



POVERTY & EQUITY NOTES

MARCH 2018 · NUMBER 3

How well do we measure non-cognitive skills in developing countries? Initial lessons from the STEP Skills Measurement Program¹

Rachid Laajaj, Omar Arias, Karen Macours, Marta Rubio-Codina & Renos Vakis

This study evaluates how well the Big 5 (B5) personality traits –commonly used to proxy for non-cognitive (or more properly socio-emotional) skills— are measured in the STEP surveys using self-reported questionnaires so far applied in 15 developing countries. Several commonly used indicators are used to assess validity and reliability of the measures, ranging from Cronbach's Alpha, Acquiescence Bias (AB, often referred to as "yea-saying", i.e., the tendency of respondents to agree with a statement when in doubt), the explanatory power of enumerator fixed effects, the number of factors identified by exploratory factor analysis and the factor structure. Results suggest that in most aspects, the indicators of internal consistency for the B5 measures are well below the norm observed in other similar studies in developed economies. Low internal consistency is partly related to the functional literacy ability of the respondents and to the lower number of items used in the STEP surveys B5 module. Yet these factors do not explain all the differences in internal consistency compared to data from the USA that is used as a

benchmark. The results suggest that the use of B5 self-reported questionnaires to measure personality traits through household surveys in developing countries has to contend with biases arising from systematic response patterns such as AB, the mediating role of enumerators, and possibly other issues such as reference biases if different groups have different standards or reference points when assessing their behavioral traits. More work is needed to further explore these issues and test potential solutions and alternative approaches to non-cognitive skills measurement in developing settings.

The Issue

Standard skills indicators leave an information gap on workforce skill characteristics, job skill requirements, and quality of worker-job matches that prevents policymakers from making informed and timely decisions on training and education for current and future employees. Thus, there is growing interest on measures of cognitive, socio-emotional (non-cognitive), and technical skills

¹ Based on work by Rachid Laajaj (Universidad de los Andes), Omar Arias (World Bank), Karen Macours (Paris School of Economics), Marta Rubio-Codina (IADB), and Renos Vakis (World Bank). This research is part of a joint work program on methodological improvements of skills measures in household surveys co-led by the Mind, Behavior, and Development Unit (eMBeD), the Skills Global Solutions Group of the Education Global Practice and Social Protection & Labor Global Practice, and the Living Standards Measurement Study team of the DEC Survey unit.

to analyze determinants of skills formation, economic decisions and labor market outcomes. But how can we do this at a national scale in the typical household surveys that development researchers collect? Existing measures have mostly been validated in developed countries, and with data collected in relatively controlled settings, far from the context typical of developing countries.

Recent work from Kenya (Laajaj and Macours, 2017) highlighted some of the challenges of collecting reliable and valid data for a set of commonly used skills measures. Using a survey with skills measurements administered to more than 900 farmers in western Kenya and then re-administered three weeks Laajaj and Macours found that cognitive skill measures were reliable and consistent, but non-cognitive skills measures (captured through a Big Five personality traits inventory) were rife with measurement error. Do these findings hold across different settings and countries?

The Data

To expand on this work, we conducted a systematic analysis of the Skills Towards Employability and Productivity (STEP) surveys, a research initiative ran by the World Bank that aims to better understand the interplay between skills and employability in developing countries. The STEP program developed household survey instruments tailored to collect data on cognitive, socio-emotional, and job-relevant skills in low- and middle-income country contexts. The module on cognitive skills measures functional literacy linked to the scale of the OECD's Programme for the International Assessment of Adult Competencies (PIAAC). For socio-emotional skills, the STEP instrument uses a variation of the B5 personality

traits inventory commonly used in the personality psychology literature.

Motivated by the results from Kenya, we examined how well the STEP surveys measure socio-emotional skills. Using data for over 50,000 thousand working-age individuals from 15 STEP surveys collected in Sri Lanka, Yunnan (China), Lao PDR, Vietnam, Philippines, Bulgaria, Armenia, Georgia, Macedonia, Ukraine, Serbia, Kenya, Ghana, Bolivia and Colombia, as well as additional data from the USA for benchmarking, the analysis compiles a set of measures of validity and reliability around the Big-5 personality traits.

Results

The findings confirm that in all the STEP countries analyzed, the reliability and validity of the Big-5 measures are below norms commonly accepted in the psychological literature. Just like the results of Laajaj and Macours (2017) for Kenya, the indicators typically used to assess the quality of the measures (such as acquiescence bias and Cronbach's alphas) are lower than the minimum standards used in psychometrics to be considered useable. For instance, a Cronbach alpha coefficient of at least 0.7 is a common threshold used to judge a measure as internally consistent, but the alpha coefficients for the STEP measures in all countries are generally in the range of 0.3 to 0.4.

In addition, the low reliability of the measures does not appear to be random. For example, reliability and validity tend to be higher in countries with higher scores on STEP cognitive (literacy) tests, although even for countries with the highest cognitive scores in our sample (specifically, Eastern European countries), reliability levels remain below US levels. Literacy scores tend to explain more variation in reliability between countries than within

country, suggesting that other factors at the country level are at play. Similarly, when restricting (in each country) the sample only to people who completed

some tertiary education, we find almost no improvement in the reliability indicators.

Figure 1: Big 5 and factors associated for each item

	Openness			Conscientiousness			Extraversion			Agreeableness			Neuroticism		
	Q1	Q2	Q3	Q1	Q2R	Q3	Q1	Q2R	Q3	Q1	Q2	Q3	Q1	Q2R	Q3R
US	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
Sri Lanka	2	2	3	2	2	2	3	3	3	4	4	4	5	1	1
Yunnan (China)	1	1	4	4	2	4	3	4	3	4	4	1	5	5	5
Lao PDR	1	1	1	4	2	1	1	3	4	4	4	4	5	5	1
Vietnam	1	1	1	3	2	3	1	3	4	4	3	4	5	5	5
Philippines	1	1	1	2	2	1	5	3	4	1	1	1	5	5	3
Bulgaria	3	1	1	2	1	4	2	3	4	1	4	4	5	5	4
Armenia	1	1	1	1	2	1	3	3	3	4	1	4	5	5	5
Georgia	1	2	2	2	4	2	1	3	4	4	2	4	5	5	5
Macedonia	1	4	4	4	2	4	3	3	3	4	4	4	5	1	5
Ukraine	3	1	1	2	2	2	3	3	3	4	1	4	5	5	5
Serbia	1	1	1	5	2	2	3	3	3	4	4	4	5	5	5
Kenya	2	2	4	2	2	2	3	3	3	4	4	4	1	5	5
Ghana	1	1	1	1	2	1	3	4	3	1	1	1	4	5	5
Bolivia	1	1	4	1	2	1	3	3	3	4	4	4	5	5	5
Colombia	3	1	1	1	2	1	3	3	3	4	1	4	5	5	5

Q1, Q2, Q3 refer to items (questions) in the STEP B5 personality scale. R: Negatively worded items.

Using Varimax-Rotated Five-Factor Structure for Big Five items, each item was assigned to the factor which had the highest load. For example, US follows exactly the expected factor structure of the Big Five, but none of the other countries have this perfect match.

Items corrected for Acquiescence Bias.

One of the biggest concerns is that the factor structure arising from the data does not conform well with the model of the Big five personality factors. That is, in most countries several of the items in the Big Five scale do not statistically align with the personality factors they purport to measure. As Figure 1 illustrates, many items are misaligned across various factors (highlighted in red). Therefore, two items of different constructs (personality factors) can be more correlated than items of the same construct. By contrast, the five factors emerge very clearly in US data, as seen in the top line of Figure 1. It may be that in developing country contexts, responses to the B5

items do not capture well the underlying personality factors or that the underlying factor structure is altogether different in these countries than it is in the United States.

Looking Ahead

The results from the analysis of the STEP surveys call for caution in interpreting existing measures and underscore the need for further research to improve nation-wide measurement of socio-emotional skills in developing countries. Taken

together with the earlier work and additional analysis from the STEP surveys not shown, the results offer five takeaways and a path for future research:

- **Check, check, check!** Before utilizing non-cognitive skills measures, it should be customary practice to assess that they behave the way they are supposed to. Most of the reliability and consistency tests done in both sets of studies are common practice in studies that use these data, are easy to do with standard statistical packages like STATA, and thus should be done regularly by applied researchers.
- **Correct for Acquiescence Bias (or yeah saying).** Corrections for AB lead to considerable improvements to reliability and validity of the measures (even if not fully to commonly acceptable levels).
- **You cannot cut corners – include sufficient number of items per construct.** The third wave STEP surveys use seven questions per personality factor instead of three in the first two phases. We find that including more questions improves on some the reliability indicators.
- **Don't assume standard scales measure the same underlying constructs everywhere or are best suited for purpose.** Start by linking the policy question at hand with the right constructs you should aim to collect. But even then, let the data tell you the latent traits measured using factor analysis. As we show, even after correcting for AB or having a larger number of items, different factor structures may emerge in developing countries.

- **Consider self-administration when feasible.** We find that enumerator fixed effects jointly significantly predict the values of the Big Five measures. As a tentative implication, when possible, we suggest considering self-administrated surveys instead of oral surveys for socio-emotional skill measures. The absence of interaction with an enumerator may also reduce issues related to properly listening, being influenced, and/or social desirability bias (trying to give an answer that gives a better impression). This is of course a challenge to implement with low literacy populations.

- **More methodological testing please!** Looking ahead, more work is needed to test potential solutions and alternative approaches to non-cognitive skills measurement in developing settings. As an example, task-based measures may be one research avenue that can minimize some of the issues discussed above. Methodological research in this area is urgently needed.

ABOUT THE AUTHORS

Rachid Laajaj Professor, Universidad de los Andes, karen.macourspsemail.eu

Omar Arias, Practice Manager, World Bank's Education Global Practice, oarias@worldbank.org

Karen Macours, Professor, Ecole d'économie de Paris, karen.macourspsemail.eu

Marta Rubio-Codina, Economist, Inter-American Development Bank (IADB),

Renos Vakis, Lead Economist, World Bank's Poverty and Equity Global Practice (GPVGE), rvakis@worldbank.org

This note series is intended to summarize good practices and key policy findings on Poverty-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank, its board or its member countries. Copies of these notes series are available on www.worldbank.org/poverty