Online Appendices for

MEASURING SKILLS IN DEVELOPING COUNTRIES

By RACHID LAAJAJ AND, KAREN MACOURS

September 2019

The online appendix can be found at http://jhr.uwpress.org

List of Appendices

Appendix A: Brief introduction to psychometric concepts and methods used

A.1 Reliability

Reliability is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under similar conditions. A measure that is very noisy is a measure with low reliability.

In classical theory, it is assumed that a person's observed or obtained score on a test is the sum of a true score (T) and a Measurement Error (E):

$$X = T + E$$

Hence the variance of X is given by:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

In this setting, reliability is the ratio of variability in item X due to variation of the true score T:

$$Reliability = \frac{\sigma_T^2}{\sigma_X^2}$$

It can thus be interpreted as the ratio of the variance of a given measure that is driven by the true variance of the score across the population, or equivalently 1 minus the share of variance explained by pure measurement error.

An estimation of the reliability can be obtained with the test-retest correlation (consistency across time).

If measurement error is classical, the test-retest correlation gives a good indication of the signal to total variance ratio. On the other hand, the test-retest correlation can under or over-state the signal to total variance ratio in case of non-classical measurement error. If the errors in measurement are positively correlated over time, for instance because both measures suffer from persistent acquiescence bias, the test-retest correlation will overstate the reliability of the data.

The Cronbach’s alpha is also an indicator of reliability, and provides a measure of consistency across items expected to measure the same latent construct. As such the Cronbach’s alpha is, however, also an indicator of validity.

A.2. Validity

Test validity is the extent to which a test measures what it is supposed to measure.

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.

Among the key indicators of validity are the following ones:

- Face validity assesses the extent to which a test is subjectively viewed as covering the concept it purports to measure. For example a question about self-confidence should seem to ask about self-confidence (hence a question with high correlation with related measures but seemingly asking something very different cannot be considered valid)
- Content validity refers to the extent to which a measure represents all facets of a given construct.
- Piloting experience and use of psychometric scales validated in other contexts
- Construct validity: Correlation with other measures intending to measure the same construct
- Predictive validity: it should predict well related behaviors that are theoretically expected to be correlated with the measure

A more detailed explanation of reliability and validity can be found in American Educational Research Association et al. (1999).

A.3. Test-retest correlation

Test-retest correlation is the correlation between measures using the same instrument, measured twice, on the same person, in similar conditions within a relatively short period of time. Temporal stability provides an assessment of the reliability of a measure. Typically a similar test is applied twice to the same population within a period short enough that that the traits that the researcher intends to measure should not have changed, but long enough that respondents do not remember their original responses. A standard period between test and retest goes from two weeks to one month.

Under classical theory assumptions, the correlation between the test and the retest can be interpreted as a direct measure of reliability as defined above ($\frac{\sigma_T^2}{\sigma_X^2}$) hence the correlation can directly be interpreted as the share of variance of the measure explained by the variance of the true ability it is intended to measure. This is equivalent to one minus the share of variance explained by pure measurement error.[1]

[1] If measurement error is classical, the test-retest correlation gives a good indication of the signal to total variance ratio. On

Intuitively high measurement error means that the true score is an imprecise measure and leads to a low test-retest correlation, hence a low reliability. A change in the true value of the skills between the test and the retest would also lower the statistic. This is a desirable property since measures that change too much in the short run would not be good for long terms outcomes. A threshold of minimum .7 test-retest correlation is often applied. All estimates are done using z-scores of the relevant constructs and subconstructs (i.e. after subtracting the mean and dividing by the standard deviation).

Crocker and Algina (2006) and Nunally, and Bernstein (1994) provide a broader explanation of classical test theory and test-retest correlation.

A.4. Cronbach's alpha

The Cronbach's alpha (Chronbach 1951) is one of the most widely used measures of internal consistency of a test.

Cronbach's alpha is mathematically equivalent to the expected value of the split-half reliability. Split-half reliability is obtained by 1) randomly splitting the items into two sets of items of equal size, 2) calculating the average of each set of items, and 3) calculating the correlations between these two sets of items. Although not calculated this way, the Cronbach's alpha is equal to the average of the correlations obtained through all the possible combinations of split-half reliability. Its value is driven by how well the items correlate among them and by the number of items.

Assume that we have a measure X made of k items: $X = Y_1 + Y_2 + \cdots + Y_k$

Its Cronbach's alpha is given by:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

Where $\sigma_{Y_i}^2$ is the variance of item i and σ_X^2 is the variance of the measure X.

The Cronbach's alpha provides an assessment of both the construct's validity and its reliability. It is said to provide a lower bound on the reliability of a test, because for the case where all items are measuring exactly the same construct, the Cronbach's alpha would only be affected by the measurement error of each item and is a pure measure of reliability. When it is not the case, then the Cronbach's alpha is also affected by the extent to which items are measuring the same latent construct.

the other hand, the test-retest correlation can under- or over-state the signal to total variance ratio in case of non-classical measurement error. If the errors in measurement are positively correlated over time, for instance because both measures suffer from persistent acquiescence bias, the test-retest correlation will overstate reliability.

Hence a low Cronbach's alpha indicates that either the items are measuring very different latent constructs (the validity is poor since the items are usually pooled with the intention to measure one latent construct) or they are measuring the same latent construct but with a lot of noise, hence the reliability is low.

A Cronbach's alpha of 0.9 tends to be required when individual decisions will be made based on a specific test (for example student's admissions, Nunnally and Bernstein, 1994; Kline, 2013), but an alpha of .7 is often considered acceptable for the purpose of statistical analysis.

Appendix B. Construction of the Improved Indexes

The "naïve score" is calculated as the simple average of items (questions) that belong to pre-determined sub-domains. The "improved" constructs relies on different corrections to extract the most relevant information from the available items: we use exploratory factor analysis to determine the number of factors in each construct, item response theory to further improve the cognitive and technical constructs, and correct for acquiescence bias in the noncognitive construct.

B.1. Correcting noncognitive items for Acquiescence Bias

Acquiescent response style refers to the tendency of an individual to systematically agree (yea-saying) or disagree (nay-saying) with questionnaire items, regardless of their content. For example, consider the two following items, with answering scales from 1 (strongly agree) to 5 (strongly disagree):

- "Are you relaxed during stressful situations"
- "Do you get nervous easily"

They both aim at measuring the same trait (Emotional Stability), but only the second one is reverse-coded in the sense that a higher degree of agreement in the response is associated with a lower Emotional Stability. Hence a respondent that strongly agrees with both statements shows a form of contradiction, indicative of a positive acquiescence bias. This response pattern ends up being a strong driver of the variation in the data.

We correct for Acquiescence Bias in all noncognitive questions answered on a Likert scale, following common practice in psychometrics (Soto et al. 2008; Rammstedt, Kemper, and Borg, 2013; and references therein). The acquiescence score is calculated at the individual level and relies on having both reverse-coded and non-reverse coded items.

We calculate the acquiescence score and apply the acquiescence bias correction using the following steps:

1) Reverse the reverse-coded items. For example, if the possible answers range from 1 to 5, then answer 1 (fully disagreeing with a reverse-coded statement), is assigned a value of 5, answer 2 is assigned a value of 4, and so on.
2) Calculate the average answer of reverse-coded items, and the average answer of non-reverse-coded items.

3) Take the difference between the average of non-reverse-coded items and the average of reverse-coded items and divide it by two.
4) To correct for the acquiescence bias, add the acquiescence score obtained in 4) to every reverse-coded item, and subtract the acquiescence score from every non-reverse-coded item.

The intuition of the AB correction is that, once the reverse-coded items are reversed, they should on average be equal to non-reversed items since they aim to capture the same skills. If we find systematic difference in the average answer of the two groups, it is likely driven by a systematic tendency to agree (or disagree) and thus needs to be corrected for.

B.2. Exploratory factor analysis to determine the number of factors in each construct

We conduct exploratory factor analysis (EFA) separately for cognitive, noncognitive and the technical skills, and determine the number of factors that should be extracted from the data. To do so, we pool all data for each domain (pooling for instance all noncognitive questions together), instead of relying on pre-determined scales. Hence, we let the data indicate the potential factor structure and related latent traits, following an approach also used by Ledesma and Valero-Mora (2007), Cunha, Heckman and Schennach (2010), and Attanasio et al (2015). The results are presented in Appendix Table A2.

For the cognitive skills, we use the score for each of the five tests as inputs in the EFA. For the noncognitive and technical skills, we use each of the questions separately. We determine the number of latent factors that can be extracted from all the measures, using four different criteria commonly used in the psychometric literature. Here we provide a very brief description of four methods and refer to Valero-Mora (2007) for a more detailed explanation of the methods and their advantages and caveats.

1) Kaiser's (1958) criterion only keeps factors with an eigenvalue higher than one;

2) Visual inspection of the Catell (1966) scree plot. Cattell's rule is such that the number of factors should be equal to the number of eigenvalues before which the smooth decrease of eigenvalues appears to level off on to the right of the plot;

3) Velicer's (1976) Minimum Average Partial minimizes the unexplained partial correlation;

4) Horn 's (1965) Parallel Analysis keeps the factors as long as they explain more than the 95th percentile of eigenvalues from randomly generated data (Cota, Longman, Holden, Fekken, & Xinaris, 1993; Glorfeld, 1995).

Valero-Mora (2007) argues that the methods of Velicer and Horn are more reliable. Given that different methods do not always lead to the same conclusions, we opted for the number of factors most commonly suggested by the methods, putting more emphasis on the last two.

B.3. More details on the factor analysis of the noncognitive skills

The factor analysis explicitly accounts for the fact that answers to items are imperfect proxies of the true underlying latent traits. Latent factor models estimate the joint distribution of the latent factors and help remove some of this measurement error. We estimate factor loadings, then rotate the factor loadings using a principal factor analysis with quartimin rotation to predict the resulting factors. [2] Table 1 presents the resulting factor loads of the acquiescence bias corrected items, sorted by dominant factor.

Strikingly, the factor analysis does not result in a clear categorization of variables into the theoretical scales and subscales. With the exception of the first factor, most factors seem to have a mix of items from different sub-constructs (in theory meant to be measuring different latent skills).[3] CESD items are a clear exception. They uniquely load on two factors, which do not include other items, and separate negative from positive attitudes.[4] On the other hand the Big Five personality trait division typically found in the psychometrics literature is not confirmed by the factor structure, with the exception of conscientiousness related items, which mostly load on the second factor.[5] The fourth factor further raises some doubts as it is uniquely composed of the reverse questions from the "causes of poverty" sub-construct, while the positive ones load on other factors.[6] This may indicate that it is at least partially driven by a response pattern rather than the actual belief about the causes of poverty. Overall these results raise concerns about whether the scales actually measure what they intend to. Despite the

[2] The quartimin rotation re-weights the factor loadings so that each variable mostly loads on one factor. That said, some variables still load on multiple factors after rotation, and no further restrictions were imposed.

[3] This means for example that a question that is expected to measure agreeableness and a locus of control question can better correlate together (and thus be assigned to the same underlying factor) than two locus of control questions.

[4] The original scale development paper for the CESD (Radloff, 1977) similarly identifies a positive subscale/factor.

[5] A similar result is found when restricting the EFA to items of the Big Five. Items meant to measure distinct personality traits are mixed into various factors (Appendix Table A3). We return to this lack of congruence in section 4.

mixing of items, we attempted to discern a dominant interpretation for each factor, indicated in the last column of Table 1.

We use the factor loadings to aggregate the different noncognitive skills. To obtain the predicted factors, and following Attanasio et al (2015), items are assigned to the factor for which they have the highest factor loadings, with factor loads of other items set to 0.[7] To analyze the test-retest, and to guarantee we are comparing similar constructs, we apply the factor loading obtained from the first survey round (the test) also to the variable values of the second survey round (the retest). When redoing the exploratory factor analysis on the retest data, the factor structure is broadly similar, justifying the use of the same factor loads for both test and retest data. To obtain the aggregated non-cognitive skills construct, we use the average of the 6 factors.

B.4. Factor structure without correcting for acquiescence bias

The improved indexes used the factor analysis of items corrected for acquiescence bias. For comparison, Appendix Table A7 summarizes the sorting of items that results from the factor loads of the items without correction for acquiescence bias. This shows far less consistency in how the items are sorted, making it difficult to attribute a dominant interpretation to the factors. Instead, specific factors appear to be pooling questions with the same answer types and phrasing. The first, fifth and sixth factors are only pooling items that are not reversed, and second, third and fourth factors are pooling reversed items together. Almost all factors that are not in a one to five Likert-scale sorted themselves together in the sixth factor. In sum, without the correction of acquiescence bias, the share of the variance in the responses driven by acquiescence bias and other response patterns overwhelms variance in responses that is driven by the latent traits that are intended to be measured. A factor analysis that is driven by phrasing rather than actual content is arguably of little interest, hence prior correction for acquiescence bias is fundamental. The findings hence suggest it is advisable to correct for acquiescence bias first, and then systematically analyze the latent factor structure through exploratory factor analysis when using noncognitive skills data. Naïve interpretation of item aggregation following pre-existing constructs without such analysis is likely to lead to erroneous conclusions regarding noncognitive skills.

---

[6] The sorting or reverse versus non reverse questions in the factors is substantially stronger when not correcting for acquiescence bias (Table A7).

[7] Setting factor loads to 0 for all loadings other than the highest one helps reducing correlation between factors. The results are qualitatively very similar when we do not apply this correction.

B.5. Item Response Theory

Item Response Theory offers a structural way of using a set of items to measure a latent ability or trait. It is based on the idea that the probability of a correct/keyed response to an item is a mathematical function of person and item parameters. For example, in the case of binary items, it considers that the probability of getting the correct answer to each item is a logarithmic function of the difficulty of the item and the latent ability of the respondent. IRT simultaneously estimates the difficulty of each item and the ability of each respondent, such that it maximizes the likelihood of the responses observed in the data.

IRT has become the standard tool for high stakes tests such as GRE or GMAT because it is believed to provide a greater precision than Classical Test Theory.

IRT requires the following assumptions:

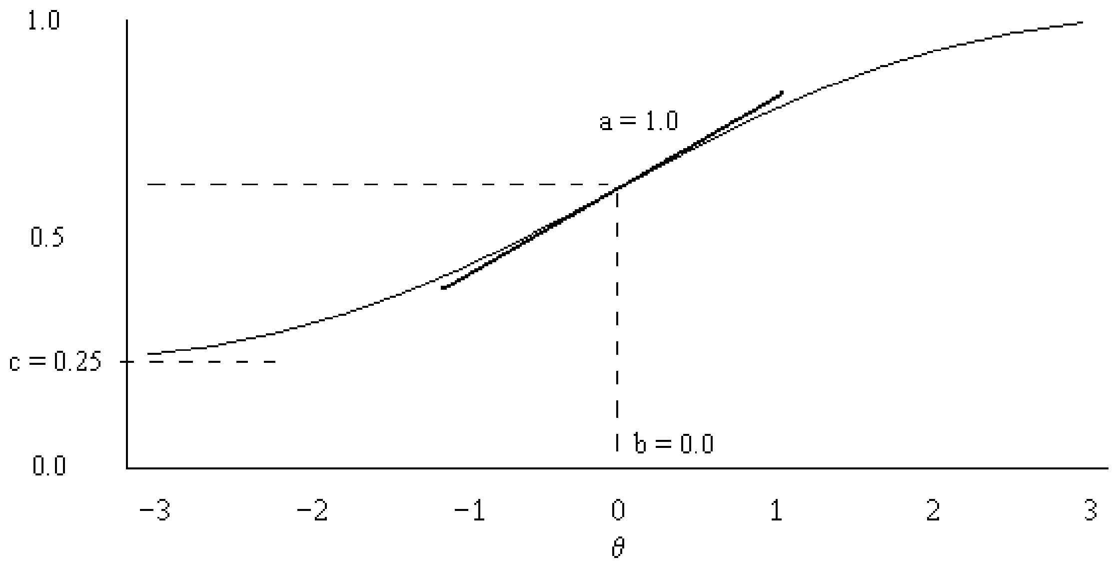
1) Uni-dimensionality[8]; assessed through factor analysis;
2) Local independence (which is violated for timed tests, or for tests for which one question affects answers to the following one(s));
3) Monotonicity (item characteristic curve well behaved – see below).

The graph below represents the Item Characteristic Curve (ICC) which is the probability of getting a correct answer to a given item, conditional on the respondent's underlying ability. In the one parameter model, also called Rasch Model, the difficulty of each item is the parameter estimated (where b-value is 0 in the graph). The two-parameter model also estimates the discriminant, which is the slope (a-value) at difficulty and can be interpreted as the effect of the underlying ability on the respondent's probability to answer the question correctly. The three-parameter model adds a pseudo guessing parameter (c-value), which estimates the probability of a respondent with lowest level of ability to obtain a correct answer.

Using these three parameters, the conditional probability of getting a correct answer to item i for an individual with underlying ability $\theta$is given by:

$$P_i(\theta) = c_i + (1 - c_i)\ \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

[8] IRT models that aim to capture multiple dimensions of ability also exist, but their convergence tends to require a large amount of data.



The Graded Response Model applies a similar logic, with multiple difficulty parameters in order to deal with ordered polytomous variables. More about GRM can be found in Van der Linden and Hambleton (2013).

IRT also allows hybrid models that combine the different types of model.

We apply IRT to cognitive skills and to technical skills to obtain the two "improved" constructs, in each case assuming unidimensionality given the result of the EFA.[9] For the technical skills, we used IRT pooling all items together and combine the Graded Response Model with the Two-Parameter Model. We do so, because we have two types of questions. The vast majority of questions are multiple choice questions where the respondent can choose only one possible answer. Based on this answer, we created a binary variable for whether the answer was correct or not. In some questions, however, it was possible to select multiple answers, in which case we created a count variable indicating the number of correct answers, but penalizing for wrong answers selected. Only 3 items were removed because they had a discriminant opposed to the expected one (meaning that respondents were more likely to have a correct answer if they had a lower predicted latent skill). Of the remaining 32 items, 28 had a significant discriminant parameter at the 5% level (and 24 items at the 1% level), indicating that most items contributed to the assessment of the latent trait.

In the cognitive sub-constructs, we used a hybrid of Three-Parameter Model and Two-Parameter Model, because for some questions the guessing parameter was found to be zero (in which case there

[9] We do not use IRT for the noncognitive skills, as the "difficulty" of each question is less applicable to noncognitive questions, and because IRT can only be used on discrete measures (after subtracting the acquiescence score are not).

is no gain from the three parameter model, which does not always converge). We applied a mixed method, given the format of the questions and the requirements of IRT. We first use IRT to calculate the subconstruct of the numeracy questions, the Raven test and reading test.[10] We then used factor analysis using these three indexes and the scores of the digit span, the reverse digit span, and the timed math test to obtain one latent factor.

For a general introduction to IRT, see Hambleton and Swaminathan (2013).

B.6. Tucker's Congruence Coefficient

The Tucker’s congruence coefficient (or simply congruence coefficient) is an index that assesses the similarity between factor structures of the same set of items applied to two different populations. One first applies a factor analysis to the two populations. In order to assess the similarity between a factor x and a factor y, after applying factor analysis to two different population, one calculates the correlation coefficient (by item) of the two vectors of factor loadings.

$$\varphi(x,y) = \frac{\sum_{i,} x_i y_i}{\sqrt{\left(\sum_i x_i^2\right)\left(\sum_i y_i^2\right)}}$$

Where $x_{i,j}$ and $y_{i,j}$ are the loadings of item i on factors x and y, respectively (each one extracted from applying the factor analysis of the same items to a different population).

$\varphi(x,y)$ can be interpreted as a standardized measure of proportionality of elements in both vectors. A coefficient that is equal to 1 corresponds to a perfectly identical factor structure between the two populations, while a coefficient equal to 0 corresponds to a factorial that is completely orthogonal.

For an order of magnitude, Lorenzo-Seva and Ten Berge (2006) indicate that a congruence coefficient over .95 implies a good similarity, and a range of [.85 - .94] shows fair similarity.

More about Tucker’s congruence coefficient can be found in Abdi (2007).

[10] IRT cannot be used on digit span, reverse digit span, and the timed math test given that its subcomponents are not independent from each other.

Appendix C: Questionnaire design and sources

A review of studies and questionnaires using different approaches to measure cognitive, noncognitive and technical skills of adults preceded the initial questionnaire design.[11] This appendix discusses the choices made to design the final instruments, including methods used to reduce the number of questions and hence the overall duration of the questionnaire. It also, provides information about the source of the different scales and tests used, and references other papers that use them.

C.1. Cognitive module

In most empirical work in development economics, a household's skill level is proxied by the education level of the household head, the maximum number of years of education in the household, or an individual's self-assessment of his literacy level. However, education is not always a strong correlate of productivity differences in agriculture. Existing literature reviews (Lockheed, Jamison and Lau, 1980; Phillips, 1994) indicate that the production increase resulting from four years of additional schooling is typically 7% to 8%. While the correlation is most often positive, in many papers it is not statistically significant. It may well be that grades attained or self-assessed literacy are not good measures of the farmers' active knowledge of reading or math. Farmers reading skills might matter e.g. for the processing of information regarding input use, and his math skills might be crucial to make optimal cost-benefit analysis. A priori it is also quite possible that it is a farmer's broader cognitive skills (such as memory, processing ability, or analytical thinking) rather than his classroom knowledge (such as reading or math) that help him adapt to the varying conditions of climate or soil. [12]

There are many tests that are designed specifically to measure cognitive skills but many are hard to apply as part of a large household survey. Not only do these tests typically require a level of standardization and quality control that goes beyond the usual training and supervision of household survey enumerators, they can also be very time-consuming, might require a standardized test-taking environment and/or specialized professional test administrators (such as licensed psychologists), might have content that is inappropriate for developing country settings, or require the use of test material that is unpractical in field circumstances. Moreover, any language-based tests are likely to suffer from lack of comparability across countries – and often also within countries - and lack of standardization

[11] We do not consider the literature on measuring skills for children and teenagers, as most instruments would not necessarily be relevant for adults See, for instance, Cueto and Leon (2012) for psychometric analysis of the skills measures in the Young Lives surveys.

[12] A useful distinction can be between *fluid intelligence* (the ability to solve novel problems) and *crystallized intelligence* (knowledge and developed skills) – Cattell (1987).

upon translation. Existing short and non-language based tests (often based on visual aids) that do not suffer from these limitations are sometimes used as alternative for inclusion in household surveys.

With the objective of measuring different aspects of adult farmers' cognitive ability, we selected five different cognitive tests: i) The 36 item Raven Colored Progressive matrices; ii) The digit span forwards and backwards; iii) A timed math test with 160 basic additions, subtractions and multiplications; iv) An oral 9-item math questions tests containing short math puzzles and increasing in level of difficulty; and v) A reading comprehension test. Table A1.A provides a detailed description for each of these tests.

Versions of the Raven and the digit span are very frequently used in surveys in developing countries (de Mel, Mckenzie and Woodruff, 2009a, 2010; Beaman and Magruder, 2012; Dupas and Robinson, 2013; Giné and Mansuri, 2014; Djankov et al., 2015); and the timed math tests has also been used before (Barham, Macours, Maluccio, 2017). Jamison and Moock (1984) used numeracy questions, literacy tests and raven tests. The specific math puzzles and the reading comprehension tests used in this paper were designed for the purpose of this experiment. The outcomes of these tests give us an observed outcome of the farmers' cognitive skills.

C.2. Noncognitive module

The choice of subscales was based on comparisons with the seminal papers in the literature on noncognitive skills, complemented with scales used in the literature on small business development in developing countries.[13] We also added measures used in the small but growing empirical literature on aspirations and locus of control in developing countries.

For each of the scales, we selected a subset of items to be included in the final survey instrument after piloting. We followed standard practices in psychology and broader insights from the psychometric literature, regarding selection of questions, question type and mode of analysis. In particular, questions of the different subscales as well as all 44 questions of the BFI were incorporated in the pilot version of the questionnaire. During piloting, a relatively large set of questions was identified with either very little variation (because everybody agreed with a certain positive statement), or a bi-modal distribution, typically in the case of reverse-coded questions. In extreme cases this led to negative correlations between variables that should capture the same latent trait. Qualitative field observations allowed to interpret the underlying answering pattern, with people either understanding the reverse question, in

[13] While most of these scales were originally designed for self-administration (i.e. respondents directly filling in answers), in Kenya they were asked by the enumerators to the respondent, reflecting how they are typically used in large household or individual surveys.

which case they often disagreed, or not understanding the reverse question, in which case they reverted to agreeing, as a fallback. Hence these distributions of the individual variables suggested a relatively high level of acquiescence bias or "ya-saying" in the study population. Variables were eliminated if they showed very little correlations with other variables belonging to the same construct, or showed very little variation.

For the Big Five personality traits, we use a version of the Big Five Inventory (BFI) written for a population with 5 years of education. The BFI is a commonly used instrument for the Big Five factor model, a higher module assumed to encompass the most generally important personality traits (John, Donahue, and Kentle, 1991; John, Naumann and Soto, 2008). The BFI has been used in the development economics literature by Dal Bo, Finan, and Rossi (2013); Callen et al. (2015). The BFI instrument has 44 items, and all 44 items were included in the pilot version of the instrument. After piloting, the number of BFI items was reduced to 23, keeping at least 3 questions for each personality trait, and a balance between the positive and reverse-coded items. The 23 items include the 10 items of the shorter BFI-10 scale (Rammstedt and John, 2007).

For Locus-of-Control we use a subset of the Levenson's (1981) "Internality, Powerful Others and Chance scales". This scale is used, for instance, by Acharya et al. (2007) and Bernard et al (2014). We also use a subset of items from the Attributions for Poverty scale (Feagin, 1972, 1975), similarly used in Bernard et al (2014).

We also added a series of locus-of-control questions with visual aids. The respondent was asked to allocate 10 beans between different answer options for three locus of control questions. The questions followed the general concepts of standardized locus-of-control instruments, and the visual aid aimed at increasing engagement and understanding of the respondents. For example a question asked the respondent to allocate 10 beans to three possible reasons for why some individuals have more economic success than others from: 1) Efforts and Decisions, 2) Luck, and 3) Birth.

The literature on the formation and predictive power of noncognitive skills in the US often uses measures of both locus-of-control and self-esteem (Heckman, Stixrud and Urzua, 2006; Heckman and Kautz, 2012; among many others). Following this literature, the questions of self-esteem used come from the Rosenberg (1965) scale. In development economics, similar measures are used in Blattman, Jamison and Sheridan (2016); Blattman and Dercon (2016); Adhvaryu, Kala and Nyshadham (2016).

A number of additional subscales were included because of their frequent use in the literature on small business development in developing countries. While this literature typically focuses on non-agricultural businesses, it seems plausible that some of the same characteristics may affect success in farming business. In particular, we followed studies analyzing what distinguishes entrepreneurs from others in developing contexts that have used measures of optimism, attitudes towards change, tenacity&organization, self-control, meta-cognitive activity, risk aversion and patience, similar to the ones we test, in addition to the BFI, internal locus of control, and self-esteem (Krauss et al. 2005; De Mel et al. 2009b, 2010; Giné and Mansuri, 2014; Djankov et al., 2015). Some of these subscales are conceptually closely related to one of the Big Five personality traits (with tenacity&organization, for instance, related to conscientiousness).

Most items were asked using a Likert scale, following the original scales. A few items were however changed to a binary scale, and some others we adapted to fit the agricultural context (see details of items in Table A1.B).

Finally, the CESD was added as it is often used in studies in developing countries, including in national representative panel surveys such as the Indonesian Family Life Survey and the Mexican Family Life Survey. It is arguably related to the Neuroticism personality trait. It consists of a set of 20 questions asking about the respondents' emotions in the last 7 days. As such it is more direct and arguably less abstract than the Likert scale questions. One additional question was added to capture perceptions of upward mobility using the same 7 day format.

C.3. Technical module

Technical skills and agricultural knowledge required for farming are likely to differ a lot from context to context. For this reason rather than replicating specific questions from prior survey instruments, we reviewed the literature to categorize existing approaches, and then designed a questionnaire that uses similar categories, but with specific questions adapted to the study population.

The majority of studies measuring technical knowledge do it with the intention of evaluating learning from a training provided to farmers, for instance through Farmer Field Schools (Godtland et al. 2004; Feder et al 2004; Maureci et al. 2007; David 2007; Buck and Alwang 2011). Consequently, they tend to provide an assessment of the practices that are taught by the intervention in order to track progresses in related knowledge. A few studies apply a broader technical knowledge assessment, including Goldstein and Udry (1999), Hanna, Mullainathan, and Schwartzstein (2014), Kondylis, Mueller and Zhu (2015) and the Ethiopia Rural Household Surveys (ERHS) of 1999.

Based on a review of questionnaires used in those studies, we classified in Appendix Figure A1 questions based on form, (i.e. how they evaluate technical skills), the technology or practices referred to when assessing skills, and the type of knowledge asked about when assessing skills.

With regard to the form, knowledge tests with a series of multiple-choice or open-ended questions are used in a number of studies, and the skill measures used in this paper are constructed from such questions.[14]

When assessing skills, the technology or practices referred to have to be a function of the crops in the region of study. Hence for a general knowledge tests, initial fieldwork and local knowledge is important to first identify which crops, technologies and practices are most common in the region and can best help distinguish farmers with best practices from less knowledgeable ones. The particular crops, technologies and practices referred to in our survey instrument were based on qualitative fieldwork prior to questionnaire design and knowledge of local agronomists.

The review of the literature further revealed a relatively large commonality regarding the types of knowledge that are being assessed (even if it were applied to different crops and practices). For example, questions often ask for mode of application, quantity and timing of inputs, all common practical issues faced by farmers with important consequences for yield. The ability to recognize deficiencies, pests etc. are also common. A potential challenge for such questions is that the optimal practice may depend on a number of other factors, making it hard to evaluate whether the answer provided by a farmer is "correct" or not. A different type of question asks for theoretical knowledge such as, for instance, the type of nutrients included in certain fertilizers. These have the advantage of having unambiguous correct answers, but one can wonder whether they capture the type of practical knowledge that matters for productivity (in case farmers, for instance, know which fertilizer to use and when, but do not know its composition). Whether such theoretical questions are good predictors of practices and productivity is part of the questions of interest for this study.

Based on the categorizing of existing questions in the literature, we designed an instrument that covered the different types of knowledge for the practices and technologies relevant in the region of study. Extensive fieldwork in cooperation with local agronomists was required to design, test and adapt the questions. As for the noncognitive module, only questions showing sufficient variation in answers during piloting were kept. This led to exclusion of certain practices (such as pest management or irrigation) as knowledge about them was extremely limited in the region.

[14] A relatively large number of questionnaires also ask farmers to self-assess their level of knowledge. Alternatively, farmers are sometimes asked what they actually do rather than what they know. The former is likely prone to subjectivity, while the later measures the combination of many other constraints (budget, time, etc.) in addition to differences in technical skills. Given these concerns, this study focuses agricultural knowledge tests.

Table A1.A: List of tests of the cognitive skill module

Oral Math questions	An oral 9-item math test containing short math puzzles and increasing in level of difficulty. Each puzzle contains one or two phrases including the question. Answers can be open or multiple choice, some questions are framed to mimic math problems farmers might need to solve in daily live but they never require actual farming knowledge.
Reading	A reading comprehension test. Farmers were given 3 small texts of 5 to 7 lines (2 in Swahili and 1 in English or vice versa). In each exercise, they were asked to read the text and then asked 4 questions about the text (2 multiple choice and 2 open). They were allowed to consult the text while answering. The texts present (fictitious) information regarding agricultural inputs and were inspired by guidelines found on input packages. No time limit was imposed.
Raven	The 36 item Raven Colored Progressive matrices, measuring visual processing and analytical reasoning and executive functioning. People are asked to select the missing puzzle piece from a bigger image, with patterns becoming increasingly difficult.
Math (timed)	A timed math tests with 160 basic additions, subtractions and multiplications. Respondents are given a sheet of paper with the questions, and get 3 minutes to complete as many as they can.
Digit Span	The digit span forwards and backwards (measuring short-term memory and executive functioning). People are asked (without visual aid) to repeat a series of numbers that the enumerator reads to them. It starts with a 3 number series with the series becoming progressively longer as long as the respondent manages to repeat the series correctly. Afterwards, they are asked to repeat different series backwards (starting from a 2 number series and again gradually increasing the length of the series).

Correlations between cognitive measures, education and literacy

	Oral Math questions	Reading	Raven	Math (timed)	Digit Span	Literacy dummy	Years of Education
Oral Math questions	1						
Reading	0.61	1					
Raven	0.52	0.51	1				
Math (timed)	0.57	0.62	0.46	1			
Digit Span	0.46	0.48	0.42	0.46	1		
Literacy dummy	0.41	0.57	0.36	0.55	0.40	1	
Years of Education	0.58	0.70	0.47	0.65	0.49	0.65	1

The first 3 sub-constructs are calculated using item response theory

Table A1.B: List of items of the noncognitive skill module

Subscale / Naïve Category	Question's short name	Question	Positive or Reversed	Answer scale
Locus of Control	LOC1	It's not always wise for you to plan too far ahead because many things turn out to be a matter of good or bad fortune	R	1 to 5
	LOC2	Your life is determined by your own actions	P	1 to 5
	LOC3	When you get what you want, it's usually the result of actions	P	1 to 5
	LOC4	You feel like what happens in your life is mostly determined by others	R	1 to 5
	LOC5	Getting what you want requires pleasing the influential people	R	1 to 5
	LOC6	Please tell me which of the two propositions you most agree with 1. Each person is primarily responsible for his/her own success or failure in life 2. One's success or failure is a matter of his/her destiny	NA	Choose 1 answer
	LOC7	Only those who inherited large farms become successful farmers	R	1 to 5
Beans (locus-of-control with visual aid)	LOC_va1	What do you think explains why some people have more ECONOMIC SUCCESS than others? The enumerator records number of beans allocated to "Effort or Decisions" (Compared to "luck or birth") [with additional explanations and visual aid]	P	Allocate 10 beans
	LOC_va2	What do you think explains why some people are more PRODUCTIVE IN AGRICULTURE than others? The enumerator records number of beans allocated to Effort or Decisions (Compared to luck or birth) [with additional explanations and visual aid]	P	Allocate 10 beans
	LOC_va3	Between effort and good decision-making, how much do you think that each one matters for being productive in agriculture [with additional explanations, visual aid and the respondent allocating beans to the possible options]	NA	Allocate 10 beans
Self-esteem	selfesteem1	You feel that you have many good qualities	P	1 to 5
	selfesteem2	All in all, you are inclined to feel that you are a failure	R	1 to 5
	selfesteem3	On the whole, you are satisfied with yourself	P	1 to 5
	selfesteem4	You certainly feel useless at times	R	1 to 5
Causes of Poverty	causepov1	Poor people are poor because they lack the ability to manage money	P	1 to 5
	causepov2	Poor people are poor no matter what they do	R	1 to 5
	causepov3	Poor people are poor because they waste their money on inappropriate items.	P	1 to 5
	causepov4	Poor people are poor because they do not actively seek to improve their lives.	P	1 to 5
	causepov5	Poor people are poor because they are exploited by rich people.	R	1 to 5
	causepov6	Poor people are poor because the distribution of land between poor and rich people	R	1 to 5
	causepov7	Poor people are poor because they lack opportunities because they come from poor	R	1 to 5
	causepov8	Poor people are poor because they lack luck	R	1 to 5
	causepov9	Poor people are poor because they are born with less talent	R	1 to 5
Attitude toward Change	att_change1	When you learn about a new farming technique, compared to most of your neighbours: 1. You are more willing to try first 2. You let others try it first	NA	Choose 1 answer
	att_change2	On the farm: 1. You prefer doing routine things 2. You prefer doing something new	NA	Choose 1 answer
	att_change3	Choose one of the following 2 options: 1. You generally prefer leaving things the way they are 2. You generally prefer changing things	NA	Choose 1 answer
	att_change4	You often go to the plots of fellow farmers to observe what they do	P	1 to 5
	att_change5	You have tried to experiment on your own plot some of the techniques learned from fellow farmers.	P	1 to 5
Tenac/Organiz	tenac1	You can think of many times when you persisted with work when others quit	P	1 to 5

/Self-cont.	tenac2	You normally don't rest until the job is fully completed	P	1 to 5
	tenac3	Your family and friends would say you are a very organized person	P	1 to 5
	tenac4	You are much happier if everything is planned well ahead of time	P	1 to 5
	tenac5	You often spend money and regret later that you spent it	R	1 to 5
	tenac6	When you see something you like, you buy it right away, rather than waiting to see how you feel about it later	R	1 to 5
Metacognitive	metacog1	You think a lot about something before taking a decision about it	P	1 to 5
	metacog2	You set goals for yourself in order to direct your activities	P	1 to 5
	metacog3	You spend a lot of time reflecting on your mistakes in order to improve your farming practices	P	1 to 5
Optimism	optim1	In uncertain times you usually expect the best.	P	1 to 5
	optim2	Things go wrong for me most of the time.	R	1 to 5
	optim3	You talk more about solutions than problems.	P	1 to 5
CESD	cesd1	During the last 7 days, how many days … were you bothered by things that usually don't bother you?	R	0 to 7
	cesd2	… did you not feel like eating? (your appetite was poor)	R	0 to 7
	cesd3	… did you feel that you could not shake off the blues even with help from your family and friends?	R	0 to 7
	cesd4	… did you feel that you were just as good as other people?	P	0 to 7
	cesd5	… did you have trouble keeping your mind on what you were doing?	R	0 to 7
	cesd6	… did you feel depressed?	R	0 to 7
	cesd7	… did you feel that everything you did was an effort?	R	0 to 7
	cesd8	… were you hopeful about the future?	P	0 to 7
	cesd9	… did you think your life had been a failure?	R	0 to 7
	cesd10	… did you feel fearful?	R	0 to 7
	cesd11	… was your sleep restless?	R	0 to 7
	cesd12	… were you happy?	P	0 to 7
	cesd13	… did you talk less than usual?	R	0 to 7
	cesd14	… did you feel lonely?	R	0 to 7
	cesd15	… people were unfriendly?	R	0 to 7
	cesd16	… did you enjoy life?	P	0 to 7
	cesd17	… did you have crying spells?	R	0 to 7
	cesd18	… did you feel sad?	R	0 to 7
	cesd19	… did you feel that people disliked you?	R	0 to 7
	cesd20	... could you not get 'going'?	R	0 to 7
	cesd21	… did you feel that you are moving forward in life?	P	0 to 7
Big 5 Agreeableness	BF_A1	You see yourself as someone who tends to find fault with others	R	1 to 5
	BF_A2	You see yourself as someone who has a forgiving nature	P	1 to 5
	BF_A3	You see yourself as someone who is generally trusting	P	1 to 5
	BF_A4	You see yourself as someone who is sometimes rude to others	R	1 to 5
Big 5 Conscientious ness	BF_C1	You see yourself as someone who does things carefully and completely	P	1 to 5
	BF_C2	You see yourself as someone who can be somewhat careless	R	1 to 5
	BF_C3	You see yourself as someone who tends to be disorganized	R	1 to 5
	BF_C4	You see yourself as someone who tends to be lazy	R	1 to 5

	BF_C5	You see yourself as someone who does things efficiently (quickly and correctly)	P	1 to 5
	BF_C6	You see yourself as someone who makes plans and sticks to them	P	1 to 5
Big 5 Extraversion	BF_E1	You see yourself as someone who is reserved; keeps thoughts and feelings to self	R	1 to 5
	BF_E2	You see yourself as someone who generates a lot of enthusiasm	P	1 to 5
	BF_E3	You see yourself as someone who tends to be quiet	R	1 to 5
	BF_E4	You see yourself as someone who is outgoing, sociable	P	1 to 5
Big 5 Neuroticism	BF_N1	You see yourself as someone who is depressed, or gets blue	R	1 to 5
	BF_N2	You see yourself as someone who is relaxed, handles stress well	P	1 to 5
	BF_N3	You see yourself as someone who doesn't get easily upset, and is emotionally stable	P	1 to 5
	BF_N4	You see yourself as someone who gets nervous easily	R	1 to 5
Big 5 Openness	BF_O1	You see yourself as someone who is clever, thinks a lot	P	1 to 5
	BF_O2	You see yourself as someone who has an active imagination	P	1 to 5
	BF_O3	You see yourself as someone who likes work that is the same every time (routine)	R	1 to 5
	BF_O4	You see yourself as someone who likes to think and play with ideas	P	1 to 5
	BF_O5	You see yourself as someone who doesn't like artistic things (plays, music)	R	1 to 5
Risk Aversion	riskav1	You never try anything you are not sure of	R	1 to 5
	riskav2	A person can get rich by taking risks	P	1 to 5
	riskav3	Imagine that you can chose between 5 games in which you will flip a coin. First I am going to explain you 5 games, and then I am going to ask you which one you would prefer to play. In the 1st game, you get 2500 Ksh if you get head, and 2500 Ksh if you get tail 2nd game 2000Ksh vs 4000 Ksh 3rd game 1500 Ksh vs 5500 Ksh 4th game 1000 Ksh vs 7000 Ksh 5th game 0 Ksh vs 10000 Ksh Which game would you pick? The question includes more explanations and a table to visualize the choices.	P	Choose 1 answer. An index is calculated based on the response.
Patience	patience1	People often make decisions that involve trading off something soon for something else later. For example, people sometimes have to choose between having some money soon, or having more money later. The next set of questions asks how you make such decisions. There are no right or wrong answers. For each pair of options please indicate which you prefer between option (1) and option (2). Would you prefer: (1) 1000 Ksh now, or (2) 900 Ksh in one month? [following questions asked as long as the respondent picks (1)]: (1) 1000 Ksh now, or (2) 1100 Ksh in one month? (1) 1000 Ksh now, or (2) 1300 Ksh in one month? (1) 1000 Ksh now, or (2) 1500 Ksh in one month? (1) 1000 Ksh now, or (2) 2000 Ksh in one month? (1) 1000 Ksh now, or (2) 2500 Ksh in one month?	NA	Choose 1 answer per question. With Visual representation. An index is calculated based on the responses

Figure A1: Classification technical skills questions

Form of Evaluation
Test: Multiple choice or open questions about best practices, assessing whether the respondent finds the right answer.
Self-assessment: subjective assessment or "do you know…"
What the farmer does: use of practices or technology
Sources of information: training received, extension, etc.

Technology or Practices
Seeds
Fertilizer (mineral / biofertilizer)
Herbicide, Pesticide or Integrated Pest Management
Irrigation
Soil management practices: - Manure, compost, use of stalk - Rotation, intercropping - Tillage
Planting practices (number of seeds, spacing, gapping, etc.)
Storage / usage / commercialization

Type of knowledge
How to apply an input: - Where to apply it - What quantity - Timing of application - Other decisions (spacing…)
Recognizing: pests, plant deficiencies, better seeds… to decide what inputs or practices to apply.
How to use a complex practice (composting, fertilizer mix...)
Theoretical knowledge (e.g. name of nutrients in mineral fertilizer)

Table A1.C: List of items of the technical skill module

Subscale	Question	Listed Answers (when not an open question)
Maize	If one wants to cover the soil with maize stalk, should you apply or leave maize stalks:	1. Between the lines 2. On the lines, as close as possible to the next crop
	When planting hybrid maize in rows, how many seeds per hole should be applied?	
	What quantity of planting fertilizer should you apply per seed of maize :	1. Less than half of a Teaspoon 2. Half of a Teaspoon 3. A full Teaspoon 4. Two Teaspoons
	Where should you apply commercial planting fertilizer for maize: [distances shown with ruler]	1. In the same hole mixed up with the soil 2. In the same hole in contact with the seed 3. 5 cm from the hole 4. 15 cm from the hole
	Imagine a maize field is inclined like this [show]. If on such a field you need to put top dressing on the ground, where do you put the fertilizer?	1. Uphill 2. Downhill 3. On the side 4. Same hole
	How many weeks after planting should you apply commercial top dressing to maize?	
	Where should you apply commercial top-dressing fertilizer for maize:	1. In contact with the plant 2. Spread closely around the plant 3. At 15 cm from the plant 4. Apply through broadcasting
Banana	When cultivating bananas, how many adult trees should be left per banana mat?	
	(for banana) How many of the youngest trees (suckers) should you leave on a mat?	
	When do you need to prune the leaves of banana trees:	1. Never 2. When the leaves start turning yellow 3. When the leaves are completely dry 4. Prune only the green leaves
	When planting bananas, what is the optimal distance between banana trees:	1. 1m x 1m 2. 2m x 2m 3. 2m x 3m 4. 3m x 3m
	If you want to keep only one of two healthy suckers, which one should you leave:	1. The one facing the sunrise 2. The one facing the sunset 3. The youngest one
	What can you do to prevent the Cigar-end disease: [show image]	1. Remove the male part 10 days after bunch formation 2. Remove the male part 10 days before bunch formation 3. Make sure that the male part does not fall 4. Increase the water provided to the tree
Soya	When planting soybean in rows, how many seeds per hole should be applied?	
	When planting soybean, what is the optimal distance between seeds:	1. 10cm x 30cm 2. 20cm x 30cm 3. 30cm x 30cm 4. 50cm x 30cm 5. 50cm x 5cm
	How is powder biofertilizer used when planting soybeans:	1. It is applied directly to the soil and then soybean is planted 2. The biofertilizer is mixed with the seed and a sticky solution if needed 3. Put the soybean first and then put biofertilizer on top of it 4. Fill a bucket of water, pour the biofertilizer in, and then the soybean is soaked in it
	How much time should be left between mixing the seeds with powder biofertilizer and planting the seeds:	1. 5 min 2. 4 hours 3. 8 hours 4. 24 hours
intercrop/ compost	Imagine that someone intercrops beans and maize in the same field. In which order should he plant:	1. Plant the maize first and then the beans 2. Plant the beans first and then the maize 3. Plant both at the same time 4. He should not intercrop maize and beans
	Among the following crop rotations, which one is best for long term soil fertility:	1. Rotate soya with soya 2. Rotate soya with maize 3. Rotate maize with millet 4. Rotate beans with soya

	How can you use Nepia grass and Desmodium to control maize stalk borer: [Answers come with corresponding images]	1. Plant Desmodium with the maize and put Nepia grass around the parcel 2. Plant Nepia grass with the maize, and Desmodium around the parcel 3. Intercrop both Desmodium and Nepia grass with the maize 4. Rotate Maize with Desmodium and Nepia grass
	Imaging you are making compost. While it is maturing, where should it be stored:	1. In a uncovered pit 2. In an uncovered heap 3. In a covered heap 4. Inside of the house
	Is it better to apply compost when it is humid or when it is dry?	1. Humid. 2. Dry
	If you want to use the waste from your own cattle to improve the fertility of the soil, is it better to:	1. Apply some manure everyday in part of the field2. Keep it covered and then apply it all at once3. Keep it uncovered and then apply it all at once
	Please tell me all the different ways you can you use to check whether the compost is ready to be applied to the field?	10 possible components were listed and multiple answers were allowed.
Fertilizer	In the cultivation of banana, which fertilizer should be applied at planting?	4 pictures of fertilizers are shown Multiple answers allowed
	In the cultivation of banana, which fertilizer should be applied at the vegetative stage?	4 pictures of fertilizers are shown Multiple answers allowed
	In the cultivation of banana, which fertilizer should be applied at flowering?	4 pictures of fertilizers are shown Multiple answers allowed
	[A picture of a fertilizer is shown] Do you think it is:	1. Planting Fertilizer 2. Top Dressing 3. Both
	[A picture of a fertilizer is shown] Do you think it is:	1. Planting Fertilizer 2. Top Dressing 3. Both
	[A picture of a fertilizer is shown] Do you think it is:	1. Planting Fertilizer 2. Top Dressing 3. Both
	[A picture of a fertilizer is shown] Do you think it is:	1. Planting Fertilizer 2. Top Dressing 3. Both
	Which ones of these fertilizers should be used on Sweet Potatoes?	4 pictures of fertilizers are shown Multiple answers allowed
	Which ones of these fertilizers provide Nitrogen?	4 pictures of fertilizers are shown Multiple answers allowed
	Which ones of these fertilizers provide Phosphorous?	4 pictures of fertilizers are shown Multiple answers allowed
	Which ones of these fertilizers provide Potassium?	4 pictures of fertilizers are shown Multiple answers allowed

Appendix D: Colombia Replication

Figure D1: Relationships between constructs and other indicators (Colombia sample)

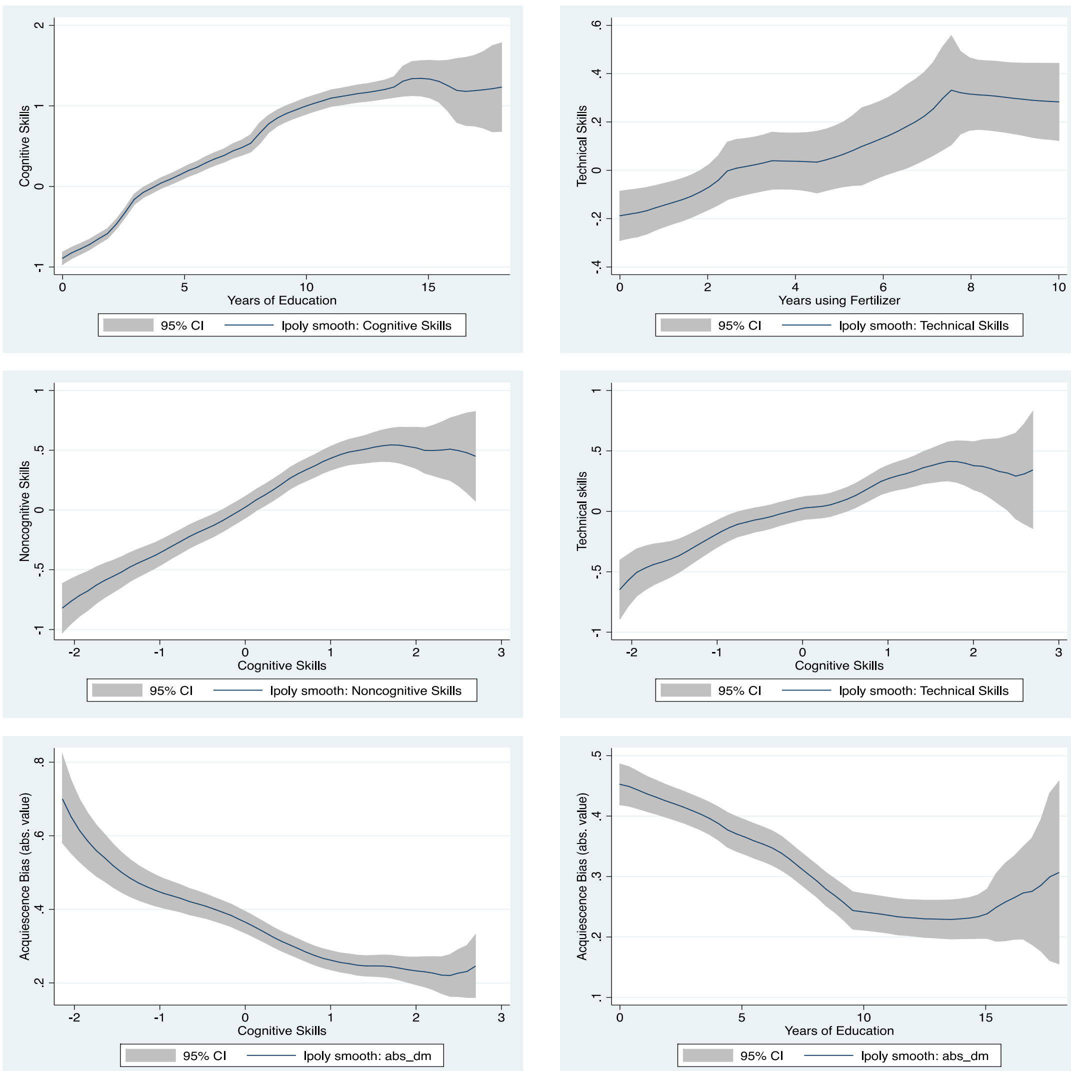


Table D1: Factor loads of noncognitive items, corrected for acquiescence bias (Colombia sample)

Factor	Label	Factor Loadings > .3
Factor 1	Conscientiousness	11 Big Five personality questions (5 conscientiousness; 3 openness; 2 Agreeableness; 1 Extraversion)
	Tenacity	2 Tenacity
	Other	1 Self-esteem; 1 Metacognitive; 1 Grit scale
Factor 2	Locus of control	4 Locus of control
	Grit scale	3 items of Grit scale
	Extraversion	3 Big Five personality questions (2 Extraversion; 1 Openness)
	Other	1 Beans; 1 Attitudes towards change; 1 Causes of poverty
Factor 3	Other	1 CESD positively phrased question 1 Big Five personality question (Conscientiousness)
Factor 4	Leadership	4 Leadership
	Other	1 Big Five personality question (Extraversion)
Factor 5	Causes of poverty	3 items of causes of poverty scale
Factor 6	Locus of control	1 Locus of control with visual aid (beans)
Factor 7	Other	1 Leadership 1 Self-esteem

Table D2: Measures of reliability and Internal Consistency (Colombia sample)

2A - Naïve Score

	Construct	Test-retest correlation	Cronbach's alpha of test	Cronbach's alpha of retest	Number of items
	Cognitive	0.78	0.62	0.65	6
	Noncognitive	0.70	0.70	0.68	15
	Technical	0.50	0.33	0.31	7
	Decomposition by subconstruct				
Cog	Arithmetic	0.69	0.60	0.65	9
	Reading	0.88	0.88	0.89	8
	Raven	0.81	0.89	0.90	36
	Math (timed)	0.93	0.98	0.98	149
	Digit Span	0.48	NA	NA	1
	Digit Span Backwards	0.50	NA	NA	1
Non Cog	Locus of Control	0.59	0.52	0.48	7
	Self-esteem	0.51	0.27	0.52	5
	Causes of poverty	0.66	0.55	0.64	4
	Attitudes towards change	0.44	0.21	0.25	3
	Organization/Tenacity/Self-Control	0.31	0.53	0.59	4
	Metacognitive ability	0.21	0.35	0.30	4
	Big 5 Agreeableness	0.46	0.43	0.43	4
	Big 5 Extraversion	0.47	0.48	0.50	4
	Big 5 Conscientiousness	0.51	0.49	0.47	6
	Big 5 Neuroticism	0.48	0.42	0.38	4
	Big 5 Openness	0.43	0.60	0.55	4
	CESD	0.47	0.65	0.68	10
	Grit Scale	0.44	0.48	0.48	5
	Leadership	0.60	0.71	0.76	5
	Ladder	0.57	0.67	0.77	2
Tech	Rotation/Intercrop	0.47	0.22	0.19	5
	Maize	0.49	0.19	0.21	7
	Roots	0.40	0.18	0.11	7
	Beans	0.31	0.01	0.12	2
	Fertilizer	0.25	0.07	0.09	5
	Organic Soil	0.34	0.06	0.14	3
	Machines/Pest	0.43	0.24	0.07	3

2B - Improved Index

	Construct	Test-retest correlation	Cronbach's alpha	Number of items
	Cognitive	0.94	0.81	6
	Noncognitive	0.64	0.69	7
	Technical	0.62	.	1
	Decomposition by subconstruct			
Cog (factor and IRT)	Arithmetic	0.73	Same as in Panel 2A	
	Reading	0.87		
	Raven	0.81		
	Math (timed)	Same as in Panel 2A		
	Digit Span			
	Digit Span Backwards			
Non Cog (factor analysis)	factor1	0.60	0.74	23
	factor2	0.35	0.64	10
	factor3	0.57	0.63	9
	factor4	0.53	0.63	9
	factor5	0.51	0.50	10
	factor6	0.67	0.49	6
	factor7	0.44	0.19	4
Tech (factor analysis)	Technical	0.62	0.44	25

Table D3: Test-retest correlations, Cronbach's alpha and influence of enumerators by subgroups (Colombia Sample)

	Test-retest correlation				Cronbach's alpha		R2 of enumerator FE		
Sample split:	Enumerator assigned for test and retest		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	0.94	0.93 (0.039)	0.83	0.81	0.46	0.51	0.05 [0.709]	0.04 [0.609]	0.07 [0.177]
Noncognitive	0.66	0.61 (0.164)	0.54	0.66	0.64	0.65	0.04 [0.099]	0.10 [0.023]	0.04 [0.705]
Technical	0.66	0.58 (0.036)	0.57	0.64	0.44	0.40	0.21 [0.000]	0.26 [0.001]	0.21 [0.187]

Note: P-value between parentheses indicate significance of the difference between same and different enumerator assigned for test and retest. R2 of Enumerator FE is the R2 of a regression of the improved construct on (randomly) assigned enumerator fixed effects. The p-values reported in parentheses and brackets were obtained by Randomization inference (10,000 repetitions).

Table D4: Correlation of different skill proxy measures with subscales measuring same domain (Colombia sample)

		Correlation with corresponding subconstruct			Test-retest correlation	
Question (CO)	Corresponding subconstruct	Self-assessment	Other village member	Average 2 village members	Asking different person about same person	Asking same person about different person
How rapidly do you understand things?	Raven	0.12	0.11	0.17	0.07	0.23
How good are you at math?	Math (timed)	0.41	0.33	0.41	0.20	0.21
How active and motivated are you?	Noncognitive skills	0.30	0.01	0.06	0.04	0.28
How much knowledge do you have in agricultural practices?	Technical skills	0.04	0.06	0.09	0.13	0.31

OTHER APPENDIX TABLES

Table A2: Number of factors to be retained according to different methods

	Number of factors recommended according to the following methods:				
	Kaiser's eigenvalue rule	Cattell's scree plot	Velicer's MAP rule	Horn's parallel analysis (p95)	Retained for analysis
Cog	1	1	1	1	1
Tech	1	1 or 3	1	8	1
Noncog naïve	7	3 or 7 or 9	4	9	7
Noncog demeaned	22	6	3	10	6
Big 5 demeaned	1	1 or 5	1	3	5

Table A3: Factor Loads of Big 5 personality traits

Question's short name	Fact Load 1	Fact Load 2	Fact Load 3	Fact Load 4	Fact Load 5
BF_C7	**0.57**	0.08	0.00	0.05	-0.01
BF_C1	**0.54**	0.00	-0.01	-0.05	0.08
BF_C8	**0.53**	0.05	0.02	0.07	-0.03
F_E8	**0.37**	-0.08	0.25	0.11	-0.08
BF_A5	**0.29**	0.10	0.19	-0.03	-0.06
BF_C4	**0.22**	0.13	0.06	0.09	0.18
BF_N2	0.07	**0.46**	-0.12	0.04	0.05
BF_O4	0.15	**0.34**	0.08	0.06	-0.16
BF_N1	0.07	**0.32**	0.09	0.06	0.10
BF_N5	-0.09	**0.31**	0.27	0.08	0.04
BF_C5	0.12	**0.30**	0.00	0.02	0.14
BF_O3	0.18	**0.27**	0.12	-0.12	-0.04
BF_E4	0.15	**0.26**	0.12	0.03	-0.23
BF_A4	0.13	-0.01	**0.41**	0.06	0.03
BF_A1	0.07	0.05	**0.31**	-0.05	0.21
BF_O8	0.17	0.20	**0.20**	0.03	-0.12
BF_O9	-0.04	0.09	**0.12**	-0.03	-0.04
BF_E2	-0.01	-0.02	0.03	**0.39**	0.01
BF_E5	0.07	-0.07	-0.10	**0.34**	0.06
BF_C2	0.04	0.09	-0.02	**0.34**	0.02
BF_N8	0.07	0.10	-0.07	**0.23**	0.05
BF_O7	-0.25	0.05	0.11	**0.18**	-0.19
BF_A8	0.04	0.09	0.15	0.09	**0.37**

All items were demeaned to correct for acquiescence bias

Table A4: Comparison of naïve scores in test and retest

	Average Naïve score				
	Test		Retest		p-value of difference
	Average Naïve score	Standard Deviation	Average Naïve score	Standard Deviation	
Cognitive	0.425	0.167	0.454	0.169	0.000
Noncognitive	3.419	0.282	3.458	0.281	0.000
Technical	0.409	0.107	0.431	0.108	0.000

Only observations available for both test and retest are kept

Table A5: Measures of reliability and validity for noncognitive measures corrected for acquiescence bias

	Construct	2A - Naïve Score			
		Test retest correlation	Cronbach's alpha of test	Cronbach's alpha of retest	Nb of items
	Noncog DE-MEANED	0.53	0.78	0.79	15
Decomposition by subconstruct:					
Noncog	Locus of Control	0.45	0.50	0.51	9
	Self-esteem	0.32	0.37	0.41	4
	Causes of poverty	0.34	0.69	0.74	9
	Attitude towards change	0.41	0.39	0.46	5
	Organization/tenacity/self-control	0.29	0.37	0.32	6
	Metacognitive ability	0.31	0.44	0.55	4
	Optimism	0.23	0.04	0.09	3
	Risk aversion	0.12	0.03	0.14	2
	Big 5 Agreeableness	0.25	0.43	0.38	4
	Big 5 Extraversion	0.23	0.32	0.26	4
	Big 5 Conscientiousness	0.33	0.57	0.57	6
	Big 5 Neuroticism	0.26	0.41	0.36	4
	Big 5 Openness	0.19	0.32	0.34	5
	CESD	0.41	0.82	0.85	21

Table A6: The effect of time (instrumented date of survey) on scores

	Test			**Retest**		
	Cognitive	Noncog	Technical	Cognitive	Noncog	Technical
I. First Stage						
Assigned Order	1.363***	1.363***	1.364***	1.718***	1.704***	1.701***
	(0.0410)	(0.0411)	(0.0410)	(0.109)	(0.109)	(0.108)
II. Second Stage						
Day of survey	0.000442	0.00233	0.00137**	-0.000373	-6.48e-05	-1.09e-05
	(0.000851)	(0.00144)	(0.000536)	(0.000698)	(0.00119)	(0.000449)
Observations	922	919	921	895	884	893

Note: First stage instruments the date (day since start of survey) with the randomly assigned order of survey assigned in planning. All regressions include randomly assigned enumerator fixed effects. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A7: Factor Loads of noncognitive items
(items taken as they are, without correction for acquiescence bias)

Factor	**Label**	**Factor Loadings >.3**
Factor 1	Acquiescence Bias Mix of items from different scales, but all positively phrased (except 1 item that is not positive nor reversed)	12 Big Five personality questions: 3 Conscientiousness; 3 Openness; 2 Agreeableness; 2 Extraversion; 2 Neuroticism 3 Metacognitive 3 Tenacity 2 Locus of Control 1 Self-esteem
Factor 2	CESD, all negative	14 CESD negatively-phrased items
Factor 3	Causes of Poverty (all reverse items) Other	6 Causes of Poverty items 1 Locus of Control
Factor 4	Negative acquiescence bias Mix of items from different scales, but all reversed (except 1 item that is not positive nor reversed)	9 Big Five personality questions (3 Conscientiousness; 2 Extraversion; 2 Agreeableness; 2 Neuroticism) 2 Self-esteem 3 Locus of Control 1 Tenacity 1 Optimism
Factor 5	CESD positive	5 CESD positively-phrased items
Factor 6	Attitudes towards change (not positive nor reversed)	2 Attitudes towards change items
Factor 7	Other	1 CESD negatively-phrased item

Table A8: Regressions of the average of log maize yield across seasons on skill constructs

	SKILLS CONSTUCTS USED AS REGRESSORS:									
VARIABLES	Naïve Score	Improved Index	Mean Naïve Score	Mean improved Index	Mean improved Index	Naïve Score	Improved Index	Mean Naïve Score	Mean improved Index	Mean improved Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cognitive skills	0.10**	0.10**	0.08*	0.06	0.15***	0.14**	0.18***	0.18**	0.20***	0.23***
	(0.041)	(0.041)	(0.043)	(0.044)	(0.042)	(0.068)	(0.067)	(0.071)	(0.072)	(0.070)
Noncognitive skills	0.13***	0.11***	0.16***	0.13**	0.17***	0.12**	0.12**	0.15***	0.13**	0.15**
	(0.040)	(0.041)	(0.046)	(0.048)	(0.050)	(0.047)	(0.053)	(0.057)	(0.061)	(0.061)
Technical skills	0.16***	0.16***	0.25***	0.27***		0.03	0.04	0.10*	0.13**	
	(0.045)	(0.043)	(0.050)	(0.048)		(0.052)	(0.048)	(0.063)	(0.061)	
Observations	900	890	900	890	890	900	890	900	890	890
R-squared	0.058	0.066	0.078	0.086	0.055	0.294	0.305	0.306	0.315	0.310
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
R2 Adj. (w/o controls)	0.0553	0.0631	0.0753	0.0825	0.0526	0.170	0.181	0.183	0.192	0.188
F Test	4.39e-09	7.80e-11	0	0	1.97e-09	0.00556	0.000910	0.000141	2.18e-05	2.01e-05
F Test Diff.	0.690	0.578	0.0815	0.0174	0.794	0.302	0.278	0.720	0.777	0.437

Note: Dependent variable is the average log of maize yields calculated over the 4 seasons (short rain 14 to long rain 16). Controls include education, literacy, gender, age and age squared of the farmer, land and cattle ownership, size and quality of the house, household size, whether the farmer is the household head, household head's gender, village fixed effects and enumerator-assignment fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9: Regressions of the average log maize yield on naïve skill sub-constructs

	(1)	(2)	(3)	(4)	(5)	(6)
Oral math questions	0.02	0.04	0.05	0.04	0.05	0.05
	(0.064)	(0.064)	(0.065)	(0.065)	(0.065)	(0.064)
Reading	0.06	0.06	0.06	0.06	0.06	0.06
	(0.062)	(0.061)	(0.060)	(0.061)	(0.060)	(0.061)
Raven	0.04	0.05	0.04	0.04	0.04	0.04
	(0.059)	(0.058)	(0.058)	(0.058)	(0.058)	(0.059)
Digit Span	0.00	0.00	0.00	0.00	0.01	0.01
	(0.041)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)
Math (timed)	0.09	0.11*	0.11*	0.11*	0.10*	0.11*
	(0.061)	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)
CESD	0.08*					
	(0.046)					
Locus of Control	0.02					
	(0.058)					
Self-esteem	-0.06					
	(0.045)					
Causes of poverty	0.07					
	(0.059)					
Attitude towards change	-0.08**					
	(0.039)					
Tenacity / Organization	0.14***					
	(0.044)					
Metacognitive	0.02					
	(0.044)					
Optimism	0.06					
	(0.043)					
Risk aversion	-0.03					
	(0.032)					
Big 5 Agreeableness	-0.02	0.04				
	(0.046)	(0.043)				
Big 5 Extraversion	0.02		0.04			
	(0.037)		(0.035)			
Big 5 Conscientiousness	-0.02			0.05		
	(0.055)			(0.045)		
Big 5 Neuroticism	0.01				0.05	
	(0.044)				(0.040)	
Big 5 Openness	0.01					0.04
	(0.044)					(0.040)
Other noncognitive	0.02					
	(0.041)					
Intercrop /Compost	0.02	0.00	0.00	0.00	0.00	0.00
	(0.044)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)
Maize	0.04	0.05	0.05	0.05	0.04	0.05
	(0.048)	(0.048)	(0.046)	(0.046)	(0.048)	(0.047)
Banana	0.04	0.05	0.04	0.04	0.04	0.04
	(0.046)	(0.045)	(0.044)	(0.044)	(0.044)	(0.044)
Soya	0.00	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.038)	(0.037)	(0.038)	(0.038)	(0.037)	(0.037)
Fertilizer	-0.02	0.00	-0.00	-0.00	0.00	-0.00
	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.047)

Observations	897	899	899	899	899	899
R-squared	0.325	0.299	0.299	0.299	0.299	0.299
R2 Adj. (w/o controls)	0.0673	0.0528	0.0560	0.0551	0.0558	0.0538
F Test (Cog)	0.307					
F Test (Noncog)	0.0437					
F Test (Tech)	0.822					
Test NC diff.	0.0299					

Note: Dependent variable is the average of log maize yields calculated over the 4 seasons (short rain 14 to long rain 16). Controls include education, literacy, gender, age and age squared of the farmer, land and cattle ownership, size and quality of the house, household size, whether the farmer is the household head, household head's gender, village fixed effects and enumerator-assignment fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10: Regressions of the average log of maize yield on improved skill sub-constructs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cognitive skills	0.17**	0.19***	0.21***	0.19***	0.18***	0.21***	0.20***
	(0.068)	(0.067)	(0.066)	(0.067)	(0.066)	(0.066)	(0.064)
Factor 1	0.05	0.07*					
(CESD)	(0.041)	(0.043)					
Factor 2	0.02		0.06				
(Conscientiousness/Tenacity)	(0.051)		(0.047)				
Factor 3	0.03			0.05			
(LOC/Metacog/Openness)	(0.047)			(0.042)			
Factor 4	0.05				0.10***		
(Causes of poverty, negative items)	(0.052)				(0.038)		
Factor 5	-0.03					-0.00	
(Attitude towards change/Beans)	(0.043)					(0.045)	
Factor 6	0.06						0.09**
(CESD positive/Confidence/Risk aversion)	(0.040)						(0.037)
Technical skills	0.04	0.05	0.05	0.05	0.04	0.06	0.05
	(0.048)	(0.050)	(0.048)	(0.049)	(0.048)	(0.049)	(0.050)
Observations	890	890	890	890	890	890	890
R-squared	0.308	0.302	0.301	0.301	0.304	0.299	0.304
R2 Adj. (w/o controls)	0.0581	0.0577	0.0604	0.0575	0.0602	0.0569	0.0577
F Test (Cog)	0.0127						
F Test (Noncog)	0.0599						
F Test (Tech)	0.456						
Test NC diff.	0.589						

Note: Dependent variable is the average rank of maize yields calculated over the 4 seasons (short rain 14 to long rain 16). Controls include education, literacy, gender, age and age squared of the farmer, land and cattle ownership, size and quality of the house, household size, whether the farmer is the household head, household head's gender, village fixed effects and enumerator-assignment fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Skills asked to a village informant: prediction of average log of maize yield

Corresponding Skill Index	Explanatory variables: Question asked to village informant	Regressions with average log of maize yield as dependent variable		
Cognitive	Level of education		0.14**	0.03
			(0.068)	(0.079)
Non-cognitive	Active/ Motivated		0.13	0.09
			(0.119)	(0.120)
Technical	Agricultural knowledge		0.36***	0.33***
			(0.097)	(0.102)
	Controls	Vil. FE	Vil. FE	All
	Observations	883	883	883
	R-squared	0.186	0.244	0.299
	F Test		5.83e-08	0.0003

Note: Skill proxies obtained through village informant (CHW), scored on scale from 1 (low) to 3 (high). In the regressions, the dependent variable is the average rank of Maize yields calculated over the 4 seasons (short rain 14 to long rain 16). Controls include education, literacy, gender, age and age squared of the farmer, land and cattle ownership, size and quality of the house, household size, whether the farmer is the household head, household head's gender, village fixed effects and enumerator-assignment fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Additional References from the appendices

Abdi, H., 2007. "RV coefficient and congruence coefficient." *Encyclopedia of measurement and statistics*, pp.849-853.

Acharya, V., A. Rajan, and A. Schoar, 2007. "What determines entrepreneurial success? A psychometric study of rural entrepreneurs in India." *mimeo*, London Business School, IFMR, MIT.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (US), 1999. "Standards for educational and psychological testing." *Amer Educational Research Assn.*

Barham, T., K. Macours, and J. A. Maluccio, 2017. "Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years." *CEPR Discussion Paper* 11937

Beaman, L. and J. Magruder, 2012. "Who Gets the Job Referral? Evidence from a Social Networks Experiment." *American Economic Review*, 102(7): 3574-93.

Buck, S. and J. Alwang, 2011. "Agricultural extension, trust, and learning: results from economic experiments in Ecuador." *Agricultural Economics*, 42(6): 685-699.

Cattell, R.B., 1966. "The scree test for the number of factors." *Multivariate Behavioral Research* 1: 629-637.

Cattell, R.B., 1987. *Intelligence: Its Structure, Growth and Action.* Amsterdam: North-Holland.

Cota, A.A., Longman, R.S., Holden, R.R., Fekken, G.C., & Xinaris, S., 1993. "Interpolating 95th percentile eigenvalues from random data: An empirical example." *Educational & Psychological Measurement, 53:* 585-596.

Crocker, L. and Algina, J., 1986. *Introduction to classical and modern test theory.* Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Cronbach, L.J., 1951. "Coefficient alpha and the internal structure of tests." *Psychometrika*, 16(3), pp.297-334.

Cueto, S and J. Leon, 2012. "Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives", *Young Lives technical note* 25.

Cunha, F., J. Heckman, and S. Schennach. 2010. “Estimating the Technology of Cognitive and Non-cognitive Skill Formation.” Econometrica 78(3):883–931

David, Soniia, 2007. "Learning to Think for Ourselves: Knowledge Improvement and Social Benefits among Farmer Field School Participants in Cameroon", *Journal of International Agricultural and Extension Education*, 14(2): 35-49

de Mel, S. D. McKenzie, C. Woodruff, 2010. "Who are the Microenterprise Owners? Evidence from Sri Lanka on Tokman versus De Soto", in: Josh Lerner and Antoinette Schoar (eds.): *International Differences in Entrepreneurship*, University of Chicago Press. Chicago.

de Mel, S., D. McKenzie and C. Woodruff, 2009a. "Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns." *American Economic Journal: Applied Economics*, 1(3): 1-32.

De Mel, S., D. McKenzie, and C. Woodruff, 2009b. “Innovative firms or innovative owners? Determinants of innovation in micro, small, and medium enterprises.” *IZA Discussion Paper* No. 3962

Djankov, S., E. Miguel, Y. Qian, T. Roland, and E. Zhuravskaya, 2005. “Who are Russia's entrepreneurs?” *Journal of the European Economic Association*, *3*(2-3): 587-597.

Dupas, P. and J. Robinson, 2013. “Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya” *American Economic Journal: Applied Economics* 2013, 5(1): 163–192

Feagin, J. R., 1972. “Poverty: We Still Believe That God Helps Those Who Help Themselves.” *Psychology Today 1*, 101–29.

Feagin, J. R., 1975. *Subordinating the Poor: Welfare and American Beliefs*. Englewood Cliffs, NJ: Prentice Hall

Feder, G., R. Murgai, and J.B. Quizon, 2004. “The acquisition and diffusion of knowledge: The case of pest management training in farmer field schools, Indonesia.” *Journal of agricultural economics*, 55(2): 221-243.

Giné, X. and G. Mansuri, 2014. “Money or ideas? A field experiment on constraints to entrepreneurship in rural Pakistan.” *World Bank Policy Research Working Paper* 6959, World Bank, Washington DC.

Glorfeld, L.W., 1995. “An improvement on Horn’s parallel analysis methodology for selecting the correct number of factors to retain.” *Educational and Psychological Measurement, 55*: 377-393.

Godtland, E.M., Sadoulet, E., De Janvry, A., Murgai, R. and Ortiz, O., 2004. “The impact of farmer field schools on knowledge and productivity: A study of potato farmers in the Peruvian Andes.” *Economic Development and Cultural Change*, 53(1): 63-92.

Goldstein, M. and C. Udry, 1999. “Agricultural innovation and resource management in Ghana.” *Final Report to IFPRI under MP17*, Mimeo. Yale University.

Hambleton, R.K. and Swaminathan, H., 2013. *Item response theory: Principles and applications*. Springer Science & Business Media.

Hanna, R., S. Mullainathan, and J. Schwartzstein, 2014. “Learning through noticing: theory and experimental evidence in farming”, *Quarterly Journal of Economics*, 1311–1353.

Horn, J. L., 1965. “A rationale and test for the number of factors in factor analysis.” *Psychometrika, 30:* 179-185.

Jamison, D.T. and P.R. Moock, 1984. “Farmer education and farm efficiency in Nepal: The role of schooling, extension services, and cognitive skills.” *World Development*, *12*(1): 67-86.

John, O.P., E.M. Donahue, R.L. Kentle, 1991. *The Big Five Inventory – Versions 4a and 54*. Berkeley CA: University of California, Berkeley, Institute of Personality and Social Research.

Kaiser, H., 1960. “The application of electronic computers to factor analysis.” *Educational and Psychological Measurement* 20: 141-151.

Kaiser, H.F., 1958. “The varimax criterion for analytic rotation in factor analysis.” *Psychometrika,* 23(3), pp.187-200.

Kline, P., 2013. *Handbook of psychological testing*. Routledge.

Kondylis, F., V. Mueller, and S. Zhu, 2015. “Measuring agricultural knowledge and adoption.” *Agricultural Economics*, *46*(3): 449-462.

Krauss, S.I., M. Frese, C. Friedrich, and J.M. Unger, 2005. “Entrepreneurial orientation: A psychological model of success among southern African small business owners” *European Journal of Work and Organizational Psychology*, *14*(3): 315-344.

Ledesma, R.D., and P. Valero-Mora. 2007. “Determining the Number of Factors to Retain in EFA: An Easy-To-Use Computer Program for Carrying out Parallel Analysis.” Practical Assessment, Research & Evaluation 12(2):1–11.

Levenson, H., 1981. "Differentiating among Internality, Powerful Others, and Chance." In H. M. Lefcourt (Ed.), *Research with the Locus of Control Construct*, pp. 15–63. New York: Academic Press.

Lockheed, M. E., T. Jamison, and L.J. Lau, 1980. "Farmer Education and Farm Efficiency: A Survey." *Economic Development and Cultural Change*, 29 (1): 37-76.

Lorenzo-Seva, U. and Ten Berge, J.M., 2006. "Tucker's congruence coefficient as a meaningful index of factor similarity." *Methodology*, 2(2), pp.57-64.

Mauceri, M., J. Alwang, G. Norton, and V. Barrera, 2007. "Effectiveness of integrated pest management dissemination techniques: a case study of potato farmers in Carchi, Ecuador." *Journal of Agricultural and Applied Economics*, 39(03): 765-780.

Nunnally, J.C. and Bernstein, I.H., 1994. *Psychometric theory* (3rd ed) New York: McGraw-Hill.

Phillips, J.M., 1994. "Farmer Education and Farmer Efficiency." *Economic Development and Cultural Change*, Vol. 43(1): 149-166.

Radloff, L.S. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." Applied Psychological Measurement 1(3):385–401.

Rammstedt, B. and O.P. John, 2007. "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German." *Journal of Research in Personality, 41*: 203– 212

Rauch, A. and M. Frese, 2000. "Psychological approaches to entrepreneurial success: A general model and an overview of findings." *International review of industrial and organizational psychology*, 15: 101-142.

Rammstedt, B., C.J. Kemper, and I. Borg. 2013. "Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study." European Journal of Personality 27(1):71–81.

Rosenberg, M., 1965. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Soto, C.J., O.P. John, S.D. Gosling, and J. Potter. 2008. "The Developmental Psychometrics of Big Five Self-Reports: Acquiescence, Factor Structure, Coherence, and Differentiation from Ages 10 to 20." Journal of Personality and Social Psychology 94:718–37.

Van der Linden, W.J. and Hambleton, R.K. eds., 2013. *Handbook of modern item response theory*. Springer Science & Business Media.