What’s good enough in SEL and Soft Skills Measurement?

4 experiences in developing assessment tools in diverse contexts
Moderator: Rebecca Pagel, USAID
Panelists:

- Smita Das, World Bank Gender Innovation Lab
- Dr. Catherine Honeymoon, World Learning
- Dr. Allyson Krupar, Save the Children US
- Dr. Elizabeth Kim, International Youth Foundation
Agenda

- Presentation of 4 experiences in developing assessment tools in diverse contexts
- Introduce Handout “Roadmap for Measuring Soft Skills and SEL”
- Panel discussion
- Q&A with audience
# International Social and Emotional Learning Assessment (ISELA)

- **6-12 years old**
- **Background information, experiences of risks in their family/community, perspective of threats in the learning environment**
- **Enumerator administered**
- **Uses:**
  - Cross-sectional use: Provides profile of children’s SEL competencies and the SEL environment
  - Longitudinally: Provides picture of how SEL competencies change over time
  - ISELA is not a diagnostic tool

<table>
<thead>
<tr>
<th>SEL COMPETENCY</th>
<th>Skill included in ISELA (name of ISELA task)</th>
</tr>
</thead>
</table>
| Self-awareness                  | Accurate self-perception (Self concept)  
    | Self-efficacy (Self-concept)                                                                                     |
| Self management                 | Stress management (Stress management)  
    | Self-motivation (Perseverance)                                                                                     |
| Social awareness                | Perspective-taking (Empathy)  
    | Empathy (Empathy)                                                                                                  |
| Relationship skills             | Relationship building (Relationships)                                                                        |
| Responsible decision-making     | Solving problems (Solving conflict)                                                                                         |
Development and Testing ISELA (n=576)

- Strong interrater reliability (>0.80 Krippendorf’s alpha)
- Good internal consistency reliability (>0.70)
- Convergent validity: We found a statistically and practically meaningful positive relationship between grade and four of the SEL skills of their score on the help construct. One-grade difference was associated with a 4-8 percentage point change in the number of items that children answered appropriately. Children’s self-reported perception of interpersonal threats was positively associated with their empathy and conflict resolution scores. A one-unit change in the number of threats that a child identified was positively associated with a 2-5 percentage point change in their SEL skills.
- Discriminant validity: Exposure to risk factors - we predicted a negative relationship between SEL skills and exposure to risk factors. However, we found no relationships between these variables when controlling for the effect of important equity factors like grade, gender, and household wealth.
IYF’s Life Skills Survey Tool (LiSST)

- Youth self-report survey with pre-post or retrospective administration
- 49 statements on life skills organized into 4 life skills domains based on IYF’s Life Skills Framework: **Positive Mindset, Interpersonal Skills, Higher Order Thinking, and Community Mindset**
- Participants rate degree of agreement with statements such as: “I react positively to suggestions from others on how I might improve myself” (PM) and “I am comfortable taking risks when trying to solve a problem.” (HOTS)
- Developed in fall 2019 and tested with a pretest population of 300+ US youth. Further pilot tested in Mexico and Tanzania in 2020.
The survey tool was pretested with **321 youth** between the **ages of 16 and 24** across the United States. Results suggested:

- **Domain scores were reliable.** All domains demonstrated high reliability (>0.80) and interitem correlations (corr > 0.30)
- **Domain scores were equitable.** No significant differences were found on average ratings across 4 domains by subgroup (gender, age, ethnicity, urbanicity, socioeconomic status) with few exceptions.
- **Domain scores were valid.** Confirmatory factor analysis (CFA) models fit the four specified domains as factors.
- **Domain scores were correlated with other characteristics:** Positive mindset was positively correlated with being employed and higher grades, and negatively correlated with being sad/depressed. Interpersonal skills and higher order thinking skills both were positively correlated with grades.
World Learning’s WorkLinks Skills and Values Assessment (WLSVA)

- **56 core questions**, divided among 3 subscales:
  - Soft skills, Earning Skills, and Civic Values
- **Validated with ages 18-29**: Algeria (N=294) and Iraq (N=70+)
  - Available in English, French, Arabic (2 versions), and Kurdish
  - 120 downloads from 40+ countries—available under CC BY-NC-SA from [https://www.worldlearning.org/what-we-do/wlsva-toolkit/](https://www.worldlearning.org/what-we-do/wlsva-toolkit/)
- **Self-administered** online (auto scoring) or paper-based
- **High internal reliabilities** (0.79-0.94), positive correlation with employment and stress management
- **Sensitive enough to detect individual change**: High test-retest reliabilities and smallest real difference of 3.8%-5.5% for subscales
WLSVA development and validation

Review of existing measures
Comprehensive review of existing soft skills measures, including detailed item analysis of 9 validated assessment tools - none fully matched our curriculum, objectives, or research in Algeria & Iraq.

Pilot testing - Twice
Initial pilot test with 128 youth, item total statistics and EFA, retaining strongest items and strongest constructs, Second test-retest pilot with 166 youth. Finalization.

2020
Translation & FGDs
Translation into Algerian French, FGDs and cognitive interviewing with youth, item adaptation.

2021
Initial usage in 5 projects
Creation of a self-scoring assessment on Kobo Toolbox. Initial use within 5 projects (2 in Algeria, 3 in Iraq) including new opportunity to assess predictive validity with employment outcomes.

COVID-19
World Bank Africa Gender Innovation Lab + Innovations for Poverty Action

Measuring 14 Socio-Emotional Skills to Unpack Which SES Matter Most for Employment

Part of a 4-RCT Study
Measure Development

The Instrument

Data Collection: Iterative Development

Process within each country

Alternative Measures

Self-Reported Scales

HH Informant Interviews

Situational Judgement Tests

Tasks/Games

Pilot in CDI (1858)

Spring 2019

Reliability

Pilot 1: Nigeria

(100)

Fall 2019

Correlation, Reliability, Concurrent

Pilot 2: Nigeria

(500)

Spring 2020

+CFA, IRT

2nd Cohort in Cote d'Ivoire (1155)

Fall 2020

Baseline in Tanzania (4800)

May 2021

+Predictive Validity

Focus Group Discussions/Interviews with Entrepreneurs

• What skills key for: success in each sector, men vs. women, household discussions
• Which behaviors associated with each skill

Cognitive Interviews

• Slow walkthrough: “what did you understand with this question” “is there a correct answer”
• Key for enumerator training/feedback, timing, order, engagement

Intensive Translation Process

• Translation, Back-translation
• Translation Ratings, Linguistic Feedback by Psychologist, Final Translation
• Glossary items

Review from Subject Matter Experts

• Psychologists
• Curricula Experts
• Education Assessment Experts
• International Organizations
• Field Coordinators
Roadmap for Measuring Soft Skills and SEL

- Determine What You Want to Measure and Identify Your Theoretical Framework
  - Consider: What are the key outcomes, relevant skills, and theories?

- Design How You Will Collect Data and Any Constraints
  - Consider: Language, literacy, numeracy, time, cultural, and contextual factors.

- Decision Point: Can You Use an Existing Measure?
  - Yes
    - Conduct Qualitative Work with Beneficiaries and Content Experts to Develop Items for New or Adapted Measure
      - Consider: How do skills manifest as behaviors?

  - No
    - Conduct Careful Translation and Adaptation
      - Consider: Comparison of translation accuracy, content, and cultural feedback.

- Test for Understanding
  - Consider: How effective are the measures in different settings?

- Create a Plan for Analysis
  - Consider: Data analysis techniques and software.

- Collect and Analyze Data
  - Consider: Data collection methods and tools.

- Check for Reliability and Validity
  - Consider: Reliability and validity of the measures.

- Link Measures and/or Theoretical Frameworks as Needed (Iterative Development)

Authors: Smita Das, Elizabeth Kim, Alyson Kupfer, Catherine Honeyman, Rebecca Page © 2021