Innovations in Early Childhood Development Assessment

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

@RTI_INTL_DEV | #RTILearns | #EarlyYears
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
BACKGROUND
APPROACHES TO MEASURING ECD

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

Precise
- Resource intensive
- Culturally specific

Programmatic Assessments
- Used in research to measure impacts of programs or policies

Imprecise
- Quick/cheap
- Cross-culturally valid

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

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

Used in research to measure impacts of programs or policies.

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

<table>
<thead>
<tr>
<th>Age</th>
<th>Assessment</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5 - 7</td>
<td>EDI</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>4 - 6</td>
<td>MELQO</td>
<td>Direct Assessment + Teacher/Parent Report</td>
</tr>
<tr>
<td>3 - 4</td>
<td>ECDI (UNICEF)</td>
<td>Parent Report</td>
</tr>
<tr>
<td>2 - 4</td>
<td>PRIDI (Inter-American Development Bank)</td>
<td>Direct Assessment + Parent Report</td>
</tr>
<tr>
<td>0 - &lt;3</td>
<td>CREDI</td>
<td>Parent Report</td>
</tr>
<tr>
<td>0 - &lt;3</td>
<td>IYCD (WHO)</td>
<td>Parent Report</td>
</tr>
<tr>
<td>0 - &lt;3</td>
<td>GCDG</td>
<td>Direct Assessment + Parent Report</td>
</tr>
</tbody>
</table>
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
## Versions of the CREDI

<table>
<thead>
<tr>
<th></th>
<th><strong>Short Form</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Population-level monitoring</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>20 items</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>6-mo age brackets</td>
</tr>
<tr>
<td><strong>Administration time</strong></td>
<td>~5 mins</td>
</tr>
<tr>
<td><strong>Domain inclusion</strong></td>
<td>One score representing all domains (&quot;overall ECD&quot;)</td>
</tr>
</tbody>
</table>

- Both versions allow for oral or written administration
CREDI DOMAINS

Motor
- Fine motor
- Gross motor

Language
- Receptive language
- Expressive language

Cognition
- Executive function
- Problem solving & reasoning
- Pre-academic knowledge

Social-Emotional
- Emotional & behavioral self-reg
- Emotion knowledge
- Social competence

Mental Health
- Internalizing
- Externalizing

Milestones / Skills

Behaviors / Attributes
<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Don't Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td>Can the child climb onto an object such as a chair or bench?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Motor</td>
<td>Can the child pick up a small object (e.g., a small toy or small stone) with just his/her thumb and a finger?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Language</td>
<td>Can the child say one or more words (e.g., names like Mama or “ba” for “ball”)?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Language</td>
<td>Can the child follow orders or instructions that have more than one part (e.g., &quot;Go get water and go to bed&quot;).</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Can the child figure out how to turn a spoon or object if you give it to him/her the wrong way around?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Can the child count up to five objects (e.g., fingers, people)?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Social-Emotional</td>
<td>When the child is upset, does he/she calm down quickly on his/her own?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Social-Emotional</td>
<td>Does the child have difficulty getting along with other children?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Social-Emotional</td>
<td>Is the child often impatient or unwilling to wait when you ask him/her to?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Mental Health</td>
<td>Is the child frequently sad, worried, or anxious?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Mental Health</td>
<td>Does the child often kick, bite, or hit other children or adults?</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>
**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

\[ N = 149 \text{ items, 21 sites, 17 countries, 14 languages, 16,029 caregivers} \]
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)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASQ:SE</td>
<td>Chile</td>
</tr>
<tr>
<td>BSID cognition</td>
<td>Pakistan</td>
</tr>
<tr>
<td>BSID cognition</td>
<td>Tanzania (Ifakara)</td>
</tr>
<tr>
<td>INTERNDA</td>
<td>Zambia (Chipata)</td>
</tr>
<tr>
<td>MacArthur Bates CDI</td>
<td>Chile</td>
</tr>
<tr>
<td>PRIDI</td>
<td>Brazil (Sao Paulo)</td>
</tr>
</tbody>
</table>
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

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

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

Maureen Black, PhD

Global Child Development Group
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

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

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

Outfit 0.64(−10.26)
Infit 0.74(−18.19)

ACTIVE

% pass

D-score (607_15_final)

study
- Bangladesh
- Chile 1
- Columbia 2
- Jamaica 1
- Jamaica 2
- Netherlands 1
- Netherlands 2
- South Africa

sa2v16: use a short sentence
n47: Says sentences with 2 words
gh2_17: Uses word combinations
dl19: Combine Words
b1m136: sentence of 2 words
b2m114: Uses a two-word utterance
30: EXP26
Equate group – variable item difficulty

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n43</td>
<td>Throws ball without falling down</td>
</tr>
<tr>
<td></td>
<td>(Eye and hand) Can throw a ball</td>
</tr>
<tr>
<td>ge2_14</td>
<td></td>
</tr>
<tr>
<td>dg23</td>
<td>Throw Ball Overhand</td>
</tr>
<tr>
<td>b3g44</td>
<td>Throws ball</td>
</tr>
<tr>
<td>b2p64</td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>GM44</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Age range (months)</th>
<th>Bayley--I, II, III&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Other Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cognition</td>
<td>Language</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MDI</td>
<td>PDI</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>18</td>
<td>0.797</td>
<td>0.503</td>
</tr>
<tr>
<td>Brazil 1</td>
<td>5-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11-19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile 1</td>
<td>6</td>
<td>0.861</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.880</td>
<td>0.361</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.835</td>
<td>0.249</td>
</tr>
<tr>
<td>Chile 2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>24-35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36-47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>18</td>
<td>0.541</td>
<td>-</td>
</tr>
<tr>
<td>Colombia 1</td>
<td>10-26</td>
<td>0.710</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td>28-45</td>
<td>0.742</td>
<td>0.840</td>
</tr>
<tr>
<td>Colombia 2</td>
<td>6-17</td>
<td>0.386</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>18-29</td>
<td>0.671</td>
<td>0.837</td>
</tr>
<tr>
<td></td>
<td>30-42</td>
<td>0.649</td>
<td>0.811</td>
</tr>
</tbody>
</table>

Weber et al., under review
**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

Measurement of Early Child Development

Population

Program Evaluation

Individual Evaluations
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

Overall Spearman’s Correlation: 0.968
N=12,271

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
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
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
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?
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 not fit in any category.
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

$pineapple\_rating\_wt$

$cat\_rating\_wt$
Results – Gender

The histogram shows the distribution of overall scores for male and female participants. The x-axis represents the overall score, ranging from 100 to 250, while the y-axis represents the frequency. The data indicates a higher concentration of scores for males in the 200-250 range compared to females, who have a more dispersed distribution with a peak around 150-200.

Overall Score

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Frequency</td>
</tr>
<tr>
<td>100-150</td>
<td>150-200</td>
</tr>
<tr>
<td>200-250</td>
<td>200-250</td>
</tr>
<tr>
<td>250-300</td>
<td>250-300</td>
</tr>
</tbody>
</table>
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.


Thank you! Asante! Muvea!

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

When you say 'I can' and expect success, you fill yourself with confidence and happiness.
Developing Early Literacy Assessments in Spanish

Karen Ford
October 10, 2018
Innovations in ECD Assessment, RTI
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
  
  \[ n = 2,000 \text{ in K-3} \]
- Exploring funding sources
IES grant to develop K and 1-3

• Item development, piloting, parallel forms
  
  \(4x/2\) yrs./\(n = 400\)

• Field testing
  
  \(4x/2\) yrs./\(n = 5,000\)
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phonological Awareness</td>
</tr>
<tr>
<td>Rhyme Awareness</td>
<td>✓</td>
</tr>
<tr>
<td>Beginning Sound Awareness</td>
<td>✓</td>
</tr>
<tr>
<td>Alphabet Recognition</td>
<td></td>
</tr>
<tr>
<td>Letter Sound Knowledge</td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>✓</td>
</tr>
<tr>
<td>Concept of Word</td>
<td>✓</td>
</tr>
<tr>
<td>Tasks</td>
<td>Domains</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td>Orthographic Knowledge</td>
</tr>
<tr>
<td>Word Recognition in Isolation</td>
<td>✓</td>
</tr>
<tr>
<td>Spelling</td>
<td>✓</td>
</tr>
<tr>
<td>Oral Reading in Context</td>
<td>✓</td>
</tr>
<tr>
<td>Alphabet Recognition</td>
<td></td>
</tr>
<tr>
<td>Letter Sound Knowledge</td>
<td></td>
</tr>
<tr>
<td>Concept of Word in Text</td>
<td>✓</td>
</tr>
<tr>
<td>Blending</td>
<td></td>
</tr>
<tr>
<td>Segmenting</td>
<td></td>
</tr>
</tbody>
</table>
PALS español: 2013 - 2018

IES grant to develop PreK

- Item development, piloting
  \( (2x/1\text{ yr.}/n = 677) \)
- Field testing
  \( (5x/3\text{ yrs.}/n = 3,900) \)
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phonological Awareness</td>
</tr>
<tr>
<td>Syllable Clapping</td>
<td>✓</td>
</tr>
<tr>
<td>Rhyme Awareness</td>
<td>✓</td>
</tr>
<tr>
<td>Beginning Sound Awareness</td>
<td>✓</td>
</tr>
<tr>
<td>Alphabet Recognition</td>
<td>✓</td>
</tr>
<tr>
<td>Letter Sound Knowledge</td>
<td>✓</td>
</tr>
<tr>
<td>Print and Word Awareness</td>
<td></td>
</tr>
<tr>
<td>Name Writing</td>
<td>✓</td>
</tr>
<tr>
<td>Language and Listening Comprehension</td>
<td></td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Summed Score</th>
<th>Scaled Score</th>
<th>Percentile Rank</th>
<th>Summed Score</th>
<th>Scaled Score</th>
<th>Percentile Rank</th>
<th>Summed Score</th>
<th>Scaled Score</th>
<th>Percentile Rank</th>
<th>Summed Score</th>
<th>Scaled Score</th>
<th>Percentile Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>0.0</td>
<td>38</td>
<td>254</td>
<td>30.1</td>
<td>76</td>
<td>360</td>
<td>63.8</td>
<td>114</td>
<td>452</td>
<td>84.9</td>
</tr>
<tr>
<td>1</td>
<td>102</td>
<td>0.1</td>
<td>39</td>
<td>255</td>
<td>31.0</td>
<td>77</td>
<td>363</td>
<td>64.4</td>
<td>115</td>
<td>455</td>
<td>85.2</td>
</tr>
<tr>
<td>2</td>
<td>105</td>
<td>0.2</td>
<td>40</td>
<td>259</td>
<td>32.4</td>
<td>78</td>
<td>365</td>
<td>64.9</td>
<td>116</td>
<td>458</td>
<td>85.7</td>
</tr>
<tr>
<td>3</td>
<td>108</td>
<td>0.3</td>
<td>41</td>
<td>262</td>
<td>34.0</td>
<td>79</td>
<td>368</td>
<td>66.6</td>
<td>117</td>
<td>461</td>
<td>86.1</td>
</tr>
<tr>
<td>4</td>
<td>111</td>
<td>0.4</td>
<td>42</td>
<td>264</td>
<td>35.0</td>
<td>80</td>
<td>360</td>
<td>66.0</td>
<td>118</td>
<td>465</td>
<td>86.7</td>
</tr>
<tr>
<td>5</td>
<td>116</td>
<td>0.6</td>
<td>43</td>
<td>267</td>
<td>36.6</td>
<td>81</td>
<td>363</td>
<td>66.7</td>
<td>119</td>
<td>468</td>
<td>87.2</td>
</tr>
<tr>
<td>6</td>
<td>121</td>
<td>0.9</td>
<td>44</td>
<td>269</td>
<td>37.5</td>
<td>82</td>
<td>365</td>
<td>67.1</td>
<td>120</td>
<td>471</td>
<td>87.5</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
<td>1.2</td>
<td>45</td>
<td>272</td>
<td>38.9</td>
<td>83</td>
<td>368</td>
<td>67.9</td>
<td>121</td>
<td>475</td>
<td>88.2</td>
</tr>
<tr>
<td>8</td>
<td>133</td>
<td>1.5</td>
<td>46</td>
<td>275</td>
<td>40.3</td>
<td>84</td>
<td>370</td>
<td>68.5</td>
<td>122</td>
<td>478</td>
<td>88.5</td>
</tr>
<tr>
<td>9</td>
<td>139</td>
<td>1.6</td>
<td>47</td>
<td>277</td>
<td>40.9</td>
<td>85</td>
<td>372</td>
<td>69.3</td>
<td>123</td>
<td>482</td>
<td>88.9</td>
</tr>
<tr>
<td>10</td>
<td>146</td>
<td>1.9</td>
<td>48</td>
<td>280</td>
<td>41.9</td>
<td>86</td>
<td>375</td>
<td>69.8</td>
<td>124</td>
<td>485</td>
<td>89.2</td>
</tr>
<tr>
<td>11</td>
<td>151</td>
<td>2.2</td>
<td>49</td>
<td>282</td>
<td>42.8</td>
<td>87</td>
<td>378</td>
<td>70.5</td>
<td>125</td>
<td>488</td>
<td>89.6</td>
</tr>
<tr>
<td>12</td>
<td>157</td>
<td>2.8</td>
<td>50</td>
<td>285</td>
<td>44.4</td>
<td>88</td>
<td>380</td>
<td>71.0</td>
<td>126</td>
<td>493</td>
<td>90.1</td>
</tr>
<tr>
<td>13</td>
<td>162</td>
<td>3.3</td>
<td>51</td>
<td>287</td>
<td>45.0</td>
<td>89</td>
<td>383</td>
<td>71.6</td>
<td>127</td>
<td>497</td>
<td>90.6</td>
</tr>
<tr>
<td>14</td>
<td>168</td>
<td>3.8</td>
<td>52</td>
<td>290</td>
<td>46.1</td>
<td>90</td>
<td>386</td>
<td>72.0</td>
<td>128</td>
<td>501</td>
<td>90.9</td>
</tr>
<tr>
<td>15</td>
<td>172</td>
<td>4.4</td>
<td>53</td>
<td>292</td>
<td>47.0</td>
<td>91</td>
<td>388</td>
<td>72.5</td>
<td>129</td>
<td>505</td>
<td>91.2</td>
</tr>
<tr>
<td>16</td>
<td>177</td>
<td>5.2</td>
<td>54</td>
<td>295</td>
<td>48.2</td>
<td>92</td>
<td>391</td>
<td>73.2</td>
<td>130</td>
<td>509</td>
<td>91.4</td>
</tr>
<tr>
<td>17</td>
<td>182</td>
<td>6.1</td>
<td>55</td>
<td>297</td>
<td>48.9</td>
<td>93</td>
<td>393</td>
<td>73.6</td>
<td>131</td>
<td>514</td>
<td>91.8</td>
</tr>
<tr>
<td>18</td>
<td>186</td>
<td>6.7</td>
<td>56</td>
<td>300</td>
<td>50.2</td>
<td>94</td>
<td>396</td>
<td>74.2</td>
<td>132</td>
<td>518</td>
<td>92.1</td>
</tr>
<tr>
<td>19</td>
<td>190</td>
<td>7.6</td>
<td>57</td>
<td>302</td>
<td>50.9</td>
<td>95</td>
<td>399</td>
<td>74.7</td>
<td>133</td>
<td>523</td>
<td>92.4</td>
</tr>
<tr>
<td>20</td>
<td>194</td>
<td>8.8</td>
<td>58</td>
<td>305</td>
<td>51.9</td>
<td>96</td>
<td>401</td>
<td>75.2</td>
<td>134</td>
<td>528</td>
<td>92.8</td>
</tr>
<tr>
<td>21</td>
<td>198</td>
<td>9.6</td>
<td>59</td>
<td>307</td>
<td>52.6</td>
<td>97</td>
<td>404</td>
<td>75.7</td>
<td>135</td>
<td>533</td>
<td>93.1</td>
</tr>
<tr>
<td>22</td>
<td>202</td>
<td>10.5</td>
<td>60</td>
<td>310</td>
<td>53.5</td>
<td>98</td>
<td>406</td>
<td>76.2</td>
<td>136</td>
<td>539</td>
<td>93.5</td>
</tr>
<tr>
<td>23</td>
<td>206</td>
<td>11.5</td>
<td>61</td>
<td>312</td>
<td>54.0</td>
<td>99</td>
<td>409</td>
<td>76.9</td>
<td>137</td>
<td>544</td>
<td>93.8</td>
</tr>
<tr>
<td>24</td>
<td>210</td>
<td>12.3</td>
<td>62</td>
<td>315</td>
<td>54.8</td>
<td>100</td>
<td>412</td>
<td>77.7</td>
<td>138</td>
<td>550</td>
<td>94.2</td>
</tr>
<tr>
<td>25</td>
<td>213</td>
<td>13.4</td>
<td>63</td>
<td>317</td>
<td>55.5</td>
<td>101</td>
<td>414</td>
<td>78.2</td>
<td>139</td>
<td>556</td>
<td>94.6</td>
</tr>
<tr>
<td>26</td>
<td>217</td>
<td>14.8</td>
<td>64</td>
<td>320</td>
<td>56.5</td>
<td>102</td>
<td>417</td>
<td>78.8</td>
<td>140</td>
<td>563</td>
<td>95.2</td>
</tr>
<tr>
<td>27</td>
<td>220</td>
<td>15.9</td>
<td>65</td>
<td>322</td>
<td>57.0</td>
<td>103</td>
<td>420</td>
<td>79.3</td>
<td>141</td>
<td>569</td>
<td>95.5</td>
</tr>
<tr>
<td>28</td>
<td>224</td>
<td>17.4</td>
<td>66</td>
<td>325</td>
<td>57.7</td>
<td>104</td>
<td>423</td>
<td>79.9</td>
<td>142</td>
<td>575</td>
<td>95.9</td>
</tr>
<tr>
<td>29</td>
<td>227</td>
<td>18.2</td>
<td>67</td>
<td>327</td>
<td>58.1</td>
<td>105</td>
<td>425</td>
<td>80.3</td>
<td>143</td>
<td>580</td>
<td>96.4</td>
</tr>
<tr>
<td>30</td>
<td>230</td>
<td>19.1</td>
<td>68</td>
<td>330</td>
<td>58.8</td>
<td>106</td>
<td>428</td>
<td>81.1</td>
<td>144</td>
<td>585</td>
<td>96.8</td>
</tr>
<tr>
<td>31</td>
<td>233</td>
<td>20.3</td>
<td>69</td>
<td>332</td>
<td>59.3</td>
<td>107</td>
<td>431</td>
<td>81.7</td>
<td>145</td>
<td>590</td>
<td>97.0</td>
</tr>
<tr>
<td>32</td>
<td>236</td>
<td>21.9</td>
<td>70</td>
<td>335</td>
<td>60.0</td>
<td>108</td>
<td>434</td>
<td>82.2</td>
<td>146</td>
<td>592</td>
<td>97.7</td>
</tr>
<tr>
<td>33</td>
<td>239</td>
<td>23.0</td>
<td>71</td>
<td>337</td>
<td>60.6</td>
<td>109</td>
<td>437</td>
<td>82.6</td>
<td>147</td>
<td>596</td>
<td>98.3</td>
</tr>
<tr>
<td>34</td>
<td>242</td>
<td>24.5</td>
<td>72</td>
<td>340</td>
<td>61.4</td>
<td>110</td>
<td>440</td>
<td>83.0</td>
<td>148</td>
<td>597</td>
<td>98.7</td>
</tr>
<tr>
<td>35</td>
<td>245</td>
<td>25.8</td>
<td>73</td>
<td>342</td>
<td>61.9</td>
<td>111</td>
<td>443</td>
<td>83.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>248</td>
<td>27.2</td>
<td>74</td>
<td>345</td>
<td>62.7</td>
<td>112</td>
<td>446</td>
<td>84.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>251</td>
<td>28.7</td>
<td>75</td>
<td>348</td>
<td>63.2</td>
<td>113</td>
<td>449</td>
<td>84.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.
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
SELF-REGULATION
attention, behavior, emotions

1. SELECT & PURSUE GOALS
2. PLAN & ORGANIZE BEHAVIOR
3. MONITOR & EVALUATE PROGRESS
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

- No passcode
- Voice instruction

Current tasks
- Driving Game
- Memory Game
- Hearts & Flowers
- Color Dot

Customizable task settings

<table>
<thead>
<tr>
<th>Task Settings</th>
<th>1: Hearts</th>
<th>2: Flowers</th>
<th>3: Mixed</th>
<th>4: Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td># practice trials</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td># correct practice trials</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td># test trials</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Stimulus present time(ms)</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Inter-stimulus interval(ms)</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>
HOW DO WE PROMOTE EXECUTIVE FUNCTIONS?
INFANCY

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

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

• **Games, puzzles**: turn taking, strategy, taking other person’s perspective, collaborating

• **Physical activities, dancing, cooking, playing music**: 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 childcare, 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
ECD intervention effect remained significant after controlling for physical growth, age 2 & age 4 cognitive skills, and maternal cognitive skills.

Obradović et al., 2016
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
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
That was a woof sound, so what animal should you touch in the Silly Sounds Game? [If correct, praise child for selecting the cat. If incorrect, provide verbal correction before clicking next]
Pick the Picture

20/52
Let's start over. Here are some new pictures. Please touch one of them.
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**
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.

<table>
<thead>
<tr>
<th>Emergent Literacy</th>
<th>Emergent Numeracy</th>
<th>Social and Emotional</th>
<th>Gross and Fine Motor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Emergent Writing</td>
<td>Writing</td>
<td>1. Conflict Resolution</td>
<td>Conflict</td>
</tr>
<tr>
<td>2. Expressive Vocabulary</td>
<td>Vocab</td>
<td>2. Emotional Awareness</td>
<td>EmoAware</td>
</tr>
<tr>
<td>5. Oral Comprehension</td>
<td>OralComp</td>
<td>5. Perspective Taking</td>
<td>Perspect</td>
</tr>
<tr>
<td>6. Print Awareness</td>
<td>Print</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Classifying &amp; Sorting</td>
<td>Sort</td>
<td>1. Copying a Shape</td>
<td>Copy</td>
</tr>
<tr>
<td>2. Comparison</td>
<td>Size</td>
<td>2. Drawing a Person</td>
<td>Draw</td>
</tr>
<tr>
<td>4. Number Identification</td>
<td>NumberID</td>
<td>4. Hopping on One Foot</td>
<td>Hop</td>
</tr>
<tr>
<td>5. Shape Identification</td>
<td>ShapeID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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

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

*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 domain bi-factor model in each country?
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

<table>
<thead>
<tr>
<th>Country</th>
<th>$\chi^2$ (df)</th>
<th>RMSEA (90% CI)</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>143.738 (58)</td>
<td>0.024 (0.019, 0.029)</td>
<td>.984</td>
</tr>
<tr>
<td>Bolivia</td>
<td>86.634 (58)</td>
<td>0.032 (0.016, 0.046)</td>
<td>.981</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>138.489 (58)</td>
<td>0.045 (0.036, 0.055)</td>
<td>.965</td>
</tr>
<tr>
<td>Uganda</td>
<td>87.968 (58)</td>
<td>0.032 (0.017, 0.045)</td>
<td>.971</td>
</tr>
<tr>
<td>Vietnam</td>
<td>116.053 (58)</td>
<td>0.039 (0.028, 0.049)</td>
<td>.969</td>
</tr>
</tbody>
</table>

*Note: Goodness of fit indicators: $\chi^2 / df < 2; \ RMSEA < .050; \ TLI > 0.90.*
No evidence of scalar invariance

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$ (df)</th>
<th>RMSEA (90% CI)</th>
<th>TLI</th>
<th>$\chi^2$ (difference)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural</td>
<td>224.940 (57)</td>
<td>.035 (.030, .040)</td>
<td>.952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DIF: No anchor set of items identified

<table>
<thead>
<tr>
<th>Item</th>
<th>Afghanistan</th>
<th>Bolivia</th>
<th>Ethiopia</th>
<th>Uganda</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CONFLIC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2 CONFLIC2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3 EMOTION1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4 EMOTION4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5 EMPATHY1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>6 EMPATHY2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>7 EMPATHY3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>8 FRIENDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>9 PERSON1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10 PERSON2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>11 PERSON3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>12 PERSON4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>13 PERSON5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>--</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Afghanistan</th>
<th>Bolivia</th>
<th>Ethiopia</th>
<th>Uganda</th>
<th>Vietnam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Literacy</strong></td>
<td>.877</td>
<td>.807</td>
<td>.917</td>
<td>.880</td>
<td>.830</td>
</tr>
<tr>
<td><strong>Numeracy</strong></td>
<td>.897</td>
<td>.690</td>
<td>.881</td>
<td>.830</td>
<td>.667</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td>.820</td>
<td>.624</td>
<td>.781</td>
<td>.846</td>
<td>.649</td>
</tr>
</tbody>
</table>
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

“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
Social Change and Human Development


---

**Sociodemographic Change**

<table>
<thead>
<tr>
<th>Rural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture Subsistence</td>
<td>Commerce</td>
</tr>
<tr>
<td>Less formal education</td>
<td>More formal education</td>
</tr>
<tr>
<td>Many children</td>
<td>Fewer children</td>
</tr>
<tr>
<td>Living with others</td>
<td>Living alone</td>
</tr>
</tbody>
</table>

**Values**

<table>
<thead>
<tr>
<th>Collectivism</th>
<th>Individualism</th>
</tr>
</thead>
<tbody>
<tr>
<td>More obedience</td>
<td>Less obedience</td>
</tr>
<tr>
<td>Age-graded authority</td>
<td>Child-centeredness</td>
</tr>
</tbody>
</table>
### Developmental/Behavioural Change

<table>
<thead>
<tr>
<th>Respect, obedience</th>
<th>Expression, curiosity, independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shyness</td>
<td>Extraversion</td>
</tr>
<tr>
<td>Gender roles ascribed</td>
<td>Gender roles chosen</td>
</tr>
<tr>
<td>Focus on others</td>
<td>Focus on self</td>
</tr>
<tr>
<td>Empathy for others</td>
<td>Internal feeling states</td>
</tr>
<tr>
<td>Less self-esteem</td>
<td>More self-esteem</td>
</tr>
<tr>
<td>Fitting in</td>
<td>Standing out, uniqueness</td>
</tr>
<tr>
<td>Cooperation</td>
<td>Competition</td>
</tr>
</tbody>
</table>
Urban Migration and Cognitive Abilities in the Gambia

Urban Migration and Cognitive Abilities in the Gambia

“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?”

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 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

SEL Study
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**
- Perspective-Taking
- Empathy
- Appreciating Diversity
- Respect for Others

**Relationship Skills**
- Communication
- Social Engagement
- Relationship Building
- Teamwork

**Self-Management**
- Impulse control
- Stress management
- Self-discipline
- Self-motivation
- Goal Setting
- Organizational skills

**Self-Awareness**
- Identifying emotions
- Accurate self-perception
- Recognizing strengths
- Self-confidence
- Self-efficacy

**Findings from Mtwarra, Tanzania**

**Bold text** = most frequently mentioned

**Social Responsibility**
- Obedient
- Respectful
- Attentive
- Disciplined
- Polite and calm
- Clean
- Religious

**Social Relationships**
- Trustworthy
- Truthful
- Cooperative
- Sociable
- Funny
- Loving others (Empathic)

**Individual Competencies**
- **Self-directed**
  - Careful
  - Persistent
  - To have goals

- **Self-belief**
  - Curious
  - Courageous/Daring
  - Self-awareness

**Valued by parents and teachers**
- Important for life in general

**Valued by teachers only**
- Important for schooling

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

- Exchange ideas with elders
- Likes to ask many questions
- Agrees to be sent on an errand
- Greets elders

Confidence
Curiosity
Obedience
Respect

More than others
The same as others
Less than others
No
### Obedience

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does child easily agree to be sent?</td>
<td>Hamisa agrees easily to be sent</td>
</tr>
<tr>
<td>Does child willingly follow instructions?</td>
<td>Juma agrees easily to be sent</td>
</tr>
<tr>
<td>When child is directed to complete a task, does s/he do it with heart and complete it successfully?</td>
<td></td>
</tr>
<tr>
<td>When child is sent to complete an errand, does s/he return on time?</td>
<td></td>
</tr>
<tr>
<td>If child is told to do a chore, does s/he do it?</td>
<td></td>
</tr>
<tr>
<td>Does child refuse to stop and continue doing something after being told not to</td>
<td></td>
</tr>
</tbody>
</table>
# Obedience

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

# Respect

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does child offer to assist/receive elders?</td>
<td></td>
</tr>
<tr>
<td>Does child kindly greet elders?</td>
<td></td>
</tr>
<tr>
<td>Does child get your permission before doing something or going somewhere?</td>
<td></td>
</tr>
<tr>
<td>If child makes a mistake, does s/he ask for forgiveness without being told to?</td>
<td></td>
</tr>
</tbody>
</table>

# Final Obedience Measure

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does child easily agree to be sent?</td>
<td></td>
</tr>
<tr>
<td>Does child willingly follow instructions?</td>
<td></td>
</tr>
<tr>
<td>When child is directed to complete a task, does s/he do with with heart and complete it successfully?</td>
<td></td>
</tr>
<tr>
<td>When child is sent to complete an errand, does s/he return on time?</td>
<td></td>
</tr>
<tr>
<td>If child is told to do a chore, does s/he do it?</td>
<td></td>
</tr>
<tr>
<td>Does child kindly greet elders?</td>
<td></td>
</tr>
</tbody>
</table>
## Persistence

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

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

### Emotional Control

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does child easily become frustrated or angry?</td>
</tr>
<tr>
<td>Does child quit working on tasks before s/he is finished?</td>
</tr>
<tr>
<td>Does child give up easily when tasks or work seem difficult?</td>
</tr>
<tr>
<td>Does child react angrily when s/he doesn’t get what she wants/ told to stop doing something?</td>
</tr>
<tr>
<td>Does child have strong emotions?</td>
</tr>
</tbody>
</table>

### Polite and Calm

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is child calm even when disturbed/irritated by others?</td>
</tr>
<tr>
<td>Does child respond nicely/politely when asked a question?</td>
</tr>
<tr>
<td>Does child speak in a soft, measured way?</td>
</tr>
<tr>
<td>Does child react angrily when s/he doesn’t get what she wants/ told to stop doing something?</td>
</tr>
<tr>
<td>Does child have strong emotions?</td>
</tr>
<tr>
<td>Is child calm even when disturbed/irritated by others?</td>
</tr>
</tbody>
</table>
### Exploratory Factor Analysis of Parents Items

<table>
<thead>
<tr>
<th>Factor</th>
<th>Original targeted constructs</th>
<th>Variance Explained</th>
<th>Cumulative Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obedient</td>
<td>Obedient</td>
<td>29%</td>
<td>29%</td>
</tr>
<tr>
<td>Curious</td>
<td>Curious, confident</td>
<td>8%</td>
<td>37%</td>
</tr>
<tr>
<td>Conscientious</td>
<td>Persistent, careful, self-directed</td>
<td>6%</td>
<td>43%</td>
</tr>
<tr>
<td>Emotional Regulation</td>
<td>Polite, persistent, obedient</td>
<td>5%</td>
<td>48%</td>
</tr>
<tr>
<td>Sociable</td>
<td>Sociable, cooperative</td>
<td>4%</td>
<td>52%</td>
</tr>
<tr>
<td>Polite</td>
<td>Polite, cooperative</td>
<td>3%</td>
<td>55%</td>
</tr>
<tr>
<td>Religious</td>
<td>Religious</td>
<td>3%</td>
<td>58%</td>
</tr>
</tbody>
</table>
Exploratory Factor Analysis of Parents Items

<table>
<thead>
<tr>
<th>Factor</th>
<th>Original targeted constructs</th>
<th>Variance Explained</th>
<th>Cumulative Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obedient</td>
<td>Obedient</td>
<td>29%</td>
<td>29%</td>
</tr>
<tr>
<td>Curious</td>
<td>Curious, confident</td>
<td>8%</td>
<td>37%</td>
</tr>
<tr>
<td>Conscientious</td>
<td>Persistent, careful, self-directed</td>
<td>6%</td>
<td>43%</td>
</tr>
<tr>
<td>Emotional Regulation</td>
<td>Polite, persistent, obedient</td>
<td>5%</td>
<td>48%</td>
</tr>
<tr>
<td>Sociable</td>
<td>Sociable, cooperative</td>
<td>4%</td>
<td>52%</td>
</tr>
<tr>
<td>Polite</td>
<td>Polite, cooperative</td>
<td>3%</td>
<td>55%</td>
</tr>
<tr>
<td>Religious</td>
<td>Religious</td>
<td>3%</td>
<td>58%</td>
</tr>
</tbody>
</table>
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$^2 = 4.48, p = .03$
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
Dr. Nkanileka Mgonda Salaam
Dr. Prosper Mosha
Florentina Nsolezi
Grace Jeremiah
Kellie Betts
Prof. Kristen Bub
Sarrynna Sou, Corina Owens

Tusome Pamoja
University of Dar es Salaam
University of Dodoma
University of Dodoma
St. Augustine University
RTI
University of Illinois, USA
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

@RTI_INTL_DEV | #RTILearns | #EarlyYears
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.