Construction and validation of the D-Score Population-Based Indicators of Early Child Development

20 September, 2016
Ann Weber, Stanford University for the Global Child Development Group
Advisory Board Members

- Represent a wide variety of backgrounds, ethnicities, and training
- Are well established investigators in ECD
- Agreed to provide their expertise to the project
- Agreed to share their longitudinal cohort data
Project Objectives

• Develop a global child development instrument to assess **0-3 year-olds** that:
  – Is reliable & valid across different cultures and income contexts
  – Is feasible for use in population-level surveys
  – Is sensitive to intervention
  – Is predictive of later school-related outcomes

• Develop a scale with a common numerical unit, or **D-score**, that can be used for quantitative comparisons within and between ages
Example of a D-score reference chart

D-score unit

Leverage existing longitudinal data

15 cohorts in 11 countries
> 16,000 children

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

ECD Measures for ages < 3 years
- Bayley I, II, III
- Griffiths
- Denver
- Dutch Scale
- Battelle
- Ages & Stages Questionnaires
- Barrera Moncada
- Other

Longitudinal data for ages > 5 years
- PPVT
- Ravens
- Strengths & Difficulties Questionnaire
- Other
Typical theoretical framework for a measurement instrument
Our framework linking items from multiple instruments to ability

Latent Ability

Scores (0/1)

Items

\[
\begin{align*}
\delta_{b1} & \quad \delta_{b2} & \quad \ldots & \quad \delta_{g1} & \quad \delta_{g2} & \quad \ldots & \quad \delta_{d1} & \quad \delta_{d2} & \quad \ldots & \quad \delta_{dz} \\
\text{Bayley III} & \quad \text{Griffiths} & \quad \ldots & \quad \text{Denver} & \quad \ldots 
\end{align*}
\]
We mapped items across instruments to Bayley III in 5 domains 
Cognition, Receptive & Expressive Language, Fine & Gross Motor

Ex: 3 language items mapped 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>
Selecting candidate items for the final instrument

• Domain knowledge expertise
  – Ease of administration and interpretation with minimal training
  – Easily understood by program personnel and policy makers

• Performance cross-culturally in our dataset
  – Item response theory (IRT): links ability to item difficulty
  – The simplest model in IRT is the Rasch model: a probability model
Testing fit to the model

Probability of “yes” response

Child ability ($\theta$)
Too much variability

Produces 2 or more vowel sounds

Probability of “yes” response

Child ability ($\theta$)
Too little variability

Probability of “yes” response

Hands you object when named

Child ability ($\theta$)
Just right

Imitates sounds that others produce

Probability of “yes” response

Child ability (θ)
Differential item function: by instrument, country, language, gender...

Understands and says “ball”

Boys

Girls

Probability of “yes” response

Child ability ($\theta$)
Overall instrument

Curves for easier items are shifted to the left

Curves for harder items are shifted to the right

Three items of varying difficulty

Child ability (θ)
Goal: 2 population-level instruments for children 0-3 years

**Monitoring & Evaluation**
- Global development score
- Few items (aim: 10-15 per child)
- Less time (aim: < 10 min)
- Less training

**Program Impact Evaluations**
- Global and domain-specific
- More items (aim: 10-15 per domain per child)
- More time (aim: < 30 min)
- More training
- Goal: direct assessment

*For both*
- Open access with no licensing fees
- Materials (if necessary) at cost (no profit)
- Training, quality assurance guidelines downloadable
Risks

• A single set of common items may not work across countries and contexts.
  – But, we have a large number of items from widely used and accepted instruments

• Our findings may not reproduce elsewhere.
  – Importantly, we will share the indicators with researchers who can provide additional evidence of the validity of the scale in their data
Strengths

The scale will...

• Be validated using data from 11 countries
• Be designed with field work in mind
• Demonstrate prediction to school-age skills
• Be used to create D-score age trajectories using population representative data
  – We will start with South African cohort data
  – Define process to extend to other countries
Acknowledgements

• Bill & Melinda Gates Foundation
• Global Child Development Group
  – Susan Walker, PI, University of the West Indies
  – Stef van Buuren
  – Sally Grantham-McGregor
  – Maureen Black
  – Orazio Attanasio
  – Hedwig Hoffstetter
  – Other Advisory Board members
• Participating Colleagues
EXTRA SLIDES
11 Countries, 15 Cohorts, > 16,000 children, & ≥ 1 phases = Dataset with >30,000 rows & > 300 items

<table>
<thead>
<tr>
<th>Country</th>
<th>Investigators</th>
<th>Bayley</th>
<th>Denver</th>
<th>Griffiths</th>
<th>Battelle</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>Hamadani*, Tofail</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>Lozoff*</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>Behrman, Bravo, Fernald*, Reynolds</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Lozoff*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia (Bogota)</td>
<td>Attanasio*, McGregor*, Rubio-Codina*</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>Attanasio*, McGregor*, Rubio-Codina*</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ecuador</td>
<td>Araujo*, Schady</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Hanlon*, Medhin</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jamaica</td>
<td>Walker*, Chang*</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jamaica</td>
<td>McGregor*, Powell</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madagascar</td>
<td>Galasso, Fernald*, Ratsifandrihamanana*, Weber*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>Verkerk, Schönbeck, Van Buuren*</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>Richter*, Cameron</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

* Advisory board member
The Rasch model is the simplest IRT model and assumes:

• The existence of single, continuous latent variable, $\theta$, for a sample of $n$ children
  ▪ larger $\theta$ = higher ability

• A difficulty parameter, $\delta$, for each of a set of $i$ items
  ▪ larger $\delta$ = more difficult

• The probability of succeeding on an item is a function of the difference between $\theta_n$ and $\delta_i$

\[ P(X_{ni} = 1 \mid \theta_n, \delta_i) = f(\theta_n - \delta_i) = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}} \]
The mathematical unit is the log odds, or logit; a common linear scale for persons and items

\[ \theta - \delta = 0, \]
\[ P = 50\% \]

\[ \theta - \delta = -1, \]
\[ P = 27\% \]

\[ \theta - \delta = 1, \]
\[ P = 73\% \]

Fig. 1. Representation of three relationships between respondent location and the location of an item.
Model features and assumptions

• **Item invariance**: the interpretation of the relative location of the items is independent of the location of the persons (1 cm represents 1 cm wherever you are on a map).

• **Specific objectivity** – if the model holds, then we do not need any particular set of persons to obtain the item difficulties, and we do not need to give every person the same set of items.

• **Conditional independence** - Each item response is independent of the others conditional on knowing the item difficulty and person ability.
  – To obtain the probability of a response pattern, we can multiply the probability for each response
  – Does not hold if items share common stimulus material.

• Different tests given at different times can be put on a common scale with linking or anchor items.
Timeline

Now: Post-harmonization of existing developmental data obtained

End of 2016: Fit Rasch model to the data and identify items for a composite scale

February 2017: Estimate an ability score for each child and the distribution of the resulting D-scores within and across the cohorts.

March 2017: Test predictive validity of the D-scores assessed in several cohorts


End of 2017: Establish process to select a set of developmental indicators for “off track” development