# Meet the new tools of the INEE Measurement Library

10 December 2020

Measurement Library

Inter-agency Network for Education in Emergencies

# Webinar Housekeeping

Participant audio and video are disabled.

Post questions using the Q&A function at any time.

This session is being recorded.

Closed captioning available in En

The recording and presentations will be shared on the INEE website – http://www.inee.org

# Agenda

1. Overview of the Measurement Library

   *Sarah Montgomery, Knowledge Management Coordinator, INEE*

2. Child-Friendly School Questionnaire (CFS-Q)

   *Michael Wu, Graduate Research Assistant, NYU Global TIES for Children*

3. Self-Regulation Assessment-Assessor Report (SRA-AR)

   *Dr. Kalina Gjicali, Research Scientist, NYU Global TIES for Children*

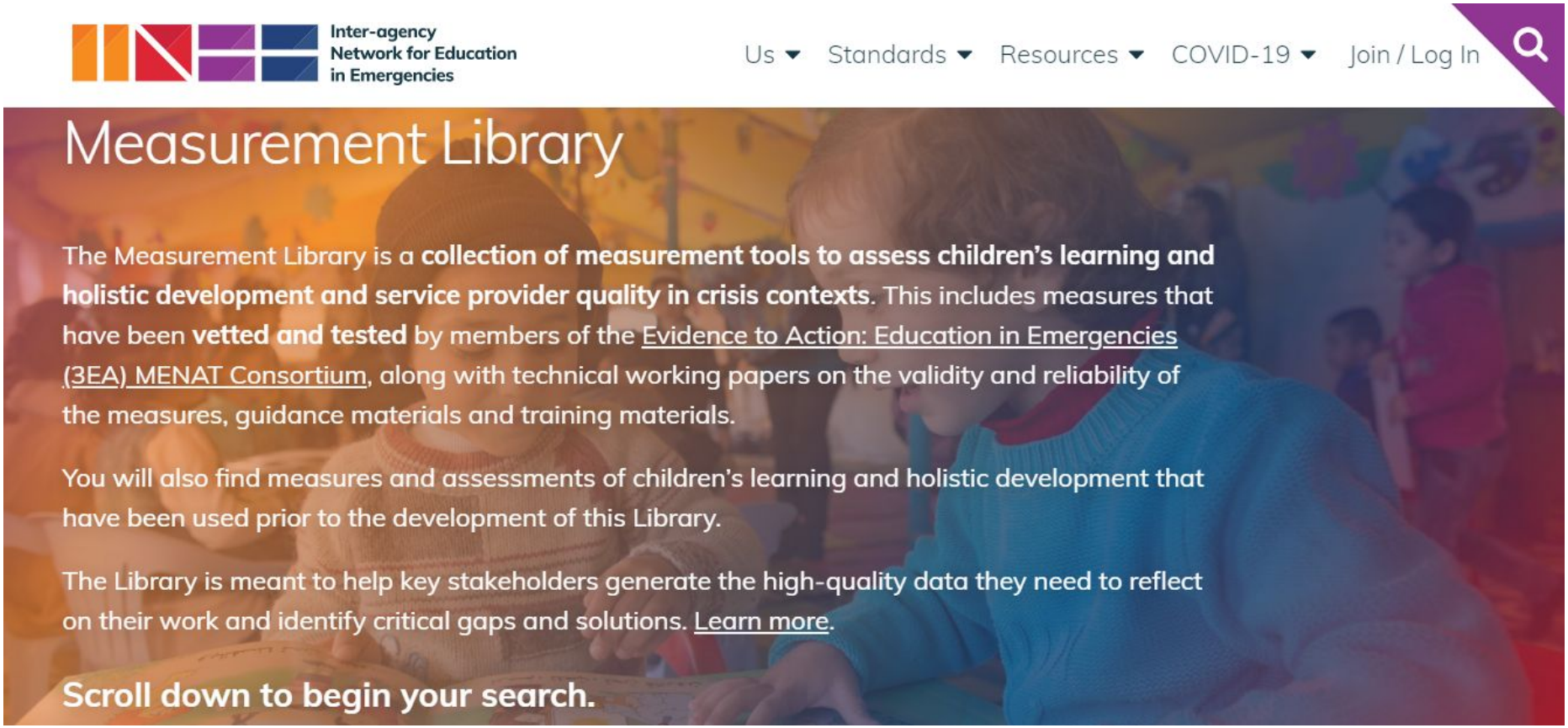
4. Teacher Classroom Observation (TCO)

   *Dr. Jeongmin Lee, Education Researcher, IRC*

5. Q&A

   *Moderator: Roxane Caires, Research Scientist, NYU Global TIES for Children*

# INEE Measurement Library



# Child Friendly School Questionnaire (CFSQ-SL): self-report perceived school climate for refugee children in Lebanon

Michael Wu, Ed.M.
Global TIES for Children

# Measuring school climate in humanitarian context

- With the influx of Syrian refugees, the number of public-school students in Lebanon nearly doubled within five years (MEHE, 2018), creating significant challenges to the delivery of quality education to both refugees and host-country children in these public schools (Dryden-Peterson et al., 2019).

- Because of the sudden changes the refugee crisis created in these schools, it is vital for us to understand how refugee children feel about the school environment.

- In 1999, UNICEF introduced the Child Friendly School (CFS) framework with the exact aim of building child-friendly school environments to "serve the whole child" (Osher et al., 2009).
  - Inclusiveness / Child-centeredness / Democratic participation

UNICEF 56 items (Osher et al., 2009)
CFS 30 items (Godfrey et al., 2012)

Good reliability and validity in Philippines, Nicaragua, South Africa

→ 3EA initial selection 34 items
Emotionally Supportive Climate scale (CFS1-10)
Challenging Student-Centered Learning Environment scale (CFS11-22)
Safe, Inclusive, and Respectful Climate scale (CFS23-34R)

→ Factor analysis → CFSQ-SL 19 items (Gjicali et al., 2020)

Tested on Syrian children in Lebanon public schools

# Evidence from what context/population?

- Bekaa and Akkar regions of Lebanon, in school year 2016-17

- 4,598 Syrian refugee children (ages 5-15)
- Enrolled in Lebanese public schools
- In grades 1st – 9th  (95% in grades 1st – 6th)
- Average of 2.6 years as a refugee
- 17% migrated in the past year

- Developed and used for program evaluation (SEL, Retention Support)




# Child Friendly School Questionnaire (CFSQ-SL): self-report perceived school climate in Lebanon

**Appx. 5 minutes**


### 1. Caring and supportive teachers (4 items)

E.g.
- Teachers at this school really care about students like me
- Teachers give students opportunities to improve their work if they do poorly on an assignment

### 2. Engaging and motivating school (7 items)

E.g.
- The subjects we are studying at this school are interesting
- Students are encouraged to work together in class

### 3. Respectful and inclusive school (5 items)

E.g.
- My teachers treat me with respect
- This school places a high value on understanding and respecting children's rights

### 4. Safe school (reverse coded; 3 items)

E.g.
- I sometimes stay home from school because I am worried about my safety
- This school is badly affected by crime and violence in the community

| Strongly disagree [1] | Disagree [2] | Agree [3] | Strongly Agree [4] |
|---|---|---|---|

# Evidence of reliability and validity

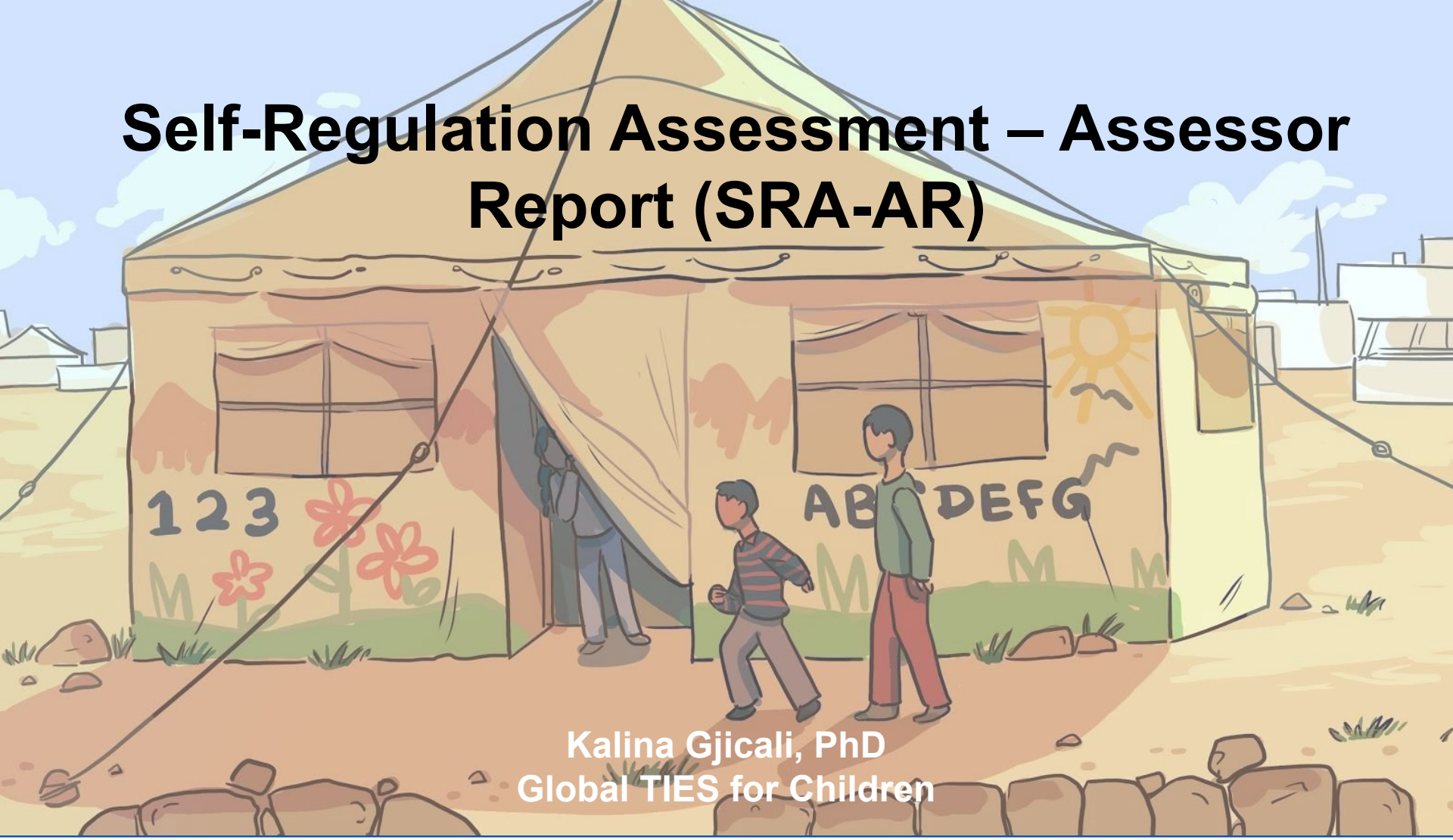
| Type of evidence | Evidence |
|---|---|
| **Structural evidence of validity** | Consistent factor structures and high factor loadings |
| **Internal consistency of subscales** | Acceptable internal consistency, but alpha is low in some subscales at certain time points |
| **Correlation across subscales** | At each wave, factor 1-3 are highly correlated with each other<br><br>At each wave, factor 4 (Safe School) has low to moderate correlation with the other factors. |
| **Correlation across time** | Low correlation across time for each subscale |
| **Evidence of measurement invariance** | Measurement invariance over time, and across treatment groups, gender groups, and age groups |

# Recommendations for Use and Adaptation of CFSQ-SL

**Careful adaptation and validation required for use with different context and population:**

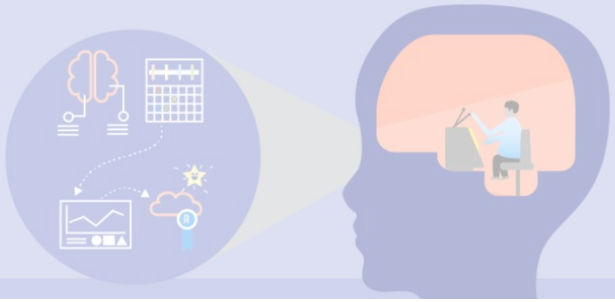
- **CONDUCT COGNITIVE INTERVIEWS** to:
  - check on whether each item is fully comprehended
  - identify what specific words and phrases in the question mean
  - identify what types of strategies are used to retrieve information and respond to items
  - identify whether children are answering the questions accurately and thoughtfully
  - assess the degree to which respondents match their internally generated answer to the response categories given by the survey question
- **USE WEIGHTED SCORES** with model-based weights (e.g., factor scores) to reduce measurement error (i.e., coefficient omega over coefficient alpha)
- **EMBED EXPLICIT PROMPTS** in the introductory statement to ask respondents to think back to their experience in school as a whole (e.g., “Think back to your time in school over the past two weeks. Some days you may have felt positive and supported, some other days may have been unpleasant. In general,…)
- **ALIGN** the measure with the program implementation timeline (e.g., Before school starts and after school routines have been established)

# Self-Regulation Assessment – Assessor Report (SRA-AR)




Kalina Gjicali, PhD
Global TIES for Children

# Self-Regulation: A Pilar of Social & Emotional Learning



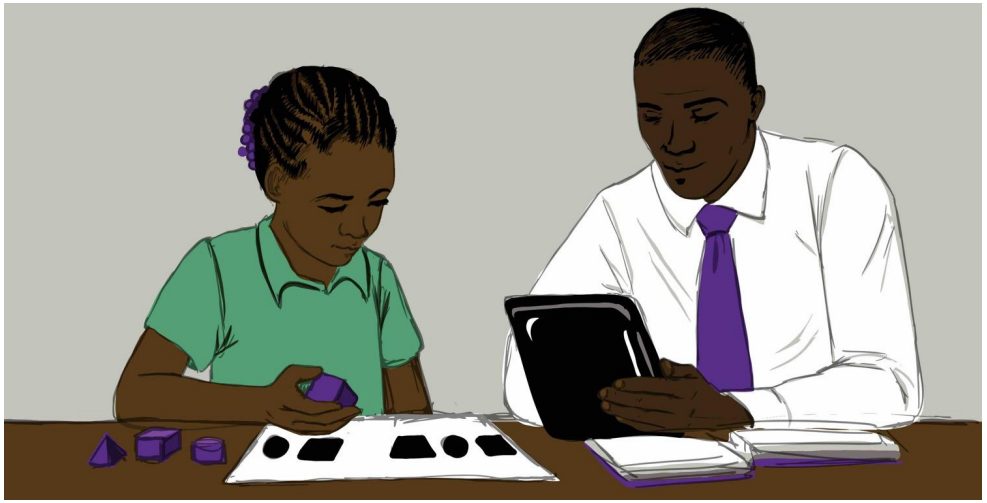
- Self-regulation is the degree to which children can modulate their behavior and/or emotional state towards a specific goal (Duncan et al., 2017; Ursache et al., 2012).

- Strong self-regulatory skills can protect children against the negative impacts of adversity (e.g., US-based studies, Blair, 2010).
  - High-risk settings in which Syrian refugee children reside and learn.

- Behavioral regulation is associated with children's later academic achievement, interpersonal skills, and mental and physical health (Pandey et al., 2018; Robson et al., 2020).

- The **SRA-AR**, is an assessor-report measure of behavioral regulation.
  - Adapted from the post-assessment, assessor-report section of the Preschool Self-Regulation Assessment-Assessor Report (PSRA-AR: Smith-Donald et al., 2007), a performance-based measure originally designed to assess self-regulation skills of preschool children in the U.S.

# SRA-AR: Measure Snapshot



| SRA-AR: Measure Snapshot | |
|---|---|
| For what? | Program evaluation |
| About what? | Behavior regulation |
| By whom? | Assessor/enumerator/data collector/interviewer |
| What method? | Observation of child |
| How? | Pencil and paper, tablet (digitized) |
| Where? | After task administration/data collection session |

Observation

13-items

Assessor-report

3-5 minutes

Program evaluation

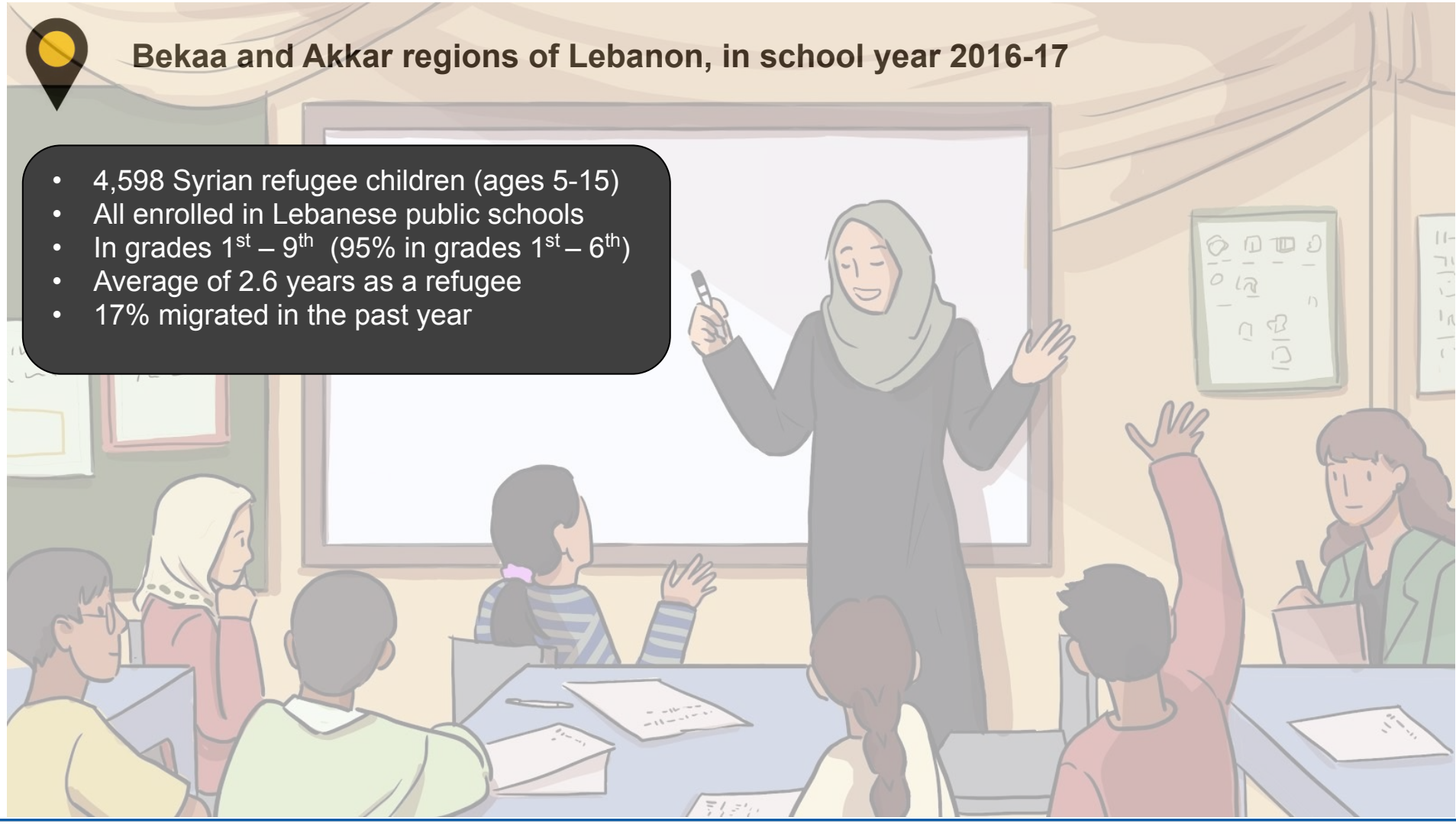
# SRA-AR: *Self-Regulation Assessment-Assessor Report*

**Observation measure developed to measure children's behavioral regulation.**

| Item | Item description |
|---|---|
| SRA1 | Pays attention to instructions and demonstration |
| SRA2 | Careful, interested in accuracy |
| SRA3 | Sustains concentration; willing to try repetitive tasks |
| SRA4 | Is careless or destructive with test materials |
| SRA5 | Can wait during and between tasks |
| SRA6 | Remains in seat appropriately during the test |
| SRA7 | Alert and interactive; is not withdrawn |
| SRA8 | Cooperates; complies with requests |
| SRA9 | Shows pleasure in accomplishment and active task mastery |
| SRA10 | Confident |
| SRA11 | Defiant |
| SRA12 | Passively noncompliant |
| SRA13 | Modulates and regulates arousal level in self |

| Scale | Label for Item SRA1 |
|---|---|
| 1 | Child spends most of time off-task, inattentive. |
| 2 | Child's attention frequently drifts and requires frequent prompts. |
| 3 | Child's attention occasionally drifts, particularly at the end of activities, but is responsive to prompts. |
| 4 | Child listens closely. Child attends to and complies with interviewer. |

# Contextual Evidence



**Bekaa and Akkar regions of Lebanon, in school year 2016-17**

- 4,598 Syrian refugee children (ages 5-15)
- All enrolled in Lebanese public schools
- In grades 1st – 9th (95% in grades 1st – 6th)
- Average of 2.6 years as a refugee
- 17% migrated in the past year

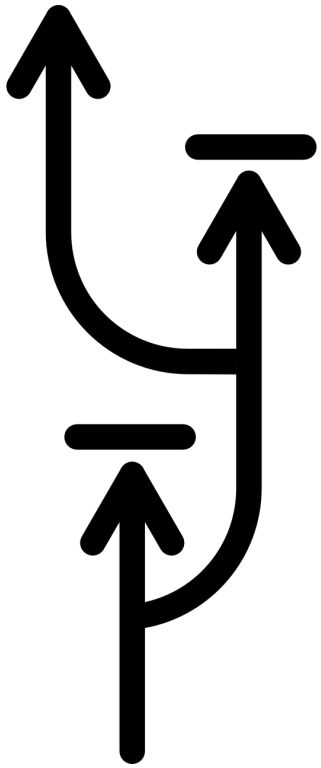
# Evidence of Reliability and Validity

| Type of evidence | Evaluation |
|---|---|
| **Structure of the Measure** | ✓ One-factor model measuring one single aspect of behavioral regulation. Great model fit. |
| **Internal Consistency** | ✓ Excellent internal consistency (Baseline α = .955; Midline α = .952; Endline α = .953) |
| **Inter-rater Reliability** | X Absence of evidence of inter-rater reliability required for observation measures. |
| **Correlational Evidence of Validity** | |
| **Correlations across time** | ✓ $r$ = 0.4 – 0.47 across waves of data (November, March, May) |
| **Correlations with related constructs** | ✓ Positive high correlations with literacy and numeracy measures ($r$ = 0.58) and moderate correlations with executive functioning measures; 0.10 < $r$ < 0.27. |
| **Measurement Invariance** | ✓ Measurement invariance over time, by treatment groups, by gender groups, and by age groups. |

# Recommendations for Use and Adaptation

**Careful adaptation and validation required for use with different context and population:**

1. **DEVELOP** A RIGOROUS ASSESSOR TRAINING PROTOCOL
    1. To promote inter-rater reliability to ensure that different reporters *see the same level* of behavioral regulation.
2. **STANDARDIZE** THE OBSERVATION SCENARIO
3. **EASILY ADAPTABLE** MEASURE FOR OTHER PURPOSES
    1. Potential use by a teacher/facilitator to track children's behavioral regulation over the course of the school year.
    2. Child self-report (self-perceptions), peer-reported perceptions, parent-reported perceptions, or teacher-reported perceptions.
        - For the purposes of measuring *perceptions of behavioral regulation* rather than observation of behavioral regulation, evidence of inter-rater reliability is not required.



# The psychometrical qualities of the Teacher Classroom Observation (TCO) assessment using data from a study with learning facilitators in Northeast Nigeria

December 10, 2020
Jeongmin Lee, PhD
Senior Research Advisor

INTERNATIONAL RESCUE COMMITTEE

# Overview

- Background of TCO
- Development of TCO
- Content of TCO 3.0
- Findings with TCO 3.0 in Nigeria
- Conclusion
- Next steps

# 1. Background of TCO

- UN's SDG 4 – "Quality education for all children"
- Teachers are the key to achieving this goal

> "Teachers and educators should be empowered, adequately recruited and remunerated, motivated, professionally qualified, and supported within well-resourced, efficient and effectively governed systems." (UNESCO, 2016, p. 15).

# 1. Background of TCO

A robust policy and program interest in teacher development and support systems

Emerging questions on program effectiveness and quality implementation

- “What strategies do teachers use to teach in difficult school environments?”
- “What aspects of teacher performance lead to improved student learning?”
- “What teacher support programs can improve effective teacher behaviors?”

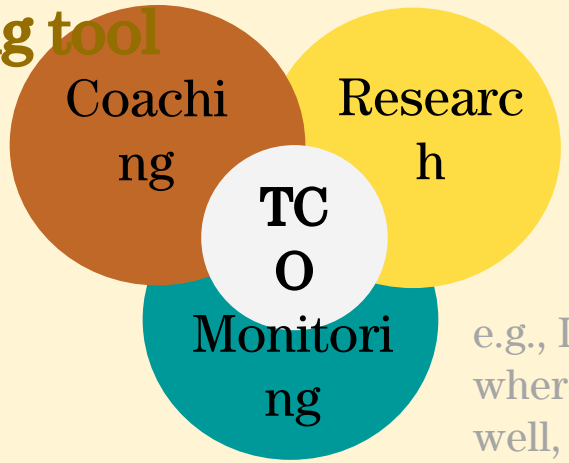
To answer these questions, we need **quality teacher measures,**

- That are reliable, valid and inexpensive to use in conflict settings
- That provide timely, actionable data to teachers and program implementers

# 2. Development of TCO

IRC's longstanding commitment to quality child education through improved teacher instruction and motivation in conflict contexts

**A hybrid evaluation, monitoring & coaching tool**

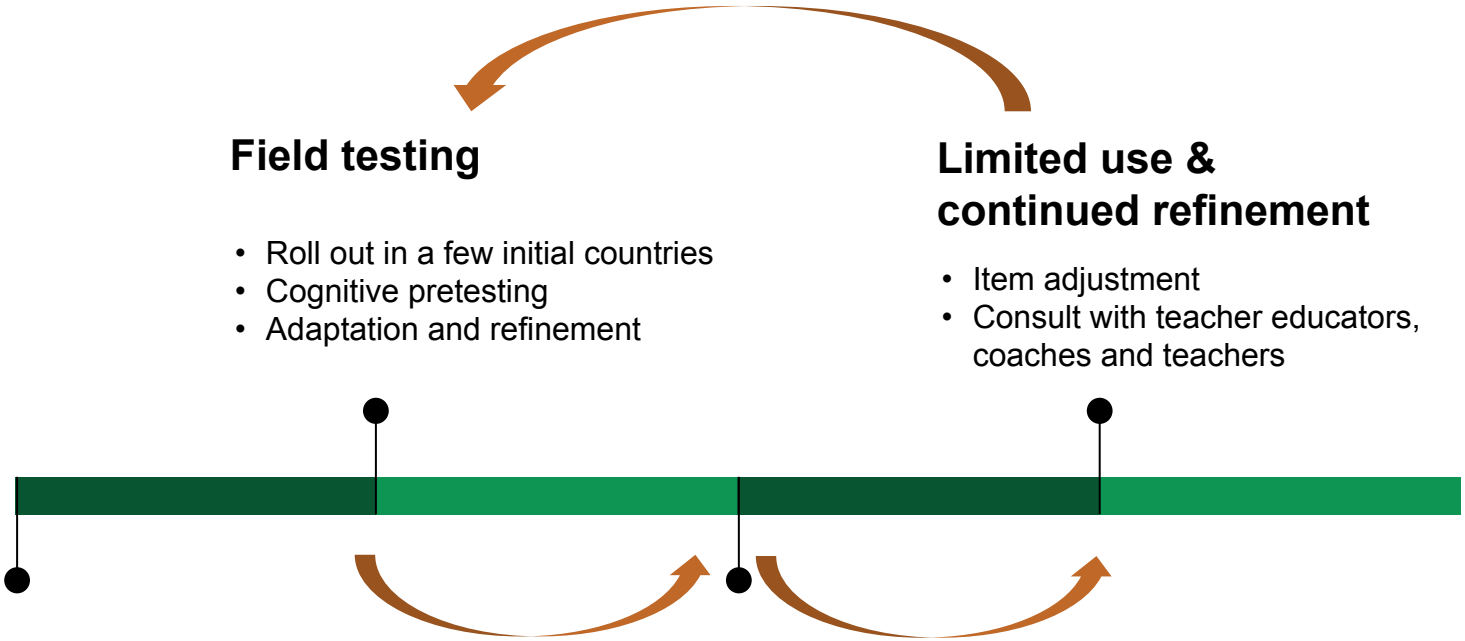



e.g., Identifying and refining coaching topics

e.g.,Testing the effects of a teacher intervention on teacher performance

e.g., Identifying areas where teachers are doing well, and areas where they need further support

# 2. Development of TCO



**Field testing**

- Roll out in a few initial countries
- Cognitive pretesting
- Adaptation and refinement

**Limited use & continued refinement**

- Item adjustment
- Consult with teacher educators, coaches and teachers

**Item development**

- Desk review of teacher literature and existing tools
- Consult with teacher educators, coaches and teachers

**Psychometric testing**

- Item characteristics
- Reliability
- Validity

# 3. TCO version 3.0

| | | | | |
|---|---|---|---|---|
| **Physical environment** | 1 | Environmental scan index | 8 items | 1/0 (presence-absence) |
| **Teacher instruction** | 2 | Time on Task | 1 Item | (Time spent on academic activities)/ (Total instructional time) |
| | 3 | Use of teaching and learning resources | 3 Items | 4-point Likert scale |
| | 4 | Use of teaching methods | 5 Items | 4-point Likert scale |
| | 5 | Promotion of student participation | 5 Items | 4-point Likert scale |
| | 6 | Use of formative assessment and feedback | 2 Items | 4-point Likert scale |

# 4. TCO 3.0 in Nigeria – Context & Data

## A non-formal education program

- Was implemented in 400 communities in Borno and Yobe states
- Accommodated 33,883 out-of-school children (ages 6-16)
- Taught literacy, numeracy and social-emotional skills

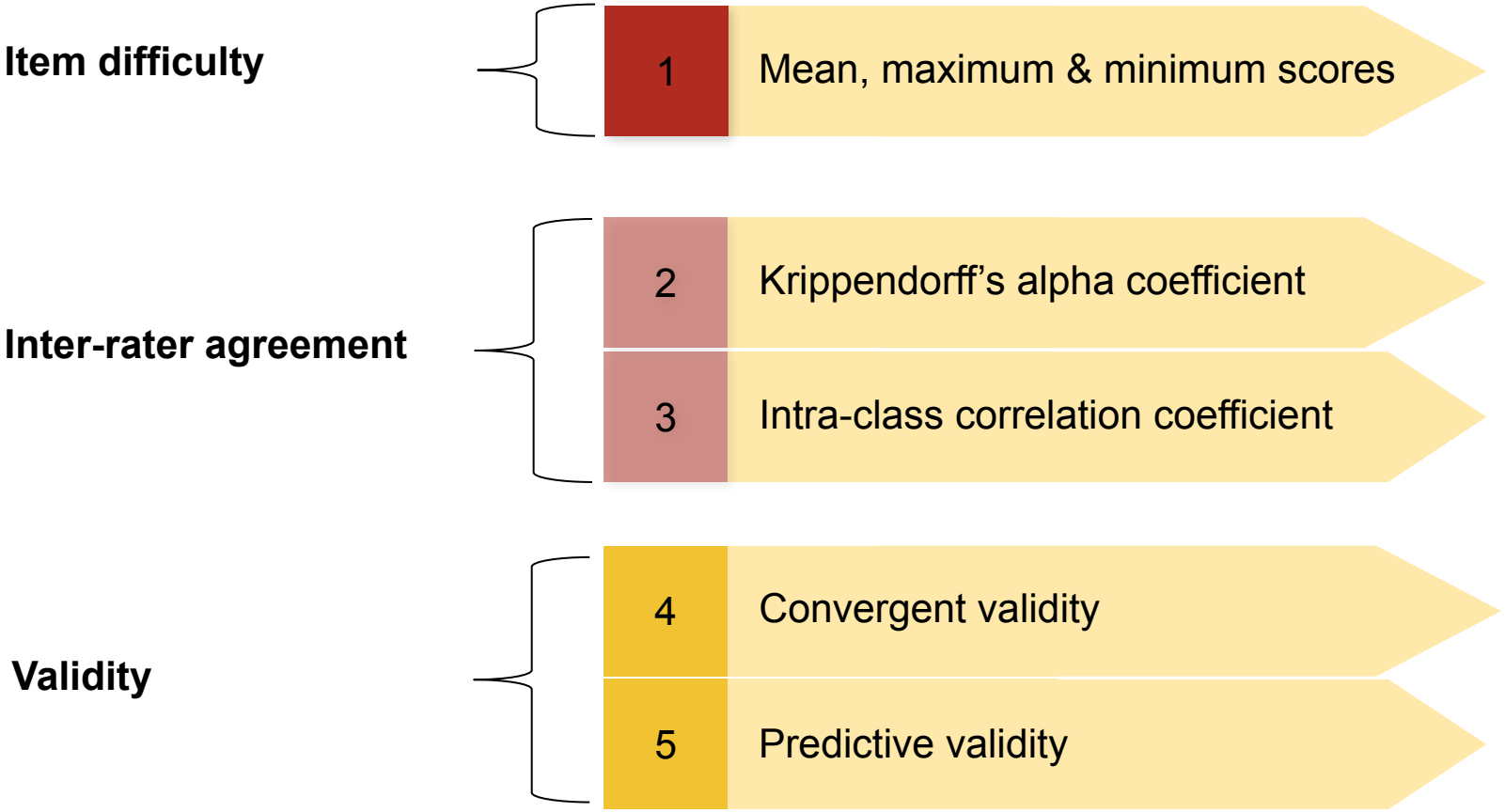
## Learning facilitators, LF

- 400 respected community volunteers
- Had proficiency in reading and math
- Attended four-day face-to-face training
- Attended monthly teacher-learning circles
- Received on-site coaching visits

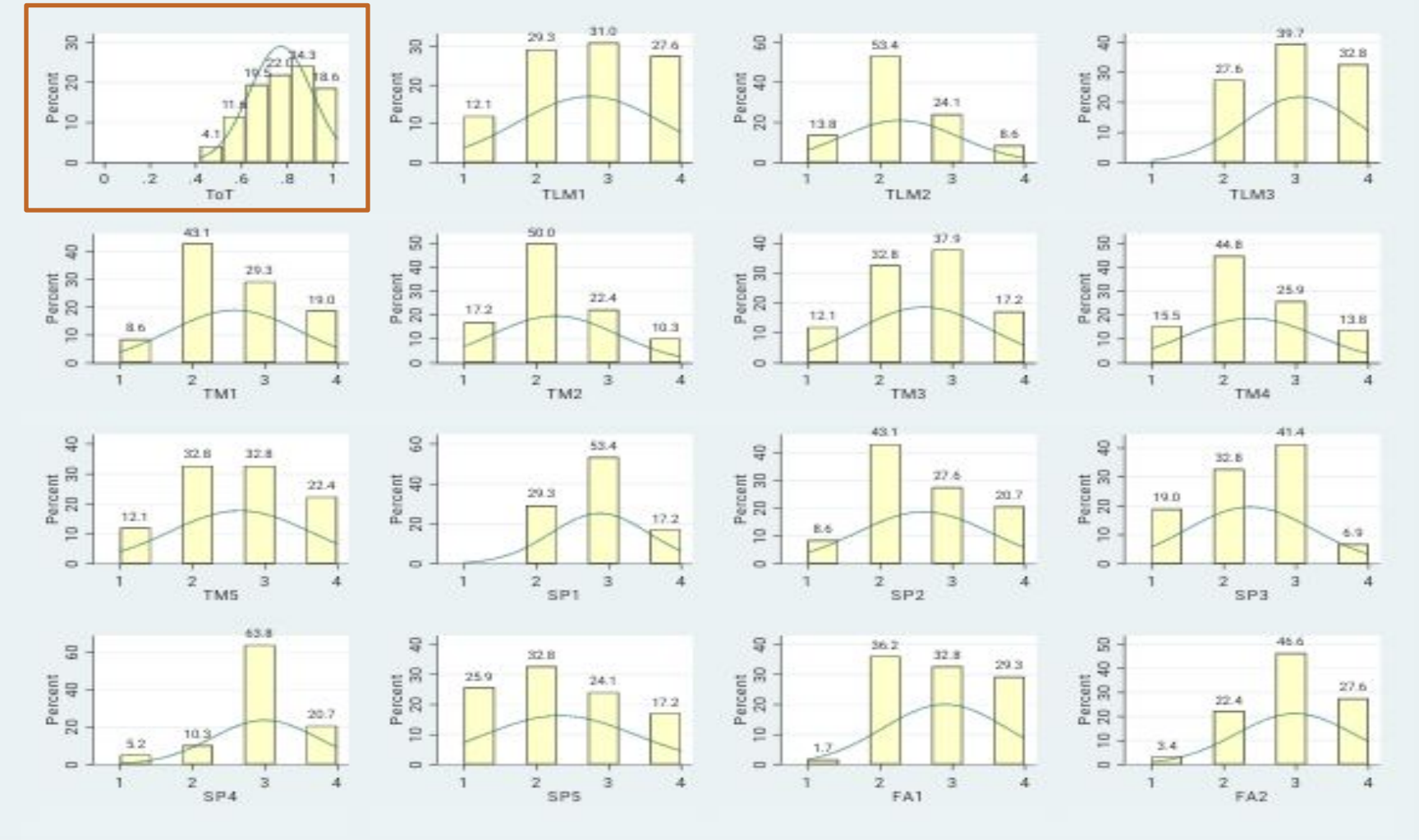
## TCO data

- From a representative sample of 80 LFs (32 male)
- Baseline & endline assessment
- 80 LFs had on average 4.3 years of teaching experience
- Each taught about 36 learners

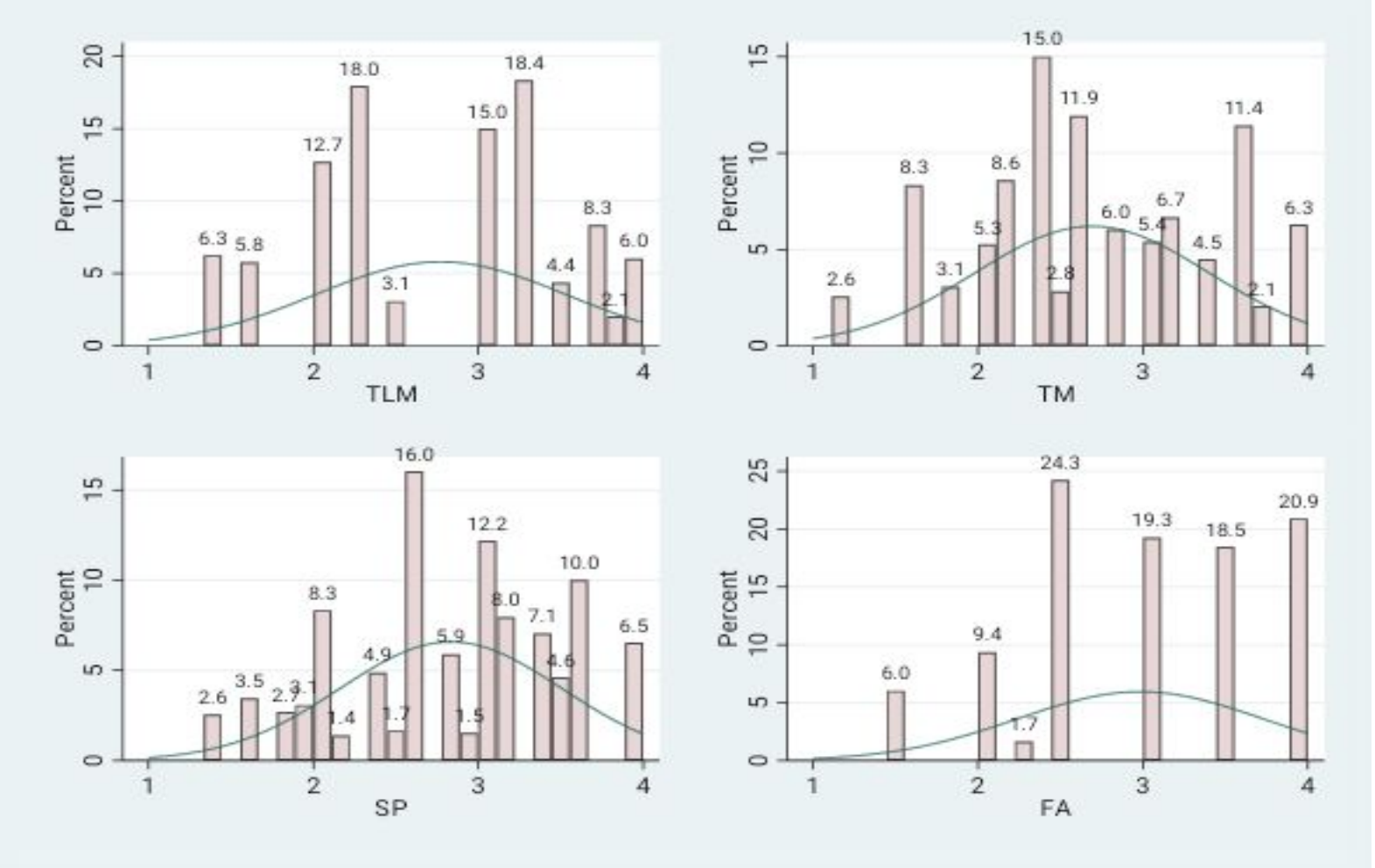
# 4. TCO 3.0 in Nigeria - Analysis



# 4. TCO 3.0 in Nigeria – Item difficulty



# 4. TCO 3.0 in Nigeria – Item difficulty



# 4. TCO 3.0 in Nigeria - Interrater agreement

| Variables | | Krippendorff's Alpha (> 0.67) | Agreement tenable? (Yes = ✓ : No = ✗ ) | ICC ( > 0.75) | Agreement tenable? (Yes = ✓ : No = ✗ ) |
|---|---|---|---|---|---|
| ***Item level*** | | | | | |
| Time on task | | 0.34 | ✗ | 0.25 | ✗ |
| Teaching-learning materials | TLM1 | 0.95 | ✓ | 0.97 | ✓ |
| | TLM2 | 0.73 | ✓ | 0.82 | ✓ |
| | TLM3 | 0.92 | ✓ | 0.96 | ✓ |
| Teaching methods | TM1 | 0.88 | ✓ | 0.93 | ✓ |
| | TM2 | 0.85 | ✓ | 0.91 | ✓ |
| | TM3 | 0.55 | ✗ | 0.73 | ✗ |
| | TM4 | 0.88 | ✓ | 0.94 | ✓ |
| | TM5 | 0.70 | ✓ | 0.83 | ✓ |
| Student participation | SP1 | 0.83 | ✓ | 0.90 | ✓ |
| | SP2 | 0.54 | ✗ | 0.69 | ✗ |
| | SP3 | 0.61 | ✗ | 0.76 | ✓ |
| | SP4 | 0.45 | ✗ | 0.59 | ✗ |
| | SP5 | 0.78 | ✓ | 0.85 | ✓ |
| Formative assessment | FA1 | 0.69 | ✓ | 0.80 | ✓ |
| | FA2 | 0.84 | ✓ | 0.93 | ✓ |
| ***Construct level*** | | | | | |
| Teaching-learning materials | | 0.96 | ✓ | 0.97 | ✓ |
| Teaching methods | | 0.85 | ✓ | 0.92 | ✓ |
| Student participation | | 0.83 | ✓ | 0.86 | ✓ |
| Formative assessment | | 0.80 | ✓ | 0.88 | ✓ |

# 4. TCO 3.0 in Nigeria – Convergent validity

| Variables | Time on task | TLM1 | TLM2 | TLM3 | TM1 | TM2 | TM3 | TM4 | TM5 | SP1 | SP2 | SP3 | SP4 | SP5 | FA1 | FA2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Item level*** | | | | | | | | | | | | | | | | |
| Time on task | 1 | | | | | | | | | | | | | | | |
| TLM1 | -0.16 | 1 | | | | | | | | | | | | | | |
| TLM2 | -0.24 | 0.24 | 1 | | | | | | | | | | | | | |
| TLM3 | -0.33* | 0.26* | 0.28* | 1 | | | | | | | | | | | | |
| TM1 | -0.20 | 0.71* | 0.18 | 0.10 | 1 | | | | | | | | | | | |
| TM2 | -0.25 | 0.36* | 0.54* | 0.41* | 0.43* | 1 | | | | | | | | | | |
| TM3 | -0.03 | 0.34* | 0.47* | 0.49* | 0.39* | 0.74* | 1 | | | | | | | | | |
| TM4 | -0.09 | 0.26* | 0.42* | 0.43* | 0.42* | 0.69* | 0.68* | 1 | | | | | | | | |
| TM5 | -0.26* | 0.55* | 0.32* | 0.58* | 0.54* | 0.44* | 0.47* | 0.36* | 1 | | | | | | | |
| SP1 | -0.23 | 0.62* | 0.31* | 0.54* | 0.55* | 0.56* | 0.57* | 0.61* | 0.55* | 1 | | | | | | |
| SP2 | -0.33* | 0.44* | 0.40* | 0.56* | 0.39* | 0.57* | 0.51* | 0.47* | 0.51* | 0.71* | 1 | | | | | |
| SP3 | -0.06 | 0.14 | 0.40* | 0.38* | 0.41* | 0.40* | 0.46* | 0.57* | 0.33* | 0.52* | 0.40* | 1 | | | | |
| SP4 | 0.10 | 0.43* | -0.14 | 0.21 | 0.21 | 0.19 | 0.23 | 0.05 | 0.20 | 0.39* | 0.21 | 0.00 | 1 | | | |
| SP5 | -0.26* | 0.58* | 0.34* | 0.55* | 0.51* | 0.55* | 0.55* | 0.50* | 0.65* | 0.74* | 0.64* | 0.48* | 0.34* | 1 | | |
| FA1 | 0.00 | 0.46* | 0.42* | 0.50* | 0.35* | 0.60* | 0.64* | 0.50* | 0.42* | 0.58* | 0.46* | 0.40* | 0.39* | 0.56* | 1 | |
| FA2 | -0.10 | 0.53* | 0.32* | 0.50* | 0.52* | 0.38* | 0.56* | 0.46* | 0.64* | 0.76* | 0.56* | 0.50* | 0.33* | 0.67* | 0.63* | 1 |

# 4. TCO 3.0 in Nigeria – Convergent validity

| ***Dimension level*** | **TLM** | **TM** | **SP** | **FA** |
|---|---|---|---|---|
| Teaching-learning materials | 1 | | | |
| Teaching methods | 0.74* | 1 | | |
| Student participation | 0.74* | 0.77* | 1 | |
| Formative assessment | 0.71* | 0.72* | 0.77* | 1 |

# 4. TCO 3.0 in Nigeria – Predictive validity

| Variables | | Literacy skills | | | Numeracy skills | | |
|---|---|---|---|---|---|---|---|
| | | Letter sound | Oral reading fluency | Reading comprehension | Addition | Subtraction | Word problem |
| ***Item level*** | | | | | | | |
| Time on task | | -9.27 | -13.44 | -9.33 | 1.65 | 0.56 | 3.35 |
| Teaching-learning materials | TLR1 | 1.84 | 1.56 | -0.34 | 1.22 | 1.04 | 0.56 |
| | TLR2 | 1.24 | 1.61 | 0.19 | 1.10 | 0.94 | 1.34 |
| | TLR3 | 1.55 | 0.72 | -0.81 | 0.25 | 0.50 | 0.69 |
| Teaching methods | TM1 | 2.98 | 3.09 | 0.84 | 1.73 | 1.63 | 0.21 |
| | TM2 | 0.51 | 2.16 | 1.26 | 0.56 | 0.70 | 1.42 |
| | TM3 | 1.38 | 1.31 | 0.64 | 0.56 | 0.49 | 0.91 |
| | TM4 | 3.05 | 3.66 | 0.92 | 1.65 | 1.51 | -0.00 |
| | TM5 | 0.98 | 1.87 | -0.21 | 1.16 | 0.98 | 2.85 |
| Student participation | SP1 | 2.64 | 1.75 | 0.24 | 0.67 | 0.68 | 1.02 |
| | SP2 | 2.58 | 2.11 | 0.52 | 0.79 | 0.74 | 2.39 |
| | SP3 | 0.03 | 1.86 | -0.00 | -0.07 | -0.01 | 0.14 |
| | SP4 | 1.82 | 3.07 | 1.46 | 0.91 | 0.76 | 0.58 |
| | SP5 | 0.77 | 1.49 | -0.60 | 0.48 | 0.60 | 2.60 |
| Formative assessment | FA1 | 0.16 | 0.17 | -0.64 | 0.16 | 0.36 | -0.08 |
| | FA2 | 2.79 | 2.65 | 0.90 | 1.28 | 1.06 | 0.99 |
| ***Dimension level*** | | | | | | | |
| Teaching-learning materials | | 2.41 | 2.10 | -0.43 | 1.43 | 1.34 | 1.33 |
| Teaching methods | | 2.49 | 3.46 | 1.03 | 1.57 | 1.48 | 1.54 |
| Student participation | | 2.45 | 3.02 | 0.34 | 0.85 | 0.86 | 2.44 |
| Formative assessment | | 1.86 | 1.78 | 0.22 | 0.90 | 0.88 | 0.58 |

# 5. Conclusion

## Summary

- Item difficulties ranged from easy to difficult
- Evidence of interrater agreement was incomplete
- Evidence of convergent validity was incomplete
- Evidence of predictive validity was incomplete

## Limitations

- Has not had an opportunity to test construct validity
  - Had a small sample of LFs
- Measurement errors inherent due to environmental conditions
  - Unavailability of repeated observations and measures
  - Scores based on just one 45-minute classroom observation
  - Environmental conditions and LF behaviors might have differed on other days
- Limited number of covariates used for predictive validity

# 6. Next steps

- Reviewing item content or scoring rubrics
  - Too easy or too difficult to demonstrate, or observe?
  - Conceptual clarity and coherence – double barreled? non-unidimensional?
  - Language translation?
  - Length of items?

- Reviewing the quality/process of enumerator training and data collection
  - Training – duration, level of understanding, practice opportunities etc.
  - Mode/process of data collection – tablet, pen & paper, video-recording etc.

- Construct validity besides other qualities
  - Testing at a larger scale – Tanzania

# Thank you

For more information about the TCO, please contact
Jeongmin Lee at Jeongmin.lee@rescue.org or
Jonah Bautista at Jonah.Bautista@rescue.org

# Thank You!

**www.inee.org/measurement-library**

*measurement.library@inee.org*