

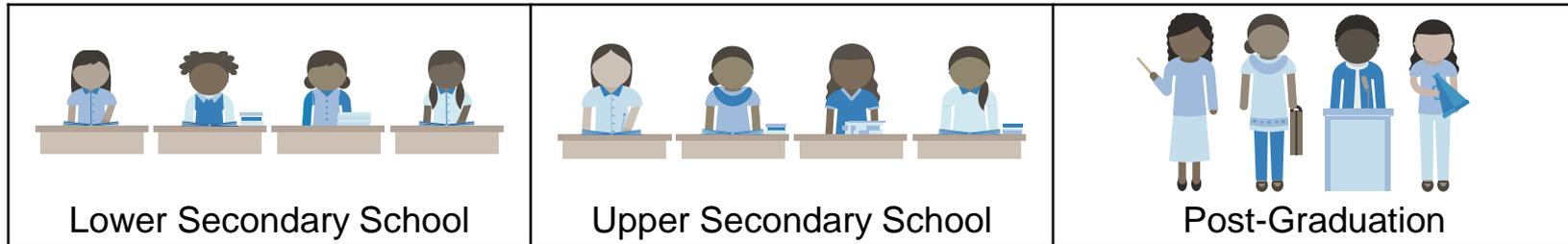


Room to Read®

Christine Beggs
Ryan Hebert



Moving from theory to practice: A framework for evaluating the impact of life skills education on life outcomes



Are Girls Gaining the Skills They Need?

Life Skills Assessment at Key Intervals & Evaluations

- Self-confidence
- Creative Problem Solving
- Perseverance
- Expressing & Managing Emotions
- Empathy
- Self-Control
- Decision-Making
- Relationship Building
- Communication
- Critical Thinking

Complements **Project Implementation Monitoring, Risk and Response** System and **Alumnae Surveys**.

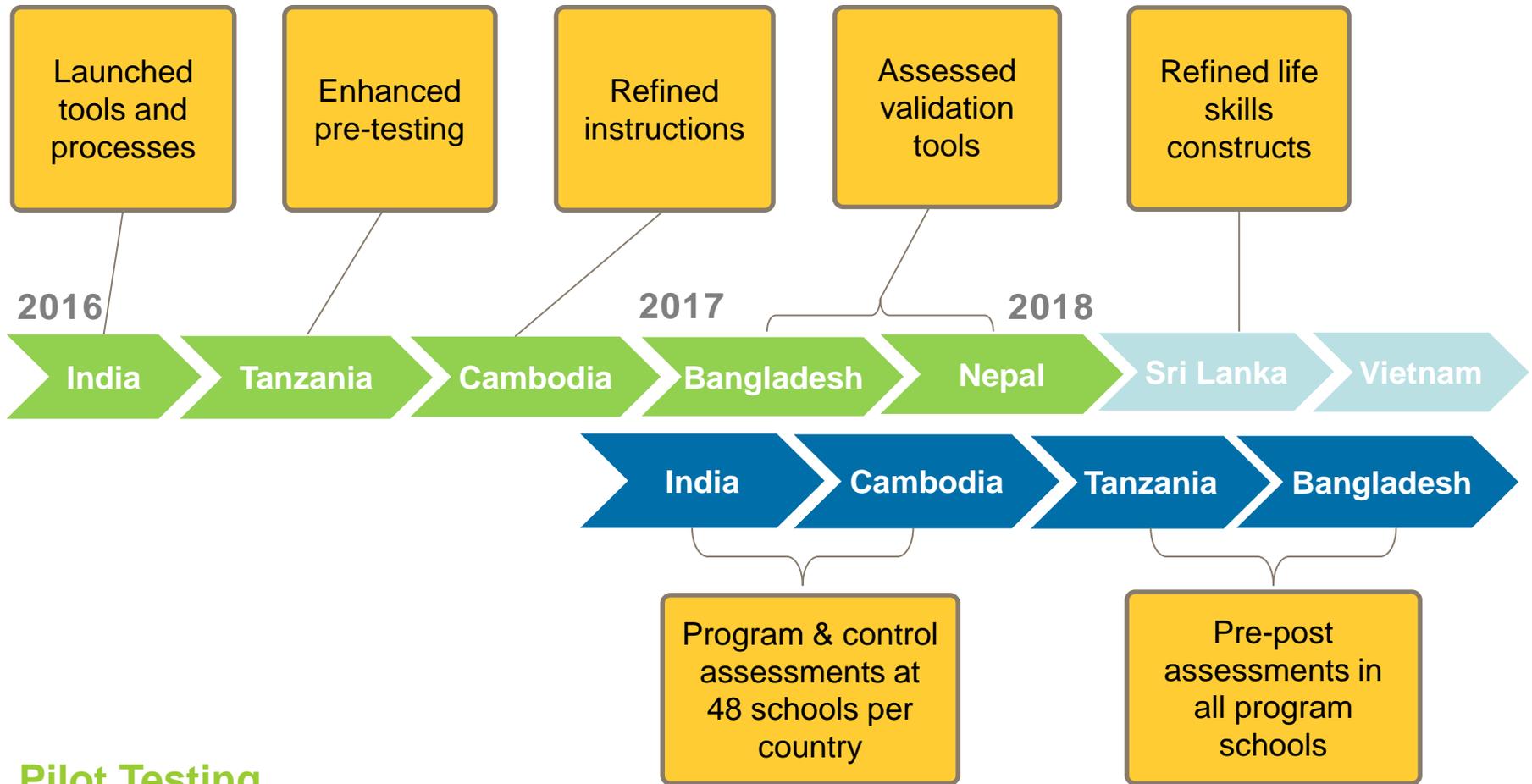
Parameters for Assessment

- Quantitative
- Comprehensive, aligned with curriculum & skill building
- Valid and reliable
- Designed to preclude “teaching to the test”
- Sensitive to changes in life skills over time
- Can be implemented across countries with adaptation
- Streamlined, not resource-intensive
- Appropriate for grades 6-12

Reference Scales and Measures for Assessment

- Positive Youth Development Survey
- AIR & CASEL, 2013
- Child Trends Flourishing Indicators Empathy Scale, Moore & Lippman 2011
- GRIT Scale, Duckworth & Quinn, 2009
- Jamaica Youth Survey, Meeks Gardner et al. 2011
- *Newly-developed measures*

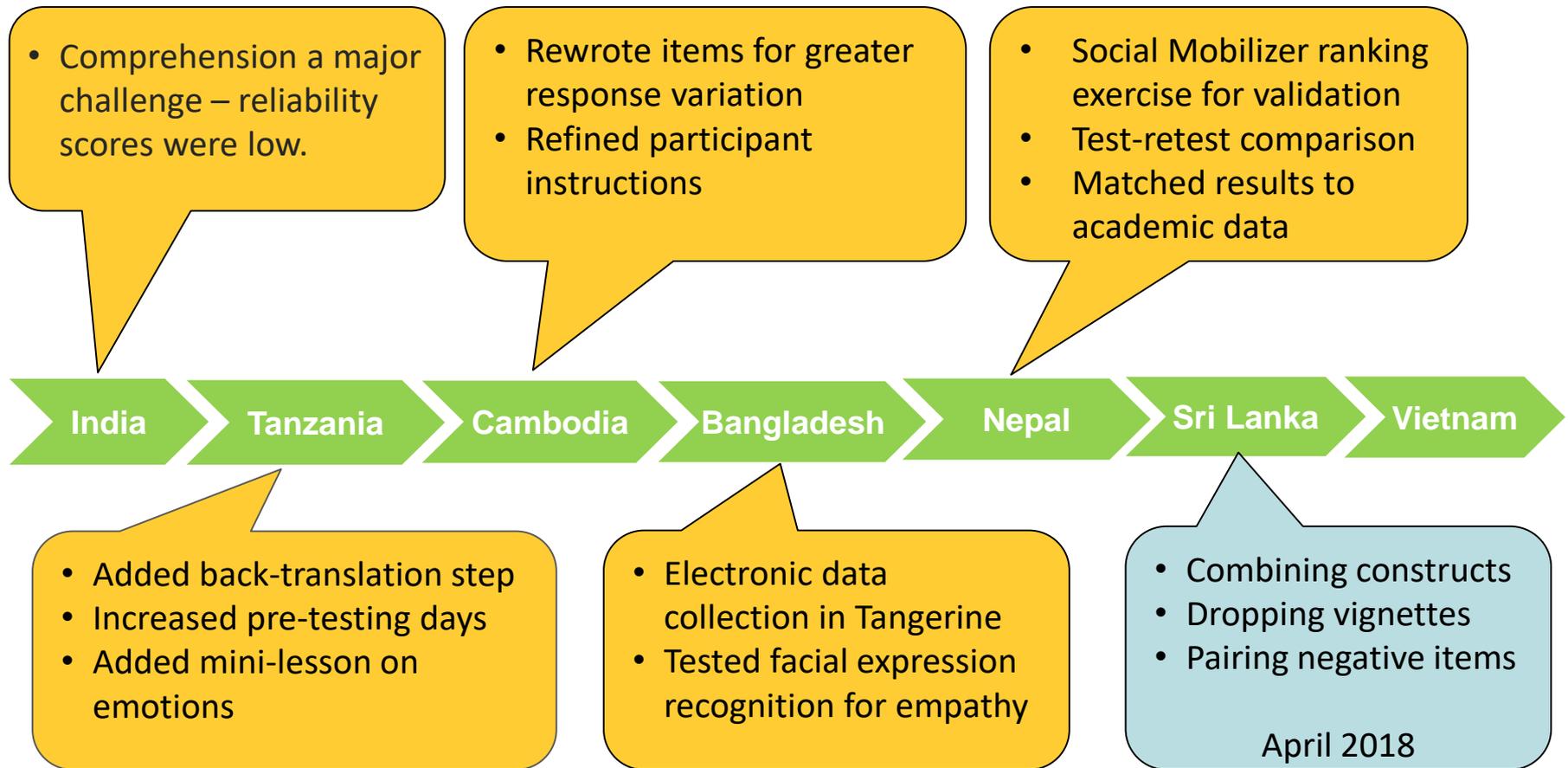
Assessment Development Process & Timeline



Pilot Testing

Baseline Assessments

Assessment Changes during piloting



Tool Contextualization

- Team Review and comments
- Translation, Back Translation and correction
- Pre-testing (face to face interview with probes)
- Country team debrief and decisions on vocabulary use/sentences simplification

Typical Tool Vulnerabilities

- Reliability, internal consistency for constructs
- Biases (reference, desirability)
- Validity
- Embedded norms, value-laden

Questions to Discuss for Each Life Skill Construct

- What does this life skill mean?
- How does this skill represent itself in the country context?
- How does the tool seek to measure this?
- Does this measurement strategy make sense? Are there critical elements of this skill that are being missed?

Questions to Discuss for Each Question Category or Activity

- Will girls understand the instructions the enumerator provides to them?
- After hearing these instructions, will the questions/activity make sense to them?
- What further explanation might be needed to make sure girls understand?

Questions to Discuss for Each Question

- How does this relate to the life skill it is intended to measure? Is it a good measure of this skill?
- Is it relevant? Does it reflect the way this skill represents itself in this country?
- Is there any cultural sensitivity that needs to be taken into account in this question?
- Can all terms in the question be easily translated into the local language without changing their meaning?
- Once translated, will girls easily understand these terms? If not, are there synonyms that can be used for those girls who need more explanation?

Item Changes after Pre-test/Pilot

Original Measure	Current Tool	Notes
New ideas and projects sometimes distract me from previous ones	(Dropped)	Low contextual relevance
Setbacks (delays and obstacles) don't discourage me. I bounce back from disappointments faster than most people	Big difficulties do not discourage me. I try again.	Overly complex phrasing
I have been obsessed with a certain idea or project for a short time but later lost interest	(Dropped)	Low factor loading
I am a hard worker	When I do something, I work hard at it	Changed from adjective to verb for simpler translation
I have difficulty maintaining (keeping) my focus on projects that take more than a few months to complete	(Dropped)	Low contextual relevance

Increasing Understanding for better reliability

For example, if I say “I like the color blue”. You should think about yourself and how this sentence applies to you. If you think it has nothing to do with you (in our example, if you don’t like blue), you should select “not true”. If you think this sentence has some element of truth for you (you find blue ok but you prefer yellow and pink a lot more), you should select “slightly true”...

PRACTICE QUESTIONS

	Statement	Not True	Slightly True	Neutral	Mostly True	Completely True	No Response
	I like ice cream.	1	2	3	4	5	
	I like spending time with my friends.	1	2	3	4	5	
	I know how to speak English.	1	2	3	4	5	
	I like waking up before five in the morning.	1	2	3	4	5	
	I find it difficult to do my homework alone	1	2	3	4	5	

Reliability as a function of grade/age, comprehension in play?

<u>Cronbach's alpha</u>	<u>Nepal Pilot</u>		
Indicator	grade 6	grade 7	grade 8
Self Confidence	0.48	0.48	0.66
Self-confidence (without negatively worded items)	0.5	0.65	0.67
Expressing and managing emotion	0.7	0.65	0.73
Empathy	0.58	0.6	0.77
Self-Control	0.58	0.68	0.66
Self-control (without negatively worded items)	0.55	0.56	0.74
Decision-Making	0.42	0.38	0.64
Perseverance	0.65	0.65	0.73
Communication	0.46	0.7	0.75
Relationship building with friends	0.66	0.7	0.78
Relationship building with adults	0.68	0.69	0.79
Relationship building 3	0.53	0.68	0.58
Average on likert items	0.57	0.62	0.71

Language & Concepts: **Concept of emotions in Tanzania** a challenge. Added an introductory module where respondents are asked to cite emotions they had over the past week, then reviewed which were emotions before survey.

Negatively worded items: a **problem everywhere** but Tanzania.

Enumerator rating for validation – expectations of behavior and interpretation of what those behaviors mean in terms of skills
- varies greatly by context.

Testing for Internal Consistency/Reliability

Alpha range from 0 to 1. the higher the alpha, the more likely the items measure the same underlying concept. Alpha between 0.65 and 0.85 are generally acceptable. Higher may mean we have redundant items, lower means we aren't measuring something consistently.

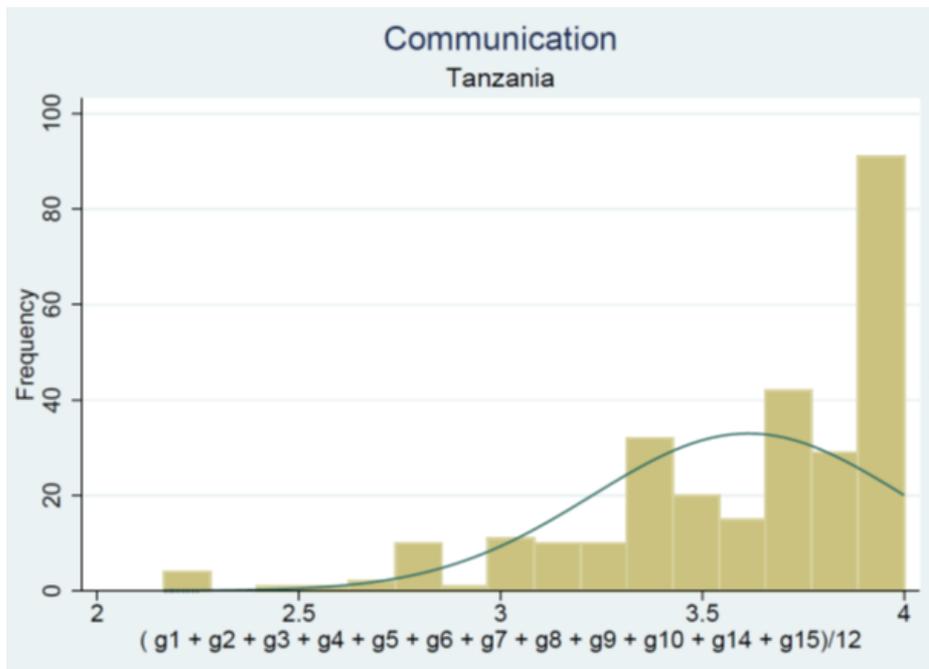
Indicator	Pilots					Baseline		Average (Tz-KH BL)
	India	Tanzania	Cambodia	Bangladesh	Nepal	India	Cambodia	
Self Confidence	0.5	0.69	0.59	0.7	0.54	0.5	0.66	0.60
Self-confidence (without negatively worded items)		0.68	0.72	0.73	0.59	0.62	0.67	0.67
Expressing and managing emotions	0.41	0.7	0.72	0.67	0.69	0.63	0.62	0.67
Empathy	0.44	0.65	0.66	0.72	0.64	0.6	0.66	0.66
Self-Control	0.4	0.73	0.67	0.62	0.54	0.54	0.72	0.64
Self-control (without negatively worded items)		0.67	0.69	0.69	0.68	0.57	0.69	0.67
Decision-Making	0.62	0.62	0.57	0.67	0.45	0.51	0.65	0.58
Perseverance	0.51	0.69	0.69	0.63	0.66	0.61	0.64	0.65
Communication	0.49	0.79	0.72	0.72	0.64	0.56	0.65	0.68
Relationship building with friends	0.65	0.81	0.79	0.83	0.7	0.76	0.74	0.77
Relationship building with adults		0.87	0.77	0.8	0.72	0.76	0.79	0.79
Relationship building 3	0.48	0.65		0.62	0.57			0.61
Average reliability on Likert items	0.50	0.71	0.68	0.70	0.61	0.59	0.67	0.66
Number of solutions given in problem solving exercises	0.67	0.82	0.69	0.52	0.54	0.74	0.49	0.63

Communication Skill Construct

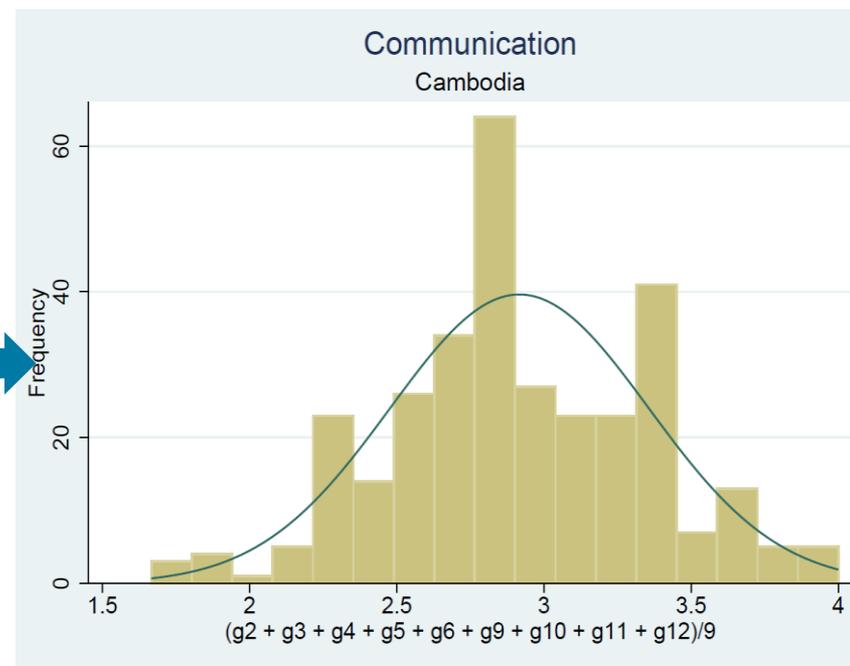
	Not True	Slightly True	Neutral	Mostly True	Completely True	No Response
I ask my teachers for help when I need it.	1	2	3	4	5	
If I get angry with someone, I can talk to them and make things better.	1	2	3	4	5	
I can work well with others.	1	2	3	4	5	
It is easy for me to join a conversation that others have already started.	1	2	3	4	5	
I get along well with my classmates.	1	2	3	4	5	
I feel confident enough to talk in front of a group.	1	2	3	4	5	
I ask questions when I do not understand what someone is telling me.	1	2	3	4	5	
I know how to get support/help if I have a problem.	1	2	3	4	5	

- **Understanding** of the scales and **response pattern in Likert scales** : people aren't familiar with rating scales, some do not understand it, some never use extreme options while other always do.
- **Desirability bias**: girls want to **please the enumerator** and make sure they give a good picture of themselves. Girls are afraid they are being judged, they fear the consequences of their interview. **Parents exhibit bias also.**
- **Reference bias** : two similar persons having different references with regards to the construct measured, and therefore rating themselves differently. **Anchoring vignettes** theoretically address this.

Reducing Desirability Bias



TANZANIA



CAMBODIA

- Clear, reassuring instructions
- Pre-testing & re-writing
- “Forced choice” questions

Reducing Desirability Bias: Forced Choice Questions

I make sure that I finish whatever I start to do before starting something new

When I see someone being picked on, I feel sorry for him or her.

I can do things as well as most other girls my age. (in and out of school)

I have a plan to achieve my goals in life

Most like me

|_____|

Least like me

|_____|

Assessing Validity – Three Approaches

Related Measures: Measure taken with the same respondent, at the same moment, measuring a similar or close construct conceptually

- Strengths and Difficulties Questionnaire
- Task based measures: mirror drawing exercise, emotions listing exercise, face test

Academic Data: Does academic performance map to skills as we might expect (e.g., perseverance)?

Third party: To triangulate girls' answer with one or several other respondents' opinion

- Teacher
- Social Mobilizer
- Parents
- Enumerator
- Peer nomination

Creative Problem Solving – Free Listing Exercise

Indira and Gita are sisters. They have one mango, and they both want it, so they are starting to fight about the mango. What can the mother of the girls do to help solve this problem? Think about every possible thing she can do. (Probe: good, what else can she do?)

Chetana and Asha are talking in class. This prevents Namrata from listening to the teacher. What can Namrata do? Think about every possible thing she can do. (Probe: good, what else can she do?)

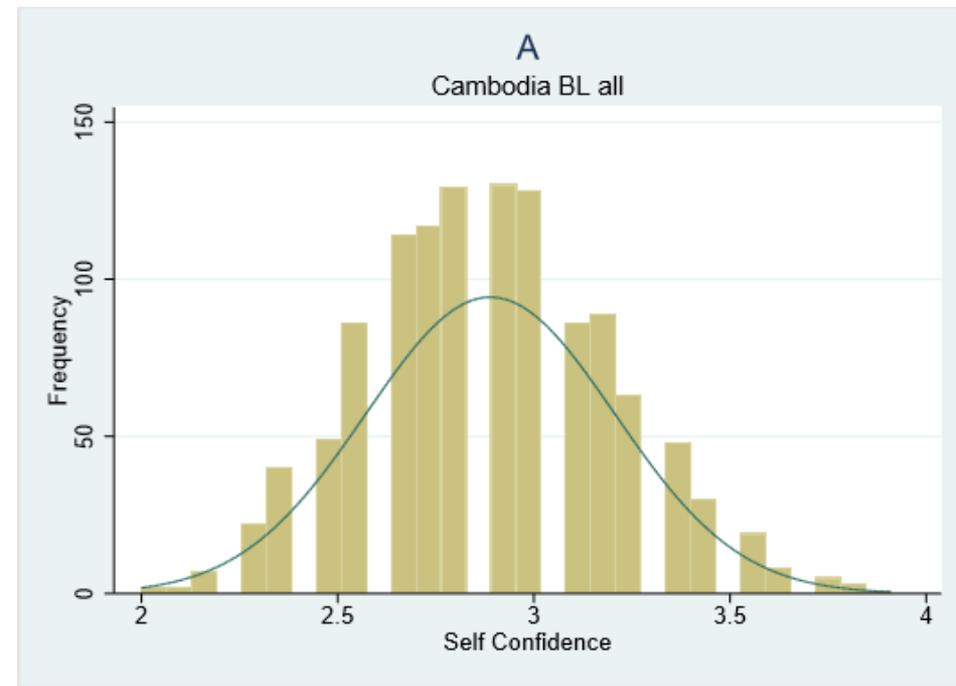
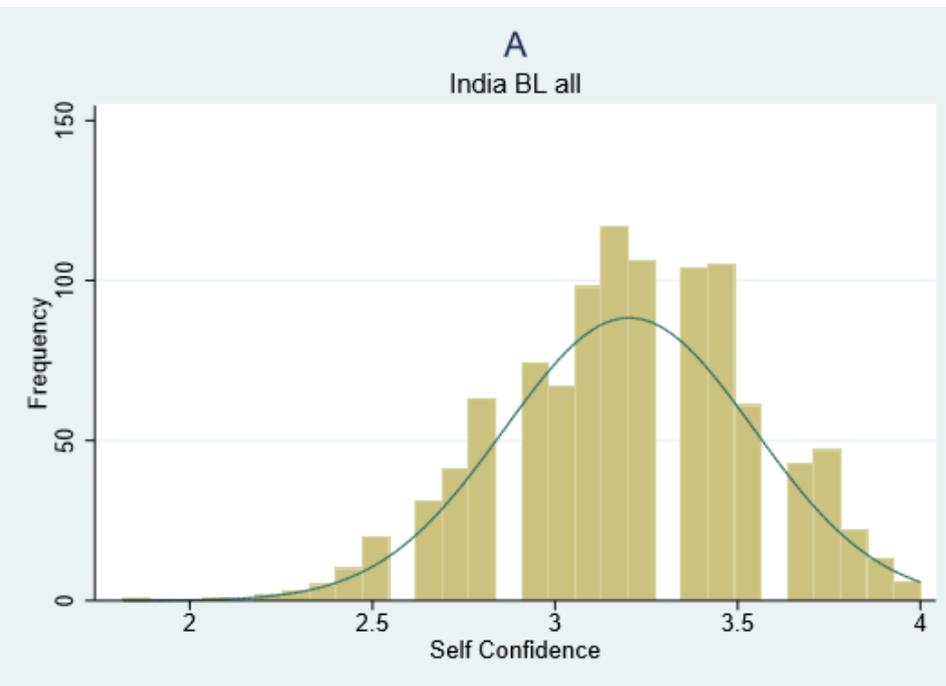
Shanti comes back home after receiving a bad grade on an exam. She is afraid her father will be angry about it. What can she do? Think about every possible thing she can do. (Probe: good, what else can she do?)

Does this measure creative problem solving, conflict resolution, decision making....some of all?

Distributions for “Self-Confidence”

SELF-CONFIDENCE: INDIA BASELINE

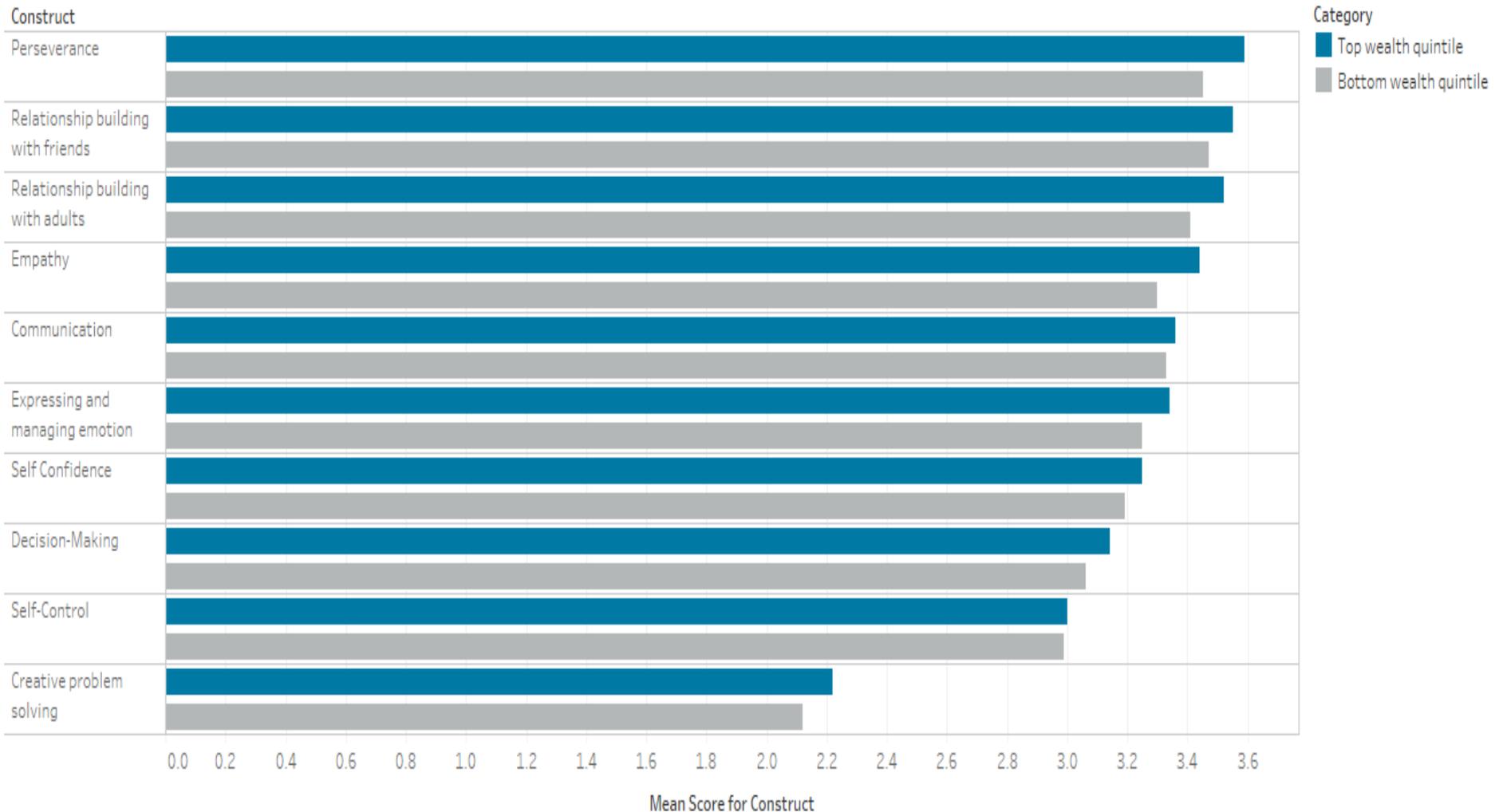
SELF-CONFIDENCE: CAMBODIA BASELINE



India Baseline: Mean Scores by Wealth

Wealth does not seem to be strong predictor of life skills score

Score Breakdown by Household Wealth - India



Still exploring and evolving the assessment:

- Measuring skills vs. **inherent traits**
- Re-thinking our **reference measures for validity testing** - more aligned with measurement range and population characteristics.
- Negatively worded items, adding paired items.
- Desirability bias – create “**social desirability score**” for each respondent and adjust accordingly.
- New item structures such as **leveled items** (maturation and evolving learning outcomes) or **situational judgement test** items – version 2.0.

A close-up photograph of four young women of diverse ethnicities, smiling and looking towards the camera. They are wearing colorful clothing: a purple and white striped shirt, a blue and yellow patterned top, a red top, and a green top. The woman on the far right is holding a small object, possibly a pen or pencil, near her mouth. The background is slightly blurred, suggesting an outdoor setting.

Thank you!

Christine.Beggs@roomtoread.org

Ryan.Hebert@roomtoread.org