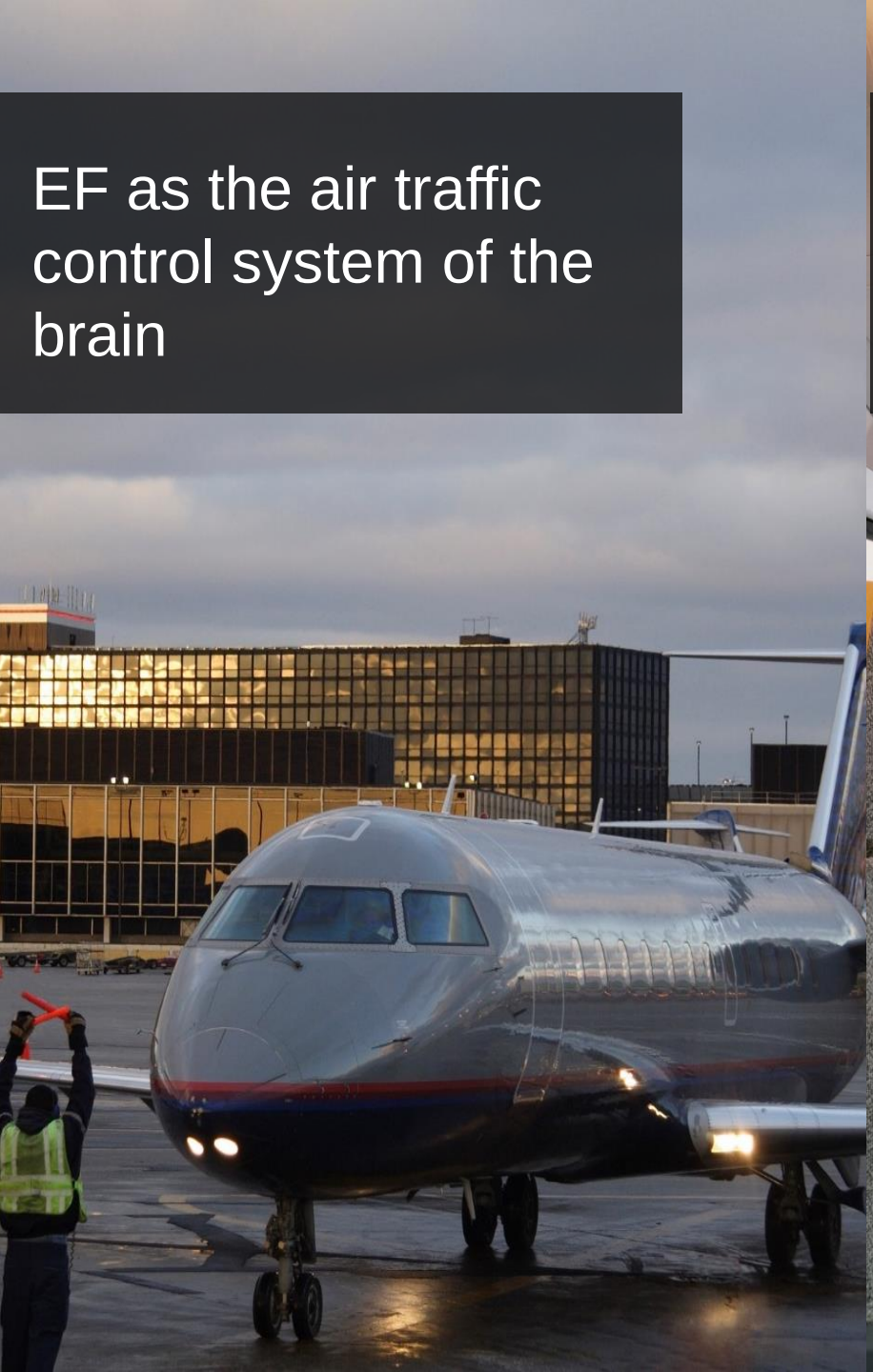
## Measuring Executive Functions

Michael Willoughby

October 10, 2018

# Defining Executive Functions (EF)

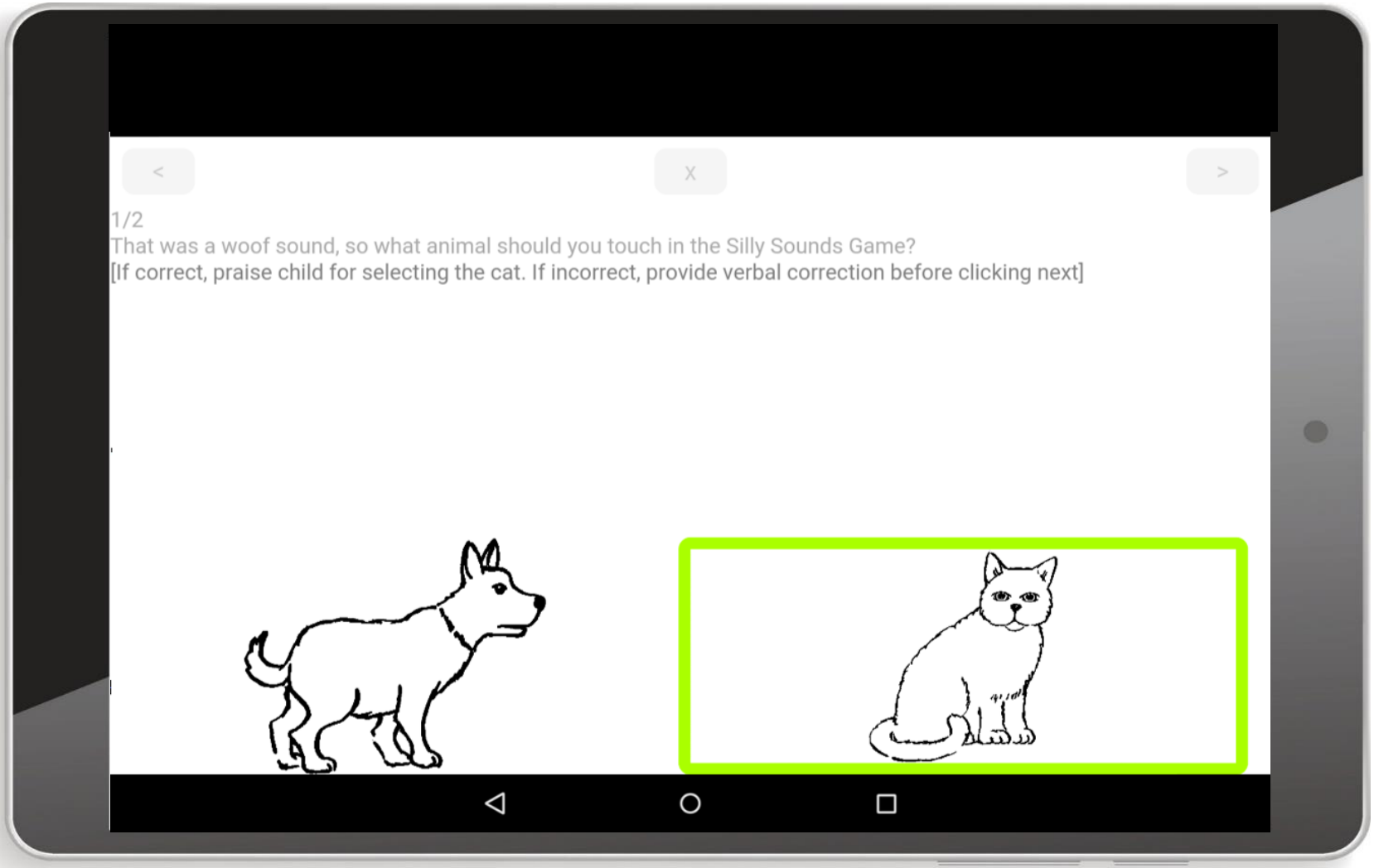
EF as the air traffic control system of the brain



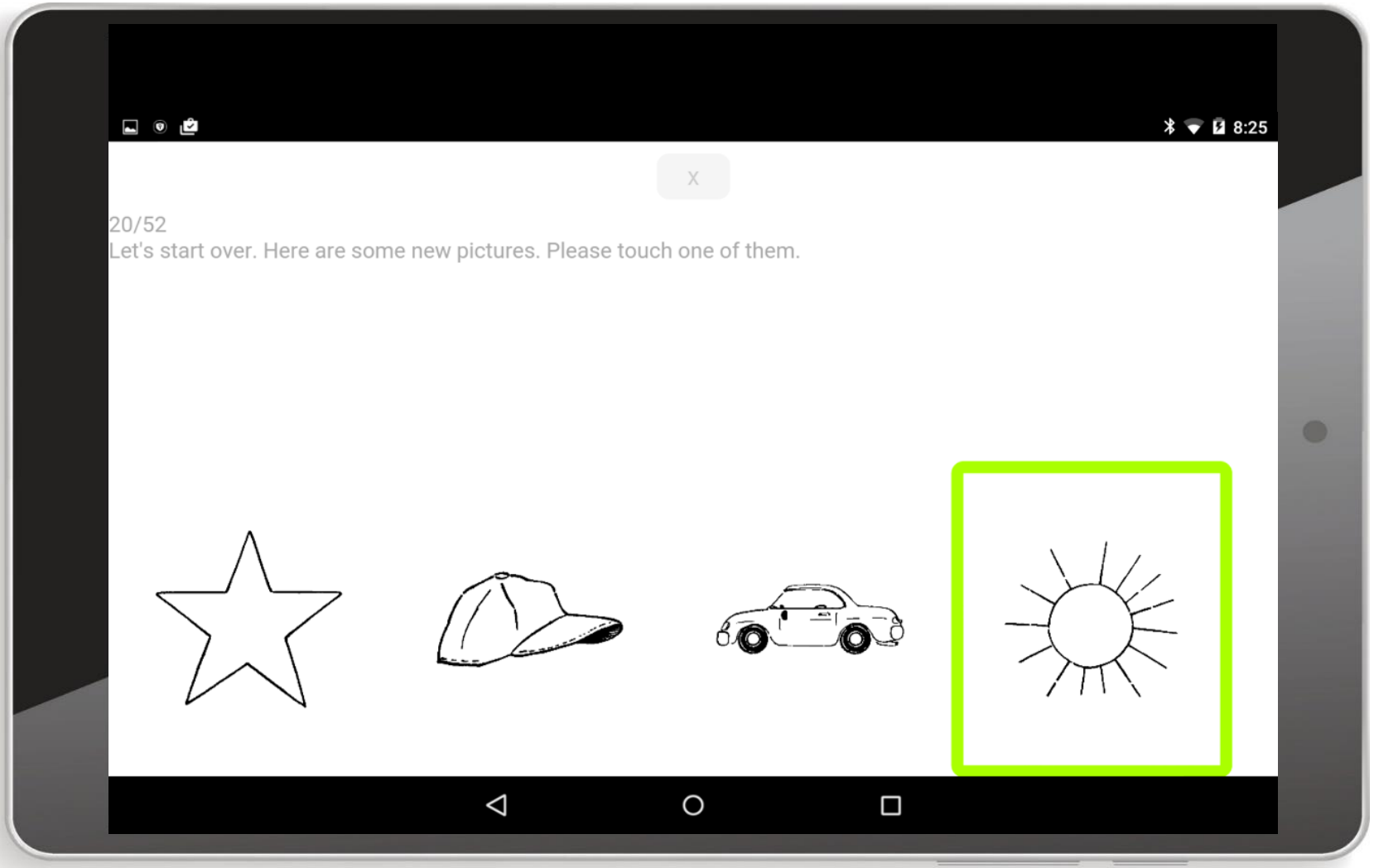
EF skills are needed when autopilot won’t work

EF skills are mediated by circuitry in the prefrontal cortex

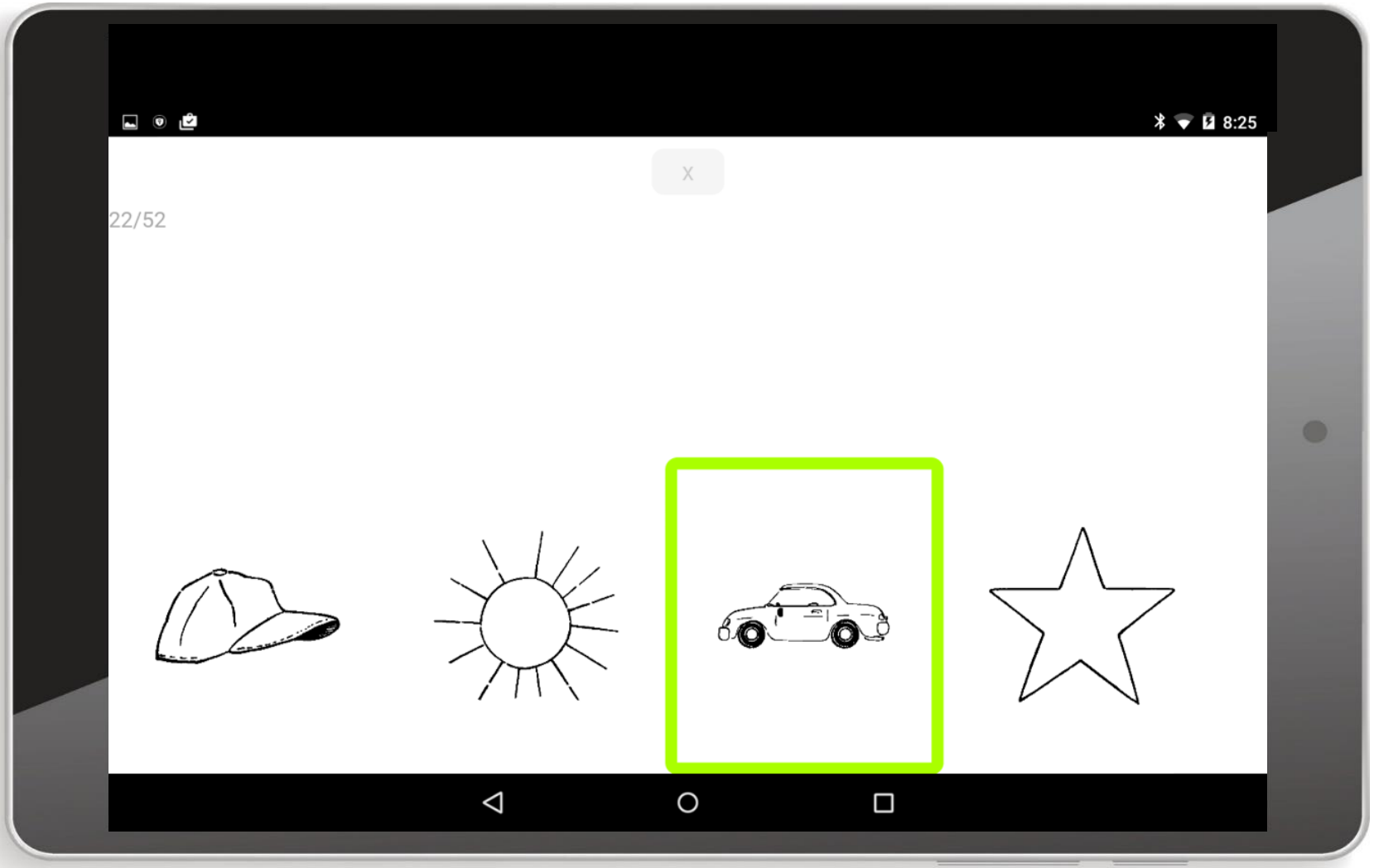
# Silly Sounds Stroop



# Pick the Picture



# Pick the Picture



# Measuring EF in Developing Country Contexts

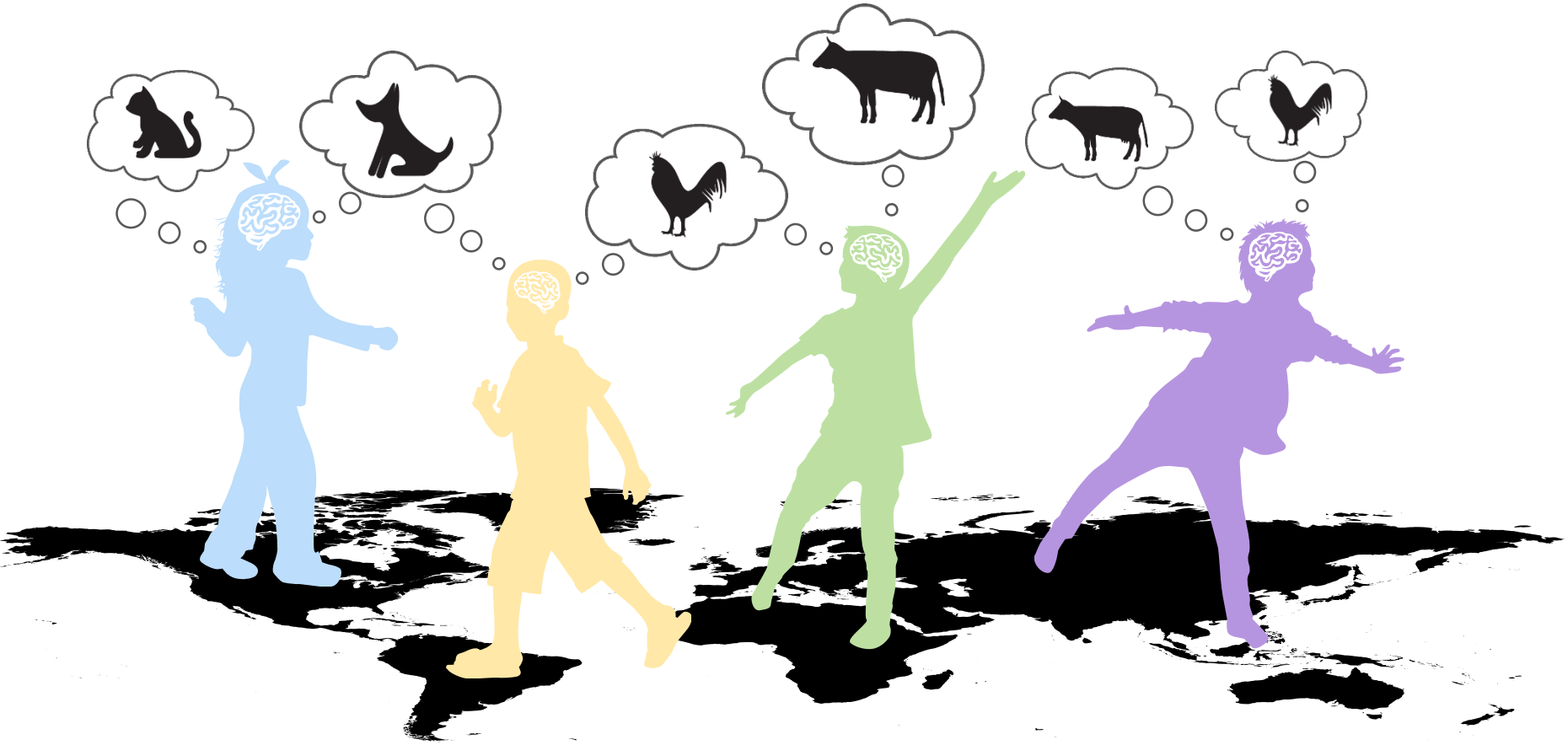
## *Theoretical* Considerations

Cross-cultural variation in young children's experience may influence child EF task performance

# Measuring EF in Developing Country Contexts

## *Practical* Considerations

# Thinking quickly vs. accurately



Specific task stimuli

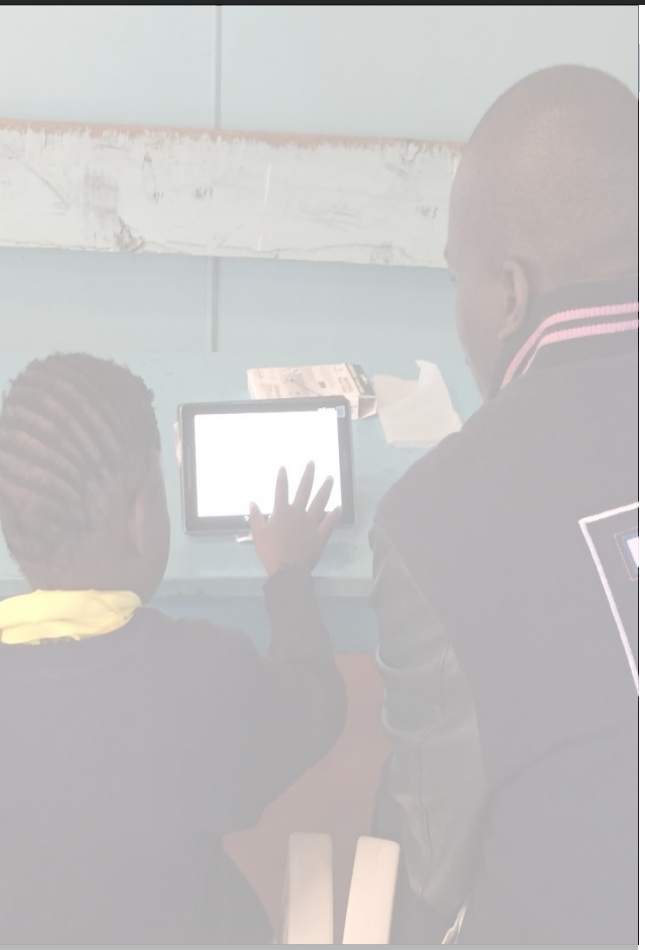
# Benefits of Using a Tablet

- Standardized administration
- Accuracy and timing data
- Facilitates large scale testing
- Experience with tablet is unnecessary

# Does Familiarity with Technology Influence Performance?

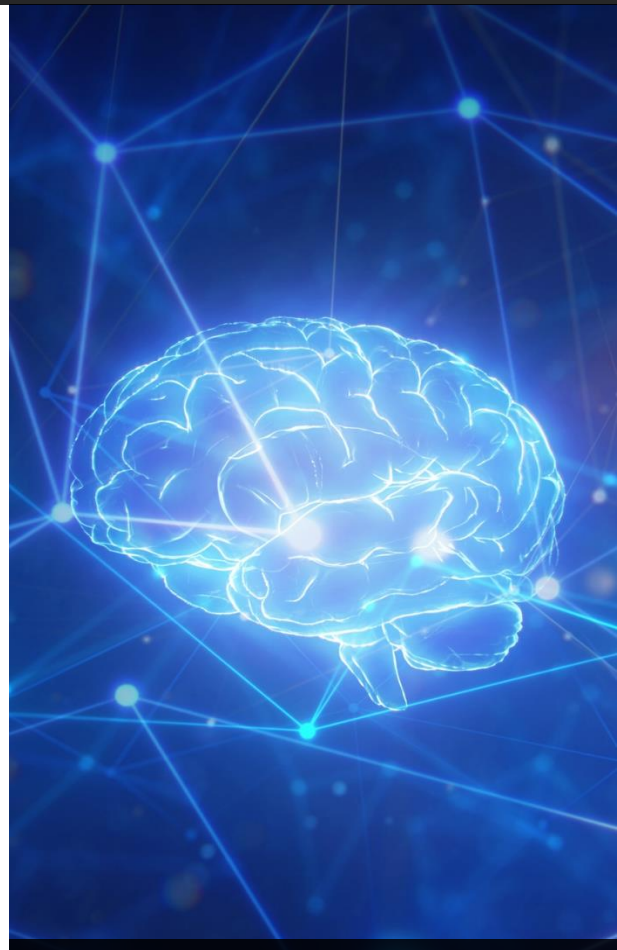
# Measurement Impurity Problem

# Measurement Impurity Problem



Assessor effects?

Performance reflects trait & state



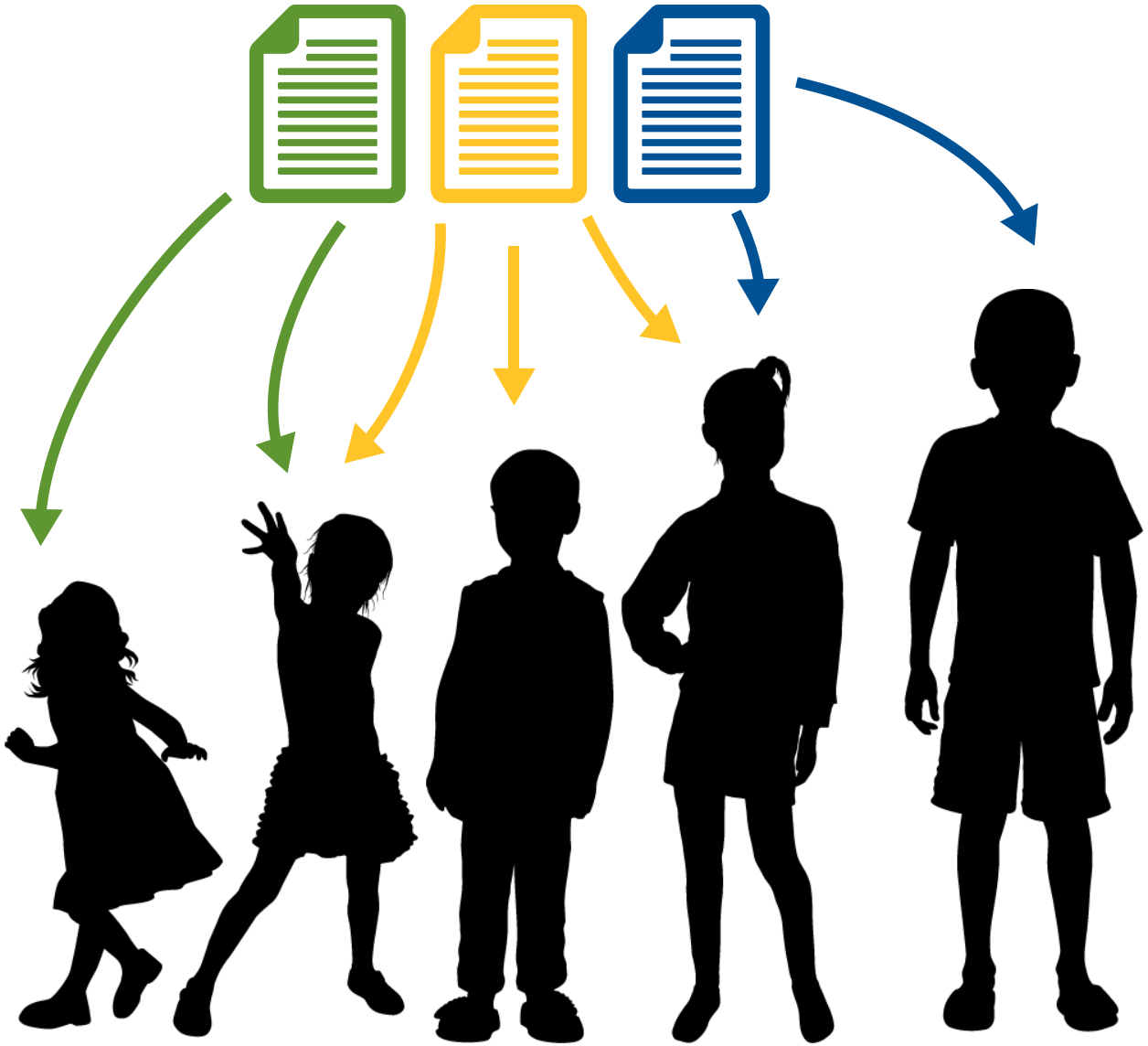
Performance reflects multiple cognitive skills

# Limited Time Problem

- Measurement vs. logistical constraints
- Prefer battery of tasks
- Limitations of single task approach

# Linkages Problem

# Linkages Problem



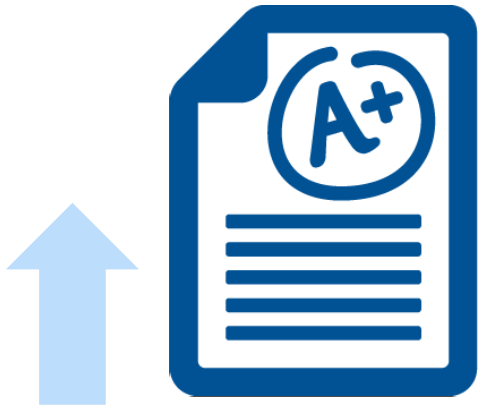
# Measure Selection Problem

- Open Access
- Iterative measure development matters
- Beware of aesthetics over psychometrics

# Inference Problem

# Inference Problem

Executive Function




School Readiness/Achievement

- Associations established
- Correlations provide weak inference
- Need experimental studies to inform policy relevant Qs.

Thank you!

mwilloughby@rti.org

# Executive Function Assessment

**Jelena Obradović**
Associate Professor, Developmental and Psychological Sciences Program
*Stanford Graduate School of Education*

**Michael Willoughby**
Fellow and Senior Research Public Analyst
*RTI International*

# Social Emotional Learning Assessment

**Matthew Jukes**
Fellow and Senior Education Evaluation Specialist
*RTI International*

**Sharon Wolf**
Applied Developmental Psychologist and Assistant Professor
*University of Pennsylvania Graduate School of Education*

# Measuring Social-emotional Skills across Cultures: Invariance of the IDELA across Five Countries

Sharon Wolf
*University of Pennsylvania*

Innovations in Early Childhood Development Assessment
RTI International, October 10, 2018

# Two parts

- **Part 1:** Measuring social-emotional skills across five countries
  - *Co-authors*: Peter Halpin, Hirokazu Yoshikawa, Natalia Rojas, Sarah Kabay, Amy Jo Dowd, Lauren Pisani

- **Part 2:** Examining the role of social-emotional skills with other domains of development
  - *Co-author*: Dana C. McCoy

# SDG 4.2.1: Tracking ECD holistically

- Target 4.2: By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are **ready for primary education**
  - **Indicator 4.2.1:** Proportion of children under 5 years of age who are developmentally on track in health, learning and **psychosocial well-being**, by sex.

- Universal indicators, or separate national standards? Can we compare indicators of "psychosocial well-being" across countries?

# To address these questions, we need…

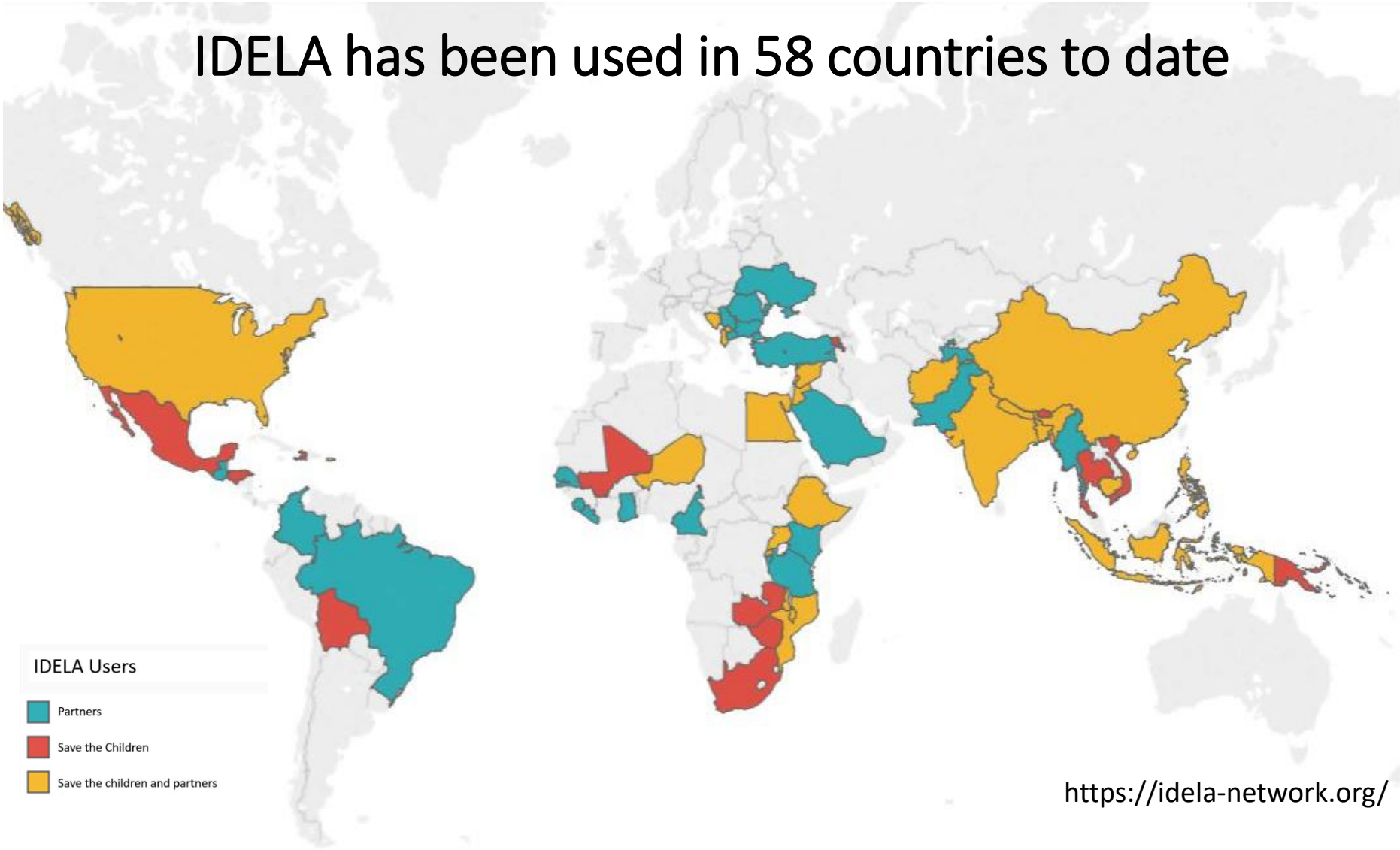
- Measures using the same tool across countries
- Construct validity within countries
- Measurement invariance across countries
- This study attempts to answer these questions using the **International Development and Early Learning Assessment (IDELA).**
- **5 countries:** Afghanistan, Ethiopia, Bolivia, Uganda, Vietnam

# Overview of the IDELA

- Play-based assessment tool designed for children in the 3.5-6 age group
- Takes about 30 minutes per child
- Includes 24 core items that cover 4 developmental domains + self-regulation
- Plus the enumerator's overall assessment of the child's approaches to learning

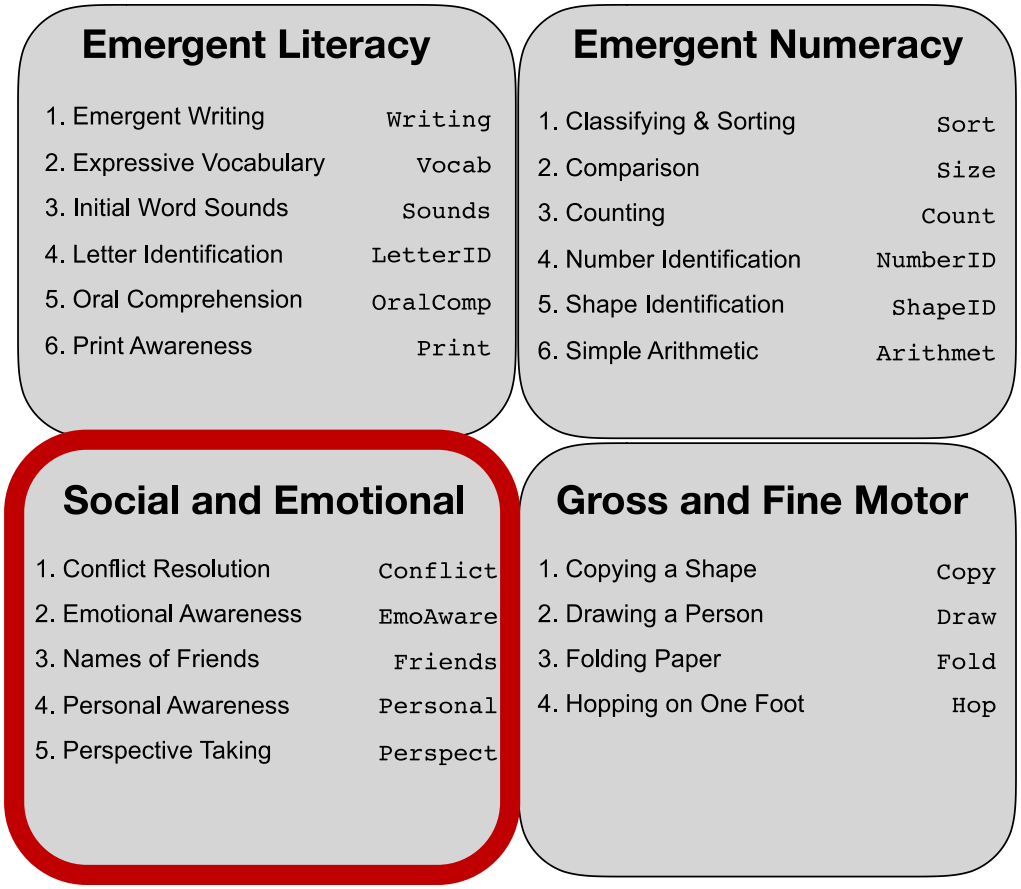
Save the Children®

# IDELA has been used in 58 countries to date




https://idela-network.org/

*While the questions are the same across countries, each country team decides on what answers are "appropriate" / correct in that context.*




Example:

**Imagine you are playing with a toy you like and another child wants to play with the same toy, but there is only one toy. What would you do in this situation?**

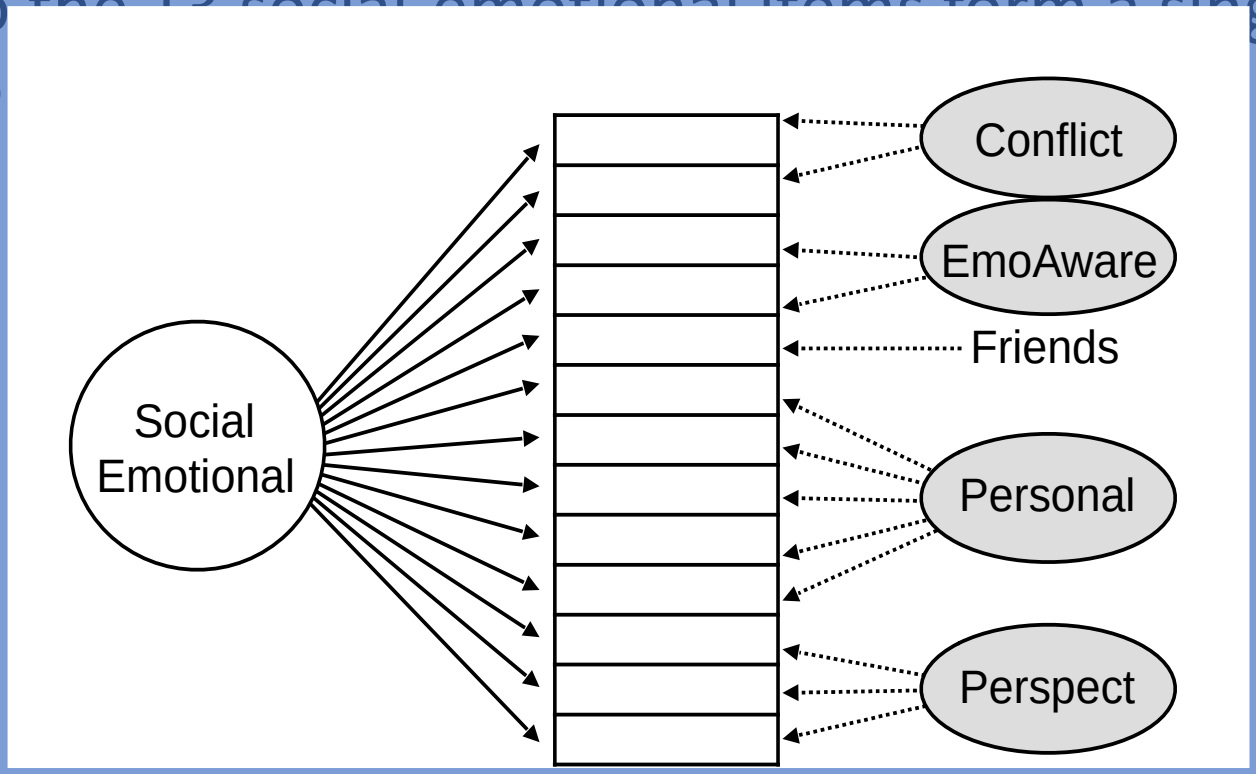
# Datasets

| Country | Sample Size | Age | % female | % ECCD | Urbanicity |
|---|---|---|---|---|---|
| Afghanistan | 2,629 | 5.4 (3 – 8) | 57% | 44.6% | Urban + rural |
| Bolivia | 480 | 4.7 (3 – 6) | 49% | 100.0% | Peri-urban |
| Ethiopia | 682 | 5.9 (4 – 7) | 52% | 76.1% | Rural only |
| Uganda | 504 | 4.6 (4 – 6) | 48% | 48.6% | Rural only |
| Vietnam | 675 | 4.3 (3 – 5) | 50% | 100.0% | Rural only |

*Note*. The sample mean and range are reported for Age. ECCD denotes enrollment in an early child care and development program.

# Analytic approach

1. Do the 13 social emotional items form a single do



# Analytic approach

1. Do the 13 social-emotional items form a single domain bi-factor model in each country?
   - ***Configural invariance:*** If yes, does this factor have a similar interpretations across groups? Necessary but not sufficient to make group comparisons.
2. Do the scores obtained from the factor provide unbiased comparisons over countries?
   - ***Scalar invariance:*** Ensures *all* test items perform equivalently (e.g., have the same difficulty) across groups. Very restrictive.
   - ***Partial invariance***: uses a small "anchor set" of equivalent items to statistically equate other items that are not directly comparable groups.
     - Differential item functioning (DIF) analysis.

# Configural invariance established across all 5 countries

Summary of goodness of fit across countries: Social-emotional factor

| Country | $\chi^2$ (df) | RMSEA (90% CI) | TLI |
|---|---|---|---|
| **Afghanistan** | 143.738 (58) | 0.024 (0.019, 0.029) | .984 |
| **Bolivia** | 86.634 (58) | 0.032 (0.016, 0.046) | .981 |
| **Ethiopia** | 138.489 (58) | 0.045 (0.036, 0.055) | .965 |
| **Uganda** | 87.968 (58) | 0.032 (0.017, 0.045) | .971 |
| **Vietnam** | 116.053 (58) | 0.039 (0.028, 0.049) | .969 |

*Note:* Goodness of fit indicators: $\chi^2$ / df < 2; RMSEA < .050; TLI > 0.90.

# No evidence of scalar invariance

| | $\chi^2$ (df) | RMSEA (90% CI) | TLI | $\chi^2$ (difference) | p-value |
|---|---|---|---|---|---|
| **Configural** | 224.940 (57) | .035 (.030, .040) | .952 | | |
| **Scalar** | | | | | |

# DIF: No anchor set of items identified

| | Item | Afghanistan | Bolivia | Ethiopia | Uganda | Total |
|---|---|---|---|---|---|---|
| 1 | CONFLIC1 | | | | | |
| 2 | CONFLIC2 | | | | | |
| 3 | EMOTION1 | | | | | |
| 4 | EMOTION4 | | | | | |
| 5 | EMPATHY1 | | | | | |
| 6 | EMPATHY2 | | | | | |
| 7 | EMPATHY3 | | | | | |
| 8 | FRIENDS | | | | | |
| 9 | PERSON1 | | | | | |
| 10 | PERSON2 | | | | | |
| 11 | PERSON3 | | | | | |
| 12 | PERSON4 | | | | | |
| 13 | PERSON5 | | | | | |
| | **Total** | | | | | |

# Conclusions – Part 1

- Robust statistical evidence supporting the generalizability of social-emotional development across countries.
- Results do not support unbiased cross-country comparisons (e.g., mean differences), and it was also not evident that any particular subset of items can serve this purpose.
- This issue is likely not particular to IDELA, but rather reflective of cultural and contextual variation in expectations about child development at the level of specific skills and competencies.

# Part 2: If we can measure social-emotional development, what is its role in children's learning?

# Correlations of SE with other developmental domains

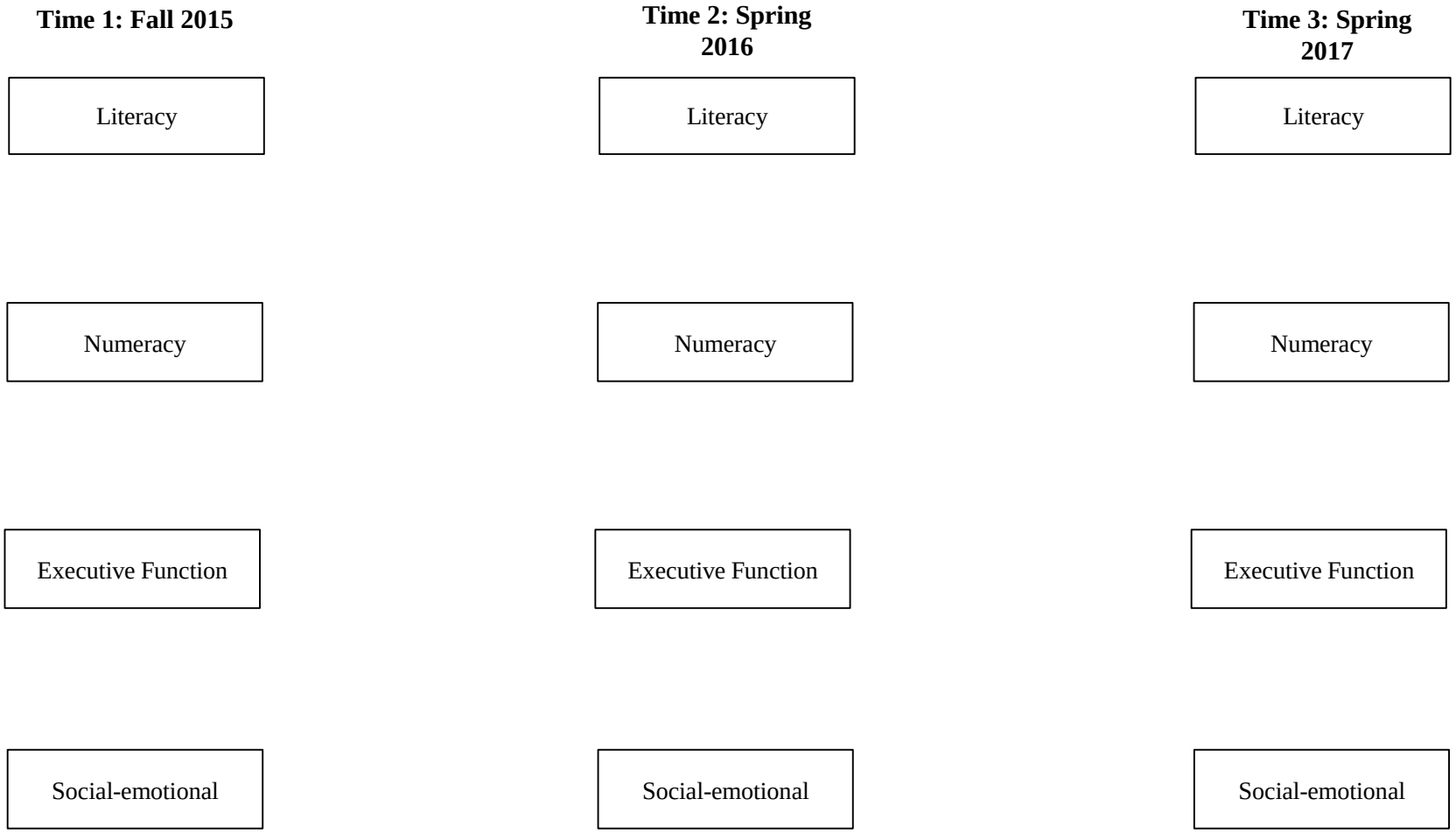
| | Country | | | | |
|---|---|---|---|---|---|
| | Afghanistan | Bolivia | Ethiopia | Uganda | Vietnam |
| **Literacy** | .877 | .807 | .917 | .880 | .830 |
| **Numeracy** | .897 | .690 | .881 | .830 | .667 |
| **Motor** | .820 | .624 | .781 | .846 | .649 |

**Time 1: Fall 2015**

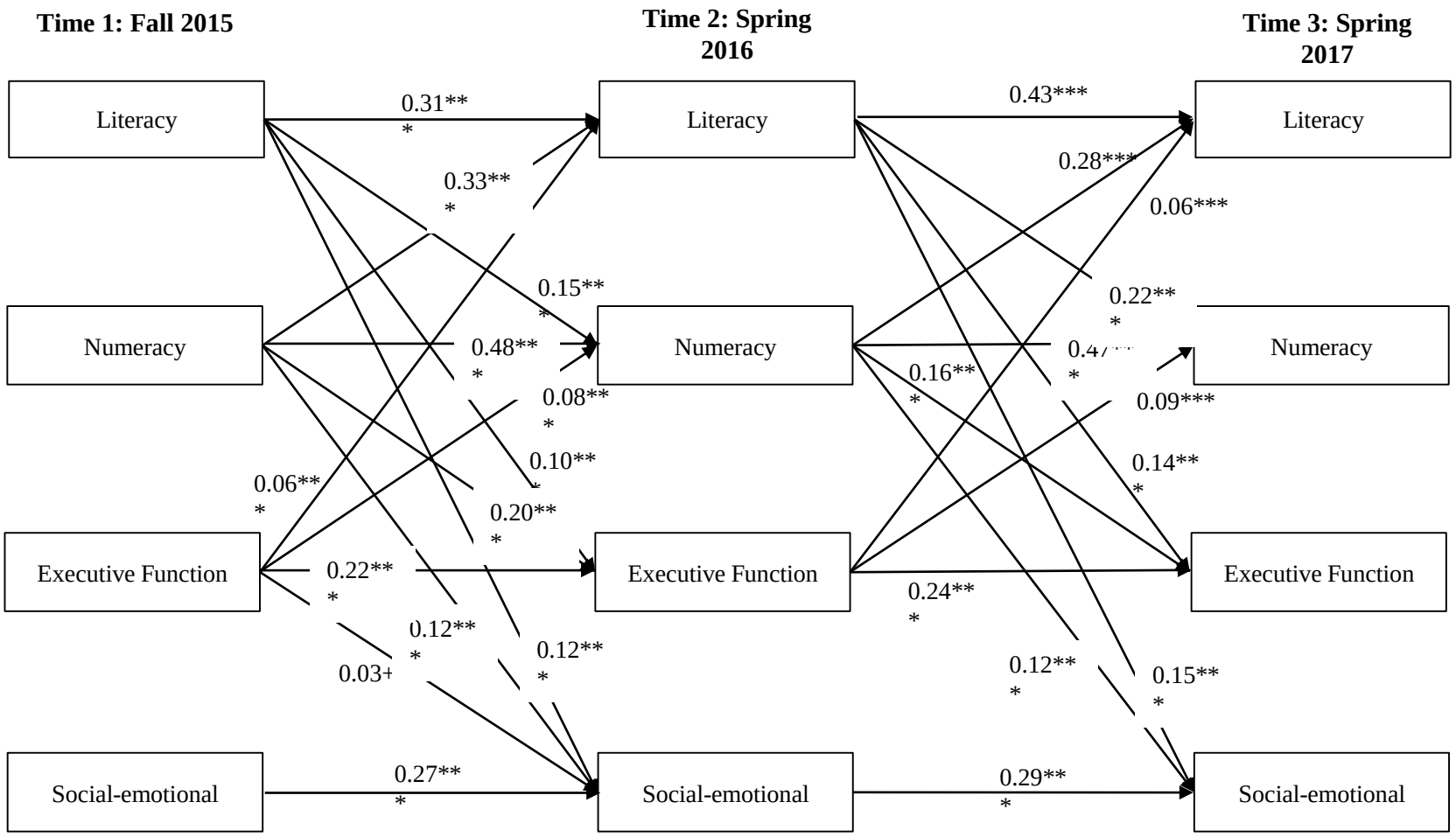
**Time 2: Spring 2016**

**Time 3: Spring 2017**

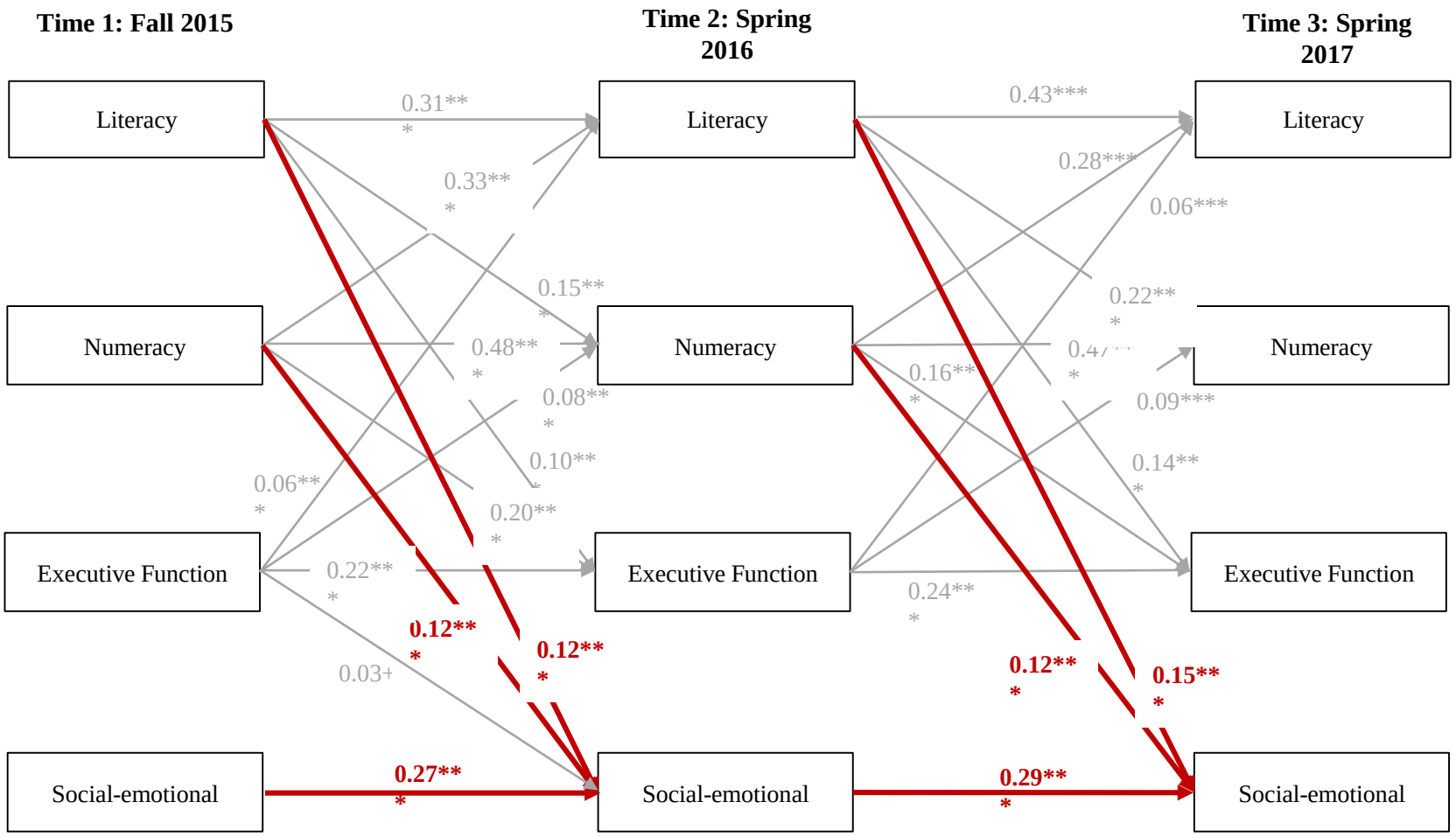
**The inter-connected role of academic and non-academic skills in Ghana; M = 3,862** (Wolf & McCoy, under review)



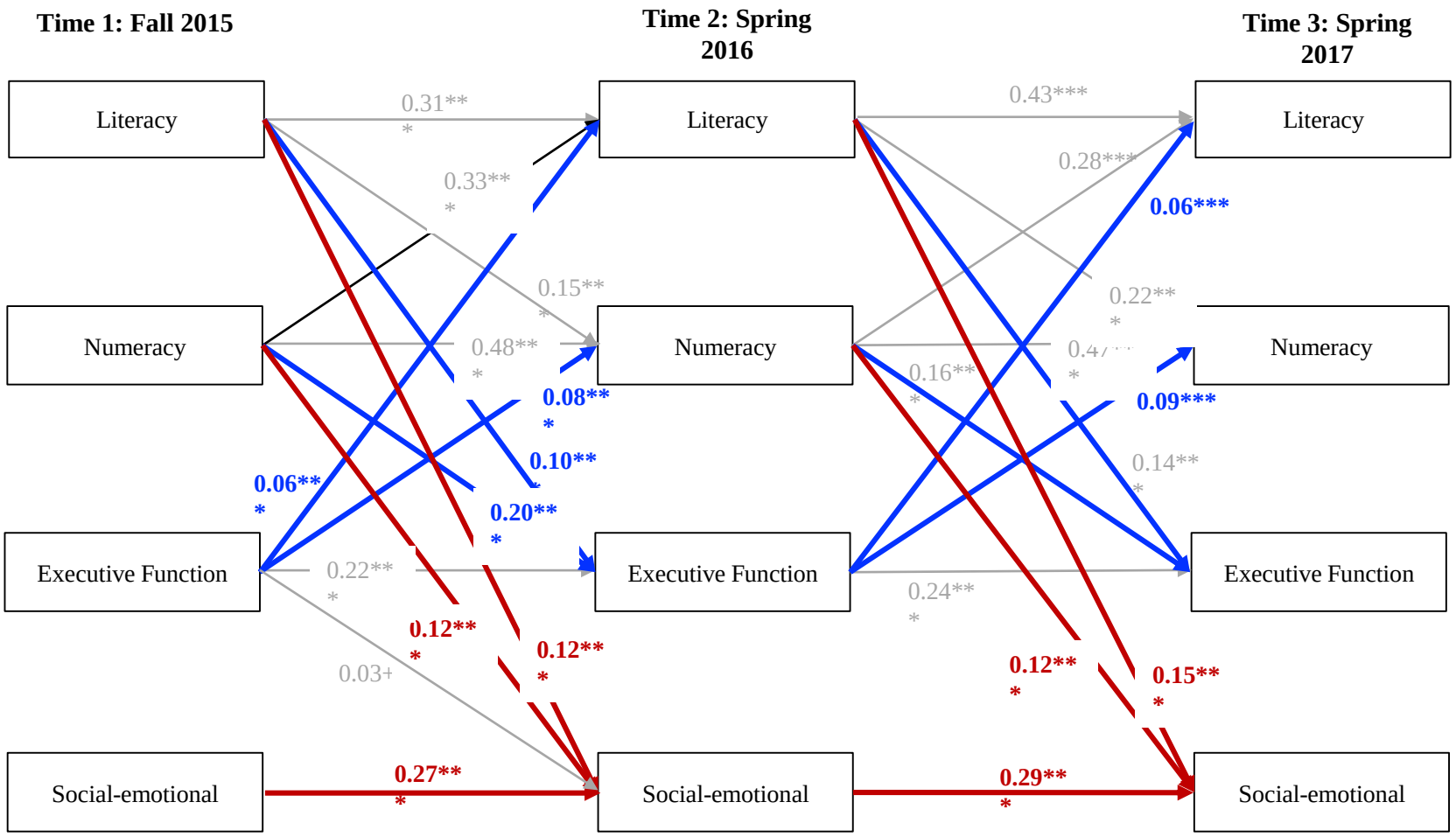

**The inter-connected role of academic and non-academic skills in Ghana; M = 3,862** (Wolf & McCoy, under review)




**The inter-connected role of academic and non-academic skills in Ghana; M = 3,862** (Wolf & McCoy, under review)




**The inter-connected role of academic and non-academic skills in Ghana; M = 3,862** (Wolf & McCoy, under review)




**The inter-connected role of academic and non-academic skills in Ghana; M = 3,862** (Wolf & McCoy, under review)

# Conclusions – Part 2

- Once we have a good measure of SE skills, what role does it play in children's development?
- Among Ghanaian preschoolers, EF plays a central role in supporting growth in literacy and numeracy skills; SE skills do not.
- But academic outcomes predict subsequent SE and EF skills.
- Examining the pattern of associations of SE skills with other domains will inform our understanding of its role in development in diverse contexts.

# Thank you

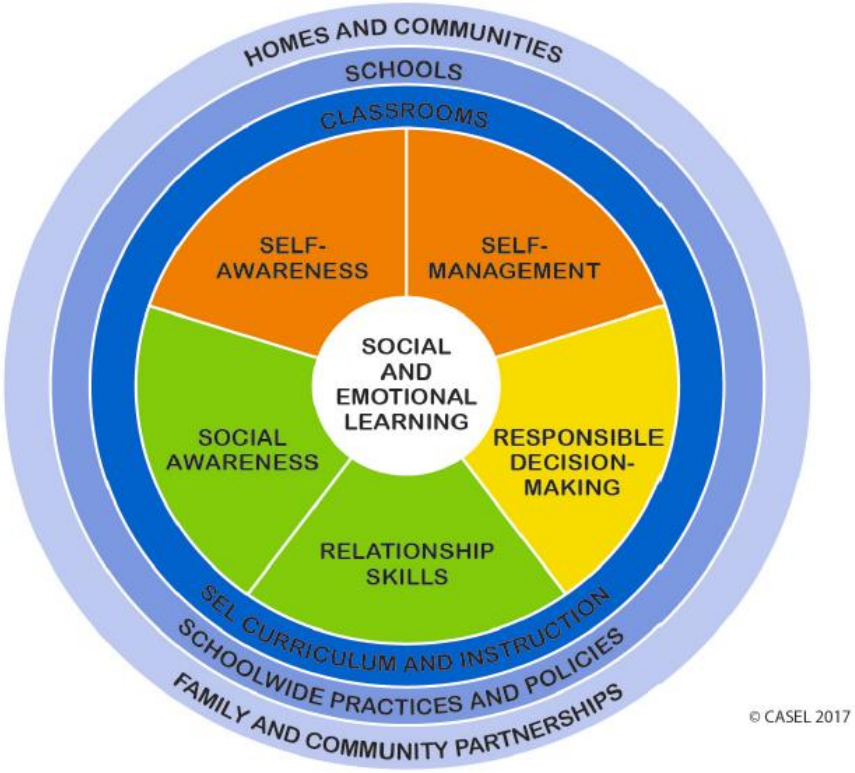
wolfs@upenn.edu

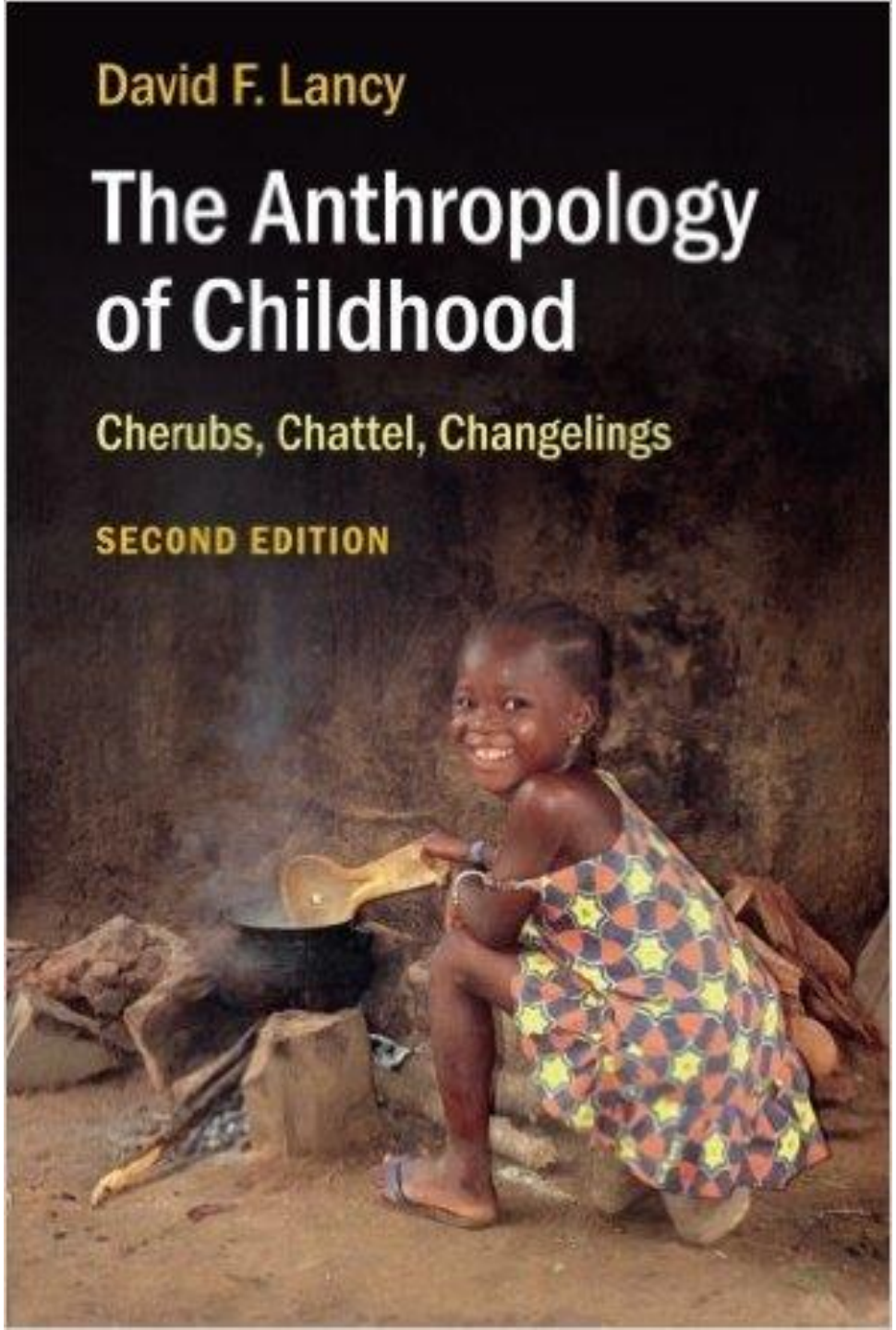
Halpin, P.H., Wolf, S., Yoshikawa, H.Y., Rojas, N., Kabay, S., Pisani, L., & Dowd, A.J. (in press). Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Developmental Psychology.*

# “Respect is an Investment” Measuring Locally Defined Social and Emotional Learning (SEL) of Young Children in Tanzania

Matthew Jukes
RTI Fellow
Sr Education Evaluation Specialist
RTI International

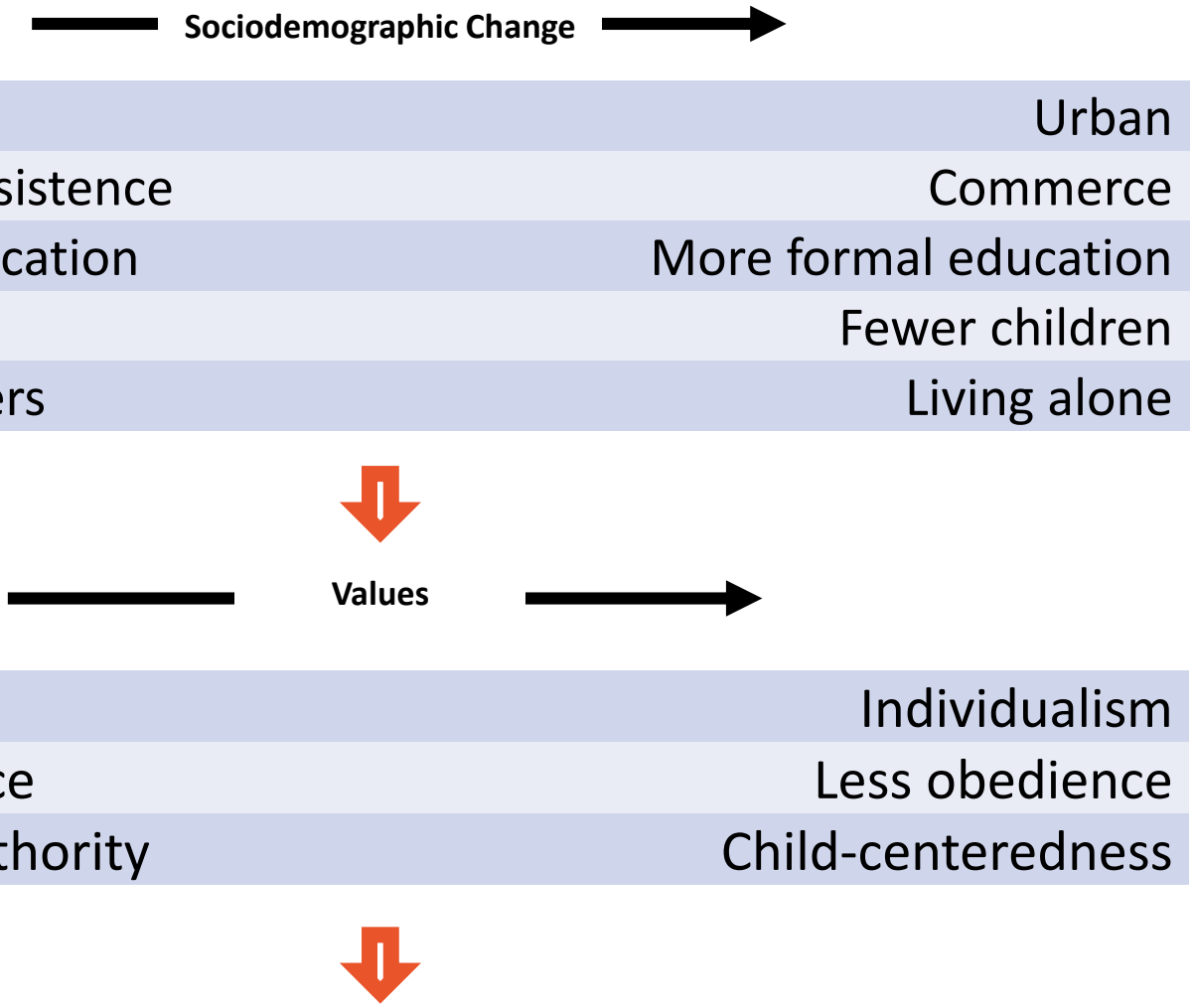
# Core Social and Emotional (SEL) Competencies




David F. Lancy
The Anthropology of Childhood
Cherubs, Chattel, Changelings
SECOND EDITION

# Social Change and Human Development

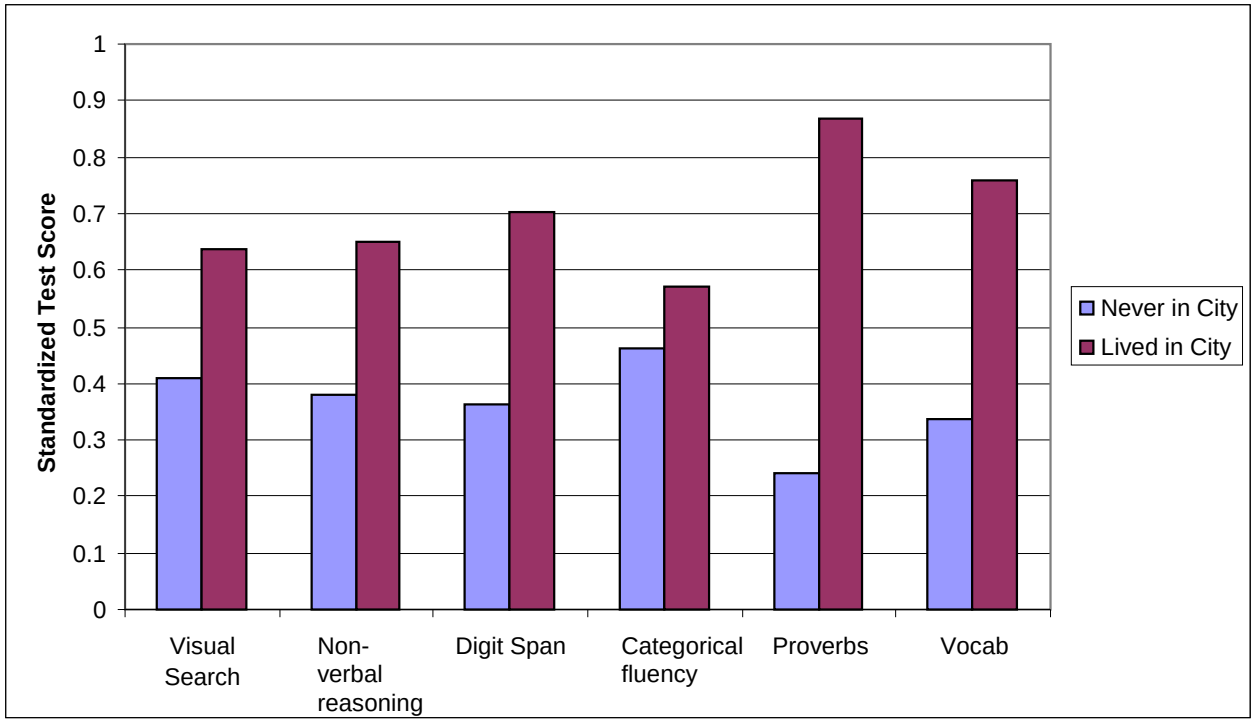
Greenfield, P. M. (2016). Social change, cultural evolution, and human development. Current Opinion in Psychology, 8, 84-92. doi:10.1016/j.copsyc.2015.10.012



**Developmental/Behavioural Change** →

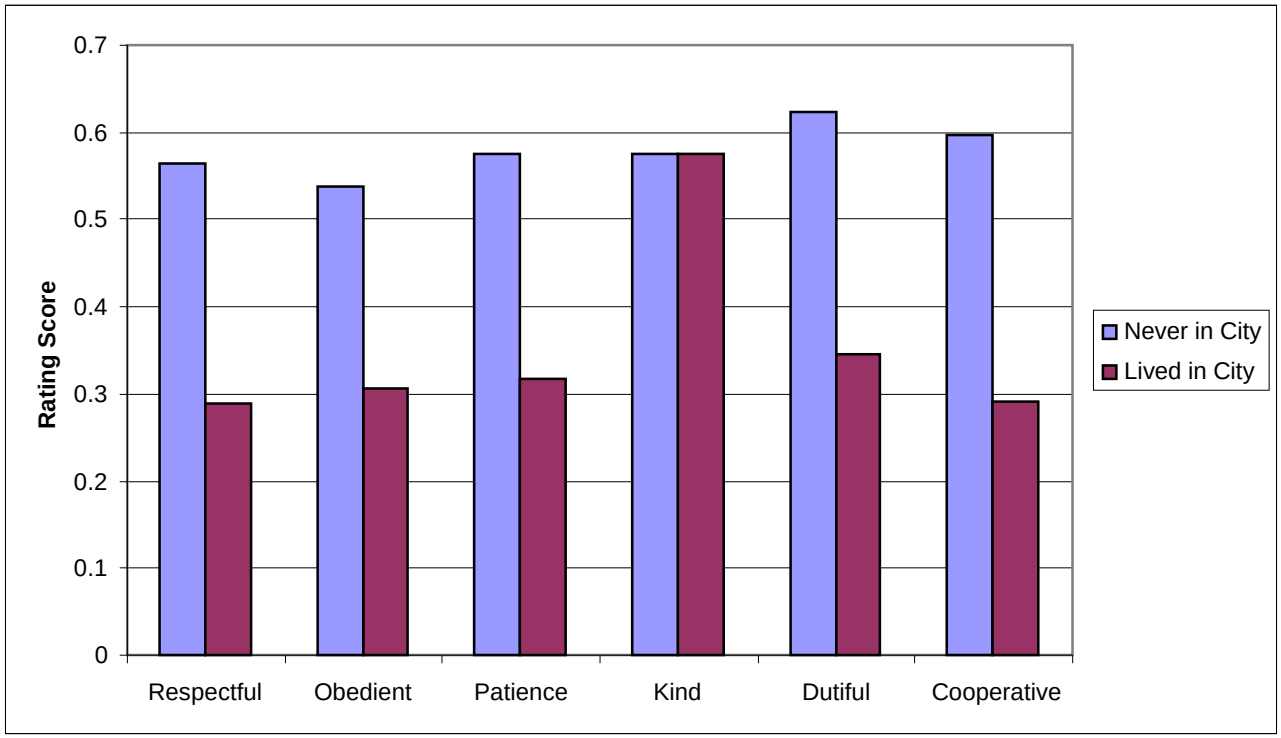
| | |
|---|---|
| Respect, obedience | Expression, curiosity, independence |
| Shyness | Extraversion |
| Gender roles ascribed | Gender roles chosen |
| Focus on others | Focus on self |
| Empathy for others | Internal feeling states |
| Less self-esteem | More self-esteem |
| Fitting in | Standing out, uniqueness |
| Cooperation | Competition |

# Urban Migration and Cognitive Abilities in the Gambia



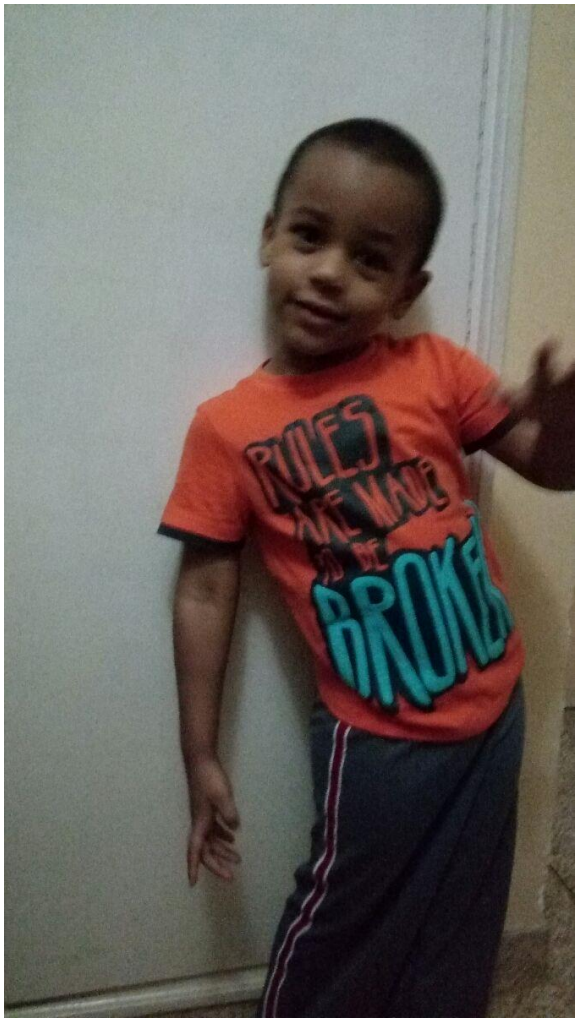

Jukes, Zuilkowski and Grigorenko (2018). Schooling, urban migration and the development of cognitive skills and social responsibility in the Gambia, West Africa. *Journal of Cross Cultural Psychology,* Vol. 49(1) 82–98

# Urban Migration and Cognitive Abilities in the Gambia




Jukes, Zuilkowski and Grigorenko (2018). Schooling, urban migration and the development of cognitive skills and social responsibility in the Gambia, West Africa. *Journal of Cross Cultural Psychology,* Vol. 49(1) 82–98

# “Respect is an investment”

# Locally Generated SEL measures

Two levels of ‘locally generated’:

“I want to measure empathy .. What are examples of empathic behaviour in Mtwara?”

“What should I measure in Mtwara? What competencies are important to people there?”

Betancourt TS, Speelman L, Onyango G, Bolton P. A qualitative study of mental health problems among children displaced by war in northern Uganda. Transcultural Psychiatry. 2009; **46**(2): 238-56.

## Questions

1. **What is the best approach to developing SEL measures in Tanzanian context?** How do you combine developmental science with local perceptions to create a culturally relevant instrument?

2. **Are the SE competencies of children in rural Tanzania characteristic of rural agricultural economies?** Do they have competencies associated with educated urban populations? Are these two sets of competencies at odds?

USAID
FROM THE AMERICAN PEOPLE

# *USAID Tusome Pamoja*

# USAID Tusome Pamoja

Grades 1-4 Reading & Math

Pre-Primary

$68 million

1.4 million children

26,000 teachers

3,025 schools, 5 regions

# Part 1 - Qualitative Study

- 4 schools in 3 districts in Mtwara Region

- 4 Focus groups – 61 male parents
- 5 Focus groups – 34 female parents
- 9 Individual parent interviews (5 female)

- 27 teacher interviews (11 female)

- 80 students in grades 1, 2 and 4
  - Drawings of positive and negative experiences of starting school
  - Individual interviews

## **Questions about child development in general** (for parents and teachers)

- What are the qualities you would like all children to develop?

- What are the characteristics you would want for your child?

    - What are the differences between a good child and a bad child?
    - Describe how you would want your child to behave?
    - What kind of qualities would make a child successful in life?

## Discipline, Attentive Listening

"A child *is* discipline" ("mtoto *ni* nidhamu")

- School 3, Parent FGD 1

"Discipline is the genesis of other children's qualities"

- School 1, Parent FGD 2

"He/she who does not listen to elders will break a leg (i. e., face difficulties)"

("Asiyesikia Mkuu, huvunjika guu")

- School 2, Parent FGD 1; School 3 Teacher 4

"Attentive listening is like a 'safe box' where all other qualities are found"

- School 1, Parent FGD 1

## Questions about qualities for school success
(for parents and teachers)

- What are the qualities that help a child to succeed at school?
- What are the differences between a child who succeeds at school and one who doesn't?
- How do these differences appear on day 1 of school?

## Curiosity and Courage/Confidence

“My son is very inquisitive, one day he asked me: Mom if all people in the world were of the same sex, would the population stop increasing?”

- School 1, Teacher 1

“In our villages curious pupils are very few, most of our people are devoted to religious teachings of Islam, there is not much attention given to such education in our families”

- School 4, Teacher 1

“a pupil who dares to follow teachers in the office or outside the classroom and asks questions is courageous”

-School 3, Teacher 1

CASEL framework

| Social Awareness | Relationship Skills | Self-Management | Self-Awareness |
|---|---|---|---|
| Perspective-Taking<br>Empathy<br>Appreciating Diversity<br>Respect for Others | Communication<br>Social Engagement<br>Relationship Building<br>Teamwork | Impulse control<br>Stress management<br>Self-discipline<br>Self-motivation<br>Goal Setting<br>Organizational skills | Identifying emotions<br>Accurate self-perception<br>Recognizing strengths<br>Self-confidence<br>Self-efficacy |

Findings from Mtwara, Tanzania

**Bold text** = most frequently mentioned

| Social Responsibility | Social Relationships | Individual Competencies | |
|---|---|---|---|
| **Obedient**<br>**Respectful**<br>**Attentive**<br>**Disciplined**<br>**Polite and calm**<br><br>Clean<br>Religious | Trustworthy<br>Truthful<br><br>**Cooperative**<br>**Sociable**<br>Funny<br>Loving others<br>(Empathic) | **Self-directed**<br>Careful<br>Persistent<br>To have goals | **Self-belief**<br>**Curious**<br>Courageous/Daring<br>Self-awareness |

Valued by parents and teachers
Important for life in general

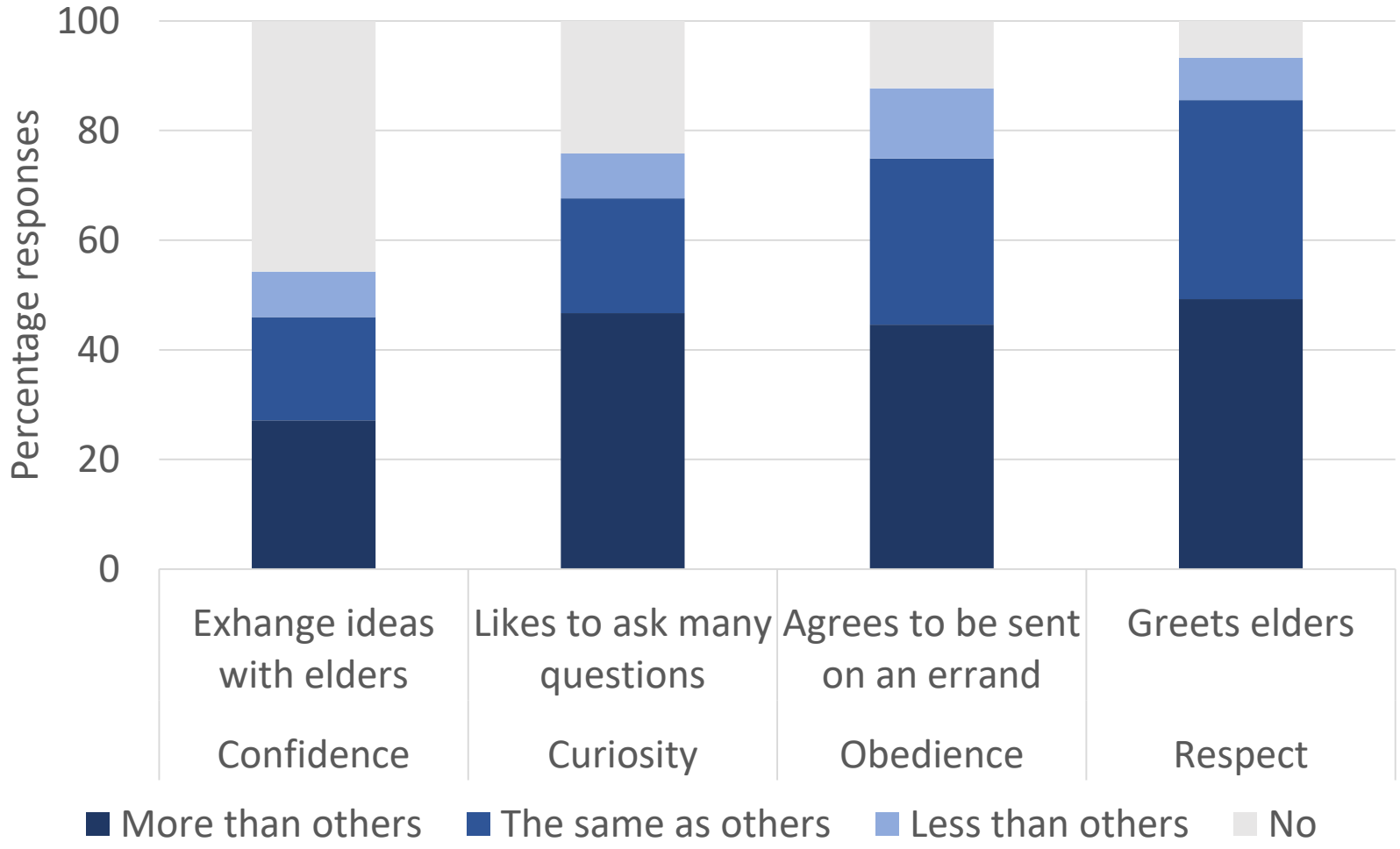
Valued by teachers only
Important for schooling

Jukes et al (2018). *Global Education Review*

## Part 2 - Psychometric Assessment

- 23 schools
- 478 students in preschool, grade 1 and grade 2
- Rated by 1 parent and 1 teacher
- 72 parent questions and 42 teacher questions

# Responses on Example Questions



## Obedience

| |
|---|
| Does child easily agree to be sent? |
| Does child willingly follow instructions? |
| When child is directed to complete a task, does s/he do with with heart and complete it successfully? |
| When child is sent to complete an errand, does s/he return on time? |
| If child is told to do a chore, does s/he do it? |
| Does child refuse to stop and continue doing something after being told not to |

Hamisa agrees easily to be sent
Hamisa willingly follows instructions

Juma agrees easily to be sent
Juma willingly follows instructions

## Obedience

- Does child easily agree to be sent?
- Does child willingly follow instructions?
- When child is directed to complete a task, does s/he do with with heart and complete it successfully?
- When child is sent to complete an errand, does s/he return on time?
- If child is told to do a chore, does s/he do it?
- Does child refuse to stop and continue doing something after being told not to

## Respect

- Does child offer to assist/receive elders?
- Does child kindly greet elders?
- Does child get your permission before doing something or going somewhere?
- If child makes a mistake, does s/he ask for forgiveness without being told to?

## Final Obedience Measure

- Does child easily agree to be sent?
- Does child willingly follow instructions?
- When child is directed to complete a task, does s/he do with with heart and complete it successfully?
- When child is sent to complete an errand, does s/he return on time?
- If child is told to do a chore, does s/he do it?
- Does child kindly greet elders?

## Persistence

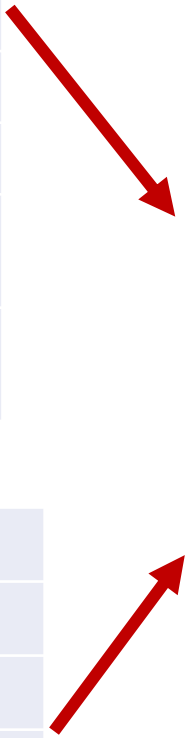
| |
|---|
| Does child easily become frustrated or angry? |
| Does child quit working on tasks before s/he is finished? |
| Does child give up easily when tasks or work seem difficult? |
| If child cannot do something, do they try again? |
| If child has chores to do, does s/he like to finish them in one go? |
| Does child continue with a task at home/school even when it is tiring |

## Persistence

- Does child easily become frustrated or angry?
- Does child quit working on tasks before s/he is finished?
- Does child give up easily when tasks or work seem difficult?
- If child cannot do something, do they try again?
- If child has chores to do, does s/he like to finish them in one go?
- Does child continue with a task at home/school even when it is tiring

## Polite and Calm

- Is child calm even when disturbed/irritated by others?
- Does child respond nicely/politely when asked a question?
- Does child speak in a soft, measured way?
- Does child react angrily when s/he doesn't get what she wants/ told to stop doing something?
- Does child have strong emotions?
- Is child calm even when distrubed/irritated by others?



## Emotional Control

- Does child easily become frustrated or angry?
- Does child quit working on tasks before s/he is finished?
- Does child give up easily when tasks or work seem difficult?
- Does child react angrily when s/he doesn't get what she wants/ told to stop doing something?
- Does child have strong emotions?

# Exploratory Factor Analysis of Parents Items

| Factor | Original targeted constructs | Variance Explained | Cumulative Variance |
|---|---|---|---|
| **Obedient** | Obedient | 29% | 29% |
| **Curious** | Curious, confident | 8% | 37% |
| **Conscientious** | Persistent, careful, self-directed | 6% | 43% |
| **Emotional Regulation** | Polite, persistent, obedient | 5% | 48% |
| **Sociable** | Sociable, cooperative | 4% | 52% |
| **Polite** | Polite, cooperative | 3% | 55% |
| **Religious** | Religious | 3% | 58% |

# Exploratory Factor Analysis of Parents Items

| Factor | Original targeted constructs | Variance Explained | Cumulative Variance |
|---|---|---|---|
| **Obedient** | Obedient | 29% | 29% |
| **Curious** | Curious, confident | 8% | 37% |
| **Conscientious** | Persistent, careful, self-directed | 6% | 43% |
| **Emotional Regulation** | Polite, persistent, obedient | 5% | 48% |
| **Sociable** | Sociable, cooperative | 4% | 52% |
| **Polite** | Polite, cooperative | 3% | 55% |
| **Religious** | Religious | 3% | 58% |

# What is the best approach to developing SEL measures in Tanzanian context?

1. Start with a theory

2. Some domains are locally generated but not in current frameworks

    - Social responsibility

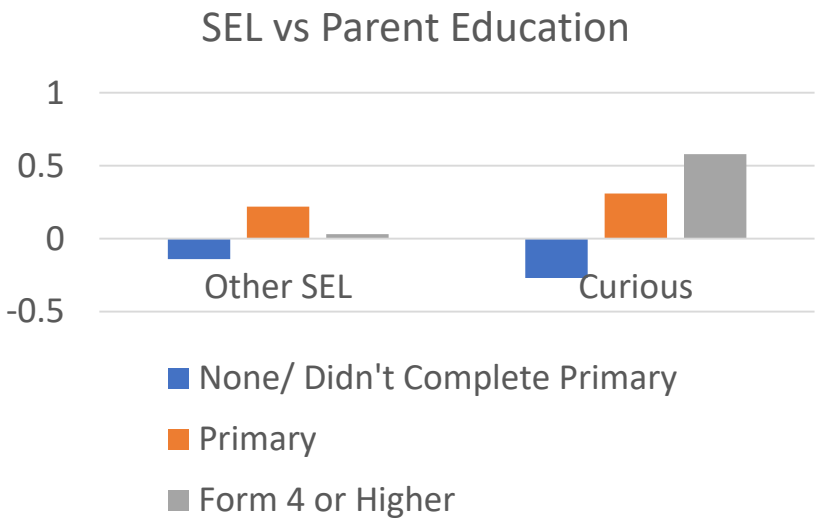
3. Some domains are in current frameworks but not locally generated

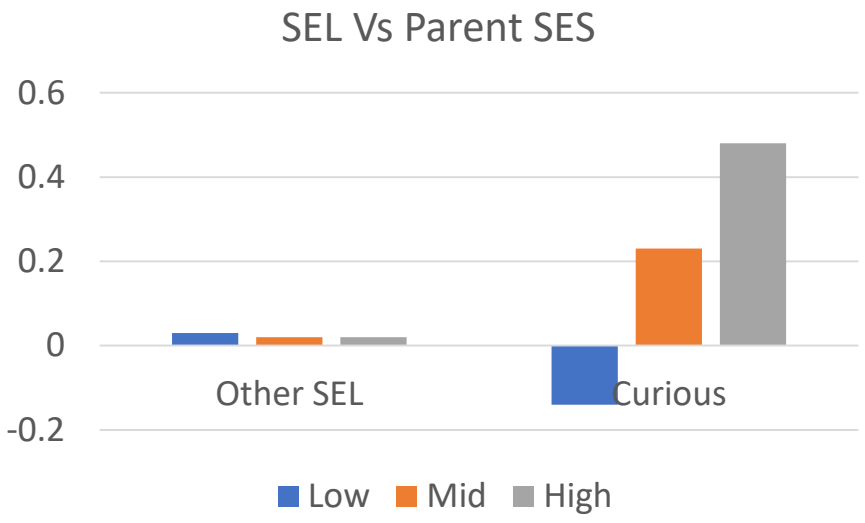
    - Controlling emotions, conscientiousness

4. Some domains in current frameworks are combined/configured differently in local perceptions

    - Curiosity and courage/confidence

# Curiosity prevalent among educated and wealthy



## Are the SE competencies of children in rural Tanzania characteristic of rural agricultural economies?

SE competencies are consistent with those of other subsistence agricultural communities

In line with theory, curiosity and confidence more common among urban children of educated parents

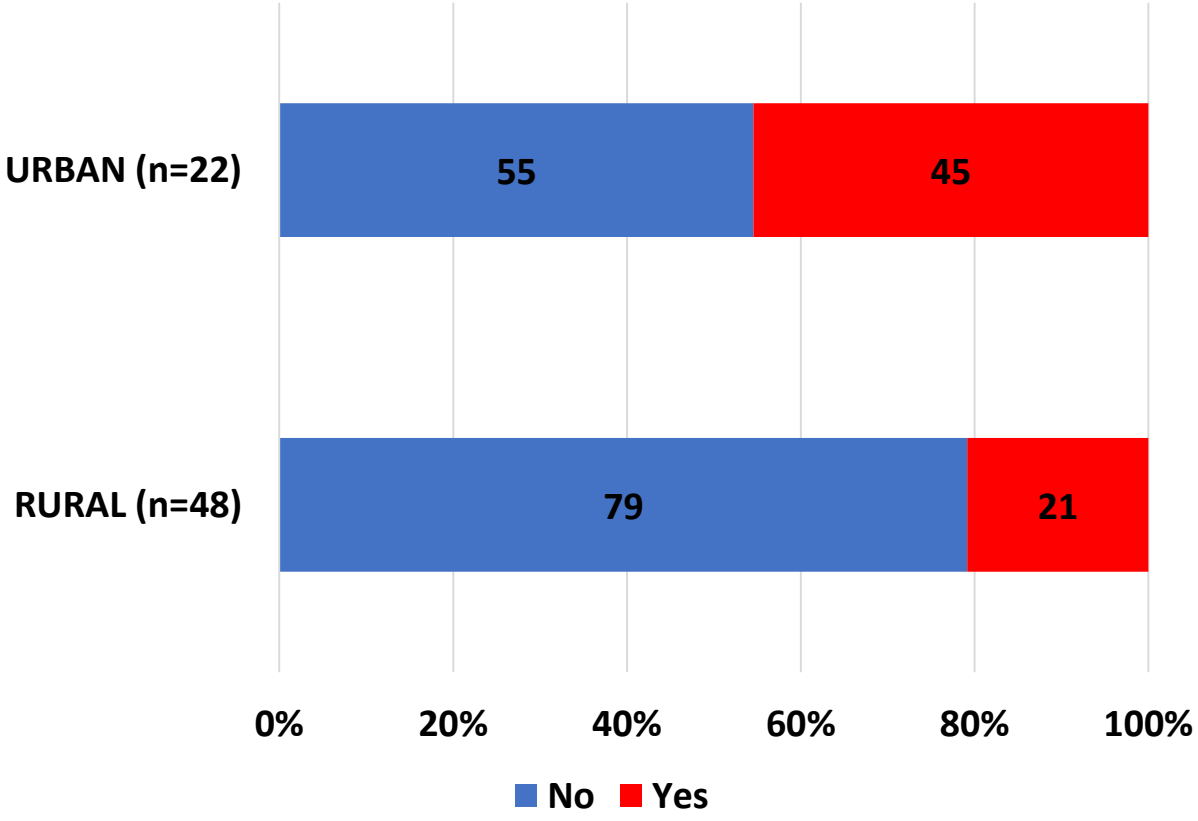
Do students with curiosity/confidence do better in school? Is pedagogy designed assuming students are curious and confident?

# Culture and Teaching Activities

Ongoing study of pupil SEL and teaching activities involving participation:

- I do / we do / you do
- Checking for understanding
- Group work

# More urban classrooms have children who ask questions or express their ideas (MELQO 2017)




Pearson $Chi^2 = 4.48$, $p = .03$

# Goals for Schooling

- Julian Huxley, first director of UNESCO (1932). ‘The dual mandate of education’

> ‘education should be adapted to the local environment of time and place, and yet give the opportunity of transcending that environment’

# Research Partners

| | |
|---|---|
| Jovina Tibenda | Tusome Pamoja |
| Dr. Nkanileka Mgonda | University of Dar es Salaam |
| Dr. Prosper Mosha | University of Dodoma |
| Florentina Nsolezi | University of Dodoma |
| Grace Jeremiah | St. Augustine University |
| Kellie Betts | RTI |
| Prof. Kristen Bub | University of Illinois, USA |
| Sarrynna Sou, Corina Owens | RTI |

Funding: USAID Tusome Pamoja, RTI ECD Strategic investment fund

## Thank you!

Please contact me - mjukes@rti.org

@matthewchjukes on twitter

Thanks to:

PORALG and MOEST

Mtwara Regional Education Office

District Education Offices of Mtwara Rural, Tandahimba and Nanyamba

Parents, students and teachers of four schools.

# Social Emotional Learning Assessment

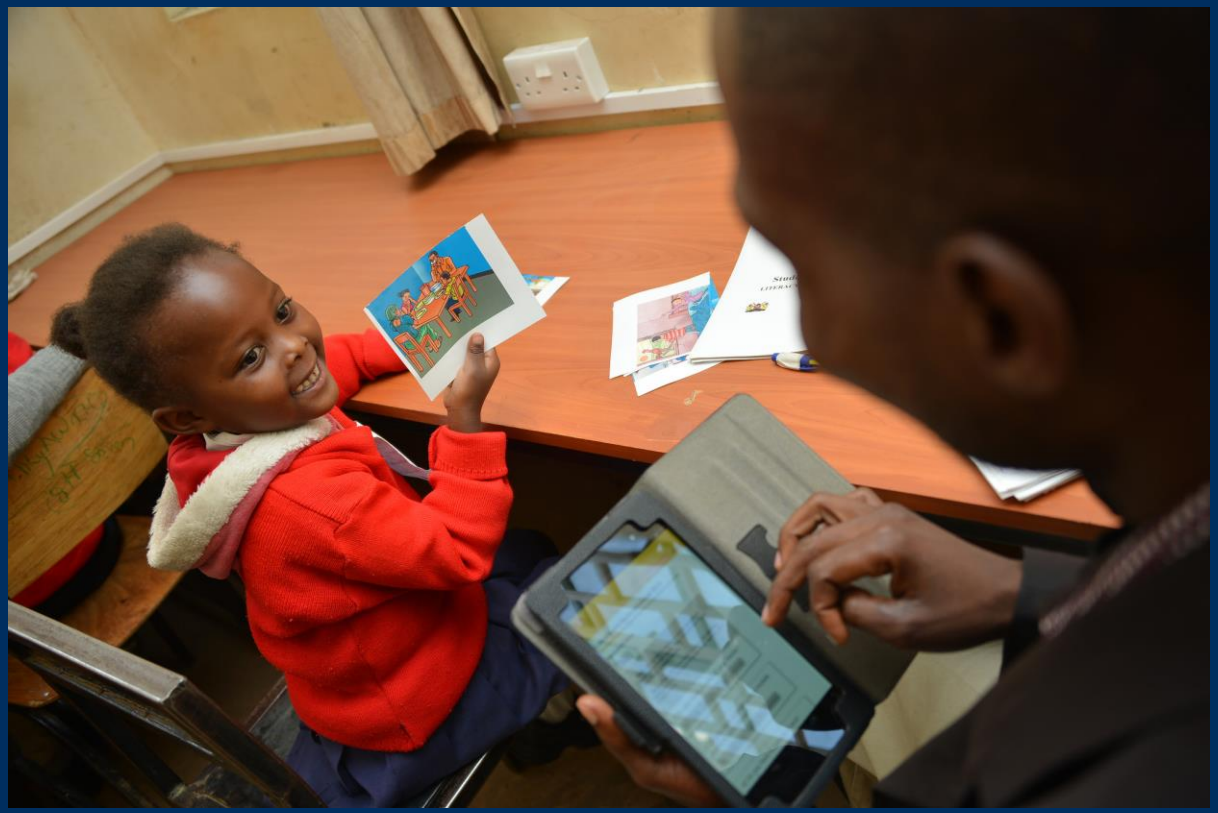
**Matthew Jukes**
Fellow and Senior Education Evaluation Specialist
*RTI International*

**Sharon Wolf**
Applied Developmental Psychologist and Assistant Professor
*University of Pennsylvania Graduate School of Education*

# Cross-Cutting Themes

- Cultural Transport of Assessments
  - Content
    - Are we aiming to measure the same domains in a new setting?
    - Do those domains manifest themselves differently in a new setting?
  - Methods
    - Can we assess domains with the same methods?
- Agenda for Research and Practice
  - What are the knowns, unknowns and pressing issues in international assessments in each domain?

# Innovations in Early Childhood Development Assessment



Wednesday, October 10, 2018
1:00 p.m. – 5:00 p.m.

@RTI_INTL_DEV | #RTILearns | #EarlyYears

RTI INTERNATIONAL