

Possible ways to think about cog. and non-cog. skills measurement

Four slides only

Comments by Luis Crouch at symposium on “Improving the Availability and Quality of Individual-level Data on Women and Youth in Living Standards Measurement Study (LSMS) surveys”

IFAD, Rome, 28-29 Nov 2018

Cognitive skills options

- Self-report, binary (trad. census)
 - Once sentence, non-binary (DHS)
 - Sentence processing (timed or un-timed)
 - Citizen-led assessments class
 - MICS cognitive module - UNICEF
 - Mini-LAMP (UNESCO Institute of Stats)
 - EGRA/EGMA class
 - STEP
 - TIMSS/PIRLS/PISA/PIAAC,
regional assessments (e.g. SACMEQ)
- } Cheap but inaccurate?
- } Sweet spot??
- } Most accurate, but too costly for HH surveys?

Non-cognitive skills

- “You get hired for your hard skills, and fired for your lack of soft ones”
- All true skills are cognitive at some level, “soft” is wrong, “21st Century” is just stupid
 - Therefore: employability or life skills a better term?
- Maybe 1/500th or 1/1000th the research as for cog.
 - Therefore no options spectrum, just things to consider
- Criteria for items/skills to measure
 - Malleable/teachable, predictive (of variety of work outcomes), measurable, quantity/quality of research behind it as per meta-research, contextually invariant (culture, gender)
 - Best one I am familiar with done for USAID by Child Trends
- Results:
 - Higher-order thinking
 - Communication
 - Sociability
 - Self-control
 - Positive self-image
- No way to measure all that in HH survey

Extra considerations

- Methodological
 - “Technique”: Tablets, SMS, gamification, wearable sensors
 - But just as important and much less common to think about: what are the “hooks” to implementation and usage in program evaluation and program management?
 - Along lines of what Gero was saying: what can make the indicator change?
 - Is there a useful “operational” dual to the construct/item/question? What is it?
 - “Hooks” to monitoring
 - This is a very under-researched area

Example of decision matrix

Item or whole tool or parts of tools	Optimality criteria						
	Cost	Usefulness of construct	Accuracy	Malleability of the construct	Hooks to implementation/program eval.	“Political” value	Others, etc.
Measure executive function with self-reports	Low	High	Low	Ok	Ok	Very high	
Using direct measurement	Very high	High	Much higher	Ok	Ok	Very high	

Notes

- Most of the cog. assessments are based on work with primary school children but most generalize to youth/adults, some are intended for youth or partially so (LAMP, PIAAC, MICS skills module)
- Some have filter questions or are designed as a set of filters, such as CLA-class; this can shorten time vastly and results automatically in useful classifications
- Some of the cognitive assessments were designed for HH surveys (MICS, STEP and the CLA-class), take advantage of that? Others assume a school (and children within) as the sampling unit.
- For some assessments the variance is very low so sample sizes can be quite small (justifying use in only a sub-sample?) and yet you get a good confidence intervals.
- Many of these have had their “concurrent validity” tested against each other, often with good results (e.g., EGRA-class and CLA-class, EGRA and MICS learning module) which boosts confidence in both
- Some don’t yield single-valued, numerical, continuous metric; some do
- Timed assessments have some inherent value (e.g., EGRA-class) because fluency → efficiency in cognitive processing
- The cognitive ones are mostly not experimental; they are ready to go; little methodological research needed; perhaps adaptation research, or research on mixing features to produce a “best” one for quick use in HH surveys (MICS learning module comes the closest?)
- Some are quite “authentic” and life and labor-market oriented (mini-LAMP best at this?)
- Big differences in time cost. MICS about 15-20 mins, the “big” ones like PISA take 2-3 hours
- “Sweet spot” is not the high point on a parabola, it is simply the most likely best compromise, because the tradeoff is pretty much linear except at the extremes
- Some designed for individual testing; others must be done in groups to be economical
- Non-cognitive have 1/500th or 1/1000th the research behind them; not ready to go; need methodological work, especially for short tests and self-report; maybe use vignettes