Cognitive, Non-Cognitive, and Technical Skills of Poor Rural Farmers: Measurement Challenges and Implications

Rachid Laajaj and Karen Macours, Paris School of Economics and INRA
Outline

- Motivation
- Research questions and study design
- Reliability – Validity - Consistency
- Predicting agricultural decision making
- Lessons learned: Practical and Conceptual
- Next steps
Why Measuring skills?

- Human capital much more than grades of schooling attained
  - What you learn in school and what is retained
  - What you learn elsewhere
  - ...

- Having good measures of skills can
  - Help to better understand poor households’ decision making
  - Allow to observe changes (impact) on outcomes that themselves might trigger longer term results
    - => can be key for dynamic understanding of poverty
Do skills matter?

- Good body of evidence in cognitive skills, mostly but not exclusively on developed countries.
  - Hanushek and Kimko (2000) use math and science test scores, and find it to predict growth much better than years of education.
  - Numerous studies establish that measured cognitive ability is a strong predictor of schooling attainment and wages, conditional on schooling (Cawley, Heckman, and Vytlacil 2001).

- Heckman, Stixrud & Urzua (2006) find that non-cognitive skills can be even more important than cognitive skills to explain success in life (income, wages, criminal behavior, teenage pregnancy …)
Adult skills and household surveys

- When analyzing household-level data we often worry about unobservables such as entrepreneurship, attitudes, knowledge, etc.
- So why not try to make those “observables”?
- Many existing measures of cognition/achievement, non-cognitive skills and approaches to measure technical skills.
- Often taken from developed country literature – e.g. psychology literature on “Big Five”

=> lack of validation in developing country settings
Skills in developing countries

- Cognitive delays from early childhood and important socio-economic gradients in cognition
- Literature has documented
  - Low levels of aspiration
  - High levels of depression
  - Lack of Self-control
  - ...
- Lack of information or ”know how” regarding agricultural practices

=> Can good measurement of adult skills allow us to better understand decision making?
Study design

- Research questions
- Defining cognitive, non-cognitive and technical skill measurements
- Survey field experiment in Kenya
  - Questionnaire design
  - Field work design
- Statistical analysis
Research Questions

- How to efficiently measure skills in the context of a household survey in rural areas in a developing country?

- Which skills matter for agronomical decision making of poor rural households?
Which skills?

- **Cognitive skills (~ IQ)**
  - Memory, processing speed, problem solving
  - “Class room” skills: Reading and math (achievement)

- **Non-cognitive skills**
  - Self-control, perseverance, self-esteem, Big Five
  - Aspirations, locus of control, ...

- **Technical skills**
  - Knowledge/Know-how
Cross-cutting challenges for skill testing in household surveys

- Many existing measures are time consuming
- Initially designed for developed country settings, lab settings, self-administered surveys, etc
- Concept often more abstract – and/or more technical – translation and understanding can become a big issue
- Standardized application of tests
- Openness of adult respondents to test-taking
- Measurement error and imperfect proxies
Methodology

- Designed an instrument with different alternative modules and approaches for each of the 3 skill types

- Conducted survey experiment in rural Kenya
  - Randomized survey instrument (finalized after extensive piloting)
  - Test-retest
  - Randomized field work implementation

- Use statistical analysis to analyze reliability and validity of the measurements
Decision making likely depends on overall cognitive ability but possibly also on math and language skills (cost-benefit calculations, reading instructions, etc).

Several tests found in household surveys – but not much standardization:

- Non-verbal cognitive ability: 36 item Raven Progressive Matrices
- Memory: digit span (forwards and backwards)
- Math
  - Oral: 9 Puzzles/Questions
  - Written: Fluency (nr of correct operations in 3 minutes)
- Reading comprehension: questions about short paragraphs in English or Swahili
Raven’s Progressive Matrices

- 60 Multiple choice questions in order of difficulty, test the reasoning ability.
Cognitive skills: e.g. digit Span Test

- Memory span: longest list of items that a person can repeat back in correct order immediately after presentation on 50% of trials. Called digit span when numbers are used. Measures the short term memory.

<table>
<thead>
<tr>
<th>ITEM</th>
<th>First Attempt</th>
<th>Second Attempt</th>
<th>Correct 1 = in the first chance, 2 = in the second chance, 3 = in none</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>43</td>
<td>16</td>
<td>1 2 3</td>
</tr>
<tr>
<td>b)</td>
<td>792</td>
<td>847</td>
<td>1 2 3</td>
</tr>
<tr>
<td>c)</td>
<td>5941</td>
<td>7253</td>
<td>1 2 3</td>
</tr>
<tr>
<td>d)</td>
<td>93872</td>
<td>75396</td>
<td>1 2 3</td>
</tr>
<tr>
<td>e)</td>
<td>152649</td>
<td>216748</td>
<td>1 2 3</td>
</tr>
<tr>
<td>f)</td>
<td>3745621</td>
<td>4925316</td>
<td>1 2 3</td>
</tr>
<tr>
<td>g)</td>
<td>82973546</td>
<td>69174253</td>
<td>1 2 3</td>
</tr>
<tr>
<td>h)</td>
<td>246937185</td>
<td>371625948</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>
“Non-cognitive” skills catch-all category

- Heckman highlights the importance of perseverance, motivation, time preference, risk aversion, self-esteem, self-control, preference for leisure
- Psychologists, and more and more economists, lots of attention on “Big Five” personality traits
- Literature on small businesses in developing countries suggests measures of entrepreneurship, attitudes to change, openness to innovation, …
- As well as aspirations, locus of control, perceptions of causes of poverty, “hope”, …

Selection of commonly used questions/modules – narrowed down after translation and piloting
Measurement Non-Cognitive skills (2)

- Traditional: 1-5 scale with statements about oneself
  - “On a scale from 1 to 5 - with 1 indicating you strongly disagree and 5 indicating you strongly agree: My life is determined by my own actions”

- 1-5 scales about causes of poverty

- Economic ladder

- Locus of control through “beans” (visual aid)

- CESD: E.g. “In the last 7 days, how many days were did you feel sad?”

- Some standardized measures: risk games, time preference game, …
The Big Five Personality traits

<table>
<thead>
<tr>
<th>Trait</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Openness</strong></td>
<td>Curious, original, intellectual, creative, and open to new ideas.</td>
</tr>
<tr>
<td><strong>Conscientiousness</strong></td>
<td>Organized, systematic, punctual, achievement oriented, and dependable.</td>
</tr>
<tr>
<td><strong>Extraversion</strong></td>
<td>Outgoing, talkative, sociable, and enjoys being in social situations.</td>
</tr>
<tr>
<td><strong>Agreeableness</strong></td>
<td>Affable, tolerant, sensitive, trusting, kind, and warm.</td>
</tr>
<tr>
<td><strong>Neuroticism</strong></td>
<td>Anxious, irritable, temperamental, and moody.</td>
</tr>
</tbody>
</table>
Measurement Technical skills

- Basic **knowledge required to perform a task**: very field specific by definition.
- Use proxies or try to obtain actual tests of relevant knowledge?
  - Self-assessment
  - Past years of experience
  - Knowledge tests => Which type of question?
    - Recognition techniques/practices, timing, knowing how to implement, ”scientific” understanding,
    - work with agronomists and soil scientists on targeted questions for main crops and practices
  - Attempt to have “unambiguous” questions with varying difficulty, mostly multiple choice, visual aids
Examples of Technical Skills

Questions

- When planting hybrid maize in rows, how many seeds per hole should be applied?
- When planting bananas what is the optimal distance between banana trees?
  - 1. 1mx1m
  - 2. 2m x 2m
  - 3. 2m x 3m
  - 4. 3m x 3m
Field context

- 960 (918) rural farmers in Siaya - Western Kenya
  - Mainly maize and other annual crops – most also have livestock
  - 50/50 men-women
  - On average 6 years of education
- Selected in 96 villages to be part of RCT on farmers’ learning and agricultural technology adoption
  - analyze role of the skills for adoption & productivity
  - Analyze heterogeneous treatment effects & importance of “observing the unobservables”
CAPI Questionnaire design

- 3 main sections (cognitive, non-cognitive, technical) – order of sections is randomized
  - + Small section with self-assessment same skills
- Within each section: order of modules (subset of questions on same domain) randomized
  - Allows analysis of survey fatigue but also order effects
- Order of answer options randomized
- “Salad bowl” in non-cognitive section: 50% of surveys questions on different non-cognitive skills are mixed
Field work implementation

- Computer based survey + some visual aids
- 2 week training on standardized application tests
- Test and retest 3 weeks apart
  - Randomly assign enumerator to farmers
  - Randomize whether same enumerator in retest

=> Enumerator effects

- Randomize the order of villages
  => Survey duration – day of the week?

=> Simultaneous household survey
What is a good measurement?

Unreliable & Unvalid

Unreliable, But Valid

Reliable, Not Valid

Both Reliable & Valid
Reliability

- Item (X) is affected by True score (T) and Measurement Error (E).
  \[ X = T + E \]
- Hence the Variance of X is given by:
  \[ \sigma_X^2 = \sigma_T^2 + \sigma_E^2 \]
- Reliability is the ratio of variability in X due to T:
  \[ Reliability = \frac{\sigma_T^2}{\sigma_X^2} \]
Measures of Reliability

- Indicators:
  - Consistency across time (pure reliability): High Test-Retest Correlation if you replicate the measure within a period short enough that it should not have changed.
  - Consistency across items: High correlation among items that intend to measure the same skill: Cronbach’s Alpha (also validity)
  - Results not subject to the conditions i.e. enumerator, order of questions or responses, mood of the day.
Validity

- Are you measuring what you intend to measure?
- Indicators:
  - Use of Validated (in other context) Psychometric scales & Piloting experience
  - Correlation with other measures (same T) (Cronbach’s Alpha when excluding and item from its group of items)
  - Should predict well related behaviors: regressions on agronomical decisions and outcomes
The Aggregation Method

- Three commonly used methods:
  - Score (naïve addition of points)
  - Principal Component Analysis
  - Item Response Theory

- PCA and IRT bring improvements when well behaved may pick the wrong factor if not careful
- IRT also brings information about the noise
ITEM RESPONSE THEORY assumes that success to a question depends on the unobserved ability of respondent and the questions’ parameters estimated simultaneously.

Source: Jean Michael Linacre in www.rasch.org
Cognitive skills: Test-retest and Internal Reliability

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Test-retest</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cog</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.63</td>
<td>0.88</td>
</tr>
<tr>
<td>Numeracy Q.</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>Math sheet</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.82</td>
<td>0.92</td>
</tr>
<tr>
<td>Digit Span</td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>
Reliability of Cognitive Skills

- Generally strong reliability both in test retest and internal.
- Number of questions can be reduced: example of the reading tests with very high correlation among the 3 exercise.
- But at some cost Raven, reducing test-retest correlation from 36 to 12 items falls from .63 to .45.
- 2 math tests highly correlated among them. Digit Span backward correlates + with Raven than forward Digit Span.
- Correlation of self-assessment with test score is .66 for reading but .31 for math.
## Non-cognitive Reliability Test

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Test-retest</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Non-Cog</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Locus of Control</td>
<td>0.42</td>
<td>0.55</td>
</tr>
<tr>
<td>Causes of Pov</td>
<td>0.40</td>
<td>0.35</td>
</tr>
<tr>
<td>Attit. Change</td>
<td>0.43</td>
<td>0.46</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>BF_Extrav.</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>BF_Agree</td>
<td>0.26</td>
<td>0.40</td>
</tr>
<tr>
<td>BF_Conscious</td>
<td>0.33</td>
<td>0.51</td>
</tr>
<tr>
<td>BF_Neurotic</td>
<td>0.26</td>
<td>0.46</td>
</tr>
<tr>
<td>BF_Open</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>CESD</td>
<td>0.42</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Non-cog Skills Main Results

- Still a very challenging measurement. Each question is very noisy, and even when aggregated, remain quite noisy.
- CESD and Locus of control (including causes of poverty) doing relatively better than others.
  - More abstract questions (e.g. “you like to learn new things”) show lower reliability than more concrete questions (e.g. “in the last 7 days, how many days…”)
- Acquiescence bias is a major issue – But Questions negatively phrased create confusion
- Non-cog multidimensional in nature…
## Technical Agricultural Knowledge Reliability Test

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Test-retest</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tech</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Intercrop &amp; Rotat.</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>Maize</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td>Banana</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>Soybean</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Composting</td>
<td>0.25</td>
<td>0.19</td>
</tr>
<tr>
<td>Min. Fertilizer Use</td>
<td>0.28</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Technical Skills Main Results

- Measurement even more challenging than non-cog.
- Requires more adaptation to local context.
- Low correlation between questions possibly pointing to idiosyncrasy of agricultural knowledge (and guessing)
- Some knowledge is too widely known (everyone gets it right) and other is too context specific: very narrow area between the 2.
- Contrast: nb of years of experience test-retest correlation is .68
- Low correlation with self-assessment (.1 to .4)
Predicting agricultural decisions

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Fertilizer dummy</th>
<th>Hybrid dummy</th>
<th>Manure-Comp dummy</th>
<th>Share plots weeded at most once</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cog</td>
<td>0.0122</td>
<td>0.00709</td>
<td>-0.00458</td>
<td>0.0550**</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.027)</td>
<td>(0.034)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Non Cog</td>
<td>0.0538***</td>
<td>0.0615**</td>
<td>0.0269</td>
<td>-0.0196</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.024)</td>
<td>(0.026)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Tech</td>
<td>0.0347**</td>
<td>0.0259</td>
<td>0.0581**</td>
<td>0.0132</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Observations</td>
<td>884</td>
<td>801</td>
<td>817</td>
<td>885</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.475</td>
<td>0.409</td>
<td>0.271</td>
<td>0.377</td>
</tr>
<tr>
<td>R2 of Skills</td>
<td>0.048</td>
<td>0.031</td>
<td>0.013</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Predictive Power of Skills

- Cognition overall has low predictive power and seems only relevant for selected agricultural decisions
  - Cognitive measures also show good correlation with education level, but have additional predictive power
- Despite measurement error, non-cognitive skill construct has predictive power for agricultural decisions
- When disaggregating, predictive power of technical skill important for agricultural decisions (e.g. fertilizer and compost practices – up to 9% variation explained).
Enumerators Matter

- Enumerator fixed effects explain up to 15% of variation especially for non-cog and Tech
- Reliability largely affected by changing enumerators (randomized)
- Questions with visual aids, open questions, more difficult questions seem to be more sensitive
- Putting fixed-effects do not solve the problem
- Important to balance enumerators (& randomize?)

<table>
<thead>
<tr>
<th></th>
<th>All Tests</th>
<th>Same enumerator</th>
<th>Different enumerator</th>
<th>All, with enum. FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>0.83</td>
<td>0.88</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>Non-Cog</td>
<td>0.53</td>
<td>0.63</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>Technical</td>
<td>0.31</td>
<td>0.45</td>
<td>0.26</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Other Practical Lessons learned (1)
Translation

- Keeping intent of the questions through translation is hard, especially for non-cognitive skills
- Abstract concepts hard to translate
- “difficult” constructions possibly enhancing acquiescence bias (and survey fatigue?)
Other Practical Lessons learned (2): Survey fatigue matters differentially

- No evidence of strong threshold effects for survey duration
- Various cognitive scores are higher when cognitive module is asked first, lower when asked at the end
- Survey fatigue of respondents or enumerators?
  - Some non-cognitive questions affected by order in survey
  - But also relatively strong effects of day-of-the-week and duration field work on several outcomes
  - Bias in both directions
Conceptual lessons learned

- Cognitive skills can be measured reliably.
  - High correlation between measures suggest including a subset or shorter tests can provide a good proxy

- Non-cognitive and Tech skills remain more challenging to measure
  - For noncog:
    - Test-retest statistics better for constructs than indiv. Questions
    - Sensitive to enumerator, order, etc.
  - For agricultural knowledge
    - Few unambiguous answers
    - Idiosyncratic

- However Non-cog and Tech seem to predict farming decisions better than cog
Next steps: Kenya

- Improve efficiency:
  - Purging constructs of Some Variables which add Noise to the aggregated Index.
  - Recommendations based on time-validity trade-offs
- Incorporating skills of other household members
- Predicting heterogeneous treatment effects
- Accounting for measurement error in estimations
Next steps: 2\textsuperscript{nd} field experiment in different setting

- Comparability and external validity of findings
- Incorporate “observable” measures even if it’s more time consuming
- Test the importance of the farmer’s perception about the consequences of their responses?
- Suggestions very welcome!
Conclusion

- Large potential for skill measurements to better understand typical unobservable determinants of poverty and the poor’s decision making
- But integration in large surveys comes with its own requirements and costs
  - Blind adoption of existing questions and scales probably not a good idea
  - Quality control is key as are realistic and controlled field protocols
  - Very low-cost alternatives may not exist for all relevant skills
- Better understanding of role of different skills arguably key for better design of skill-enhancing interventions => lot’s to learn
Thank you!

- rachid.laajaj@psemail.eu
- karen.macours@psemail.eu