

What do we know about (data quality of) skill measures in large household surveys?

Karen Macours

Paris School of Economics and INRA

Based on joint work with Rachid Laajaj, Universidad de Los Andes

Motivation

- Cognitive, noncognitive and technical skills believed to be important
- (1) ideas for improvements at the margin to existing questionnaires, to make sure we harmonize and employ good practices when we know what these are
- (2) ideas to run a few methodological experiments to understand and validate new methods or methods that have maybe been used in specialized surveys on skills, but that will need to be adapted for inclusion in a different type of survey

=> take some stuff of the wish list?

Challenges for skill testing in household surveys

- Many **scales** initially **designed for developed country** settings, lab settings, etc.
- Hard to guarantee **standardized application** of tests
- Lower **education** and **abstract** concepts
- **Translation**
- **Openness** of adult respondents to **test-taking**
- **Expectations** about **why** survey is done may affect answers
- Use of **enumerators** instead of self-administered surveys
- Measures that are **time** consuming

This presentation

- What is the **reliability and validity of skill measures** from **household surveys in developing countries**?
- What can be done to **improve** existing measures?
 - ⇒ Evidence of survey randomized survey experiment in **rural Kenya and Colombia**
 - ⇒ Zoom in (out) on personality traits: **23** low and middle income countries

Context and Data Collection

- Context:
 - 918 farmers in 96 villages in Siaya - Western Kenya
 - Mainly maize & other annual crops. Most have livestock
 - About 50/50 men-women, 6 years of education
 - Language: 97% Luo and 3% Swahili
- Implemented:
 - Review of instruments & work with local agronomists
 - Extensive piloting & 2 week training given by PIs
 - Test of Skills Measurement & Retest after 3 weeks
 - Random order of sections, of questions, of answers, of villages, enumerators
 - 4 follow-ups on agricultural practices and production
- Replication in Colombia Sucre 804 hh
 - In Spanish, about 93% males and 4.5 years of education

Which skills do we measure?

- **Cognitive skills (~ IQ)**

- Memory (Forward and backward Digit Span), problem solving (Raven matrices)
- “Class room” skills: Reading and math (achievement)

- **Non-cognitive skills (Socio-emotional skills)**

- Locus of control, self-esteem, Self-control, perseverance, Big Five, aspirations, CESD (depression)

- **Technical skills**

- Knowledge/Know-how. We worked with agronomists on targeted questions for main crops and practices

Cognitive skills: Test-retest and Internal Reliability

Indicator	Test-retest	Cronbach's Alpha
All Cog	0.83	0.82
Raven	0.63	0.88
Numeracy Q.	0.60	0.70
Math sheet	0.69	0.99
Reading	0.82	0.92
Digit Span	0.52	0.47

Format Non-Cognitive skills questions

- Traditional : 1-5 scale with statements about one-self
 - “On a scale from 1 to 5 - with 1 indicating you strongly disagree and 5 indicating you strongly agree : “You see yourself as someone who tends to be very talkative”
 - Big Five (BFI) but also standardized subscales for lower order constructs (locus-of-control, Rosenberg self-esteem,...)
 - 1-5 scales about causes of poverty
- CESD : E.g. “In the last 7 days, how many days did you have crying spells?”

The Big Five Personality traits

Trait	Description
O penness	Curious, original, intellectual, creative, and open to new ideas.
C onscientiousness	Organized, systematic, punctual, achievement oriented, and dependable.
E xtraversion	Outgoing, talkative, sociable, and enjoys being in social situations.
A greeableness	Affable, tolerant, sensitive, trusting, kind, and warm.
N euroticism	Anxious, irritable, temperamental, and moody.

Non-cognitive Reliability Test

Indicator	Test-retest	Cronbach's Alpha
All Non-Cog	0.53	0.75
Locus of Control	0.49	0.56
Self esteem	0.32	0.28
Causes of Pov	0.40	0.82
Attit. Change	0.37	0.37
Org-Tenac-Contr.	0.26	0.42
BF_ Extrav.	0.23	0.33
BF_Agree	0.25	0.39
BF_Conscious	0.33	0.51
BF_Neurotic	0.26	0.31
BF_Open	0.15	0.37
CESD	0.41	0.82

Towards less Naïve Measures of the Skills

We apply some corrections used in psychometrics:

- **Factorial Analysis** to re-group and re-weight questions
- **Correct Acquiescence Bias** in Non-cog questions
tendency to say yes, even to contradictory questions
 - Calculate **acquiescence score** : averaging between the mean of the positively-coded items and the mean of reverse-coded items
 - Subtracted from all answers
- **Item Response Theory** for cognitive and technical tests

Factorial analysis

- Factorial analysis cognitive and technical **points to 1 cognitive and 1 technical factor**
- Factorial analysis of naïve NonCog gives worrying results:
 - **Pools Items in non coherent groups** (except CESD).
 - 1st Factor is the Acquiescence Bias (if not corrected)
- Factorial analysis of only BF measures
 - Don't coincide at all with Big Five!

Statistics of improved measures

- IRT for cognitive and technical
- Factors after correcting AB for non-cognitive

	Test-retest correlation		Nb of Indexes
	Naïve	Improved	
Cognitive	0.83	0.86	1
Non-Cog	0.53	0.70	6
Technical	0.30	0.41	1

Test-retest demeaned non-cog factors vary between .28 and .53

Predictions of Productivity and Agricultural Decisions

- Clearly **only correlations**, no causality! (measurement exercise)
- Correlated for **3 possible reasons** (all useful):
 - Skills affect agriculture (decisions or productivity)
 - Agricultural experience affects skills
 - Omitted variables

We also run regressions with **large set of controls** (including education, assets) to test additional predictive power ? Do we capture “unobservables”?

Predicting Agricultural Productivity

Table 4: Regressions of the average rank of maize yield across seasons on skills

VARIABLES	SKILLS CONSTUCTS USED AS REGRESSORS:				
	Naïve Score	Improved Index	Mean Naïve Score	Mean improved Index	Mean improved Index
Cognitive skills	2.87*** (0.753)	2.27*** (0.736)	2.25*** (0.820)	1.44* (0.803)	3.04*** (1.143)
Noncognitive skills	4.03*** (0.723)	3.90*** (0.701)	4.85*** (0.858)	4.42*** (0.846)	4.27*** (0.900)
Technical skills	3.37*** (0.828)	4.40*** (0.829)	5.27*** (1.026)	6.54*** (0.937)	3.41*** (1.042)
Observations	903	893	907	906	906
R-squared	0.121	0.145	0.143	0.171	0.443
Controls	No	No	No	No	Yes ¹⁵

Enumerators Matter

Cognitive skills matter

Table 8: Test-retest correlations, Cronbach's alpha and influence of enumerators by subgroups

Sample split:	Test-retest correlation				Cronbach's Alpha		R2 of Enum FE*		
	Enumerator assigned for test and		By Cognitive skill		By Cognitive skill		By Cognitive skill		
	Same	Different	Below median	Above median	Below median	Above median	All	Below median	Above median
Cognitive	0.89	0.84	0.67	0.68	0.54	0.48	0.04	0.05	0.08
Non-cognitive	0.60	0.55	0.50	0.46	0.63	0.67	0.09	0.16	0.11
Technical	0.50	0.38	0.25	0.44	0.46	0.55	0.07	0.15	0.12

Note: R2 of enumerator FE is the R2 of a regressions of the improved construct on (randomly) assigned enumerator fixed effects.

Summary Kenya Evidence Generalizable?

- Survey experiment in rural Kenya shows:
 - Cognitive skill measures reliable and consistent
 - Technical skills measures very noisy
 - Socio-emotional skills measures suffer from both random and systematic measurement error.
- ⇒ Do these findings hold across different settings and countries?
 - ⇒ 2nd experiment in Colombia
 - ⇒ “cross-country” analysis

Similar analysis on data from Sucre Colombia

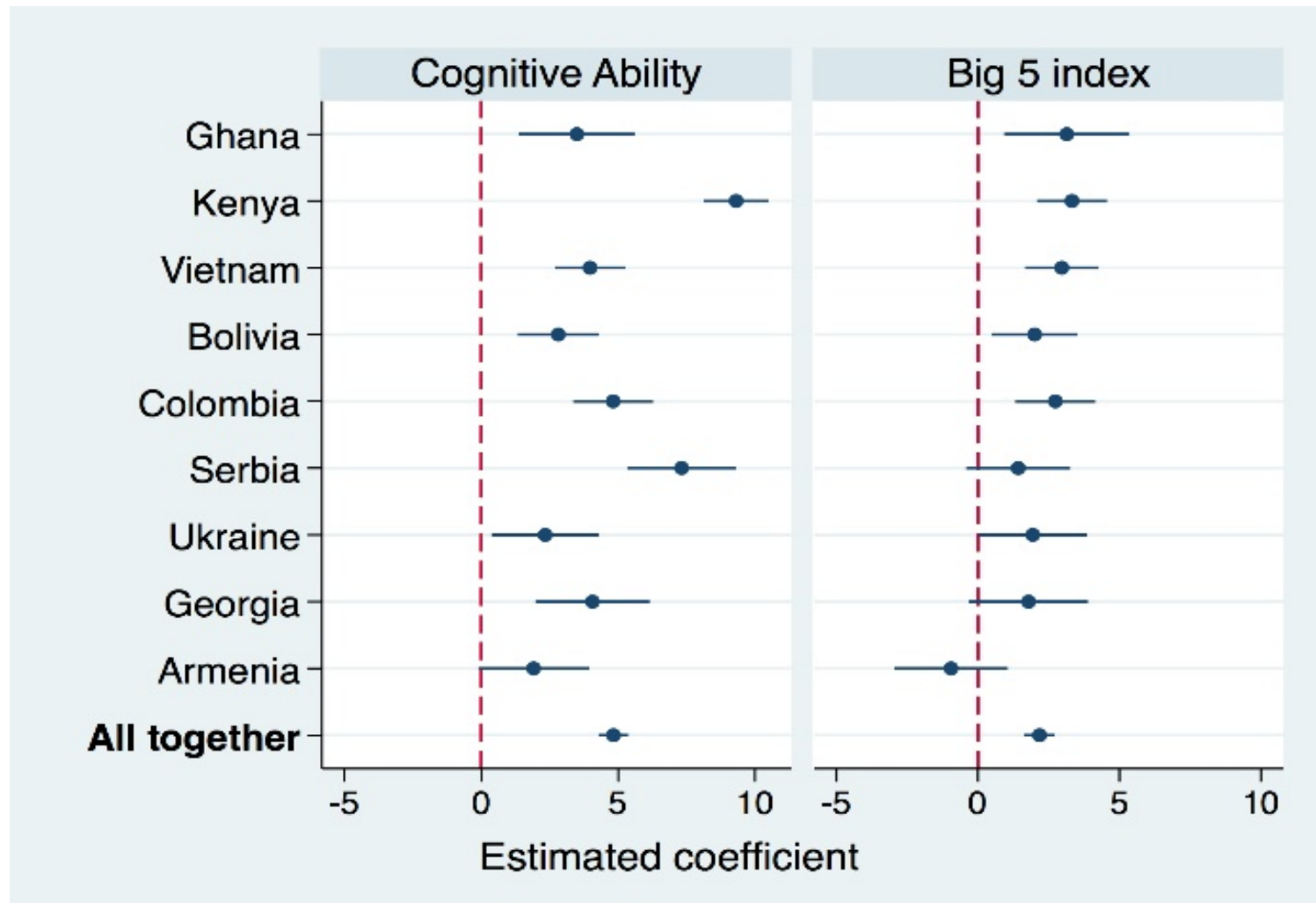
Naive Score	Test-retest correlation	Cronbach's Alpha of test	Nb of items	Enumerator Assigned (Test-retest correlation)		R2 of Enum FE	Acquiescence Bias
				Same	Different		
Cognitive	0.84	0.65	5	0.84	0.83	0.05	-
Noncog	0.70	0.70	15	0.73	0.66	0.05	0.37
Technical	0.50	0.33	7	0.51	0.50	0.14	-

Improved Score	Test-retest correlation	Cronbach's Alpha of test	Nb of items	Enumerator Assigned (Test-retest correlation)		R2 of Enum FE
				Same	Different	
Cog (IRT and factor)	0.94	0.81	6	0.94	0.93	0.06
Noncog (Factor)	0.66	0.54	7	0.68	0.65	0.22
Technical (IRT)	0.62	-	1	0.66	0.58	0.04

Widening the evidence base

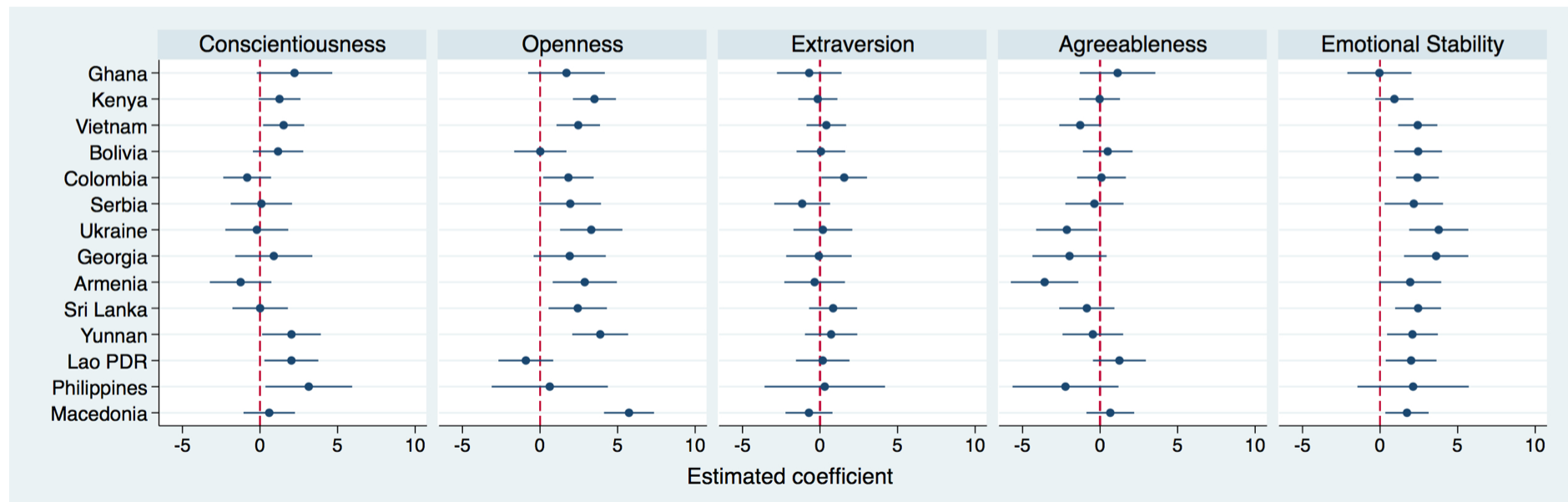
- Validity Big-Five in the Skills Towards Employability and Productivity (STEP) surveys
 - data for over 20K thousand working-age individuals from 14 STEP surveys (from Africa, Asia, ECA, Latin America)
 - **Mostly urban, education levels between 6 and 14 years**
 - **module on cognitive skills** measures functional literacy
 - variation of **the B5 personality traits** inventory commonly used in the personality psychology literature.
- Compare with internet data from 198 K respondents in same 15 countries
- And with Big Five from top pubs (54 K, 12 countries)
- Joint with Rachid Laajaj, Daniel Alejandro Pinzon Hernandez, Omar Arias, Marta Rubio-Codina, Renos Vakis

Skills and income



Are Big-Five related to income?

Figure 1: Prediction of income with the Big Five Personality Traits and Cognitive Ability



Are we measuring what we intent to measure?

- Confirmatory factor analysis: are measures consistent with the five-factor structure of the Big Five?
- **US data:** always pull in the same factor the items that in theory were meant to measure the same latent skill.
- **STEP data:** In many cases a factor pulls together items from what was believed to be different sub-constructs

Factor to item (mis)match: STEP

	Congruence coefficient	Openness			Conscientiousness			Extraversion			Agreeableness			Emot. Stability		
		O1	O2	O3	C1	C2R	C3	E1	E2R	E3	A1	A2	A3	ES1R	ES2	ES3R
Ghana	0.68	4	1	1	1	2	4	3	3	3	1	1	1	5	5	5
Kenya	0.71	2	1	4	2	2	1	3	3	3	4	4	4	5	5	5
Sri Lanka	0.71	2	1	1	1	2	1	3	3	1	4	4	4	5	5	5
Yunnan	0.75	1	1	4	2	2	2	3	3	3	1	4	4	5	5	5
Laos	0.70	1	1	3	4	2	1	1	3	4	4	4	4	5	5	5
Vietnam	0.69	1	1	1	2	2	2	4	3	3	4	3	1	5	5	5
Philippines	0.59	3	1	1	4	2	4	3	3	1	3	2	4	4	5	5
Bolivia	0.84	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Colombia	0.72	3	2	4	2	1	2	3	3	3	4	4	4	5	5	5
Macedonia	0.67	1	1	4	1	1	1	3	3	3	4	4	4	5	2	5
Serbia	0.79	1	1	1	2	2	4	3	3	3	4	4	4	5	5	5
Ukraine	0.81	2	1	1	2	4	2	3	3	3	4	1	1	5	5	5
Georgia	0.77	3	1	1	1	2	2	3	1	3	4	4	1	5	5	5
Armenia	0.82	1	1	1	1	2	2	3	3	3	4	4	1	5	5	5
Average	0.73															

Factor to item (mis)match: other

	Congruence coefficient	Openness			Conscientiousness			Extraversion			Agreeableness			Emot. Stability		
		O1	O2	O3	C1	C2R	C3	E1	E2R	E3	A1	A2	A3	ES1R	ES2	ES3R
D1	0.78	2	1	1	2	2	2	3	3	2	4	5	4	5	5	5
D2	0.69	2	3	5	2	1	2	3	3	3	4	4	4	5	1	5
D3	0.92	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
D4	0.60	3	4	4	3	2	4	3	3	4	1	1	4	5	1	5
D5	0.67	2	1	1	1	2	4	3	3	1	4	4	4	5	5	5
D6	0.68	2	1	1	2	2	3	2	3	5	1	4	4	5	5	5
D7	0.66	1	1	1	2	4	4	2	3	3	4	5	4	5	5	2
D8	0.71	2	1	1	5	2	5	3	3	3	2	5	4	5	4	2
D9	0.73	1	1	2	2	4	2	3	3	3	1	4	4	5	5	5
Average	0.71															

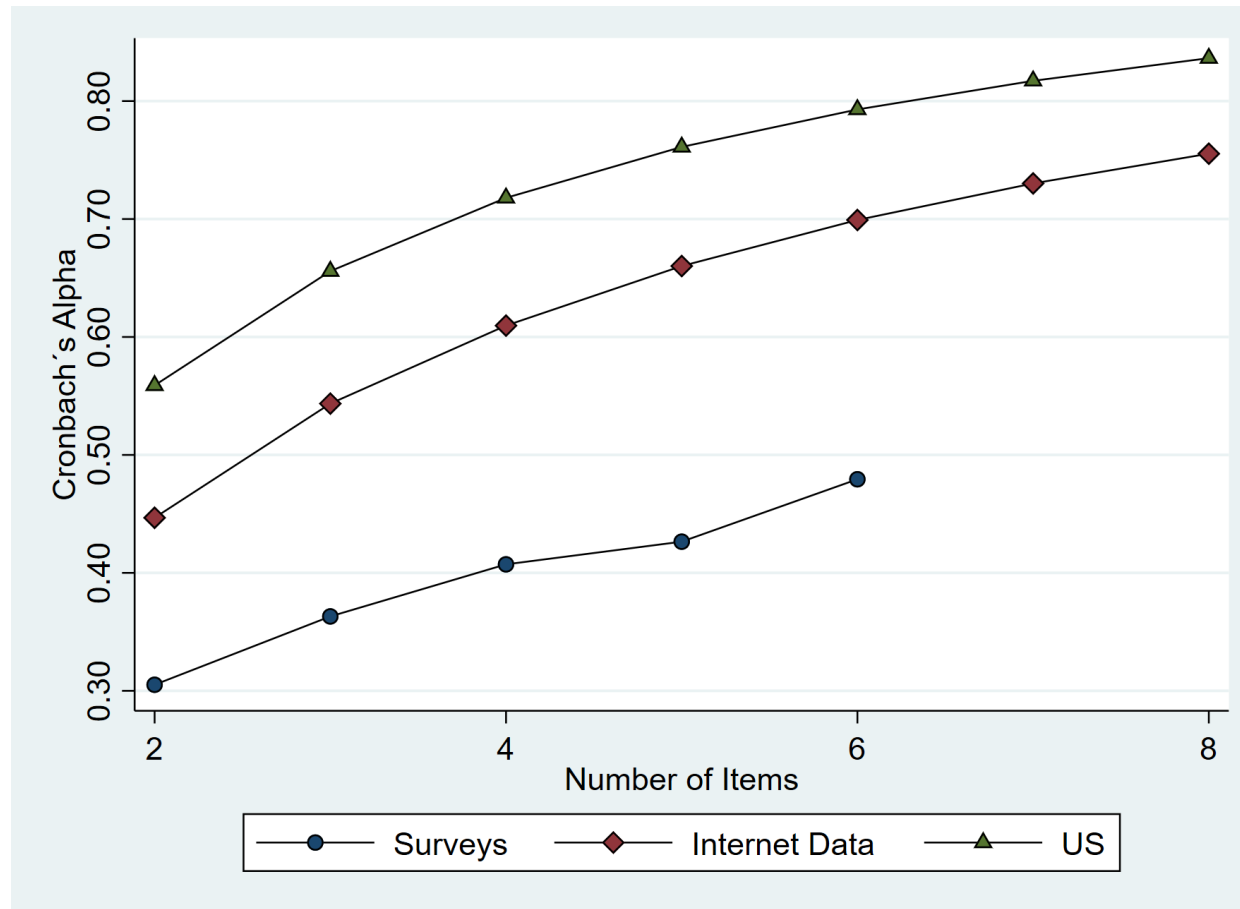
Comparing with personality test internet data

- <http://www.outofservice.com> run personality test for over 10 million people
- Exists for same countries
- **Many differences with STEPS:**
 - A **self-selected** population with access to internet (highly educated)
 - **Motivation** to do it correctly since most are doing it for the purpose of seeing the results of the personality test
 - **Self-administrated**
 - **Data cleaning** > We use the raw data, before corrections

Factor to item (mis)match: internet

	Congruence coefficient	Openness			Conscientiousness			Extraversion			Agreeableness			Emot. Stability		
		O1	O2	O3	C1	C2R	C3	E1	E2R	E3	A1	A2	A3	ES1R	ES2	ES3R
Ghana	0.97	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Kenya	0.86	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Sri Lanka	0.64	3	1	2	2	2	2	3	3	3	2	1	2	5	5	5
China	0.98	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Laos	0.81	1	1	5	2	2	2	3	3	3	4	4	5	5	5	5
Vietnam	0.98	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Philippines	0.85	1	1	1	2	4	2	3	3	3	4	4	4	5	5	5
Bolivia	0.94	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Colombia	0.97	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Macedonia	0.76	1	1	1	2	2	2	3	4	3	4	4	4	5	5	5
Serbia	0.98	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Ukraine	0.98	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Georgia	0.90	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Armenia	0.97	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Average	0.90															

Effect of number of items



Lessons Learned

- Cognitive skills can be measured reliably (if enough items)
- Technical skills very noisy but predictive and coherent.
 - Addressing measurement error helps (IRT and/or more T)
- Non-cognitive skills are more challenging to measure
 - Non-classical measurement error & answering pattern
 - Factor structure hard to identify
 - Not clear that standardized scales from developed country settings capture intended non-cognitive skills
 - Same issues for younger population and more educated ones
 - But do have predictive power
- Testing internal consistency can be done with most data and should be required before showing results

More methodological experiments ?

- Cognitive measures
 - Shorter modules
- Noncognitive measures?
 - Survey measures:
 - Capture something, but hard to know what exactly, and clearly not a particular noncognitive skill
 - No clear improvement from possible “fixes” (such as anchoring, binary choice, self-administration, frequency) in Colombia
 - “Observable” alternatives
 - Games? Experimental decision-making, ...?
 - Text analysis and scoring of vignettes?

Thank you!