Cognitive and Socio-Emotional Skills in Low-Income Countries: Construct and Predictive Validity[1]

Alice Danon (Harvard University)

Jishnu Das (Georgetown University)

Andreas de Barros (MIT)

Deon Filmer (World Bank)

Abstract

We assess the construct and predictive validity of cognitive and socio-emotional skills in Pakistan using two innovations in measurement and sampling. First, we developed and implemented a battery of tests to capture cognitive and socio-emotional skills among young adults. We measured socio-emotional skills using both self-reported and task-based instruments and psychometrically verified the validity of the different components. For cognitive skills, we measured standard literacy and numeracy as well as skills useful for everyday life. We demonstrate the reliability and construct validity of these measures compared to previous attempts in the literature. Second, we constructed a panel that follows respondents from their original rural locations in 2003 to their locations in 2018, a period over which 38% of respondents left their native villages. We show that the predictive validity of our skills measures is mediated by the migration decision. Among male *migrants*, labor earnings are strongly correlated with years of schooling, but not socio-emotional skills. Among male *non-migrants*, wages are associated with socio-emotional skills, but not years of schooling. These associations are consistent with similar data from rural Cambodia, a region with similar levels of schooling but different patterns of migration and labor force participation.

[1] We thank participants at the World Bank's conference on linkages between schooling and labor markets and at the Methods and Measurement Conference, Northwestern University. Financial support for this research was made available through RISE grant PO13412 and The World Bank's Strategic Research Priorities Fund. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent. Please address correspondence to adanon@g.harvard.edu, jishnu.das@georgetown.edu, debarros@mit.edu and dfilmer@worldbank.org.

I. INTRODUCTION

Two strategies widely believed to improve living standards for populations living in rural areas are education and migration. A vast literature associated with each of these demonstrates that educated rural households are better able to avail of new technologies; that consumption growth is higher among households who migrate and, in recent experimental studies, that secondary schooling and migration (whether international or within-country) both increase wages.[2] These results have been demonstrated across a number of different settings and as longer-term panels built around experiments become available, further rapid progress is being made on each of these issues.

In this paper, we continue to study the link between education and outcomes (labor earnings) among (young) adults but move away from the necessary and well-understood questions of identification. Motivated by research that demonstrates (a) the importance of socio-emotional skills in the United States (Heckman 2007) and (b) the difficulty of measuring these skills in low-income countries, we focus instead on the *measurement* of both cognitive and socio-emotional skills (Lajaaj and Macours, 2021; Valerio et al., 2016). To allow for potential links between migration and skills, we study an unusual sample of children who grew up in rural Pakistani villages and were first surveyed in 2003 when they were between the ages of 5 and 15 and then re-surveyed between 2017 and 2018, regardless of where they were living at that point in time. At this point, 38% had migrated from their native homes so that migration appears as an endogenous response to the skills that respondents have acquired, and, in turn, may lead to differential returns to these skills. Using this sample, we demonstrate the internal validity and reliability of our measurements. We then examine the predictive validity of our measures, by examining correlations between our measures and years of schooling as well as labor earnings. Finally, we document that the patterns we uncover in Pakistan are also found in a similar sample in Cambodia.

We present three sets of results, related to the construct validity of our measures and their correlation with years of schooling and with labor earnings. We first follow Lajaaj and Macours (2021) and

[2] The canonical papers in this literature include Foster and Rosenzweig (1996) on the relationship between returns to schooling and technological change in an agrarian economy; Duflo, Dupas and Kremer (2021) on the experimentally estimated returns to secondary schooling in Ghana; McKenzie, Stillman and Gibson (2010) on the returns to international migration between Tonga and New Zealand; Bryan, Chowdhury and Mobarak (2014) on the experimental returns to seasonal migration in Bangladesh and Beegle, Weerdt and Dercon (2011) on consumption growth among migrant and non-migrant households in the Kagera region of Tanzania.

document that our measures of socio-emotional skills or SEMS satisfy several desirable psychometric properties. These include a fairly high Cronbach's α-statistic, which is a measure of the internal consistency of the measures and a factor structure that corresponds exactly to the constructs that are being measured. Nevertheless, there is considerable room for improvement. For instance, using repeated administrations of the tests we compute the test-retest reliability and even though we chose constructs with higher reliability, our measures are still below a desirable cutoff of $\varrho=0.7$. Similarly, even though we correct for acquiescence bias, or the tendency of respondents to agree with statements, it remains a concern in our sample.

We then examine associations between our skills measures and years of schooling. At the very least, we would expect a substantial correlation between cognitive skills and years of schooling. We should also expect some correlation between SEMS and years of schooling as one of the key motivations for the focus on SEMS in the U.S. literature was the careful documentation that variation in cognitive skills captured only a small portion of the returns to years of schooling (Bowles and Gintis, 1976; Heckman and Rubenstein, 2001).

In our sample, there is indeed a robust correlation between years of schooling and cognitive skills with every year of schooling associated with a 0.17sd increase in the tested subjects of English, Mathematics, and Urdu in Pakistan, and very similar results in Cambodia. Nevertheless, for every year of schooling, there is considerable variation in test scores. For instance, the top 5% of children who have completed Grade 5 (but no further) report test scores that are higher than the bottom 5% of children who have completed Grade 10 (but no further). It is this variation that will allow us to estimate the predictive validity of cognitive skills for labor earnings while conditioning on the years of schooling.

In contrast, for socio-emotional skills or SEMS, mean skill levels are similar to those in rich-country populations, which suggests that at the population level our measures are not indicative of substantial deficits. In addition, although SEM skills and years of schooling are positively correlated, the correlation of 0.03sd for every year of schooling is much smaller than for cognitive scores. Also, unlike cognitive skills where we are more confident that the direction of causality is from schooling to test scores, it is likely that children with higher SEMS scores (more grit, more perseverance) were also those who chose to continue in school longer. In fact, parallel research by Barrera-Osorio, De Barros and Filmer (2018) leverages an experimental design to show that there is zero causal impact of years of schooling in SEM skills in Cambodia.

Our third set of results turns to the predictive validity of our skills measures for labor force participation and earnings and how these correlations are mediated through migration. Labor force participation among men is 79%, compared to 6% for women, consistent with the literature on very low female labor force participation rates in South Asia (Field and Vyborny, 2016). Labor force participation increases with years of schooling for women but declines for men. This could reflect different search patterns as men with more years of schooling may be `waiting it out' for a better job, while women in our sample appear to have a limited window prior to marriage during which they work. However, conditioning on years of schooling, neither cognitive nor socio-emotional skills are correlated with labor force participation for either sex.

In our sample, each year of education is associated with 3.2-3.6% higher labor earnings for men and a much higher 22-22.5% for women, which partly reflects higher LFP. If we restrict our sample to those who are working, the estimates are 3.9-4.6% for men and 7-8.3% for women. One explanation why the association with labor earnings is smaller for men than the 10-12% usually found in the literature is that 79% of the men who are still enrolled in education (college) are not yet in the labor force. Pakistan, like many other countries, has seen a sharp increase in the return to college education and a decline in the return to primary or secondary education (Montenegro and Patrinos, 2014). If all the returns to education come from those who are currently enrolled in college, it is too early in our sample to pick this up. Together with the fact that many educated men still appear to be searching for employment, the lower estimate we find can be regarded as the correlation of labor earnings with years of schooling among the selected sample of those with less than college education.

We then show that, in contrast to labor force participation, both cognitive skills and SEMS are highly predictive of labor earnings for men, conditional on years of schooling. For men, point estimates for test scores range from $6 to $8 per standard deviation and for SEMS from $15 to $17 per standard deviation, depending on the specification. For women, point estimates lack statistical precision given the size of the sample. However, this average return to skills hides substantial variation by the migration status of the respondent. In our data, 43% of women and 35% of men are no longer living in their original villages. Female migration is almost entirely due to marriage (of those who have migrated, 91% are married and only 4% are working) while male migration mostly reflects work opportunities.[3] Migration and occupations for men are linked—those living in the village are equally likely to be engaged in daily labor, work in a salaried occupation or be self-employed (or in their family's business) and less likely to be engaged in agriculture. Men who have migrated are more likely to be salaried at the expense of the other occupations.

Among the men still in their native villages, we find a precisely estimated *zero* returns to years of schooling and a strong correlation between SEMS and labor earnings. The latter result is robust to multiple specifications and every sub-sample in our data. On the other hand, among men who have left the village, the most robust result is a significant correlation between labor earnings and years of schooling. The returns to cognitive skills and SEMS vary according to the specification with stronger correlations in median relative to mean regressions.

Our striking result is that for the two-thirds of men who are still in their birth village, socio-emotional skills are correlated with labor earnings, but years of schooling are not. To understand if this is a facet of this particular sample and respondent age, we therefore turned to a second dataset from Cambodia with similar features. Like in Pakistan, this sample comes from a rural area (poorer than in Pakistan) and includes children who have been tracked and re-surveyed in adulthood. Here, we again find that the correlation with years of schooling is close to zero and although the results are more imprecise, the association with SEMS is positive.

We view our paper as contributing to a wider discussion on the role of schooling, the measurement of skills, and skills formation in low-income countries where migration is an important part of people's lives. Our measurements provide a picture of the labor market where the potential returns to skills are intertwined with migration decisions, but with the substantial caveat that at this stage we are presenting correlations that await further validation in causal analysis. The remainder of the paper is as follows. In Section II, we discuss the data and the context. Then, we present results on construct validity in Section III and on predictive validity in Section IV. We conclude with a discussion in Section V.

[3] This result mirrors Beegle et al.'s (2011) previous study of migration from an initially rural sample in Kagera; one difference is that 10.5% of the men are working outside the country, with more than 90% in the Arab countries of Saudi Arabia, UAE, Bahrain, Qatar and Oman.

II. SAMPLE AND DATA

II.1. SAMPLE SELECTION

The data come from the Learning and Education Achievement in Punjab Schools (LEAPS) project, which is a longitudinal study of education in Pakistan. In 2003, the LEAPS project randomly sampled 112 villages from three districts in the province of Punjab, from a list frame of villages with at least one private school. These villages were richer and larger than the average village, but close to 70% of the population of Punjab was living in such villages at the time of the first survey. As part of the survey, 1,807 households were surveyed in these villages with information on 5,865 children between the ages of 5 and 15 in 2003. These households were then revisited four times between 2004 and 2011.

Between 2016 and 2018, we attempted to contact and resurvey the children in the 2003 sample, and the data from this resurvey are used in this paper. Over two years of tracking and re-surveying (sometimes waiting 18-20 months for a person to return home from their work location outside Pakistan), we were able to complete in-person surveys for 75.1% of the sample and we have information either through phone surveys or from a third-party respondent for another 9.4%. For these `indirect' surveys, we do not have skill measurements although we have data on many of our main outcomes including years of education, earnings, and migration. Thus, of the 5,865 children, we have some information on 4,956 children or 84.5% of the original sample at the individual level. The implied annual attrition rate of (just above) 1% compares favorably to 10-year panels with the highest retention (Outes-Leon and Dercon, 2009).

Appendix A and Table A1 detail the tracking process, the different instruments used and the types of attrition in the data. Of the 909 individuals on whom we have no data, 43 had died, 186 were living in four villages that fell into a military zone that our team was unable to access, 395 respondents refused to participate despite multiple attempts and 285 could not be located. There are an additional 550 individuals for whom we have indirect information either through third-party surveys or phones. Attritors are less likely to have ever been married and more likely to be living outside their original district and outside the country (Table A2, Panel B). They were also poorer than other households and more likely to have a father living abroad in 2003 (Table A2, Panel D). Respondents with indirect and phone surveys (no skills measures)[4] were again more likely to be living outside the village at the time of the survey and more likely to be working (Table A2, Panel A). We account for these differences using a variety of weighting schemes described in Appendix B.

II.2. SAMPLE CHARACTERITICS

We now discuss the basic characteristics of our sample, focusing both on intergenerational changes and on migration. A key point discussed by Foster and Rosenzweig (1996) is that the returns to schooling depends on growth in the economy; where growth is `stagnant', the returns to schooling may be low. To understand the overall dynamism in this context, Table 1 summarizes the characteristics of our sample of respondents (Panel A) and compares them to their *parents* in 2003 (Panel B). Both men and women in our sample have over 8 years of education compared to just above 3 years for the parents; 73% of our sample says that they can read compared to 37% among the parents. Both of these statistics reflect the well-known and often dramatic improvements in schooling participation over the last two decades (World Bank, 2018).

Interestingly, the changes are just as large in broader social and occupational regimes. For instance, the age of marriage for women will be *at least* 22, compared to 19.7 for their mothers and the share of men working in agriculture has plummeted from 33% among the fathers to 7% among the sons. There have also been improvements in living standards: 96% of our sample reports toilets on their premises in 2018 and 98% report access to electricity compared to 58% and 88% respectively in 2003. However, one statistic that remains low and unchanged across generations at 5% is female labor force participation. Multiple authors have commented on the low and declining labor force participation among South Asian women, and our data are consistent with the findings from that literature. See Afridi, Dinkelman and Mahajan (2018) for India; Field and Vyborny (2016) and Subramanian (2020) for Pakistan. Concretely, one implication for our analysis is that most of the variation in female labor force outcomes will be in the participation decision, rather than wages conditional on participation.

Like in Beegle, De Weerdt and Dercon (2011), part of this dynamism may be due to migration, which has emerged as a key feature of our respondents' lives. In our sample, 38.7% of men and women no

[4] They also include 15 respondents who answered the survey in person: 4 respondents did not finish the survey and 11 respondents for whom there was a bug with the test on tablets.

longer reside in their birth village (Figure 1). Of the original sample of men, 10.5% now live outside Pakistan, mostly in Arab countries, 16% live in Pakistan outside their native district and 65% remain in their native village with the rest migrating within their native district. The farther men are from their village, the more likely they are to be salaried and the less likely they are to rely on daily wages, their own or family business and/or agriculture (Table A3). Earnings are also higher for those who migrate—median/mean monthly income for respondents in their original village is USD $115/$139, compared to $173/$192 for respondents living outside the district and $337/$380 for those living outside the country (Figure A1). Another way to look at migration and incomes is to note that 54% of all the male income generated in this sample comes from the 35% of men who have left their birth village, and 27% comes from those who have left the country. Migration among women, on the other hand, is closely tied to marriage and the practice of virilocal residence. Among those who have migrated, 90% are married compared to 31% among those still residing in the village. Most migration is within the same district and only 0.5% or 11 women are now living outside Pakistan.

III. MEASUREMENTS OF SKILLS: INSTRUMENTS AND RELIABILITY

In our 2018 survey, we measured two types of skills that, following the literature, we refer to as cognitive and socio-emotional skills. We describe the instruments we used to measure each category and the reliability of the associated measures in turn. Table A4 provides a summary of the instruments used to measure the different sets of skills.

III.1. COGNITIVE SKILLS

For cognitive skills, we first used the same tests that had been used previously in the LEAPS project; an advantage of doing so is that in future research we can link the scores on a common scale going back as far as 2003, when the children in our sample were in Grade 3. These tests are norm-referenced (rather than criterion-referenced) and were designed to cover a wide range of topics. Andrabi et al. (2002) and Bau, Das and Yi Chang (2021) have previously shown that the LEAPS tests satisfy both the requirements of horizontal and vertical linking, which means that the function relating the latent variable, knowledge, to the likelihood of answering a question correctly is stable across test takers and over time.[5] In the limited number of items where vertical linking does not function well, eliminating the unstable items does not lead to any appreciable difference in the tests.

Given that these tests were originally designed with primary-school age children in mind, we worried that ceiling effects would censor the cognitive skills distributions in the resurvey. We therefore worked with an educational organization to also design an adaptive test administered on tablets. Each test started with a set of simple questions with the difficulty of subsequent items increasing or decreasing depending on whether the respondent answered correctly or not. For instance, a respondent who had completed 5th grade and then dropped-out may face a set of questions on two-digit division and fractions; if they were not able to answer them, the next set of questions may have multiplication and single-digit division. The idea was that an adaptive test would allow us to capture a wider range of skills and therefore provide more accuracy. Appendix C.1 presents the progress and placement logic of the test.[6]

However, and contrary to our expectations, ceiling effects in the LEAPS test were small with 11.2% of children achieving the maximum in English, 2.7% in Mathematics, and 13.1% in Urdu. In contrast, the adaptive test, which was designed to classify respondents into 6 levels (Level 1 corresponding to Early Primary and Level 6 to College), elicited little meaningful variation that we can exploit: 43%, 64%, and 77% of respondents were classified as Level 1 for Urdu, Mathematics and English respectively and the rest were largely classified as Level 2.[7] Nevertheless, to capture any additional information from the adaptive test, we aggregate all the items from the test on paper and on tablet for each subject into Urdu, Mathematics and English scores using Item Response theory with a two-

[5] Here, we do not vertically link the scores as we use results from an additional adaptive test as well. Andrabi et al. (2002) discuss test construction and assess the psychometric properties of the original test administered in 2003 sample. Bau, Das and Yi Chang (2021) further assess vertical linking in the LEAPS test and demonstrate that there was limited differential item functioning between 2003 and 2011. That is, the function relating the latent variable (ability or knowledge) to the likelihood of answering the question correctly remained stable across the years.

[6] The mapping between level and grades is as follows: Level 1: Nursery, Grades 1 to 3 (early primary); Level 2: Grades 4 and 5 (late primary); Level 3: Grades 6 to 8 (middle school); Level 4: Grades 9 and 10 (high school); Level 5: Grades 11 and 12 (intermediate); Level 6: College.

[7] In English and Mathematics, fewer than 3% of respondents were placed in Levels 3 to 6. In Urdu, 12% are placed at Level 3, and 5-7% at each subsequent level.

parameter logistic (2PL) model.[8] Formally, the item characteristic curve is given by the 2-parameter logistic:

$$P_j(\theta) = \frac{1}{1 + \exp\{-a_j(\theta - b_j)\}}$$

where $b_j \equiv \theta^* | P_j(\theta^*) = \frac{1}{2}$ is the difficulty parameter, which is the ability level at which the child will answer the question correctly half the time and $a_j \propto \frac{\partial P_j(\theta)}{\partial \theta}$ at $\theta = b_j$, is the discrimination parameter, which specifies the steepness of the item characteristic curve at the point that the ability of the child is equal to the difficulty of the question (b_j). The joint estimation of θ and these parameters follows the standard maximum-likelihood procedure in IRT using the IRT command, *OpenIRT*, developed by Zajonc for STATA and discussed in Das and Zajonc (2010). We then create a single cognitive skills index as the average of the three subject scores and assess model fit by comparing actual with predicted item responses based on estimated item parameters and the model assumptions of the 2PL model (Appendix C.2, Figures A3 to A5 for the LEAPS test and A6 to A8 for the adaptive test). For most items, the actual and predicted responses match closely, although there are several items in the adaptive test (for instance, Items 62, 96 and 107 in Urdu) where the fit is poor. Re-estimating the model after eliminating these poorly fitting items does not alter the overall estimated score.

One concern with these assessments, which are closely tied to what children are supposed to learn in school, is that they may not adequately reflect functional literacy and numeracy (Banerjee et al. 2017). We therefore also designed an assessment to capture proficiency in everyday arithmetic and literacy skills. The math questions asked respondents to first read an electricity bill and compute the correct amount given arrears (easier) and then recompute the correct amount given electricity consumption and non-linear pricing (harder). A third math question assessed competency in a marketplace transaction where respondents purchase multiple items and collected change. In order to assess literacy skills, we asked respondents to read a number of messages written in Urdu, and in Roman Urdu (Urdu but using roman language script). We also assessed whether the individual knew how to use a phone by asking them to save a contact on a phone. Appendix C.3 details these items; we note that someone who successfully completes these tasks can be arguably thought to be `functionally'

[8] We exclude items that less than 50 respondents answered as well as items that less than 5% or more than 95% of respondents got the correct answer.

literate. We aggregate these different questions into a single index using principal factor analysis; for brevity, we refer to this as a `life skills' index.

III.2. SOCIO-EMOTIONAL SKILLS

The most complex part of our skills measurement component was our assessment of socio-emotional skills or SEMS. Measuring SEMS in low-income countries has proven difficult with evidence of non-classical measurement error using self-reported instruments (Laajaj and Macours, 2021). Consequently, the instruments included in the Pakistan survey were developed through an iterative process that started with data collection in Cambodia for a related project in 2017; we will present comparative results from that study in Section III.2.2. Like in Pakistan, the sample in Cambodia also consisted of young adults from rural regions, and the data collection incorporated a comprehensive assessment of socio-emotional skills. However, despite our efforts to mitigate the kind of issues later discussed by Laajaj and Macours (2021), the self-reported scales included in that survey displayed limited internal consistency as detailed in Appendix D.

We built on that experience when designing and implementing our SEMS assessment in Pakistan in several ways. First, in addition to self-reported measures, we developed bespoke applications on tablets under the assumption that they would be less subject to biases arising from social desirability, or the tendency to over-report socially valued attitudes, and acquiescence, or the tendency to agree with yes/no questions, regardless of their content. We then conducted a pilot with 403 respondents and (a) included debrief sessions to gauge our respondent's understanding of the material and (b) randomly re-surveyed half (201) two weeks later to assess the reliability of our measures in repeated administrations of the same test.

Following Laajaj and Macours (2021) and the psychometrics literature, we then used three main criteria to select the instruments in our survey: face validity, predictive validity, and reliability. The pilot first allowed us to assess the face validity of our instruments to ensure that the questions we asked were perceived as measuring the concepts we intend to measure. Following a literature in the U.S. that establishes a link between SEMS and earnings, we were also able to assess predictive validity by calculating the bivariate correlations of each score with years of schooling and earnings (Cunha and Heckman, 2010 and Brunello and Scholotter, 2011).

Finally, we computed two types of reliability estimates: test-retest reliability and internal consistency. Multiple measurements allow us to estimate the test-retest reliability (the correlation of the same measures in repeated administrations). Under the assumption of classical measurement error, the test-retest correlation provides an estimate of the share of the variance of a measure that is explained by the true latent trait we are trying to capture, rather than by measurement error. Specifically, if the measured value X is the true value X^* plus a measurement error ε, $X = X^* + \varepsilon$, then the test-retest correlation is an estimate of the reliability defined as $Relability = \frac{\sigma_{X^*}}{\sigma_X}$. Generally, a value of 0.7 or higher is considered to be a reliable measure.

To assess internal consistency, we first computed Cronbach's α-statistic, a measure used in psychometrics that indicates the inter-correlation of the items on a scale, commonly interpreted as the extent to which the items of a scale measure the same underlying concept. Cronbach's α-statistic is computed as $\alpha = \frac{K}{K-1}(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2})$ where K is the number of items in the scale, σ_X^2 is the variance of the observed total test score, and $\sigma_{Y_i}^2$ is the variance of responses to item i for the current sample of persons. The statistic is a ratio of variances and therefore lies between 0 and 1; a rule of thumb is for a measure with high internal consistency is that Cronbach's α should be above 0.7 (Nunally and Bernstein, 1978). We follow this heuristic with two notes of caution: Cronbach's α may be high as a result of systematic response biases that lead to a high inter-item correlation, even after correcting for acquiescence bias, and the statistic mechanically increases as the number of items in a scale increase. These results from the pilot and additional details on the tools used to assess the validity of our measures are reported in Table 2 and Appendix D. Based on these results, we retained two self-reported scales and two tasks (administered on tablets).

The first self-reported 10-item scale measures grit – the combination of passion and perseverance for long-term goals—and was developed by Duckworth and Quinn (2009).[9] The second self-reported scale measures the "Big Five", a taxonomy of traits that encompasses five dimensions of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To measure these traits, we used the short 15-item Big Five Inventory (Lang et al., 2011), which consists of three items for each of the five personality traits. All items are answered using a 5-point Likert scale format

[9] There was a mistake in the translation of the tool from English to Urdu so that only 9 items were implemented.

ranging from 1 ("Disagree strongly") to 5 ("Agree strongly"). Following Laajaj and Macours (2021), we applied acquiescence bias correction on the items to correct for the tendency of respondents to agree with a statement (see Appendix D.1 for the procedure).

The two self-reported scales show fairly high internal consistency with Cronbach's α just above 0.7 for the Grit scale and just below 0.7 for the Big Five (Table 2). Cronbach's α's for the Big Five subscales range between 0.53-0.68, even though each scale only comprises 3 items. As a comparison, in rural Kenya, Laajaj and Macours (2021) found Cronbach's α's for the same constructs ranging from 0.31 to 0.51, with 4 or 5 items per sub-scale. Also, in contrast to Laajaj and Macours (2021), the skills factor structure is closely reproduced in our data (Table 3). Exploratory factor analysis identifies five factors, corresponding to the Big 5 personality traits and the items measuring grit and conscientiousness, two closely related constructs, load on the same factor.[10] Finally, acquiescence bias in our sample is 0.26 to 0.31, which is slightly below the 0.37 reported by Laajaj and Macours (2021) from Colombia. Overall, we interpret these results as showing better construct validity for our self-reported measures of SEMS compared to previous studies. Nevertheless, even with this extensive process, Cronbach's α were just around the limit of acceptance and the test re-test correlations in the pilot for the chosen items were low, suggesting that even if we are measuring the "right" constructs, there is still considerable measurement error.[11]

The second set of SEMS measures we use was administered on tablets. Contrary to commonly held beliefs among education researchers, but similar to Boon-Falleur et al. (2020), two task-based measurements of grit did not work well, with either a low test-retest correlation of 0.27 or a level of Mathematics ability that was not suitable for our respondents (single- and two-digit addition for a task designed to measure grit by Alan, Boneva and Ertac, 2019). We dropped these two measures from our final assessment and retained two other tasks.

One of the two tasks we retained was the GoNoGo task, used to measure impulse control, which had the highest test-rest correlation of 0.78 (Table 2). The participant is presented with a square on the screen for a very short period. If the square is of any color but black, the participant must touch the

[10] One of 24 items does not load on the expected factor. We drop this item for the rest of the analysis but including it does not affect our results.

[11] Interestingly, except for the extroversion subscale of the Big Five, Cronbach's α for the re-test is higher than the one from the test suggesting that repeated exposure may improve comprehension.

screen as quickly as possible. If the square is black (the "no go" stimulus), the respondent must inhibit their response. A total of 72 trials are completed (48 Go and 24 NoGo trials) and the main outcome we use is the average response time. The second was the Balloon Analogue Risk Task (BART), which measures risk-taking behavior by asking participants to maximize the amount of money they can win from the game. On each trial, respondents are presented with a balloon that they can pump. Each pump earned them (real) money but increased the likelihood that the next pump would "pop" the balloon, in which case they lost the accrued money for that balloon. If they instead chose to stop pumping the balloon, they collected their accrued money and moved to the next trial. The main outcome is the average number of pumps on the balloons that did not explode, and respondents earned on average PKR 322 or $3 from the game. While the BART displayed limited reliability (the test-retest correlation was 0.36), it was easy to understand for respondents and provided a measure of risk aversion. As with other skill categories, we aggregate the items from the self-reported scales and the scores from the tasks on tablets using principal factor analysis. We also verified that two additional versions of the index – one in which we only keep the self-reported scales, and one in which we drop the items that have poor properties as per the factor analysis—resulted in similar aggregate scales with correlations of 0.96 and 0.99 with our preferred SEMS index.

In summary, a key lesson from Laajaj and Macours (2021) is that measurement tools developed in high-income countries may have poor reliability and validity in low-income countries. Despite our extensive and iterative approach to building the SEMS measurements, an easy response to this challenge remains elusive; for instance, task-based measures, which may seem *ex-ante* attractive are not necessarily more reliable when education levels are low with high variability. Instead, our assessment, which builds on Laajaj and Macours's (2021), suggests that more enumerator training, more piloting to aid tool selection and better translations can help mitigate measurement error and response bias to a limited degree. The final assessments we employ in our survey perform better on multiple measures of validity but still have low levels of test-retest reliability (at least in the pilot). Even if this measurement error is classical, it will remain an important source of attenuation bias when these measures are used as dependent variables.

IV. MEASUREMENTS OF SKILLS: LEVELS AND PREDICTIVE ABILITY

Having documented the extent of reliability in our sample, we now turn to the question of predictive validity. We are interested in assessing the extent to which schooling is linked to the production of these skills and how cognitive and SEM skills in turn predict wages (for men) and labor force participation (for women). We are particularly interested in assessing these correlations conditional on years of schooling in order to assess the extent to which our measured skills mediate any correlation between labor earnings and years of schooling in the data. We discuss our results in two parts. We first focus on the distribution of skills in the population and its correlation with years of schooling; in particular, if measured properly, we would expect the cognitive and life skills to be strongly correlated with years of schooling. We then examine correlations between the skills we have measured and labor market outcomes among the young adults that comprise our sample, paying close attention to the role of the substantial migration we have documented previously in this sample.

IV.1. LEVEL OF SKILLS AND CORRELATION WITH SCHOOLING

Young adults, most of whom are between the ages of 19 and 28 in our sample, can count and identify numbers (79%) and a majority can add 3-digit numbers but only 5% could express the simple fraction 7/3 as $2\frac{1}{3}$ from among multiple options (Table 4). For the vernacular, Urdu, the majority can write simple words but cannot fill in a blank in a story by selecting the correct word. For English, respondents can match pictures to words, but cannot write simple sentences, for instance, using the word "deep" ("The water is deep" would be graded correctly). As reported previously, the low levels of skills in the population implied that in our adaptive testing, 43%, 64% and 77% of young adults in the sample were classified as Level 1 (Grades 1 to 3) ready for Urdu, Mathematics and English respectively, and only 26%, 34% and 20% made it to Level 2 (Grades 4 to 5) curricular levels.

We next assess the associations between our three main skills measures—cognitive skills, life skills and SEMS—and years of schooling, with individual and parental characteristics as controls. One concern is that each of these skills measures may be measuring the same underlying attribute and therefore providing little additional information. In fact, years of schooling, cognitive skills and life skills are all highly corelated with correlation coefficients ranging between 0.77 and 0.81 in our sample. However, SEM measures are capturing a different part of the skills set with lower correlation coefficients of 0.14 with cognitive skills to 0.18 with life skills (Appendix Table A6). This low correlation is similar to that reported by for AFQT scores and social skills in the United States as well as reviews of the literature, as discussed in Deming (2017).

Figure 2 then shows the relationship between years of schooling and cognitive skills and Table 5 presents the regression equivalent for the full set of skills. In Table 5 we present each regression with and without village fixed-effects as villages where the quality of schooling is higher (perhaps because returns are higher) will both have greater years of schooling and higher cognitive skills. There are several noteworthy patterns.

First, and reassuringly, more years of schooling are associated with higher cognitive skills (Figure 2 and Table 5), with coefficient estimates for each additional year stable at 0.17sd in specifications with and without village fixed-effects. As Figure 2 shows there is also considerable variation in cognitive skills for every level of schooling, a feature of these data that we will exploit when examining earnings below. Figure 2 also suggests that children learn less in college (0.28sd increase) compared to school, where moving from primary to middle or middle to secondary is each associated with a 0.7sd increase in cognitive skills. This is puzzling since gains during the college years reflect a combination of the causal impact of college, the selection into college and any depreciation in cognitive skills after leaving schools. Selection effects should be stronger for those going to college and children who only completed Grade 5 in our sample would have left school 10 years prior to our survey—we may have expected their skills to depreciate. If we believe that college adds value, both the selection and the depreciation effects should have led to steeper increases in cognitive skills in the college years.

One possibility is that these are the children for whom ceiling effects are important, but this appears not to be the case. The average cognitive skills index ranges from 0.9sd to 1.72sd between the start and end of college, which is far below the ceiling of +5sd imposed by the item-response scoring. Further, among those who have completed or are currently enrolled in college, only 1.2% report a ceiling across all tests.

We therefore investigated this further, first focusing on a sub-sample of children who have been tested multiple times and find that for this group, test scores were highest in 2011, when they were 17 years old, and then declined by 2018. Surprisingly, although youth who went on to college increased their test scores on every question, there were still basic concepts that they did not understand in each of the three subjects (Table A5). We then looked at depreciation by comparing the cognitive skills index of children who report the same number of schooling years but differ in age—those who are older would have graduated sooner (Table 5). We find that depreciation is small in our data: conditional on years of schooling, the cognitive skills index of a respondent who is a year older (and therefore left school one year earlier compared to another child with the same years of schooling) is 0.009sd lower. The data therefore suggest that the children who are enrolled in college are not the most selected and college attendance may not significantly increase cognitive skills. The low value-added of colleges has been noted previously by Loyalka et al. (2021) and the lack of selectivity in terms of test scores (as opposed to family background) has been documented by Bau, Das and Yi Chang (2021).

Second, like with cognitive skills, respondents' performance on the life skills questions was poor but positively associated with years of schooling (Table 4, Part 1, Panel B and Table 5). For Mathematics, 80% can read an electricity bill and calculate how much money they owe to the electricity company, but 50% have difficulty with non-linear pricing in their utility bill, and 36% cannot compute the correct change from a market transaction for five items (with five different quantities and prices). For reading (particularly the English alphabet) the picture is nuanced: 55% can read a complicated text ("*Peace be upon you. How are you and how is everyone at home?*") in Roman Urdu, which is Urdu written in the Roman script in texting language, but apparently cannot read the word "dog" in the Roman script. Perhaps the classification of the former as Urdu, not English, allows them to discern the question differently. For Urdu, we find that 73% can read a complex text accurately (in the Urdu script). Thus, it seems that functional reading skills are higher than what a test would suggest. Even so, the two measures have a correlation coefficient of 0.77 and every year of schooling is again associated with a 0.17sd increase in "life skills" (Tables A6 and Table 5).

Third, unlike for cognitive and life-skills, the correlation between SEM skills and schooling remains unclear for two reasons. At the population level, we draw attention to the fact that, unlike cognitive skills, where there is a very clear sense of deficits with regard to expectations (or compared to richer countries), SEMS skills in our data are actually comparable to those found in other countries (Table 4, Part 2). While cognitive skills in the population are consistent with the narrative that countries with lower GDP are also those with lower human capital, population-level measurements of SEM skills therefore do not support this conclusion. At the *individual* level, every additional year of schooling is associated with only a .033sd increase in SEMS (Table 5). If this coefficient reflects the causal contribution of schooling, it would still suggest that a year of schooling contributes 5 times as much to the develop of cognitive compared to SEM skills. But even this small correlation potentially reflects reverse causality; in fact, Barrera-Osorio, De Barros and Filmer (2018) demonstrate experimentally that there is zero causal impact of schooling on SEM skills in the Cambodian sample that we discuss later.

Fourth, there are important differences in the correlations between the different skills, gender and age. Women report higher cognitive skills, but lower SEM and life skills; the latter is particularly interesting given that life skills and years of schooling are positively correlated for this population. Further, cognitive skills appear to depreciate slowly with age while SEM skills increase. These patterns all suggest that cognitive and socio-emotional skills are very different measures of the abilities that people bring to the labor market, with different process of skills acquisition during and after the schooling years. Finally, the inclusion of village fixed-effects (Table 5, even columns) explains very little of the variation in any our skills measures with virtually no change in the R-squared or the coefficient on years of schooling. This is an independent and remarkable finding as there are large differences in consumption aggregates across these villages, ranging in 2003 from PKR 31,105 at the 10th percentile to PKR 81,718 at the 90th.

IV.2. PREDICTIVE VALIDITY OF SKILLS MEASURES: LABOR MARKET OUTCOMES

Having shown the correlation between years of schooling, cognitive and SEM skills in our sample, we now turn to the correlation between these skills and labor earnings. Given dramatic differences in labor force participation of 85% among men and 5% among women, we present specifications relating labor market outcomes to schooling and skills separately for men and women. We consider three main outcomes: labor force participation, earnings, and migration, with details regarding the measurement of wages presented in Appendix E. For skills, we treat years of schooling, cognitive skills and SEM skills as conceptually separate; we do not include life skills because they are highly correlated with cognitive skills and where the patterns differ (for women) the samples are generally too small to be able to pick up these nuanced differences. Here, we present results for our preferred specification, which excludes anyone currently enrolled, and includes age and district fixed-effects. Then, in Section III.2.2., we present a number of specifications in order to assess the robustness of our correlations to different samples, attrition weights and the treatment of agricultural income.

IV.2.1. Main results

We present three sets of descriptive correlations, using both mean and median regressions. The first two columns (Tables 6 to 8) for each outcome estimate:

$$y_i = \alpha + \beta s_i + \sum_{j=1}^{N} \gamma_j AgeFE_j + \varepsilon_i \quad (1)$$

where y_i is our outcome of interest; s_i is the years of schooling, and $AgeFE_j$ are age fixed-effects, one for each age *j*. The first column includes all the individuals while the second only those for whom we have skills measures. We then include measures of cognitive and socio-emotional skills. The specification for the third column is:

$$y_i = \alpha + \beta s_i + \delta\, Cog_i + \gamma\, SEMS_i + \sum_{j=1}^{N} \gamma_j AgeFE_j + \varepsilon_i \quad (2)$$

where Cog_i and $SEMS_i$ are the cognitive and socio-emotional skills indexes. Finally, for men, we look at associations between earnings and skills depending on where the respondent lives in a fully interacted specification (the sample of women working outside the village is too small to estimate a similar specification):

$$y_i = \alpha + \beta_1\, s_i + \delta_1\, Cog_i + \gamma_1\, SEMS_i + \varphi\, Out_i + \beta_2\, s_i * Out_i + \delta_2\, Cog_i * Out_i + \gamma_2\, SEMS_i * Out_i + \sum_{j=1}^{N} \gamma_j AgeFE_j + \varepsilon_i \quad (3)$$

Where Out_i is an indicator for the respondent living outside the village. In each of the mean regressions, standard errors are always clustered at the village level.

There are four main patterns. First, labor force participation (LFP) increases with years of schooling for women but declines for men (Table 6). As the specifications control for age fixed-effects, these coefficients do not reflect the fact that respondents with fewer years of schooling will have left school earlier and therefore have been in the labor market longer. Instead, the negative years-of-schooling coefficient likely reflects greater search durations for men, including preparing applications for public sector jobs or waiting for job offers from outside the village and country. In contrast, women in our sample are limited in terms of geographical mobility and the years they can work as participation plummets with marriage (Afridi, Dinkelman and Mahajan, 2018). What is striking for women is the size of the coefficient relative to the baseline female LFP: Among women with primary schooling or lower, female LFP is 2.39%, but among those with post-secondary education (including those who are currently enrolled), it rises to 17.6%, which is likely the maximal LFP for this cohort, as participation declines precipitously after marriage (Field and Vyborny, 2016). This difference in female LFP compares, for instance to an effect size of 4.9 percentage points in a program that is regarded as successful in improving women's labor market engagement in a similar context (Bandiera et al. 2020). For both men and women, cognitive skills are not correlated with LFP, but for men, a standard deviation increase in SEMS is associated with a 6-percentage point increase in LFP (Table 6).

The second result confirms that more years of education are associated with higher wages for both men and women. As we are interested only in the predictive validity of our skills measure, we do not regard this as an estimate of the Mincerian return and include our entire sample to capture both the wage and participation effects of higher skills for men; for women, we report associations with the full sample and for the sample of working women only. For men, each year of education is associated with $5.2 higher monthly wages, which translates to 3.4% of the $155 monthly wages reported in our sample[12] (Table 7). This is smaller than the 10-12% usually found in the literature using similar OLS specifications (Montenegro and Patrinos, 2014) and could reflect the fact that labor earnings are much higher for those attending college in Pakistan. This is a group that whose labor earnings we do not observe yet as our sample is too young to observe labor earnings after college completion and those who are studying in college report not working.

Despite the fact that the correlation between labor earnings and years of schooling does not reflect the experience of those who have studied at the college-level, for men we find that cognitive and socio-emotional skills are highly predictive of labor earnings. Our estimates suggest that, *conditional* on the years of schooling, every standard deviation increase in cognitive skills is associated with a $6.3/$8.3 increase in mean/median monthly earnings and a standard deviation increase in SEMS scores is associated with $14.6 (mean) to $16.5 (median) monthly higher earnings (Table 7). Further,

[12] If we consider the median regressions instead, each year of education is associated with $4.1 higher monthly wages, which also translates to 3.4% of the $120 median monthly wages reported in our sample

once we include cognitive skills and SEMS in the same regression as dependent variables, the coefficient on years of schooling reduces substantially, confirming that the correlation with schooling captures, in part, the predictive value of higher cognitive and socio-emotional skills.

For women, each additional year of education is associated with a $1.7 monthly increase in wages, which is 24% of the sample mean of $7.2 (Table 8). Once we condition on working women only (N=111), each additional year of education is associated with a $8.5 monthly increase from a baseline of $107, or an 8% increase[13]. In this case, the very small sample size of working women leads to a high degree of imprecision in the association between earnings and skills, with coefficients changing signs across mean and median specifications and always statistically insignificant at conventional levels of confidence.

Third, for men the associations with years of schooling, cognitive skills and SEMS are mediated by the migration status of the respondent (for women the sample of women working outside the village is too small). One channel through which skills and migration are linked is through the higher propensity to migrate for those with higher skills; this is shown in Appendix Table A7. For men, although years of schooling is not associated with migration, higher cognitive and SEM skills are both associated with a significantly greater likelihood of leaving the village (but not the country).

A second channel through which migration affects the returns to skills is more surprising and suggests that how skills are rewarded in the labor market depends on whether people are working within the village or outside. Specifically, for those who have chosen not to leave the village by the time of the resurvey, there appears to be a precisely estimated *zero* correlation with years of schooling and a strong correlation to SEMS in both mean and median specifications (Table 8, Columns 4 and 8). For those who have migrated, the results are reversed—the association with SEMS is statistically insignificant (although more imprecise) and the association with years of schooling is higher, ranging in monthly earnings from $4.8 to $10 in median and mean specifications. Finally, the association with cognitive skills is more sensitive to the specification used: for men who stayed in the village, we find a correlation of zero when looking at the median, and a positive (but imprecise) correlation when looking at the mean. For men who left the village, there is a large positive correlation with cognitive skills in the

[13] If we consider the median regressions instead, each year of education is associated with $5.9 higher monthly wages, which to translates to 9.4% of the $62.5 median monthly wages reported in the sample of working women.

median specification, but zero in the mean. Overall, the most consistent patterns are that SEMS are highly correlated with labor earnings for those in their native villages, while years of schooling are not. In contrast, for those residing outside the village, years of schooling are strongly associated with wages and the correlation with SEM skills is lower.

Fourth, there has been some discussion in the literature on precisely which SEMS are rewarded in the labor market (Díaz, Arias and Tudela, 2012; Valerio et al., 2016). To examine this question, we used the factor loadings from factor analysis to aggregate the variables into five factors corresponding to conscientiousness/grit, openness to experience, agreeableness, extroversion, and emotional stability. We then estimated Equation (3) using each measure separately, with results presented in Tables A8 and A9. Overall, we do not find clear differences in these associations. For men, there is some evidence that the SEMS with the higher correlations are grit/conscientiousness, and emotional stability and extroversion less so, but with the exception of the very low coefficient on extroversion, these differences do not point to a particularly strong correlation in the labor market for one particular skill. For women, again precision is very low, although even with the small sample the correlation with measures of grit remains positive and statistically significant (Table A9).

IV.2.2. Robustness checks

We present additional robustness checks in three parts. First, we examine the robustness of our estimates to different specifications, samples, income measures and attrition weights. Then, we present a back-of-the-envelope calculation on potential attenuation bias from measurement error. Finally, we present results from our sample in Cambodia to assess the structure of the correlations in a similar study (children born in rural areas then followed 8 years later), but a different context.

In order to check the robustness of our estimates to different samples and specifications, we estimated an additional 54 specifications. We investigated whether our estimates were affected by (a) the inclusion/exclusion of current students in the sample; (b) the inclusion/exclusion of village fixed-effects as a proxy for labor market returns in the region, (c) sensitivity to extreme values of income, (d) different ways of accounting for attrition. It is difficult *ex ante* to argue that one particular specification is definitely preferred to another in a predictive exercise such as ours—even the inclusion of students could be justified if respondents are enrolling in colleges (some of which are distancelearning courses) while waiting for jobs, or applying for certain positions (see Jeffrey, 2010). Instead, we opt for a transparent approach and assess whether the correlations we have found previously are consistent across many different robustness checks.

Rather than present each of these as separate tables, we plot all the estimates in specification curves, one each for years of schooling, cognitive skills and SEM skills for men within and outside the village in Figures 3 to 5. Here, the coefficients are plotted on the top panel and the linear combination of different restrictions is shown as black dots in the bottom panel. While there are certain specifications where the results become imprecise, the general theme of our regressions holds: SEM skills are strongly associated with wages within the village, while years of schooling are not, and years of schooling are correlated with wages for those outside the village while SEM skills are not.

A second exercise was to understand the extent of attenuation bias in our estimates for SEM skills. We used the test-retest reliability to obtain an estimate of the measurement error and rescaled the estimates using the standard formula for attenuation bias[14]. Doing so suggests that measurement-error corrected estimates would be \$42.9 instead of the \$14.6 we report in the main results. Our estimated reliability of 0.34 from the average of test-retest correlations during the pilot is likely an underestimate of reliability in the final sample as Cronbach's α is 30% higher on average suggesting lower measurement error in the final data collection. However, even if we adjust for this, and estimate reliability to be 0.44, which is now likely an upper bound, the corrected coefficient would be \$33.

A third exercise sought to understand whether these patterns are specific to our particular sample. For this, we incorporated further data from the second study site in Cambodia. The structure of the sample and the survey is very similar, with children first surveyed in 2008 and then re-contacted and resurveyed in 2016/17, allowing us to look at links between education and skills in an originally rural sample. Unfortunately the migration status is harder to determine and the measures of socio-emotional skills are less reliable as we discuss below and in Appendix D.

[14] For classical measurement error, the observed $\hat{\beta} = \frac{\sigma_{X^*}}{\sigma_{X^*}+\sigma_{\varepsilon}}\beta$ with $\frac{\sigma_{X^*}}{\sigma_{X^*}+\sigma_{\varepsilon}}$ being the reliability, that we approximate by the test-retest correlation.

To measure socio-emotional skills, we used the same Grit and "Big Five" self-reported scales as in Pakistan. We also measured growth mindset – the belief that we can get smarter through hard work and practice – using a 4-items scale. Then, we asked four questions about locus of control. Internal locus of control measures the degree to which people believe that they have power over the outcome of events in their lives, as opposed to external forces beyond their control. Finally, we administered the Strength and Difficulty Questionnaire (SDQ), a brief behavioral screening questionnaire aimed at measuring two main constructs: behavioral difficulties and pro-social behavior.

The reliability and validity of these measures are lower than in Pakistan. After correcting for acquiescence bias, the Cronbach's α's of the different constructs range between 0.08 and 0.71, with only one of 11 measures passing the 0.7 threshold (Table 2). Moreover, as in Pakistan, the test-retest correlations are low, ranging from 0.14 to 0.43. Finally, when conducting factor analysis, the skills factor structure is not reproduced. Only three factors are retained, and one single factor encompasses a wide range of items aimed at measuring different concepts.

Keeping these limitations around reliability and construct validity in mind, Tables A10 to A13 examine the predictive ability of these different skills measures. Like in Pakistan, there is a generally low level of cognitive skills in these youth populations. For instance, only 34% of respondents correctly answer the following item: "3 ox carts carry 1,000 kg of rice. How many kg of rice can 9 ox carts carry?". We also find strikingly similar associations between years of schooling and cognitive skills. Every additional year of schooling is associated with a 0.17sd increase in cognitive skills in Pakistan and a 0.18sd increase in Cambodia (Table A10). The results from Cambodia also confirm that the correlation between SEM skills and years of schooling is weaker. An additional year of schooling is associated with a 0.05sd increase in SEMS and in related experimental work on the value of additional years of schooling, Barrera-Osorio, De Barros and Filmer (2018) show that there is no causal link between schooling and SEM skills in their sample. Thus, the results replicate across two very different rural regions of the world, with an association between schooling and cognitive skills, but weaker associations between schooling and SEMS.

Turning to the predictive value in earnings regressions, we first highlight that the Cambodian sample is very different in occupational structure, with 84% of respondents reporting agriculture as their main occupation, and LFP rates of 95% for both men and women. The earnings regressions lack precision in many specifications, but if there are any associations between earnings and skills, they are entirely for SEM skills, with zero or even negative associations between earnings and years of schooling or earnings and cognitive skills. In fact, if we only focus on the median regressions, we do not see any correlations between earnings and either skills or schooling for men. In the mean regressions, we find (imprecise) negative associations to cognitive skills and schooling and positive associations with SEMS (Table A12). As in Pakistan, the positive correlations with SEMS are only for men who chose to stay in their original village. For women, we find positive associations with SEMS in both the median and mean specifications, although the mean estimates are imprecise (Table A13). Schooling also appears to be more strongly correlated with earnings for women than for men, with a positive correlation for cognitive skills among women who have remained in the village.

V. DISCUSSION

An established literature in the United States shows that the returns to schooling arise, in part, from the link between earnings and test scores (Altonji and Pierret, 2001). However, as was pointed out by Bowles and Gintis (1976) and Heckman and Rubinstein (2001), it is also the case that individuals with seemingly identical years of schooling and/or cognitive skills are compensated very differently in the labor market. To explain this earnings residual, they argued that schools must produce other skills that are recognized and rewarded in the labor market. The measurement of, and returns to these socio-emotional skills or SEMS, has been an active area of research for at least the last two decades with Deming's (2017) recent contribution suggesting that the returns to some of these skills have increased over time.

Understanding whether the consistently positive Mincer returns to years of schooling in low-income countries similarly reflects the production of cognitive and socio-emotional skills in schools has proven difficult, even before addressing difficult identification problems. For instance, Glewwe, Huang and Park (2011) report zero labor market returns to both cognitive and SEM skills in China, and Laajaj and Macours (2021) have documented severe issues with the measurement of SEMS in low-income contexts. The World Bank, through its STEPS surveys has collected data on the same socio-emotional skills among urban adults in seven countries, but are unable to find any systematic patterns. After conditioning on education and estimating separate regressions for each of the seven skills measured in these surveys, Valerio et al. (2016) find 4 coefficients that are positive and statistically significant at the 90% level of confidence, 3 coefficients that are *negative* and statistically significant and

42 coefficients that are small and imprecise. As the standard errors are not adjusted for multiple-hypotheses tests, it is likely that even the statistically significant results are due to chance. They write *"It must be noted that the noncognitive skills measures are a function of scores on three to five items each. We believe the limited number of items for each (noncognitive skill) scale could be limiting the reliability of these measures and obscuring the true relationship between noncognitive skills and earnings."* (page 26)

It is in this context that our measurements of test scores and SEMS in Pakistan and Cambodia add value. We are first able to show that a lengthy tool development process can improve the construct validity and reliability of SEM measures in low-income countries. We are then able to document the predictive value of these skills for labor earnings and highlight the critical role of migration in a sample that started off in rural areas, but branched into multiple locations over 15 years. Nevertheless, there is considerable room for improvement both for the measures we have used and in basic research on what tools are relevant in contexts such as ours. In this discussion, we highlight three implications of our results and the questions it raises.

IMPLICATION 1: SIZEABLE CORRELATIONS. A one standard deviation increase in test scores/SEMS is associated with a \$6-\$8/\$15 increase in monthly wages. If we interpret these effects as causal and, sticking to the lower end of these estimates, we would conclude that a program that can boost test scores by 1sd by the time children leave school, will increase wages by \$6 every month or \$72 per year. A simple calculation helps contextualize this number in terms of actual school budgets. Specifically, if this is the annual increase in labor earnings for a working life of 40 years, using the World Bank recommended discount rate of 5% yields an additional lifetime discounted income of \$1300 (World Bank 2015). If we assume that the investment will only kick in 10 years after the additional spending, this would still lead to an additional benefit of \$802 per child, or more than five times the average annual spend of \$132 per child in 2017-2018 (AEPAM, 2019). For a program that increases SEMS by a similar amount, governments should be willing to spend twice as much. Yet, any spending close to that is regarded as next to impossible in a country like Pakistan; we believe that part of this is because the benefits of increasing cognitive and socioemotional skills on labor earnings has never been clearly shown. Of course, the difficulty will be in assessing the extent of omitted variable bias in correlations such as these—so an urgent step given the magnitude of these coefficients is to assess the plausibility of these estimates as reflecting causal links from other contexts and studies. If these estimates hold, it could lead to an important reassessment of how much governments should invest to increase test scores in low-income countries.

IMPLICATION 2: BOWLES and GINTIS REVISITED. Bowles and Gintis (1976) argued that a central function of schools was to produce workers for the capitalist factory system and instill in them the new skills that were required to operate assembly lines at the turn of the 20th century. This may include cognitive skills, but it also included skills such as punctuality, discipline, and respect for hierarchy. In their formulation, the Mincer returns to years of school captured the imparting of these skills to their students. In current terminology, their question could be reframed as assessing the extent to which socio-emotional skills causally mediate the returns to schooling. Interestingly, even at the level of correlations, compulsory secondary education in the U.S. from the early 1900s implied that studies could only examine variation in cognitive skills as there is arguably "too little" variation in years of schooling to exploit (the relevant margin is college versus school, but this introduces multiple additional complications and is therefore treated as a composite difference).

In contrast, in our data, like in many low-income countries, there is sizeable variation in years of schooling, from zero to college completion along with considerable variation in test scores for each year of schooling. Exploiting this variation shows that, just as Bowles and Gintis (1976) predicted, including test scores and SEMS as additional dependent variables reduces the coefficient on years of schooling (by half in our case). So, Mincer returns are indeed capturing (in part) the skills that we have directly measured. Yet, two problems remain. First, as shown in Table 5, *schools* appear to be quite marginal in producing the SEMS skills we have measured, so our results sit uneasily with Bowles and Gintis' preferred explanation that the Mincer returns to years of schooling really capture the ability of schools to socialize and prepare children for factory production in a capitalist system. Second, we can explain at most 13% of the variation in labor earnings for men, and 26% for women. So, even with a comprehensive skills measurement component, much of what our respondents earn appear to reflect other considerations. Another way to see this is to note that the extent to which the coefficient on years of school will decline once we include other skills depends on the covariance of the measured skill with labor earnings and with years of schooling. Using the coefficients from Table 5 and Table 7 for all respondents regardless of migration status suggests that, for men, cognitive skills account for 27% of the estimated correlation of $5 for each year of education and SEM skills account for another 9%, for a total of 36%. This still leaves two-thirds of the correlation between labor earnings and years of schooling as an unexplained residual. If instead we focus on non-migrants, the results are incongruent with Bowles and Gintis (1976) as there is no positive return to schooling to begin with. On the other hand, among migrants the correlation between years of schooling and labor earnings is higher, but its mediation through the skills we have measured is lower. These results raise important questions regarding the types of skills that schools impart to their students that are then rewarded in labor earnings; if earnings indeed reflect skills, there is ample room for better measurement in future studies.

IMPLICATION 3: MIGRATION AND SKILLS. Our final point takes the question of how labor markets reflect skills and relates to a lively debate on migration, labor earnings, and skills. This literature is concerned with the returns to migration. The main challenge is that people who migrate may have systematically different skills so that wages for migrants reflect both the causal impact of migration and the differences in skills. Although studies have designed clever natural and randomized experiments to get around this problem, we are not aware of a literature that directly assesses the skills of migrants and non-migrants to understand how these are rewarded in the labor market at different locations. The combination of our sampling strategy and skills measures allows us to present the correlational evidence on this, with several noteworthy findings.

First, it is true that those with higher test scores and SEMS are also more likely to migrate, at least within the country if not outside. Second, it is also true that including test scores and SEMS lowers the difference in labor earnings of migrants versus non-migrants. However, a remarkable—and new—finding in this sample, which is then corroborated in Cambodia, is that the association with SEMS is positive (and large) *only* for non-migrants and the correlation with years of schooling is positive (and large) *only* for migrants. This suggests that the returns to different skills vary by the migration status of the respondent. One possibility for why this may be so, raised by Carranza et al. (2020) in South Africa, is that this reflects a lack of information among employers who therefore use observable signals (years of schooling) instead. If so, one immediate implication is that the type of information on skills that Carranza et al. (2020) provide to workers and employers may have enormous value for migrants as years of schooling explains only 15% of the variation in the SEM skills we measure. A second implication is that under the current regime, if individuals estimate the returns to years of schooling by looking at the labor earnings of those around them, non-migrants will believe these returns to be zero. This is a direct verification of Jensen's (2010) point in the Dominican Republic, again demonstrating a potential role for information in schooling decisions.

VI. CONCLUSION

We sought to measure cognitive and socio-emotional skills in a sample of young adults in Pakistan. The sample is distinguished by the fact that there were all observed first in rural areas and then followed 15 years later to their current residence. Considerable migration in this population provides insights on the link between skills, migration, and labor earnings. Our results suggest that a number of additional survey measures can help improve the measurement of socio-emotional skills in low-income countries, and that both cognitive and socio-emotional skills are correlated with (a) years of schooling and (b) correlated with labor earnings for men. In addition, there are important differences in the structure of these correlations depending on the residence of the respondent.

We, therefore, made progress in addressing several puzzles around the measurement of skills in low-income countries, but with substantial room for further improvements. In our data, cognitive and socio-emotional skills account for one-third of the correlation between years of schooling and labor earnings and the correlation between socio-emotional skills and years of schooling is small. We have therefore not been able to measure the skills that are highly correlated with years of schooling *and* highly rewarded in the labor market. It could be that the unexplained variation in labor earnings reflects wedges that lead to inefficient labor markets, like in Carranza et al. (2020). Or, it could be that there are genuinely other measures that employers look for and that schools are designed to produce (the `hidden curriculum' in Bowles and Gintis, 1976), but these will require contextual measurements built from the ground up in low-income countries.

References

Afridi, F., Dinkelman, T., & Mahajan, K. (2018). Why are fewer married women joining the work force in rural India? A decomposition analysis over two decades. Journal of Population Economics, 31(3), 783-818.

AEPAM (2019). Public financing in education sector, Academy of Educational Planning and Management Ministry of Federal Education and Professional Training. Report. Access: http://library.aepam.edu.pk/Books/Financing%20in%20Education%20Sector%202019.pdf

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. The Quarterly Journal of Economics, 134(3), 1121-1162.

Andrabi, T., Das, J., Khwaja, A., Vishwanath, T., & Zajonc, T. (2002). Test feasibility survey Pakistan: Education sector. Cambridge, MA: Harvard Kennedy School Working Paper.

Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T., & Zajonc, T. (2007). Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate. World Bank, Washington, DC.

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. The quarterly journal of economics, 116(1), 313-350.

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020). Women's empowerment in action: evidence from a randomized control trial in Africa. American Economic Journal: Applied Economics, 12(1), 210-59.

Banerjee, A. V., Bhattacharjee, S., Chattopadhyay, R., & Ganimian, A. J. (2017). The Untapped math skills of working children in India: Evidence, possible explanations, and implications. Unpublished manuscript.

Barrera-Osorio, F., De Barros, A., & Filmer, D. (2018). Long-term impacts of alternative approaches to increase schooling: evidence from a scholarship program in Cambodia. World Bank Policy Research Working Paper, (8566).

Bau, N., Das, J., & Chang, A. Y. (2021). New evidence on learning trajectories in a low-income setting. International Journal of Educational Development, 84, 102430.

Beegle, K., De Weerdt, J., & Dercon, S. (2011). Migration and economic mobility in Tanzania: Evidence from a tracking survey. Review of Economics and Statistics, 93(3), 1010-1033.

Boon-Falleur, M., Bouguen, A., Charpentier, A., Algan, Y., Huillery, E., & Chevallier, C. (2020). Measuring socio-emotional skills in schools: simple questionnaires outperform behavioral tasks.

Bowles, S., & Gintis, H. (1976). Schooling in capitalist America: Educational reform and the contradictions of economic life. Basic Books.

Brunello, G., & Schlotter, M. (2011). Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems.

Carranza, E., Garlick, R., Orkin, K., & Rankin, N. (2020). "Job Search and Hiring With Two-Sided Limited Information About Workseekers' Skills.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. Econometrica, 82(5), 1671-1748.

Cunha, F., & Heckman, J. J. (2010). Investing in Our Young People. NBER Working Paper No. 16201. National Bureau of Economic Research.

Das, J., & Zajonc, T. (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. Journal of Development Economics, 92(2), 175-187.

Deming, D. J. (2017). The growing importance of social skills in the labor market. The Quarterly Journal of Economics, 132(4), 1593-1640.

Díaz, J. J., Arias, O., & Tudela, D. V. (2012). Does perseverance pay as much as being smart? The returns to cognitive and non-cognitive skills in urban Peru. Unpublished paper, World Bank, Washington, DC.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. Journal of personality and social psychology, 92(6), 1087.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). Journal of personality assessment, 91(2), 166-174.

Duflo, E., Dupas, P., & Kremer, M. (2021). The impact of free secondary education: Experimental evidence from Ghana (No. w28937). National Bureau of Economic Research.

Field, E., & Vyborny, K. (2016). Female labor force participation in Asia: Pakistan country study. Manila: Asian Development Bank.

Foster, A. D., & Rosenzweig, M. R. (1996). Technical change and human-capital returns and investments: evidence from the green revolution. The American economic review, 931-953.

Glewwe, P., Huang, Q., & Park, A. (2011). Cognitive skills, non-cognitive skills, and the employment and wages of young adults in rural China (No. 321-2016-11018).

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. Proceedings of the national Academy of Sciences, 104(33), 13250-13255.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. American Economic Review, 91(2), 145-149.

Jeffrey, C. (2010). Timepass: Youth, class, and time among unemployed young men in India. American Ethnologist, 37(3), 465-481.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. The Quarterly Journal of Economics, 125(2), 515-548.

Laajaj, R., & Macours, K. (2021). Measuring skills in developing countries. Journal of Human resources, 56(4), 1254-1295.

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. Behavior research methods, 43(2), 548-567.

Loyalka, P., Liu, O. L., Li, G., Kardanova, E., Chirikov, I., Hu, S., ... & Li, Y. (2021). Skill levels and gains in university STEM education in China, India, Russia and the United States. Nature human behaviour, 1-13.

McKenzie, D., Stillman, S., & Gibson, J. (2010). How important is selection? Experimental vs. non-experimental measures of the income gains from migration. Journal of the European Economic Association, 8(4), 913-945.

Montenegro, C. E., & Patrinos, H. A. (2014). Comparable estimates of returns to schooling around the world. *World Bank policy research working paper*, (7020).

Nunally, J. C., & Bernstein, I. H. (1978). Psychometric theory.

Outes-Leon, I., & Dercon, S. (2009). Survey attrition and attrition bias in Young Lives.

Pierre, G., Sanchez Puerta, M. L., Valerio, A., & Rajadel, T. (2014). STEP skills measurement surveys: innovative tools for assessing skills.

Subramanian, N. (2020). Workplace Attributes and Women's Labor Supply Decisions: Evidence from a Randomized Experiment. Mimeo, Duke University.

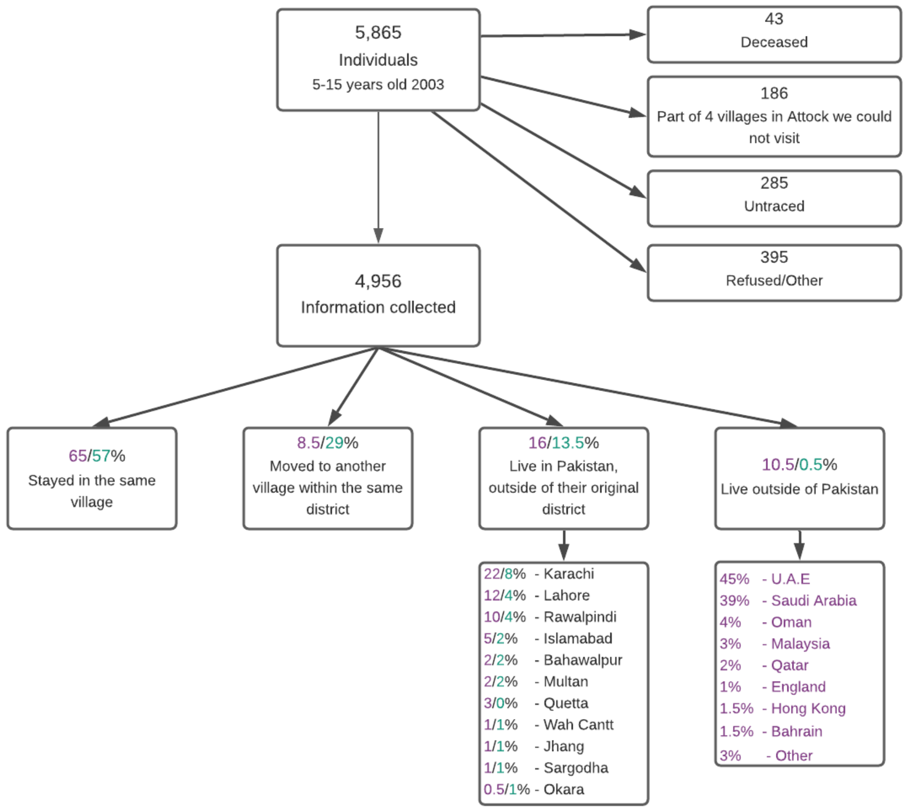
Valerio, A., Sanchez Puerta, M. L., Tognatta, N., & Monroy-Taborda, S. (2016). Are there skills payoffs in low-and middle-income countries? Empirical evidence using STEP data. World Bank Policy Research Working Paper, (7879).

World Bank. 2015a. Technical Note on Discounting Costs and Benefits in Economic Analysis of World Bank Projects. Washington, DC: World Bank.

World Bank. 2018. World Development Report 2018: Learning to Realize Education's Promise. Washington, DC: World Bank.

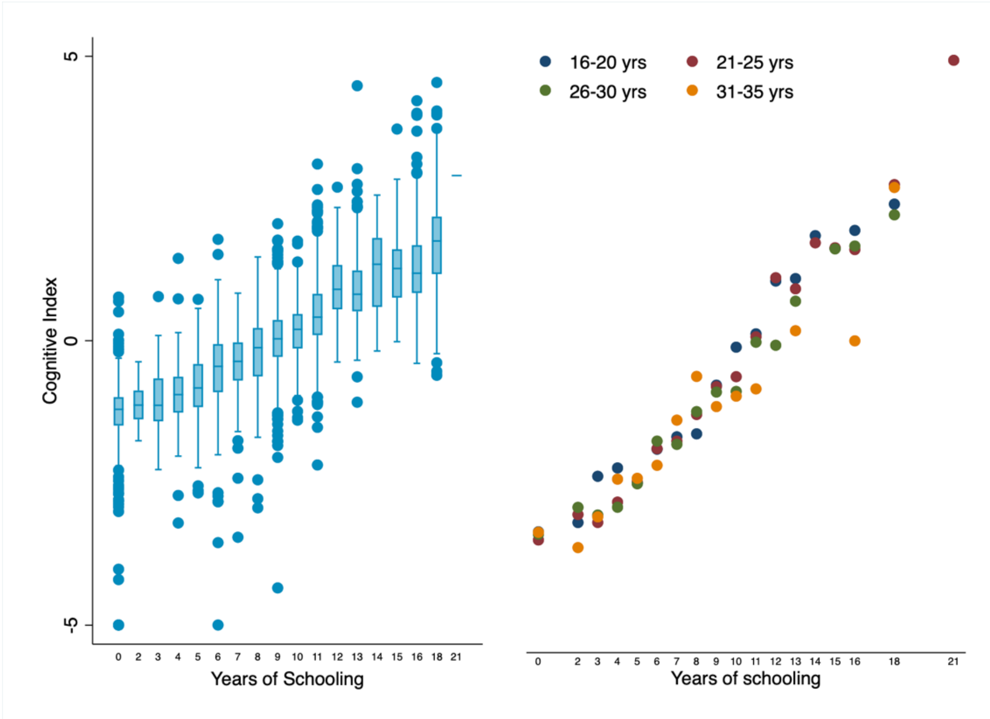
MAIN FIGURES

Figure 1. Migration patterns



Notes. This figure shows where respondents in our sample lived at the time of the follow-up survey in 2018. Numbers for men are in purple and for women in blue.

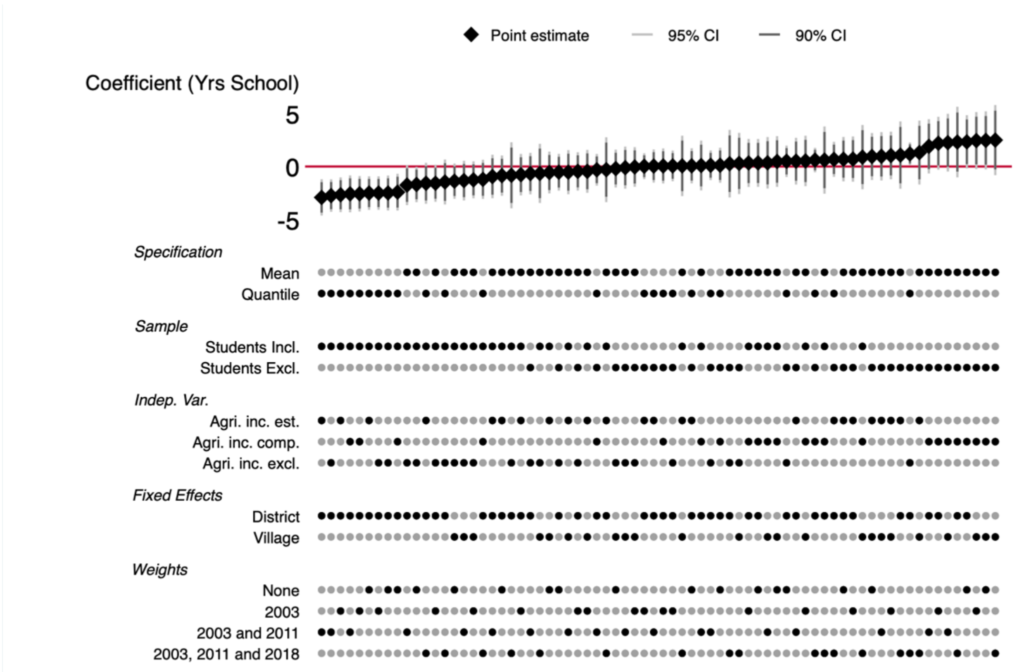
Figure 2. Years of Schooling, Age, and Cognitive Skills Formation



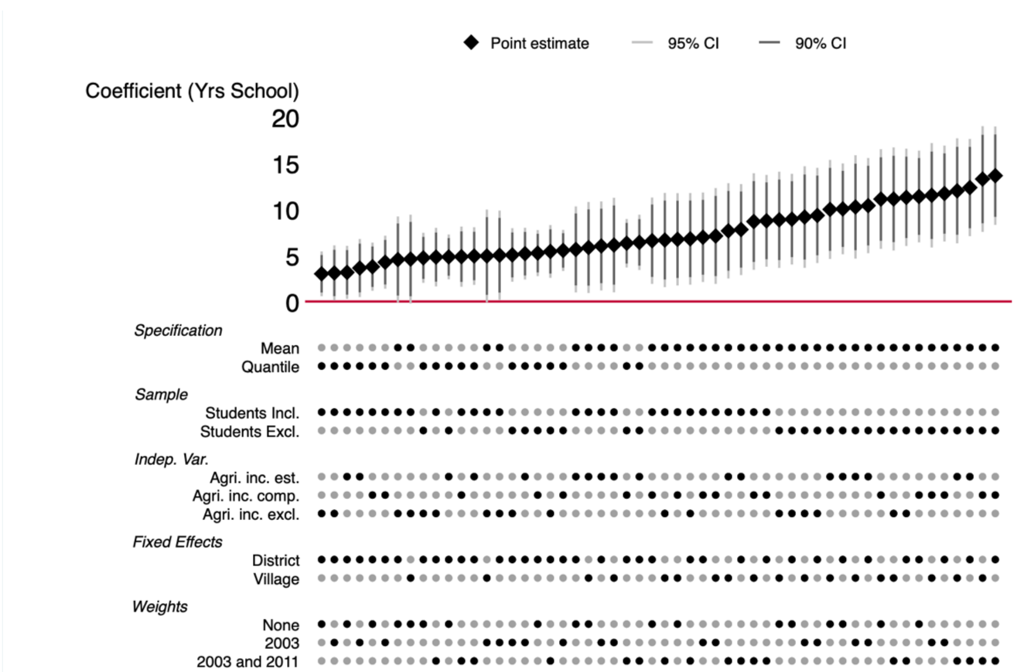
Notes. This graph shows the relationship between schooling and cognitive skills formation. The Cognitive Index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper-based test and the computer-adaptive tablet-based test (leaving out items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents answered correctly). The left panel shows, for each year of schooling, the distribution of the cognitive index. The right panel shows, for each year of schooling the average cognitive skills for respondents who are between 16-20 years old, for respondents who are between 21-25 years old, for respondents who are between 26-30 years old, and for respondents who are between 31-35 years old. The sample includes all men and women who have cognitive skills measures in our sample (4,401 respondents).

Figure 3. Specification Curves – Years of Schooling

Panel A. Men within village



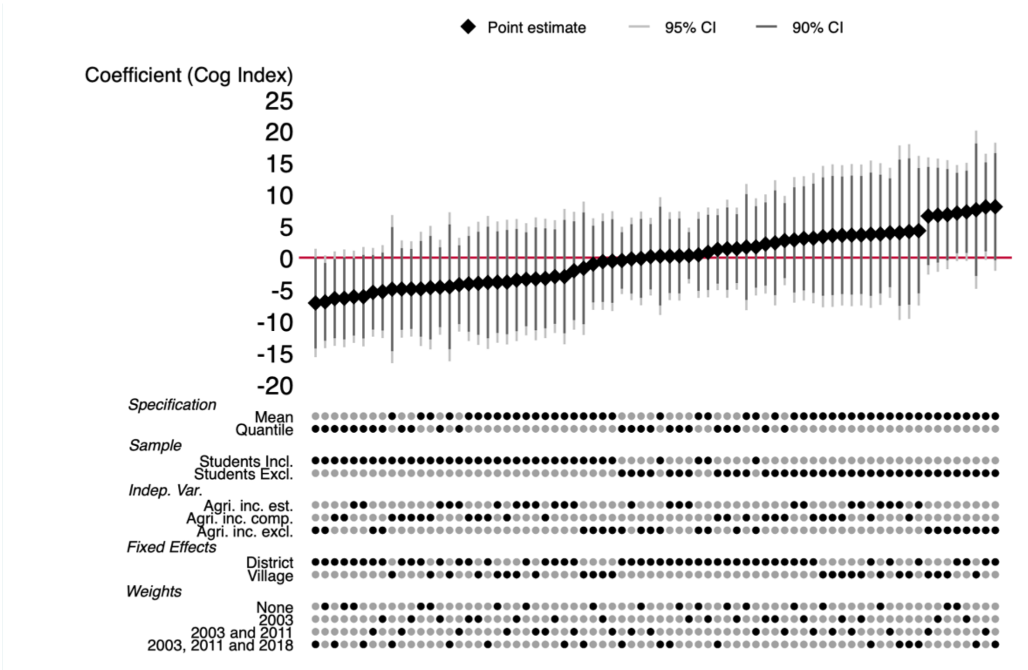
Panel B. Men outside of village



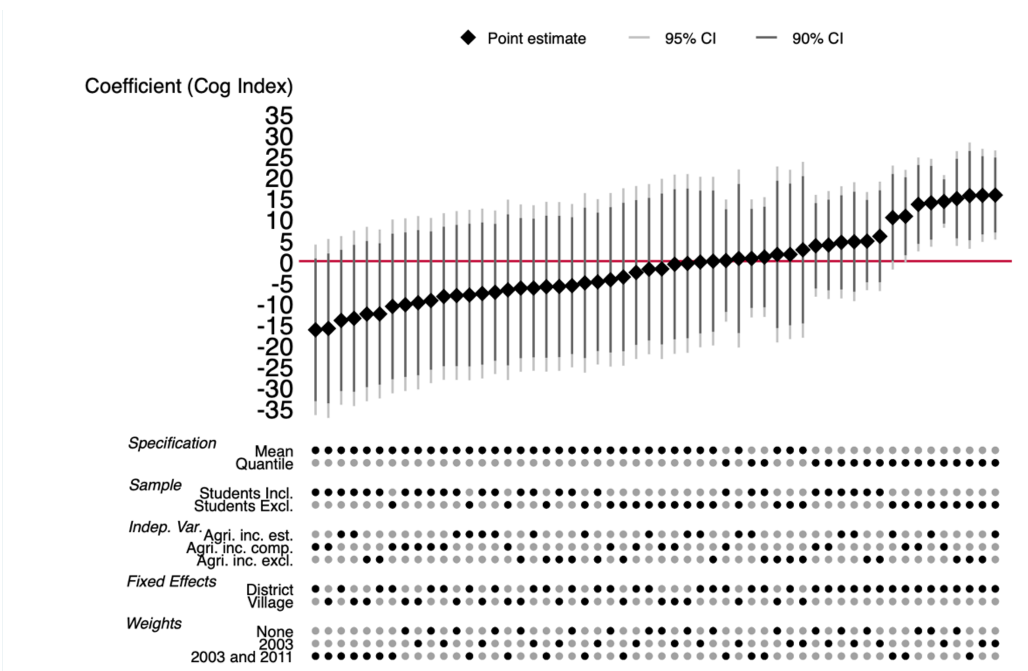
Notes. Each dot in the top panel of the graph depicts the years of schooling coefficient from the fully interacted specification for men within the village (Panel A) and men outside the village (Panel B). The dots vertically aligned below indicate the analytical decisions behind those estimates. A total of 54 specifications were estimated.

Figure 4. Specification Curves – Cognitive Skills

Panel A. Men within village



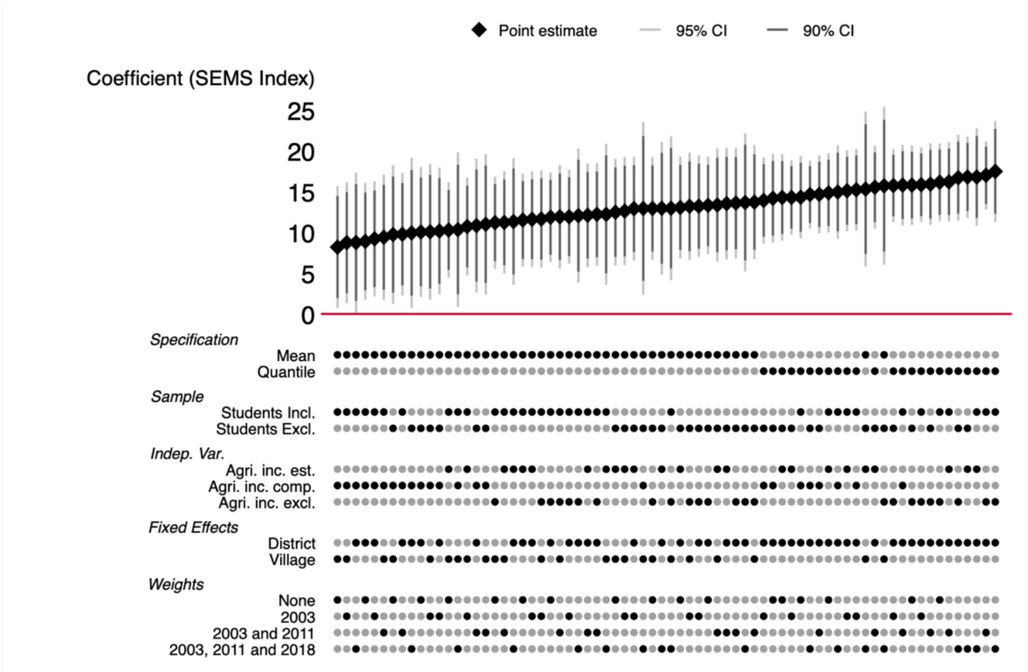
Panel B. Men outside of village



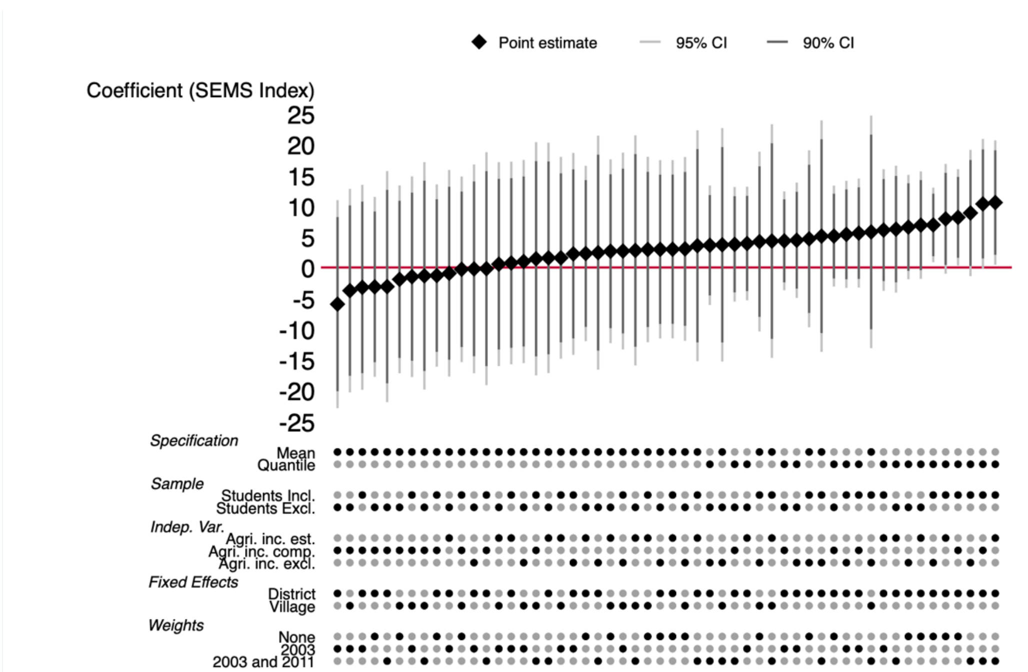
Notes. Each dot in the top panel of the graph depicts the cognitive skills index coefficient from the fully interacted specification for men within the village (Panel A) and men outside the village (Panel B). The dots vertically aligned below indicate the analytical decisions behind those estimates. A total of 54 specifications were estimated.

Figure 5. Specification Curves – Socio-Emotional Skills

Panel A. Men within village



Panel B. Men outside of village



Notes. Each dot in the top panel of the graph depicts the socio-emotional skills index coefficient from the fully interacted specification for men within the village (Panel A) and men outside the village (Panel B). The dots vertically aligned below indicate the analytical decisions behind those estimates. A total of 54 specifications were estimated.

MAIN TABLES

Table 1. Summary Statistics

	All			Men/Fathers			Women/Mothers			Difference
	Mean	SD	Obs	Mean	SD	Obs	Mean	SD	Obs	Women-Men
Panel A: Individual - 2018										
Age	23.77	3.62	4956	23.8	3.54	2596	23.73	3.7	2360	-0.07
Years of schooling	8.55	4.89	4955	8.92	4.35	2595	8.13	5.39	2360	-0.79***
Respondent can read	0.73	0.44	4402	0.74	0.44	2222	0.73	0.45	2180	-0.01
Ever Married	0.45	0.5	4956	0.34	0.47	2596	0.58	0.49	2360	0.25***
Age at first marriage	20.95	3.43	2248	22.37	3.31	874	20.05	3.2	1374	-2.31***
Has children	0.33	0.47	4956	0.22	0.41	2596	0.45	0.5	2360	0.23***
Working (students excl.)	0.47	0.5	4455	0.85	0.36	2351	0.05	0.22	2104	-0.79***
Main work is farming (students excl.)	0.04	0.19	4455	0.07	0.26	2351	0	0	2104	-0.07***
HH has toilets on premises	0.96	0.19	4404	0.97	0.18	2223	0.96	0.2	2181	-0.01
HH has access to electricity	0.98	0.14	4404	0.98	0.13	2223	0.98	0.15	2181	-0.00
Lives in same village than in 2003	0.61	0.49	4956	0.65	0.48	2596	0.57	0.5	2360	-0.08***
Panel B: Household - 2003										
Parent years of schooling	3.1	4.28	8780	4.74	4.75	4201	1.6	3.12	4579	-3.15***
Parent can read	0.37	0.48	8945	0.53	0.5	4219	0.23	0.42	4726	-0.30***
Parent age at first marriage	22.07	4.82	7522	24.6	4.8	3648	19.69	3.42	3874	-4.92***
Parent main work is farming	0.18	0.39	8964	0.33	0.47	4230	0.05	0.21	4734	-0.29***
Parent has toilets on-premises	0.58	0.49	9699	0.58	0.49	4847	0.58	0.49	4852	-0.00
Parent has access to electricity	0.88	0.33	9685	0.88	0.33	4840	0.88	0.33	4845	-0.00

* p<.10 ** p<.05 *** p<.01

Notes. This table shows the sample's characteristics for the whole sample, for men, and women. Panel A shows the characteristics of respondents in 2018. There is one respondent for whom the years of schooling information is missing because it was indicated as "Don't know" in the survey. The variable "Respondent can read" is coming from the functional literacy and numeracy ("life skills") section and therefore only filled for the respondents who were surveyed in person and have completed surveys, that is 4,402 respondents. It indicates that the respondent was able to read and understand a greeting text message in Urdu script. "Age at first marriage" is filled for all respondents who have ever been married. It was recorded as such for respondents who completed the direct version of the questionnaire and live with relatives and imputed from the age at which the respondent got married with the current spouse for the rest. "HH has toilets on-premises/access to electricity" indicates that the respondent lives in a household that has a toilet on-premises/access to electricity. This variable is filled for all in-person direct versions of the questionnaire, that is 4,404 respondents. Panel B shows the same characteristics collected in 2003 for the respondents' parents. These variables were filled for individuals whose household was surveyed that year and who were living with their parents. Father and mother age at first marriage variables are coming from the 2011 datasets and only filled for respondents who lived with their father and/or mother at the time. The table also shows the differences in the means across men/fathers and women/mothers. Standard errors are clustered at the village level. There are 108 clusters in the sample.

Table 2. Reliability of socio-emotional skills measures

Construct	Instrument	Mode	Country	Measures corrected for acquiescence bias (1) Alpha Test	(2) Alpha Re-test	(3) Test-Retest	Raw measures (4) Alpha Test	(5) Alpha Re-test	(6) Test-Retest	N	Nb. Items
Grit	**Grit Scale**	Self-reported	Cambodia	0.39	-	-	0.16	-	-	3282	8
			Pakistan	0.75	-	-	0.68	-	-	4395	9
			Pakistan Pilot	0.57	0.72	0.2	0.47	0.63	0.2	397	10
	Alan & Ertac Grit Task "Additions Game"	Task-based	Pakistan Pilot	-	-	-	-	-	0.42	402	-
	Frustration Task	Task-based	Pakistan Pilot	-	-	-	-	-	0.27	402	-
Openness to Experience	**Big Five**	Self-reported	Cambodia	0.08	0.22	0.24	0.4	0.35	0.38	3287	3
			Pakistan	0.53	-	-	0.63	-	-	4475	3
			Pakistan Pilot	0.31	0.31	0.29	0.42	0.54	0.27	403	3
Conscientiousness	**Big Five**	Self-reported	Cambodia	0.48	0.39	0.43	0.37	0.43	0.42	3286	3
			Pakistan	0.57	-	-	0.4	-	-	4475	3
			Pakistan Pilot	0.62	0.71	0.13	0.44	0.58	0.09	402	3
Extroversion	**Big Five**	Self-reported	Cambodia	0.41	0.03	0.14	0.2	0.01	0.13	3286	3
			Pakistan	0.68	-	-	0.64	-	-	4475	3
			Pakistan Pilot	0.47	0.44	0.29	0.4	0.35	0.3	403	3
Agreeableness	**Big Five**	Self-reported	Cambodia	0.45	0.44	0.38	0.31	0.47	0.38	3287	3
			Pakistan	0.6	-	-	0.42	-	-	4475	3
			Pakistan Pilot	0.5	0.73	0.13	0.35	0.59	0.11	403	3
Emotional Stability	**Big Five**	Self-reported	Cambodia	0.04	0.11	0.23	.	.	0.24	3289	3
			Pakistan	0.6	-	-	0.64	-	-	4475	3
			Pakistan Pilot	0.39	0.51	0.41	0.39	0.47	0.43	402	3
Big Five	**Big Five**	Self-reported	Cambodia	0.52	.	0.36	0.43	0.18	0.32	3279	15
			Pakistan	0.64	-	-	0.56	-	-	4475	15
			Pakistan Pilot	0.62	0.71	0.29	0.53	0.66	0.25	401	15
Locus of Control	**Locus of Control**	Self-reported	Cambodia	0.58	-	-	0.2	-	-	3287	4
			Pakistan Pilot	0.52	0.61	0.45	0.29	0.31	0.45	403	4
Growth Mindset	**Growth Mindset**	Self-reported	Cambodia	0.4	-	-	0.4	-	-	3284	4
Behavioral Difficulties	**SDQ**	Self-reported	Cambodia	0.71	-	-	0.64	-	-	3280	20
Pro-social behavior	**SDQ**	Self-reported	Cambodia	0.31	-	-	0.63	-	-	3283	5
Impulsiveness	**Barratt Impulsiveness Scale (BIS)**	Self-reported	Pakistan Pilot	0.71	0.77	0.4	0.64	0.71	0.38	397	30
Risk-taking behavior	**Balloon Analogue Risk Task (BART)**	Task-based	Pakistan Pilot	-	-	-	-	-	0.36	402	-
Self-Control	**GoNoGo**	Task-based	Pakistan Pilot	-	-	-	-	-	0.78	402	-

Notes. This table reports estimates of the reliability of socio-emotional skills measures. Our survey instruments were developed through an iterative process that started with data collection in Cambodia for a related project in 2017. For quality check purposes, we randomly re-surveyed 13% of the sample in Cambodia. This re-survey was a subset of the original instrument that included the "Big Five" scale. The estimates from the Cambodia project are displayed in the rows "Cambodia". We then conducted a pilot in Pakistan in 2018 with 403 respondents and randomly re-surveyed half (201) two weeks later. Estimates from the pilot in Pakistan are displayed in the rows "Pakistan Pilot". Finally, we show estimates from the full data collection in Pakistan in the row "Pakistan".

The instruments included self-reported scales and tasks administered on tablets. For each instrument, we present the construct the instrument intends to measure as well as the mode of administration. We use the term "self-reported" as the respondents answered the items but an enumerator was reading the item out loud given that some respondents were illiterate. Grit is defined as "the combination of passion and perseverance for long-term goals" (Duckworth and Quinn, 2009). We used three instruments to measure Grit. First, we used a 10-items version of the self-reported grit scale in Pakistan and an 8-items version in Cambodia. Then, during the pilot in Pakistan, we also used two tasks to measure Grit. The Frustration task consists of a split-screen interface with the option to complete a difficult mirror-tracing task or play some games. It lasts 5 minutes and the outcome is the proportion of time spent doing tracing rather than playing the games. In the Alan and Ertac Grit task (Alan, Boneva, and Ertac, 2019), respondents are presented with a grid that contains different numbers where the goal is to find pairs of numbers that add up to 100. There is one easy game and one difficult game, the latter of which provides a higher reward. At the end of each round, feedback is given, and individuals choose for the next round which type of task they want to do. The outcome is the probability of choosing the difficult game in all rounds. The second self-reported scale measures the “Big Five”, a taxonomy of traits that encompasses five dimensions of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To measure these traits, we used the short 15-item Big Five Inventory (BFI-S), which consists of three items for each of the five personality traits. In Cambodia, we also measured growth mindset – the belief that we can get smarter through hard work and practice – using a 4-items scale. We also measured respondents’ locus of control – the degree to which people believe that they have power over the outcome of events in their lives, as opposed to external forces beyond their control – using a four-item scale. The Strength and Difficulty Questionnaire (SDQ) is a brief behavioral screening questionnaire. The Barratt Impulsiveness Scale (BIS) is a 30-item scale aimed at measuring impulsiveness. The GoNoGo task was used to measure impulse control: The participant is presented with a square on the screen for a very short period. If the square is of any color but black, the participant must touch the screen as quickly as possible. If the square is black (the “no go” stimulus), the respondent must inhibit their response. A total of 72 trials are completed (48 Go and 24 NoGo trials) and the main outcome we use is the average response time. The Balloon Analogue Risk Task (BART) measures risk-taking behavior by asking participants to maximize the amount of money they can win from the game. Respondents earned on average 322 PKR (approximately 3 USD at the time) from the game. On each trial, they were presented with a balloon, which they can pump. Each pump earned them money but increased the likelihood that the next pump would "pop" the balloon, in which case they lost the accrued money for that balloon. If they instead chose to stop pumping the balloon, they collected their accrued money and moved to the next trial. The main outcome is the average number of pumps on the balloons that did not explode.

Following Lajaaj and Macours (2021), we applied an acquiescence bias correction on the self-reported items to correct for the tendency of respondents to agree with a statement. The procedure is detailed in Appendix D. The left panel shows the reliability estimates for the measures corrected for Acquiescence bias (when applicable) and the right panel shows the estimates for the raw scales.

We show two types of reliability estimates. In columns (1), (2), (4), and (5), we display the Cronbach's alpha statistic which is a measure of internal consistency reliability. Internal consistency is the extent to which all the items in a scale reliably measure the same attributes or the interrelatedness of scale items. The Cronbach's alpha statistic is a ratio of variances and therefore lies between 0 and 1. A Cronbach’s above 0.7 indicates acceptable internal consistency. We show the Cronbach's alpha for the test (columns 1 and 4), and the re-test (columns 2 and 5). In columns (3) and (6), we show the test-retest correlations, which measure how correlated the responses of the same individuals to the same instrument are at two different points in time. 13% of respondents (429 respondents) were randomly re-surveyed within the same week in Cambodia. 50% of respondents (201 respondents) were randomly re-surveyed two weeks later in the Pakistan pilot.

In the last two columns, we show the number of respondents for each instrument. In Cambodia, 3,294 respondents answered the survey. Some respondents answered "Don't know" or refused to respond to some items, leading to observations ranging from 3,279 to 3,289. In Pakistan, the Grit scale was administered to all respondents who answered the direct version of the questionnaire in person, so 4,406 respondents. Among them, four respondents did not complete the questionnaire fully and others refused or said "Don't know" for some of the items. 4,395 answered all the items. The Big-Five scale was administered to all respondents who answered the direct version of the questionnaire, whether it was in person or over the phone. This leads to a sample of 4,485 respondents. Among them, four respondents did not complete the questionnaire fully and others refused or said don't know for some of the items. 4,475 respondents answered all the items.

Table 3. Factor Analysis of Self-Reported Measures

Skill	Label	Factor	Uniqueness
Grit scale			
Grit	New ideas and projects sometimes distract me from previous ones	1	0.7149
Grit	Setbacks don't discourage me. I don't give up easily	1	0.5782
Grit	I often set a goal but later choose to pursue a different one	1	0.526
Grit	I have difficulty maintaining my focus on project that take more than a few months to complete	1	0.769
Grit	I finish whatever I begin	1	0.6359
Grit	My interests change from year to year.	1	0.6477
Grit	I am diligent. I never give up.	1	0.4094
Grit	I have been obsessed with a certain idea or project for a short time but later lost interest	1	0.545
Grit	I have overcome setback to conquer an important challenge.	1	0.7781
Big-Five Scale			
Openness	is original, comes up with new ideas	2	0.4713
Openness	values artistic, aesthetic experiences	2	0.6354
Openness	has an active imagination	5	0.377
Conscientiousness	does a thorough job	1	0.4286
Conscientiousness	tends to be lazy	1	0.703
Conscientiousness	does things efficiently	1	0.599
Extroversion	is talkative	4	0.5105
Extroversion	is outgoing, sociable	4	0.2589
Extroversion	is reserved	4	0.2196
Agreeableness	is sometimes rude to others	3	0.5878
Agreeableness	has a forgiving nature	3	0.4117
Agreeableness	is considerate and kind to almost everyone	3	0.3341
Emotional Stab	worries a lot	5	0.69
Emotional Stab	gets nervous easily	5	0.2792
Emotional Stab	remains calm in tense situations	5	0.3463

Notes. This table shows the outcome from exploratory factor analysis performed on the items from the Grit and Big-Five self-reported scales in the Pakistan data collection. Factor analysis is used to analyze patterns of correlations between the items variables to infer their relationship to an unknown variable—here, an index of socio-emotional skills. We perform a principal factor estimation and retain 5 factors. For each item, we indicate the main factor the item loads on. Items measuring the same skill are expected to load on the same factor. We also show the uniqueness of the item, which is the percentage of variance for the item that is not explained by the common factors. A high uniqueness could indicate either measurement error, or that the item is measuring something different than the other items of the scale.

The first self-reported 10-item scale measures grit – the combination of passion and perseverance for long-term goals—and was developed by Duckworth and Quinn (2009). There was a mistake in the translation of the tool from English to Urdu so that only 9 items were implemented. The second self-reported scale measures the "Big Five", a taxonomy of traits that encompasses five dimensions of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To measure these traits, we used the short 15-item Big Five Inventory (BFI-S), which consists of three items for each of the five personality traits. All items are answered using a 5-point Likert scale format ranging from 1 ("Disagree strongly") to 5 ("Agree strongly"). Following Lajaaj and Macours (2021), we applied an acquiescence bias correction on the items to correct for the tendency of respondents to agree with a statement. The procedure for acquiescence bias correction is detailed in Appendix D.

Table 4. Skills on young adults in Pakistan (Part 1)

Subject	Who knows	What is the question	Fraction answering correctly	N
Panel A. Paper Test				
Mathematics	70% or more get	Tick box next number that matches the number of objects	79 %	4139
Mathematics	around 50% get	678+923	56 %	4139
Mathematics	30% or less get	7/3=__	5 %	4139
Urdu	70% or more get	Match picture: Book	78 %	4139
Urdu	around 50% get	Join letters and write word: m-a-l-k	48 %	4139
Urdu	30% or less get	Fill blank in the story by selecting the correct word	29 %	4139
English	70% or more get	Match picture: Banana	77 %	4139
English	around 50% get	Missing letter to match picture: Flag	53 %	4139
English	30% or less get	Use word in sentence: deep	16 %	4139
Panel B. Life skills assessment				
Numeracy	70% or more get	Read elec bill. How much money do you need to pay for the month of November?	80 %	4402
Numeracy	around 50% get	Multiplication per bracket - Elec bill. How much money do you have to pay?	49 %	4402
Numeracy	30% or less get	Multiplication: Kv* cost per Kv - Elec bill. How much money do you have to pay	16 %	4402
Literacy	70% or more get	Respondent was able to understand the greeting message in Urdu	73 %	4402
Literacy	around 50% get	Respondent was able to understand the greeting message in Roman Urdu	55 %	4402
Literacy	70% or more get	Respondent was able to save contact on mobile phone	69 %	4402

Notes. This table shows a sample of questions for each of the three subjects tested on paper: Mathematics, Urdu, and English (Panel A) as well as a sample of questions for functional numeracy and literacy skills tested using the life skills assessment (Panel B). For each category, we show a question for which 70% or more of the respondents got it right, a question for which around 50% of the respondents got it right, and a question for which 30% or less got it right. For the life skills literacy questions, we do not have questions where less than 30% of respondents got it right. For the paper test in Urdu, the question shown is the question with the lowest rate of correct answers in the test. For each question, we indicate the proportion of respondents who answered the question correctly and the number of respondents. The number of respondents for all life skills questions is 4,402 as all respondents who answered the survey in person and completed it answered them. For the cognitive test on paper, the sample was the same, except that at the beginning of the fieldwork, the survey firm forgot to implement this paper test. They went back to survey these respondents when we found out about the issue but 263 respondents could not be found at that time, therefore the number of observations is 4,139.

Table 4. Skills on young adults in Pakistan (Part 2)

Skill	Sample	Mean	Standard Deviation	Min	Max	N	Source
Panel A. LEAPS Sample							
Grit	LEAPS	3.45	0.71	1	5	4395	LEAPS Data
Openness to Experience	LEAPS	10.26	3.19	3	15	4475	LEAPS Data
Conscientiousness	LEAPS	12.52	2.17	3	15	4475	LEAPS Data
Extroversion	LEAPS	9.92	3.26	3	15	4475	LEAPS Data
Agreeableness	LEAPS	13	2.07	3	15	4475	LEAPS Data
Emotional stability	LEAPS	6.83	3.24	3	15	4475	LEAPS Data
Panel B. Other samples							
Grit	Adults aged 25 and older (US)	3.65	0.73	1	5	1545	Duckworth et al. (2007)
Grit	West point cadets 2010 (US)	3.75	0.54	1	5	1308	Duckworth et al. (2007)
Grit	Adults between 15-64 years old	2.72	0.6	1	4	3843	STEP Data Kenya
Grit	Adults between 15-64 years old	2.98	0.61	1	4	3978	STEP Data Macedonia
Openness to Experience	Adults between 15-64 years old	3	0.56	1	4	3844	STEP Data Kenya
Openness to Experience	Adults between 15-64 years old	3.28	0.55	1	4	3979	STEP Data Macedonia
Conscientiousness	Adults between 15-64 years old	3.22	0.52	1.333333373	4	3844	STEP Data Kenya
Conscientiousness	Adults between 15-64 years old	3.05	0.5	1	4	3979	STEP Data Macedonia
Extraversion	Adults between 15-64 years old	2.85	0.59	1	4	3845	STEP Data Kenya
Extraversion	Adults between 15-64 years old	3.02	0.61	1	4	3979	STEP Data Macedonia
Agreeableness	Adults between 15-64 years old	2.86	0.57	1	4	3843	STEP Data Kenya
Agreeableness	Adults between 15-64 years old	3.28	0.59	1	4	3978	STEP Data Macedonia
Emotional stability	Adults between 15-64 years old	2.69	0.5	1	4	3843	STEP Data Kenya
Emotional stability	Adults between 15-64 years old	2.09	0.66	1	4	3979	STEP Data Macedonia

Notes. This table shows for each socio-emotional skill the mean, standard deviation, the minimum, and the maximum. Panel A shows the descriptive statistics for the LEAPS Sample while Panel B shows results from other samples reported in the literature. The results using STEP data were computed by the authors using the STEP surveys conducted by the World Bank in 2013. See Pierre et al. (2014) for details about STEP surveys. In the LEAPS sample, the grit scale was administered to all respondents who answered the direct version of the questionnaire in person, so 4,406 respondents. Among them, four respondents did not complete the questionnaire fully and others refused or said don't know for some of the items. 4,395 answered all nine items. The Big-Five scale was administered to all respondents who answered the direct version of the questionnaire, whether it was in person or over the phone. This leads to a sample of 4,485 respondents. Among them, four respondents did not complete the questionnaire fully and others refused or said don't know for some of the items. 4,475 respondents answered all 15 items.

Table 5. Relationship between schooling and skills formation

	(1) Cognitive Skills	(2) Cognitive Skills	(3) Life Skills	(4) Life Skills	(5) Socio-Emotional Skills	(6) Socio-Emotional Skills
Years of schooling	0.169***	0.168***	0.165***	0.165***	0.033***	0.033***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Respondent Age	-0.009***	-0.009***	0.001	0.001	0.007**	0.008*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Sex of the respondent = 1, Female	0.124***	0.129***	-0.156***	-0.156***	-0.622***	-0.616***
	(0.022)	(0.022)	(0.020)	(0.020)	(0.033)	(0.035)
Mother highest grade	0.011**	0.012**	0.001	-0.001	-0.001	-0.001
	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.006)
Father highest grade	0.010***	0.010***	-0.001	-0.000	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
HH SES in 2003	0.015***	0.011*	-0.001	-0.004	0.007	0.010
	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)	(0.009)
Constant	-1.330***	-1.367***	-1.323***	-1.379***	-0.214*	-0.056
	(0.076)	(0.074)	(0.087)	(0.083)	(0.113)	(0.107)
Observations	4,399	4,399	4,402	4,402	4,395	4,395
Adjusted R-squared	0.659	0.665	0.668	0.677	0.153	0.171
Village FE	No	Yes	No	Yes	No	Yes
District FE	Yes	No	Yes	No	Yes	No
N_Clusters	108	108	108	108	108	108

* p<.10 ** p<.05 *** p<.01

Notes. The dependent variable in columns (1) and (2) is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (leaving out items that less than 50 respondents answered, and those that less than 5% or more than 95% of respondents got the correct answer). The dependent variable in columns (3) and (4) is a life skills index computed using principal factor analysis on 17 real-life literacy and numeracy questions. The dependent variable in columns (5) and (6) is a socio-emotional skills index computed using principal factor analysis on the Big Five items, Grit items, BART, and GoNoGo scores. The sample for each regression is the sample of respondents who answered the direct version of the questionnaire, in person. The sample is 4,406 respondents. There are four incomplete questionnaires, leading to a sample of 4,402 respondents. Some respondents refused to answer or responded "Don't know" to some of the items, leading to a sample size of 4,399 for the cognitive skills index and 4,395 for the socio-emotional index. Regressions in columns (2), (4), and (6), control for village fixed effects, while regressions in columns (1), (3), and (5) control for district fixed effects. Standard errors are in parentheses (clustered at the village level). The R-squared shown is the adjusted R-squared.

Table 6. Schooling, Skills and Labor Force Participation

	Men				Women			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Working	Working	Working	Working	Working	Working	Working	Working
Years of Schooling	-0.022***	-0.024***	-0.017***	-0.0076**	0.012***	0.012***	0.0097***	0.0099***
	(0.0018)	(0.0020)	(0.0031)	(0.0032)	(0.0012)	(0.0013)	(0.0021)	(0.0022)
Cognitive Skills			-0.046***	-0.016			0.016	0.013
			(0.013)	(0.014)			(0.011)	(0.012)
SEMS			0.054***	0.056***			0.0016	0.00068
			(0.0097)	(0.0098)			(0.0062)	(0.0059)
Constant	0.17***	0.20***	0.18***	1.05***	-0.058***	-0.062***	-0.048***	-0.070***
	(0.024)	(0.028)	(0.033)	(0.032)	(0.0062)	(0.0067)	(0.012)	(0.022)
Observations	2595	2217	2217	1978	2360	2174	2174	1925
Adjusted R-squared	0.14	0.15	0.16	0.060	0.060	0.060	0.060	0.060
Mean Dependent	0.79	0.76	0.76	0.83	0.06	0.07	0.07	0.06
Sample	All	Skills measures	Skills measures	Skills measures, no students	All	Skills measures	Skills measures	Skills measures, no students
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, and labor force participation for men and women in Pakistan. We report mean regressions estimates in all the columns. The dependent variable "Working" is an indicator variable equal to 1 if the respondent is currently employed. The cognitive skills index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills (SEMS) index is computed using principal factor analysis on the Big-Five items, Grit items, BART, and GoNoGo scores. The sample for column (1) is all men in the sample, that is 2,596 men. Among them, one answered "Don't know" to the years of schooling variable, leading to a sample of 2,595. The sample for column (5) is all women in the sample, that is 2,360 women. For columns (2) and (3)/(6) and (7), the sample is men/women who answered the direct version of the questionnaire (they have skills measures). For columns (4) and (8), the samples further exclude any individual who is currently enrolled. Standard errors are in parentheses (clustered at the village level). All the regressions include age and district fixed effects. The R-squared shown is the adjusted R-squared.

Table 7. Schooling, Skills, Migration, and Earnings for Men in Pakistan

	Median Regressions				Mean Regressions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Monthly income	Monthly income	Monthly income	Monthly income	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)
Years of Schooling (a1)	4.12***	3.37***	1.10*	0.063	5.22***	4.55***	2.88**	0.69
	(0.50)	(0.46)	(0.60)	(0.84)	(0.93)	(0.86)	(1.15)	(1.06)
Cognitive Skills (a2)			8.31***	0.26			6.28	2.79
			(2.77)	(3.59)			(4.74)	(5.00)
SEMS (a3)			16.5***	14.3***			14.6***	13.0***
			(1.86)	(2.71)			(3.28)	(3.14)
Out Village (a4)				32.9**				8.35
				(13.2)				(22.8)
Interaction YrsSchooling and Out Village (b1)				4.78***				9.36***
				(1.44)				(2.51)
Interaction Cog and Out Village (b2)				13.6**				-4.71
				(6.35)				(11.5)
Interaction SEMS and Out Village (b3)				-7.37				-10.8
				(5.17)				(7.68)
Constant	15.0	14.4	42.1	36.3	27.1***	33.7***	56.3***	46.1***
	(91.8)	(64.3)	(89.7)	(46.7)	(8.44)	(7.99)	(9.71)	(10.0)
Observations	2340	1978	1978	1978	2340	1978	1978	1978
R-squared	0.034	0.030	0.043	0.092	0.060	0.040	0.050	0.13
Median/Mean Dependent	120.19	115.38	115.38	115.38	154.70	134.71	134.71	134.71
Sample	All	Has skills measures	Has skills measures	Has skills measures	All	Has skills measures	Has skills measures	Has skills measures
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
a1+b1=0				4.84***				10.05***
				(1.24)				(2.49)
a2+b2=0				13.89***				-1.92
				(5.33)				(10.22)
a3+b3=0				6.89				2.27
				(4.45)				(7.19)

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, migration, and earnings for men in Pakistan. The dependent variable "Monthly income" in columns (1) to (4) is the raw monthly income while the dependent variable "Monthly income (TC)" in columns (5) to (8) is top coded at 100,000 PKR per month (961.5 USD). The cognitive skills index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills index (SEMS) is computed using principal factor analysis on the Big-Five items, Grit items, BART, and GoNoGo scores. The "Out of village" variable is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household (their natal village most of the time). The sample for columns (1), and (5) are all men in the sample who are not currently enrolled, that is 2,351 men. Among them, one answered "Don't know" to the years of schooling variable, and 10 answered "Don't know" to the wage variable, leading to a sample of 2,340. For the rest of the columns, the sample is only those who answered the direct version of the questionnaire, in-person (they have skills measures): 1,978 men. All the regressions include age and district fixed effects. Robust standard errors are in parentheses. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.

Table 8. Schooling, Skills and Earnings for Women in Pakistan

	All			Working Women					
	Mean Regressions			Median Regressions			Mean Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)	Monthly income	Monthly income	Monthly income	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)
Years of Schooling	1.73***	1.91***	1.35***	5.92***	4.81***	5.25**	8.47***	8.51***	8.14*
	(0.28)	(0.31)	(0.43)	(1.47)	(1.60)	(2.31)	(1.96)	(1.95)	(4.12)
Cognitive Skills			3.13*			-3.22			1.74
			(1.83)			(9.75)			(18.2)
SEMS			0.80			-7.45			3.51
			(0.91)			(7.99)			(7.99)
Constant	-11.9***	-13.1***	-8.22*	1.48	19.2	9.56	-28.6	-27.6	-23.2
	(2.37)	(2.49)	(4.28)	(37.4)	(43.6)	(48.4)	(31.0)	(31.8)	(47.3)
Observations	2104	1925	1925	112	111	111	112	111	111
R-squared	0.060	0.060	0.070	0.22	0.22	0.22	0.26	0.26	0.25
Median/Mean Dependent	7.22	7.84	7.84	62.50	57.69	57.69	107.33	107.43	107.43
Sample	All	Has skills measures	Has skills measures	All	Has skills measures	Has skills measures	All	Has skills measures	Has skills measures
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, and earnings for women in Pakistan. The dependent variable "Monthly income" in columns (4) to (6) is the raw monthly income while the dependent variable "Monthly income (TC)" in columns (1) to (3) and (7) to (8) is top coded at 100,000 PKR per month (961.5 USD). The cognitive skills index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills index is computed using principal factor analysis on the Big-Five items, Grit items, BART, and GoNoGo scores. The sample for column (1) are all women in the sample who are not currently enrolled, that is 2,104 women. For columns (2) and (3), the sample is only those who answered the direct version of the questionnaire, in-person (they have skills measures): 1,925 women. The sample for columns (4) and (7) is all women in the sample who are working, that is 112 women. For columns (5), (6), (8), and (9), the sample is only those who answered the direct version of the questionnaire, in-person: 111 women. All the regressions include age and district fixed effects. Robust standard errors are in parentheses. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.

APPENDICES

Appendix A. Tracking and Survey Instruments

The sample for this long-term tracking exercise consisted of all individuals between 5 and 15 years old who were part of the 1,807 households sampled for the first LEAPS round, in 2003. This leads to a sample of 5,865 respondents. We had last attempted to survey these respondents in 2011, before we started a tagging and tracking exercise in 2016. Importantly, the 2011 survey was conducted at the household level and we collected information about individual respondents through household rosters. In 2016, we were interested in tracking individuals.

The first step in the tagging and tracking process was to go back to the latest households' addresses we had for these respondents. We were not able to locate the households of 285 respondents (Table A1). 79% of them were part of households that had already attrited by 2011.

For the households we could find, we implemented the following protocol. If the target individual had migrated, we would collect their address and contact information. We attempted to survey all the respondents in person, either by visiting them at their new address for those who had migrated within the country or waiting for them to visit their relatives for those who had migrated abroad. For any individual whom we were able to meet in person, we would administer five different instruments:

- Questionnaire
- Cognitive skills assessment on paper
- (Adaptive) Cognitive skills assessment on tablet
- Socio-emotional skills assessment on paper
- Socio-emotional skills assessment on tablet

For individuals we were not able to meet in person, we administered a shorter survey over the phone. The survey was similar to the questionnaire administered in person but shorter. We could not conduct any cognitive, or socio-emotional assessment over the phone[15]. We are therefore missing skills information for all the individuals surveyed over the phone (Table A1).

Finally, there were some respondents we were not able to meet in person or survey over the phone. These respondents fell mostly into two categories: women who got married and were living with their family-in-law and our team was not authorized to visit (30%), and men working in neighboring Arab countries for whom we did not have contact information (30%). Whenever possible, we collected information about these respondents from another person in their original household (parents or siblings, in 80% of the cases). We called this survey mode the "indirect" version of the survey as someone else was *indirectly* giving information about the respondent of interest. This questionnaire was short and designed to collect the main variables of interest that a

[15] Crawfurd et al. (2021) report differential item functioning across in-person and phone surveys for a cognitive assessment in Sierra Leone. Moreover, the adaptive test and part of the socio-emotional skills assessment had to be implemented on tablets.

close relative would know about such as educational attainment, employment status, marital status, and number of children. Table A1 summarizes the number of respondents for each survey type and the corresponding numbers in terms of the type of information we have.

Appendix B. Attrition Weights

To account for attrition, we apply inverse-probability weights (IPW) following Wooldridge (2010). Importantly, we predict the probability that the respondent has skills measures rather than the probability of having any data (see Appendix A and Table A1 for more details). We follow this approach as our main results relate skills to labor market outcomes so that individuals with missing skills measures are effectively treated as attritors in these regressions.

We implement two alternative approaches to modeling this probability. We start with a basic model that uses only correlates of attrition from the very first round of data collection of the LEAPS project, in 2003. We then add variables indicating if the respondent had already attrited in 2011 and 2016 (when we started the tracking process). We describe these two approaches in turn.

Model 1. Probability of having full data (including skills measures) as a function of variables collected in 2003

We start with a list of 35 variables from 2003 and select variables that are predictive of having skills measures using Lasso. This procedure leads us to keep 31 variables including the respondent's sex, age, housing characteristics (toilet on-premises, electricity access, etc.), parental education and occupation, whether the household head could read and write, and the language of the interview. We then use a probit model to predict the probability that we have full data for the respondent.

Model 2. Probability of having full data (including skills measures) as a function of variables collected in 2003 and indicators of attrition in 2011 and 2016

We use the same specification as in Model 1, to which we add two variables: a dummy indicating that the individual was already an attritor in 2011 and one indicating that the individual was an attritor in 2016, at the early stage of the tagging and tracking process.

Appendix C. Measurement of Cognitive Skills

C.1. Adaptive Cognitive Test

We worked with an organization to design an adaptive test that was administered on tablets to be able to capture the skills of our diverse pool of respondents. The organization we partnered with developed 324 items ranging from early primary level to college level. The test classified respondents into 6 levels corresponding to different grades. The mapping between level and grades is as follows:

- Level 1: Nursery, Grades 1 to 3 (early primary)

- Level 2: Grades 4 and 5 (late primary)
- Level 3: Grades 6 to 8 (middle school)
- Level 4: Grades 9 and 10 (high school)
- Level 5: Grades 11 and 12 (intermediate)
- Level 6: College

The items of the tests were designed to capture the following learning domains: (1) mastery over concepts and definitions (ex: "what is a pronoun?"), (2) application mastery (ex: "add 2+2") and (3) evaluation mastery (ex: "two boys meet two girls, one boy leaves; how many children are left?"). As we expected many of the respondents to have been out of school for a long time, items were designed to test general mastery as opposed to specific terms or formulae. All the items were multiple-choice questions with four possible answer choices and one correct answer.

The logic of the test was as follows:

- Everyone started at the same level – Level 2 for Urdu and Mathematics and Level 1 for English – and answered a batch of 6 questions.
- If the respondent got 5-6 questions right, they moved to the next higher level (or, if at Level 6, remains at Level 6).
- If the respondent got 3-4 questions right, they stayed at the same level
- If the respondent got 0-2 questions right, they moved to the next lower level (or, if at Level 1, remains at Level 1).

Then, the placement logic of the test was:

- The first time that a respondent completed three batches of 6 questions at any Level:
    - If the Level was Level 1, and the last score was 0-2 questions right, the respondent was placed at Level 1
    - If the Level was Level 1-6, and the last score was 3-4 questions right, the respondent was placed at that level
    - If the Level is Level 6, and the last score was 5-6 questions right, the respondent was placed at Level 6
    - If the Level is Level 1-5, and the last score was 5-6, the respondents was moved to the next higher level and the test continued
    - If the Level is Level 2-6, and the last score is 0-2, the respondents was moved to the next lower level and the test continued
- The first time that a respondent completed three batches of questions at any Level and any time after that at the next higher-Level scored 0-2 questions right, the respondents was placed at the Level where the respondent had completed three batches of questions.

To complete the placement, the minimum batches of questions were 3 and the minimum number of questions was 18. To complete the placement, the maximum batches of questions was 17 and the maximum number of questions was 102. The test took 15 minutes to complete on average. Examples of progress and placement logic are provided below.

Example 1.

Batch	Level	# of items correct
1	2	3
2	2	4
3	2	4

The respondent answered 18 questions and was placed at Level 2 (the respondent completed three batches of questions at Level 2, and the last score was 3-4 questions right).

Example 2.

Batch	Level	# of items correct
1	2	5
2	3	2
3	2	4
4	2	5
5	3	2

The respondent answered 30 questions and was placed at Level 2 (the respondent completed three batches of questions at Level 2, and the first time after that they got to Level 3, they only got 2 questions right).

Example 3.

Batch	Level	# of items correct
1	2	2
2	1	6

3	2	0
4	1	4
5	1	4

The respondent answered 30 questions and was placed at Level 1 (the respondent completed three batches of questions at Level 1, and the last score was 3-4 questions right).

C.2. Item Response Theory

We aggregate items from the adaptive test on tablet and the LEAPS paper test for each subject into Urdu, Mathematics and English scores using item response theory. Item Response Theory is a set of mathematical models that describe the relationship between an individual's latent "trait", θ, and their manifestations (performance on a test). They establish a link between that latent trait, the properties of the items in the scale, and how individuals respond to these items.
In the two-parameter IRT model, the likelihood of answering a question correctly is determined by the ability of the respondent, θ, and two items parameters – difficulty (labeled a) and discrimination (labeled b). In this model, the probability that a respondent answers a given question *j* correctly is modeled as:

$$P_j(\theta) = \frac{1}{1 + \exp(-a_j(\theta_j - b_j)}$$

The difficulty parameter, a, represents the ability level at which 50% of respondents get the item right. For an item with a difficulty parameter of 1, a respondent that has an ability level of 1 SD over the mean has a 50% chance of getting the item right. The discrimination parameter, a, captures how quickly the likelihood of success changes with respect to ability.
The Item Characteristic Curve (ICC) depicts the likelihood of a correct answer, $P_j(\theta)$, as a function of θ. The higher the individual's ability, the higher is the probability of a correct response. We plot the Item Characteristic Curve (solid line) and the actual pattern responses against 40 quantiles of θ for the paper test items for each of the three subjects in Figures A3 to A5. We observe a tight fit between the predicted responses based on the ICC and the actual responses in the data for all the items of the LEAPS test. Then, we produce the same plots for 25 randomly selected items from the adaptive test for each of the three subjects, in Figures A6 to A8. For the adaptive test, the fit between the predicted and actual responses varies depending on the subject and specific items.

C.3. Functional Literacy and Numeracy Assessment

We also designed an assessment to capture proficiency in everyday arithmetic and literacy skills. The assessment was divided into three sections.

The first part asked the respondents to read an electricity bill and answer the following questions:

- How much money do you need to pay for the month of November?
- How much money will you pay if you pay the amount after the due date?
- Imagine the following: The meter reading on your electricity bill for the month of March is 2500 KV units. Each kV unit costs Rs. 2. Please note that you have to pay a late fee of Rs.500 if you miss the due date. How much money do you have to pay for the month of March to cover your electricity bill if you pay before the due date? After the due date?
- Now, imagine a scenario where the electricity bill is charged by the kV bracket breakdown. The meter reading is 2500 kV units. The first 500 KV will be charged at Rs. 10 and any KV units more than 500 will be charged at Rs. 20, as shown in the table below:

kV	Cost per unit (in Rs.)
0-500	10
501 and above	20

Please note that you have to pay a late fee of Rs.500 if you miss the due date. How much money do you have to pay for the month of March to cover your electricity bill if you pay before the due date? After the due date?

The second part of the assessment asked respondents to read text messages written in Urdu, and in Roman Urdu (Urdu but using roman language script). We asked them to read one greeting message, one conversation, one advertisement text, two messages from the school, and four emergency text messages in Urdu and Roman Urdu. The messages in Roman Urdu were:

Greetings messages

Peace be upon you. How are you and how is everyone at home?
[Urdu: Asalam O alikum kya haal hai app ka aur ghar mein sub ka kya haal hai?]

Conversation

Person 1: How are you? [Urdu: Kya haal hai?]
Person 2: I am fine. [Urdu: Mein theek hoon]
Person 1: Anything new? [Urdu: Koi nayi tazi?]
Person 2: Exams are going on at school. [Urdu: Papers chal rahay hain school mein.]
Person 1: To obtain something and to be successful, you will often face problems in life, but success comes to those who work hard and do not get scared. Work hard, well done. [Urdu: Kuch hasil kar ney aur maqaam bananay kay liye zindagi mein mushkilaat aati hain aur kamyabi aun ko milti hai jo dat jatay hain aur dartnay nahi. Mehnat karo Shabaash.]
Person 2: Thank you. [Urdu: Shukariya.]

Advertisement

Great news! Before the 30th of this month, recharge your balance and get 1000 minutes and 2000 texts absolutely free. To get more information, dial 1212.

[Urdu: Shandar Khabar! Iss mahinay ki 30 tareekh se phelay apnay balance ko recharge ki jiye aur payi 1000 minutes aur 2000 sms bilkul muft. Mazeed malomaat kay liye 1212 dial karien.]

School Text messages

1. Your child has not come to school today. Is everything ok? [Urdu: Apka bacha aj school nahi aya. Sub khariat hai?]
2. Your child has taken a holiday again today and today was his exam in Urdu. Is everything ok? [Urdu: Apkay bachay ne aj phir chuti kar li hai aur aj aus ka urdu ka imtehaan thaa. Kya sub khairiat hai?]

Emergency text messages

1. I am going to be home late today. Please don't worry. [Urdu: Mein aaj ghar per dair se aaon ga. Pareshaan mat hona.]
2. Friend, I am stopped outside the village and my motorcycle has broken down. Can you pick me up? [Urdu: Yaar mein gaon se bahir ruka huwa hoon aur meri motorcycle kharab hogayi hai. Kya mujhay lenay aasaktay ho?]
3. Tomorrow there will be no water in the houses of this village from morning till evening, so please make your arrangements before hand. [Urdu: Kal gaon kay gharon mein subah se sham tak paani nahi aye ga tou isliye phelay se intezaam kar lain.]
4. Brother, the flour has finished so on your way back can you pick up some flour because we need to make rotis for dinner. [Urdu: Bhai ata khatam hogaya hai wapisi per aata letay ana khanay kay liye roti banani hai.]

We then asked similar messages in the Urdu script.

Finally, the last part mimicked a market transaction. We asked:

Imagine you go to a shop have Rs.300 with you. You get the following items from the shopkeeper and give him 3 notes Rs.100 rupees each.

1 kg rice	Rs. 30
1 kg potatoes	Rs. 20
1 kg sugar	Rs. 20
Surf	Rs. 25
Cooking oil	Rs. 100

How much money should the shopkeeper return to you if you purchase all the items at the same shop?

Appendix D. Measurement of Socio-Emotional Skills

The design of the socio-emotional skills assessment was the result of an iterative process, that started with data collection in Cambodia for a related project in 2017. The data from this project suffers from some shortcomings that we detail in section D.3. below. Nevertheless, the data are useful for two reasons. First, we used what we learned when designing and piloting our instruments in Pakistan. In particular, this is what motivated us to conduct a pilot that (1) was large-scale, (2) took place far ahead of the data collection, (3) included a test-retest component. Second, we can check whether the general pattern of results is consistent across Cambodia and Pakistan.
We start by describing the methods we use to assess the reliability and validity of our instruments and then our process of selection for the instruments we kept in the full data collection.

D.1. Acquiescence bias correction

We assessed the quality of our measures by evaluating their reliability and validity. Before assessing these two aspects, we followed Laajaj and Macours (2021) and corrected the self-reported items for acquiescence bias (the tendency to agree rather than disagree to questions). We apply the following procedure[16]:

1. Compute the average score on reverse-coded items, and the average score on non-reversed-coded items.
2. Take the average of the two averages obtained in the first step.
3. Subtract the scale mid-point (for instance 3, if the possible answers are on a 5-point Likert scale format ranging from 1 ("Disagree strongly") to 5 ("Agree strongly")). This step gives an estimate of acquiescence bias.
4. Subtract the acquiescence bias score obtained in the third step to every non-reverse-coded item and add it to every reverse-coded item.

D.2. Measures of reliability and validity

D.2.1. Reliability

A measure is said to have a high reliability if it produces similar results under consistent conditions. We provide two types of reliability estimates: internal consistency and test-retest reliability.

[16] This procedure is equivalent to the procedure described by Laajaj and Macours (2021).

Internal consistency

Internal consistency is the extent to which all the items in a scale reliably measure the same attributes, or the interrelatedness of scale items. To assess internal consistency, we compute the Cronbach's alpha statistic (Cronbach, 1951). Cronbach alpha is computed as

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

where K is the number of items in the scale, σ_X^2 is the variance of the observed total test score, and $\sigma_{Y_i}^2$ is the variance of responses to item i for the current sample of persons. It measures how correlated the items of a scale are and is also a direct function of the number of items in the scale.
The statistic is a ratio of variances and therefore lies between 0 and 1. A Cronbach's above 0.7 indicates acceptable internal consistency but a Cronbach's alpha above 0.9 may indicate item redundancy (Oxford Mind & Behaviour team[17]).
The Cronbach's alpha provides an assessment of the reliability of the scales as well as its construct validity. If there is a lot of classical measurement error (low reliability), the items of a scale will be poorly correlated, and the Cronbach's alpha will be low. However, even with no measurement error, if the items do not measure the same underlying construct, their correlation will also be low, leading to a small Cronbach's alpha. Finally, non-classical measurement error can lead to an artificially high Cronbach's alpha. If all the items suffer from systematic response bias, they may be highly correlated resulting in a high Cronbach's alpha. Although we corrected for acquiescence bias prior to estimating the Cronbach's alpha for the different scales, this remains a potential limitation.

Test-retest reliability

The test-retest reliability measures how correlated the responses of the same individuals to the same instrument are at two different points in time. The idea is that if the instrument measures the true ability we intend to measure, this ability should not vary over a short period of time (usually between two weeks to one month), and the two measures should be highly correlated. On the other hand, if we capture mostly (classical) measurement error, the test-retest correlation should be low.
The test-retest correlation is a measure of reliability. Under classical measurement error, the observed value of the variable X is equal to the true value of X plus a purely random component. We can write the measured value of X as the sum of the true value X^*plus a measurement error ε:

$$X = X^* + \varepsilon$$

Where $E(\varepsilon) = 0$ and $cov(\varepsilon, X) = 0$.
Then, the variance of X is equal to:

$$\sigma_X = \sigma_{X^*} + \sigma_\varepsilon$$

The test-retest correlation is a measure of reliability, defined as the share of the variance of X driven by the true variance of the variable, as opposed to measurement error:

[17] https://mbrg.bsg.ox.ac.uk/method/measuring-non-cognitive-skills-psychometric-validation-scales

$$Relability = \frac{\sigma_{X^*}}{\sigma_X} = \frac{\sigma_X - \sigma_\varepsilon}{\sigma_X} = 1 - \frac{\sigma_\varepsilon}{\sigma_X}$$

A test-retest correlation above 0.7 is generally considered high.
If measurement error is non-classical, errors could be correlated over time. In this case, we would overestimate the reliability of a measure. This would be the case for instance if the answer patterns suffer from systematic acquiescence or social desirability bias.

D.2.2. Validity

An instrument is said to be valid in a specific context if it measures what it is supposed to measure. We investigate three aspects of validity: face validity, predictive validity, and content validity.

Face validity

Face validity ensures that the questions asked are perceived as measuring the concepts the instrument intends to measure. In other words, when we ask respondents a question aimed at assessing their emotional stability, they should subjectively perceive it as such. We assessed respondents' understanding and perception of the questions through debriefing sessions during the pilot.

Predictive validity

Predictive validity ensures that the measures are correlated with the variables we would expect according to theory or existing empirical evidence. For instance, in theory, we would expect grit to be correlated with educational attainment and fewer career changes (Duckworth, 2007). Similarly, locus of control should be positively correlated with desirable labor market outcomes. People with stronger internal locus of control perform more complex activities and have better job performance (Judge and Bono, 2001).

Content validity

Content validity refers to the extent to which the items of a test represent all facets of a given construct. For instance, the content of an instrument aiming at measuring the Big 5 personality traits is said to be valid if it captures the five dimensions of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To assess content validity, we rely on exploratory factor analysis. Exploratory analysis is used to analyze patterns of correlations between the items variables to infer their relationship to an unknown variable—here, an index of socio-emotional skills. To determine the number of factors, we use two criteria commonly used in the literature: the Kaiser criterion (1958), where one keeps only the factors with eigenvalues[18] higher than 1 and the scree plot criterion (Cattell, 1996) where you only keep the factors up until the line (which plots the eigenvalues) becomes flatter. We perform exploratory factor analysis on all the items from the self-reported scales, corrected for acquiescence bias. In Table 3, we show the results from the factor analysis in the full data collection. For each item, we indicate the main factor the item loads on. Items measuring the same skill/construct are expected to load on the same factor. We also show the uniqueness of the item, which is the percentage of variance for the item that is not explained by the common factors. A high uniqueness could indicate either measurement error, or that the item is measuring something different than the other items of the scale.

D.3. Selection process

Here, we describe the iterative process that led us to select the two self-reported scales and two tablet-based tasks included in our data collection.

D.3.1. Cambodia project

In Cambodia, the study took place in the context of a randomized control trial aiming at measuring the impact of primary-school scholarships on schooling attainment and labor market outcomes. In 2008, the Cambodian government offered scholarships to students as they were beginning the fourth grade of primary school. The new study tracked and attempted to survey 3,825 children who were in Grade 3 or 4 in 2008 from three poor and remote Cambodian provinces that constituted the original experiment population. Data collection for the baseline in Cambodia took place from December 2008 to January 2009, and data collection for the latest round of follow-up took place from December 2016 to May 2017. These students had been re-surveyed once before, in 2011. The team was able to survey 3,294 respondents or 86% of the target sample. Attrition is therefore comparable to that of Pakistan where we collected information on 84.5% of the sample. A shortcoming of the data collection in Cambodia is that we did not systematically collect the current location of respondents. We are missing this information for 18% of the surveyed sample.

The survey content was very similar to that of Pakistan. In particular, the team implemented both a cognitive and socio-emotional skills assessment and collected data on labor market outcomes and family formation.

The cognitive assessment was a computer-adaptive math test, in which respondents answered ten questions from a larger pool of 23 items. These items are aggregated in a cognitive skills index with a two-parameter Item Response Theory model.

To measure socio-emotional skills, the team used the same Grit and "Big Five" self-reported scales as in Pakistan. We also measured growth mindset – the belief that we can get smarter through hard work and practice – using a 4-items scale. We also measured respondents' locus of control – the degree to which people believe that they have power over the outcome of events in their lives, as

[18] The higher the eigenvalue the higher the percentage of the total variation in the variables that is explained by the factor.

opposed to external forces beyond their control – using a four-item scale. Finally, we administered the Strength and Difficulty Questionnaire (SDQ), which is a brief behavioral screening questionnaire. All items are answered using a 5-point Likert scale format ranging from 1 ("Disagree strongly") to 5 ("Agree strongly"). Cronbach's alphas for these different measures, before and after acquiescence bias corrections, are displayed in Table 2. Only the SDQ passes the 0.7 threshold. The Cronbach's alphas for the other measures range from 0.04 to 0.58. Moreover, we randomly re-surveyed 13% of randomly selected respondents within the same week, using a sub-set of items. We can therefore compute the test-retest correlations for the Big-Five constructs. These are low, averaging at 0.3 (Table 2), reflecting a large amount of measurement error. Moreover, when conducting exploratory factor analysis, the skills factor structure is not reproduced. Only three factors are retained, and one single factor encompasses a wide range of items aimed at measuring different concepts.

D.3.2. Pakistan pilot

Given the results from Laajaj and Macours (2021) and those from the data collection in Cambodia, we decided to conduct a large pilot before our data collection in Pakistan. We conducted the pilot in the district of Okara, between February and March 2018. Interested participants were invited to a central location, resulting in a total sample size of 403 individuals. Then, two weeks after the completion of the first round, we tracked and re-surveyed 201 respondents, randomly selected from this group. The survey was again completed with this group. This second survey allows us to compute the test-retest reliability for our instruments.

On top of instruments included in the final data collection and described in Section II.2. of the paper, we included two additional self-reported scales and tasks on tablets. The first self-reported scale was the same locus of control scale as in Cambodia. The second was the Barratt Impulsiveness Scale (BIS), which is a 30-item scale aimed at measuring impulsiveness. Then, we also included two tasks aimed at measuring grit. The first was the Frustration task. It consists of a split-screen interface with the option to complete a difficult mirror-tracing task or play some games. It lasts 5 minutes and the outcome is the proportion of time spent doing tracing rather than playing the games. Then, we also used to Alan and Ertac Grit task (Alan, Boneva and Ertac, 2019). In this task, respondents are presented with a grid that contains different numbers where the goal is to find pairs of numbers that add up to 100. There is one easy game and one difficult game, the latter of which provides a higher reward. At the end of each round, feedback is given, and individuals choose for the next round which type of task they want to do. The outcome is the probability of choosing the difficult game in all rounds. The Barratt Impulsiveness Scale showed satisfactory reliability measures: in particular, the Cronbach's alpha was 0.71 and the test-retest correlation was 0.4 (which is low but among the highest in the Pakistan pilot). However, it took a long time to administer this scale so that we preferred to keep the Grit and Big Five scales that were also used in Cambodia and found more reliable during the Pakistan pilot. We dropped the Frustration task because respondents had difficulty understanding it, and the size of the screen was not well suited for the task. Finally, as a result of the important heterogeneity in education levels in our sample, the Alan and Ertac Grit task

(Alan, Boneva, and Ertac, 2019) was found to measure a combination of cognitive ability and grit rather than grit itself in this setting. We decided not to keep it.

Appendix E. Wage measurements and distributions

In this section, we describe how we collected information on respondents' wages in the different versions of the survey (i.e. in-person, phone, and indirect), and how we aggregated these responses in different ways.

We started by collecting respondents' employment status for their two main activities. For each activity, respondents could fall into five employment categories: daily wages, salaried, self-employed, family business, and agriculture. For salaried respondents, we simply collected their monthly wage. For respondents with daily wages, who were self-employed, or working in a family business, we collected monthly earnings for the past two months to account for income volatility and average the two measures. Agricultural income was collected in multiple ways. For respondents who answered the survey in-person, we collected detailed information on outputs quantities, prices and input costs. Using these data, we calculated a "computed agricultural income". For all respondents, we checked if they had been engaged in agriculture for more than four weeks during the last year with the goal of selling their production, and if so, how much money they earned from it. We divide their answer by 12 and call this variable the monthly "estimated agriculture income". Moreover, to be sure to accurately capture the earnings of women who may be informally working in their village, we also asked if they were engaged in tutoring, sewing, or any other activities against payment during the last month.

We compute three versions of the wage variable. For each version, we aggregate the monthly incomes from the different activities the respondent was engaged in. For respondents who completed the in-person version of the questionnaire, we also add the extra income earned by women. The three versions are:

- Version 1: We use the "estimated agriculture income" for everyone, including respondents who answered the direct version of the questionnaire and are doing agriculture as their main activity. We replace the wage of respondents who don't earn any money by zero.
- Version 2: We use the "computed agriculture income" for respondents who filled the in-person survey and are doing agriculture as their main activity. For respondents who filled the indirect or phone version of the questionnaire, as well as for other respondents who have an "estimated" agriculture income but are not engaged in agriculture as their main activity, we use the "estimated agriculture income". We replace the wage of people who don't earn any money by zero.
- Version 3: We do not include agriculture income in this version, as there are a lot of outliers in these incomes. We replace the wage of people who don't earn any money by zero.

We create a capped version of each variable with a cap at 100,000 PKR (approximately 961 USD). We convert the raw and the capped variables to USD. Figure A2 shows the distribution for the raw and capped incomes for male respondents. In the paper, for mean regressions, we always use the capped income, and for median regressions, the raw income.

Appendix – References

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. The Quarterly Journal of Economics, 134(3), 1121-1162.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate behavioral research, 1(2), 245-276.

Crawfurd, L., Evans, D. K., Hares, S., & Sandefur, J. (2021). Teaching and testing by phone in a pandemic.

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

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. Journal of personality and social psychology, 92(6), 1087.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. Journal of applied Psychology, 86(1), 80.

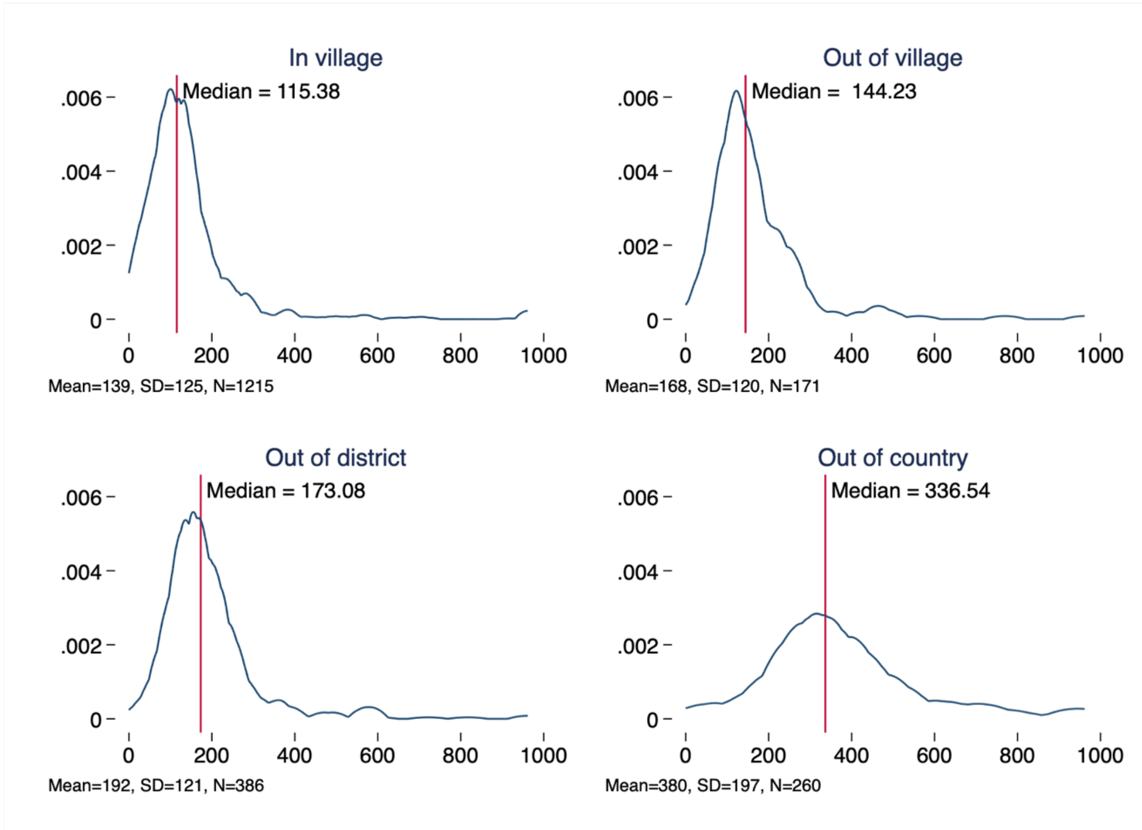
Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23(3), 187-200.

Laajaj, R., & Macours, K. (2021). Measuring skills in developing countries. Journal of Human resources, 56(4), 1254-1295.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT pre

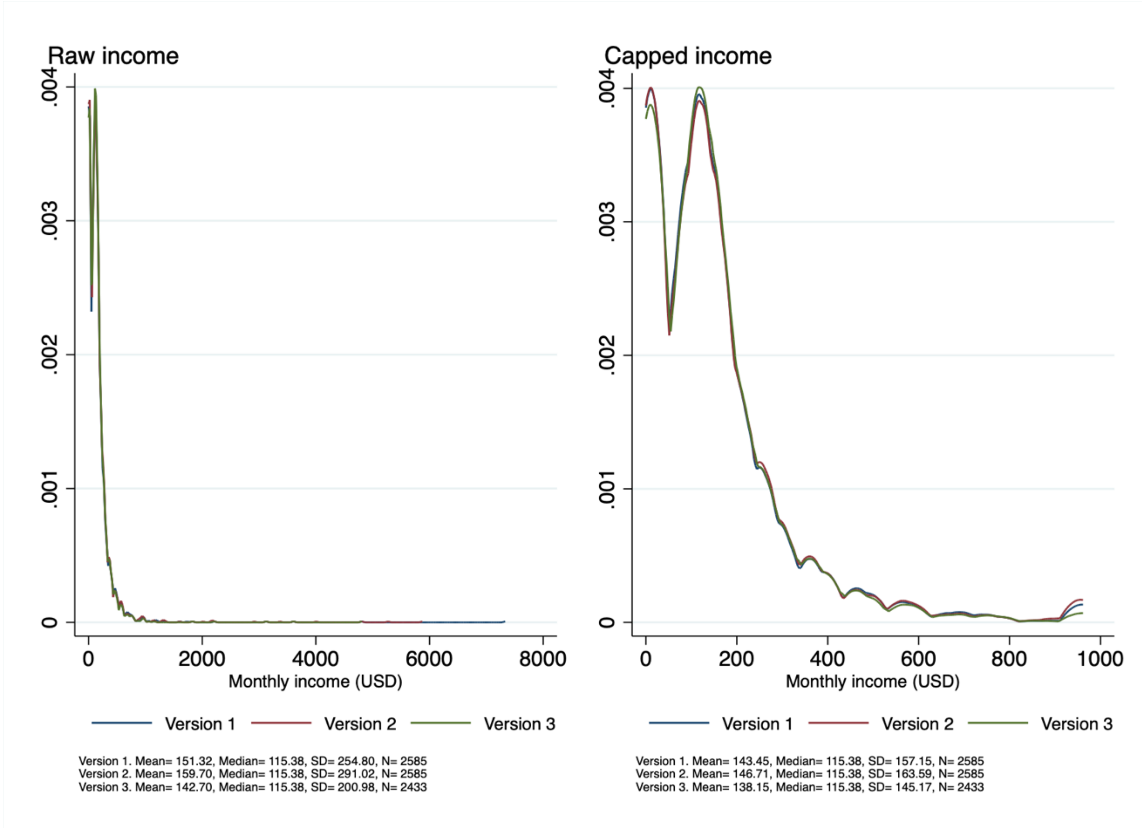
APPENDICES – FIGURES

Figure A1. Monthly earnings of males – by Location



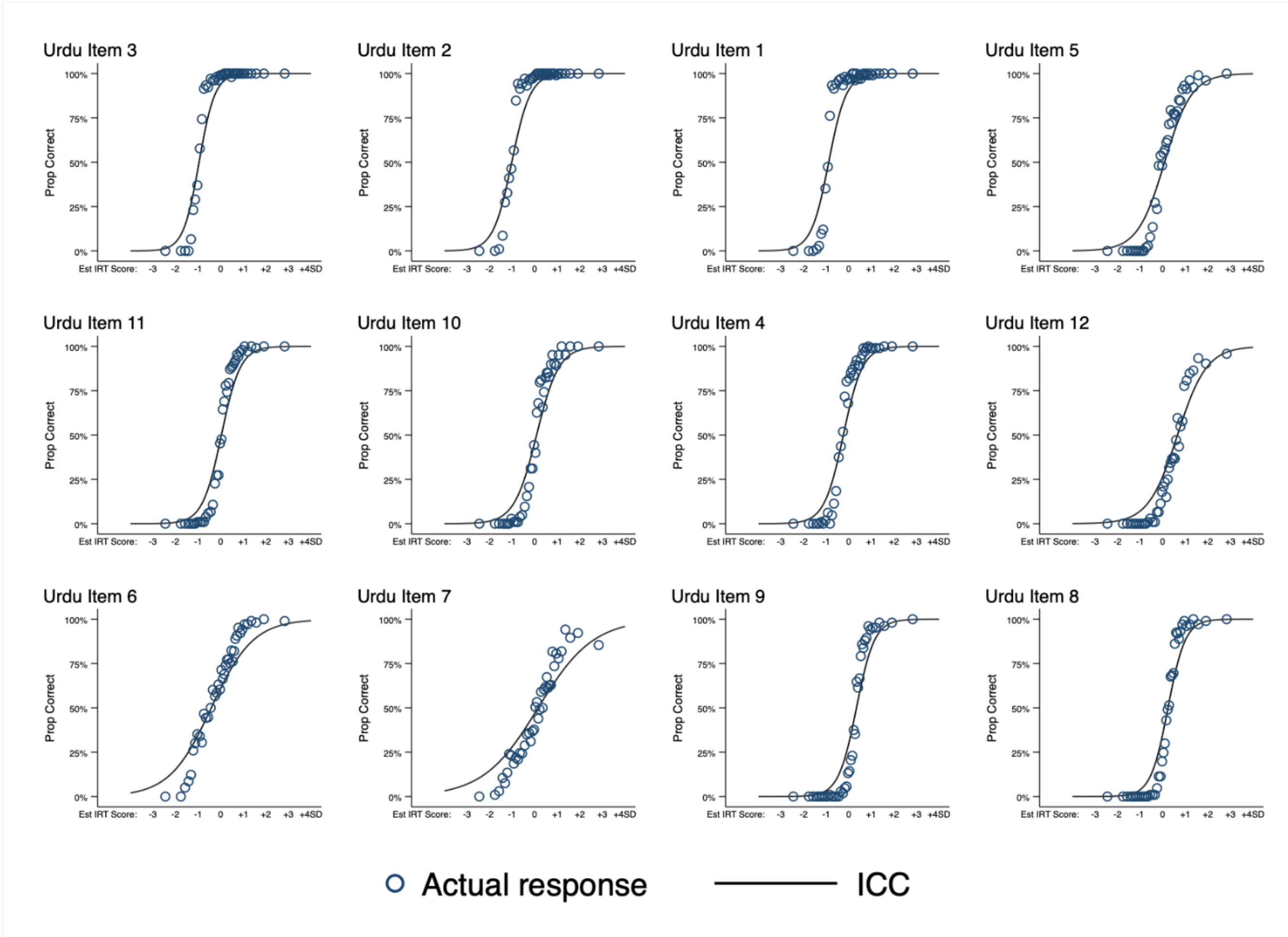
Notes. This figure shows the distribution of monthly earnings for male in our sample depending on where they live in 2018. Red lines are median. The sample includes all men currently enrolled in school.

Figure A2. Distribution of males' monthly earnings



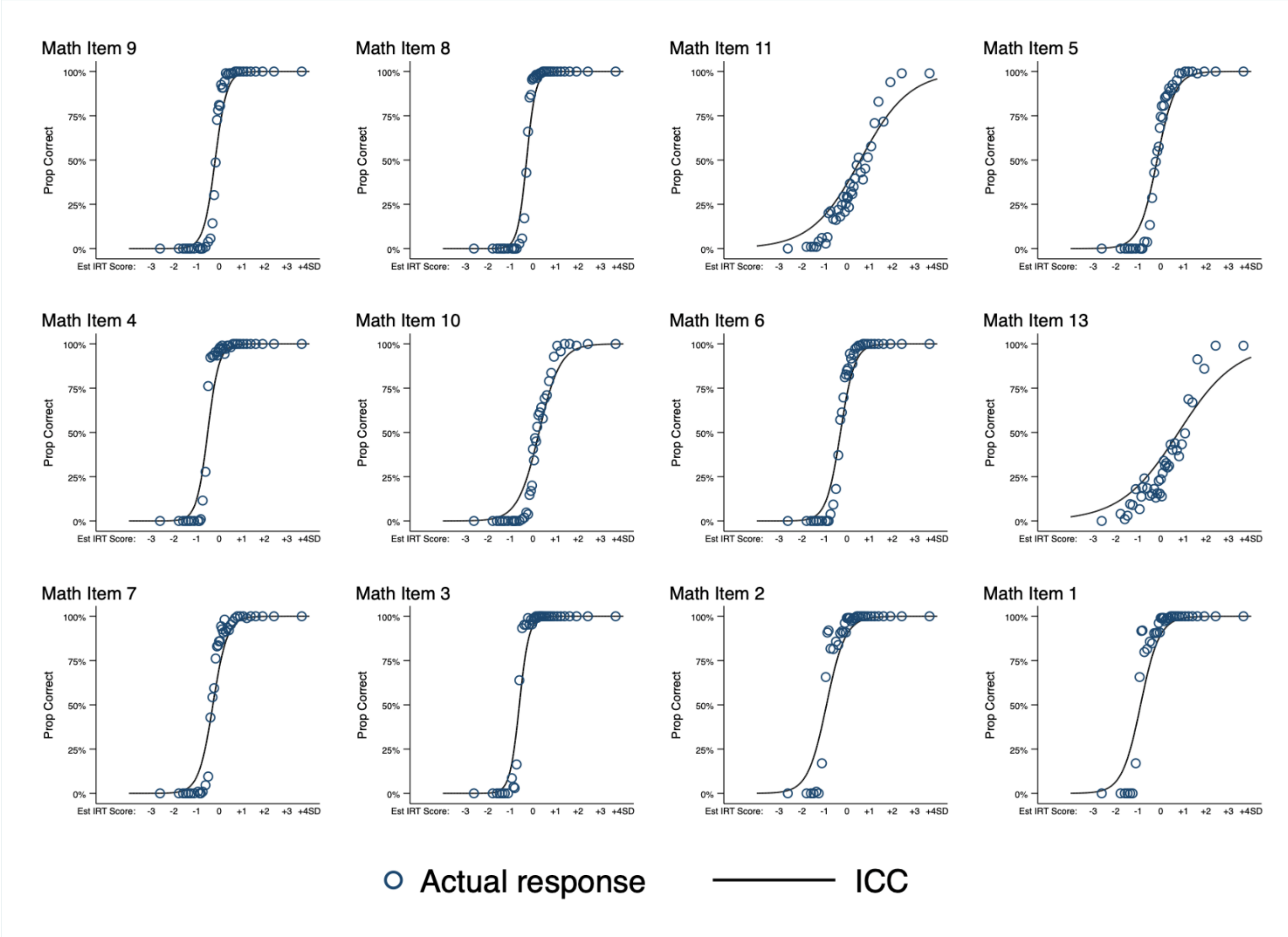
Notes. This figure shows the distribution of monthly earnings for males in our sample using different ways of aggregating respondents' income.

Figure A3. Item Characteristics Curves and Actual Response Patterns – Paper Test, Urdu



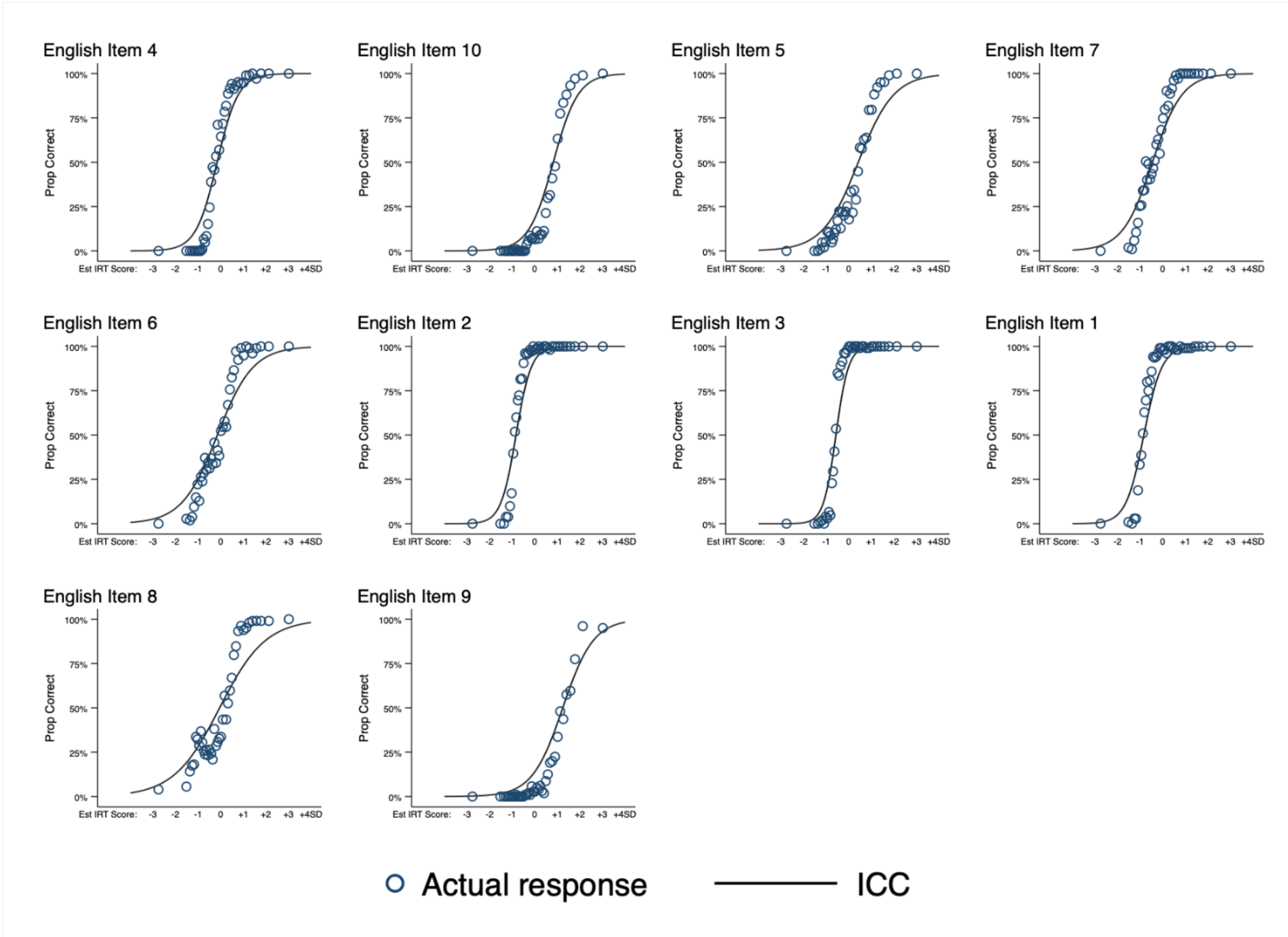
Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

Figure A4. Item Characteristics Curves and Actual Response Patterns – Paper Test, Mathematics



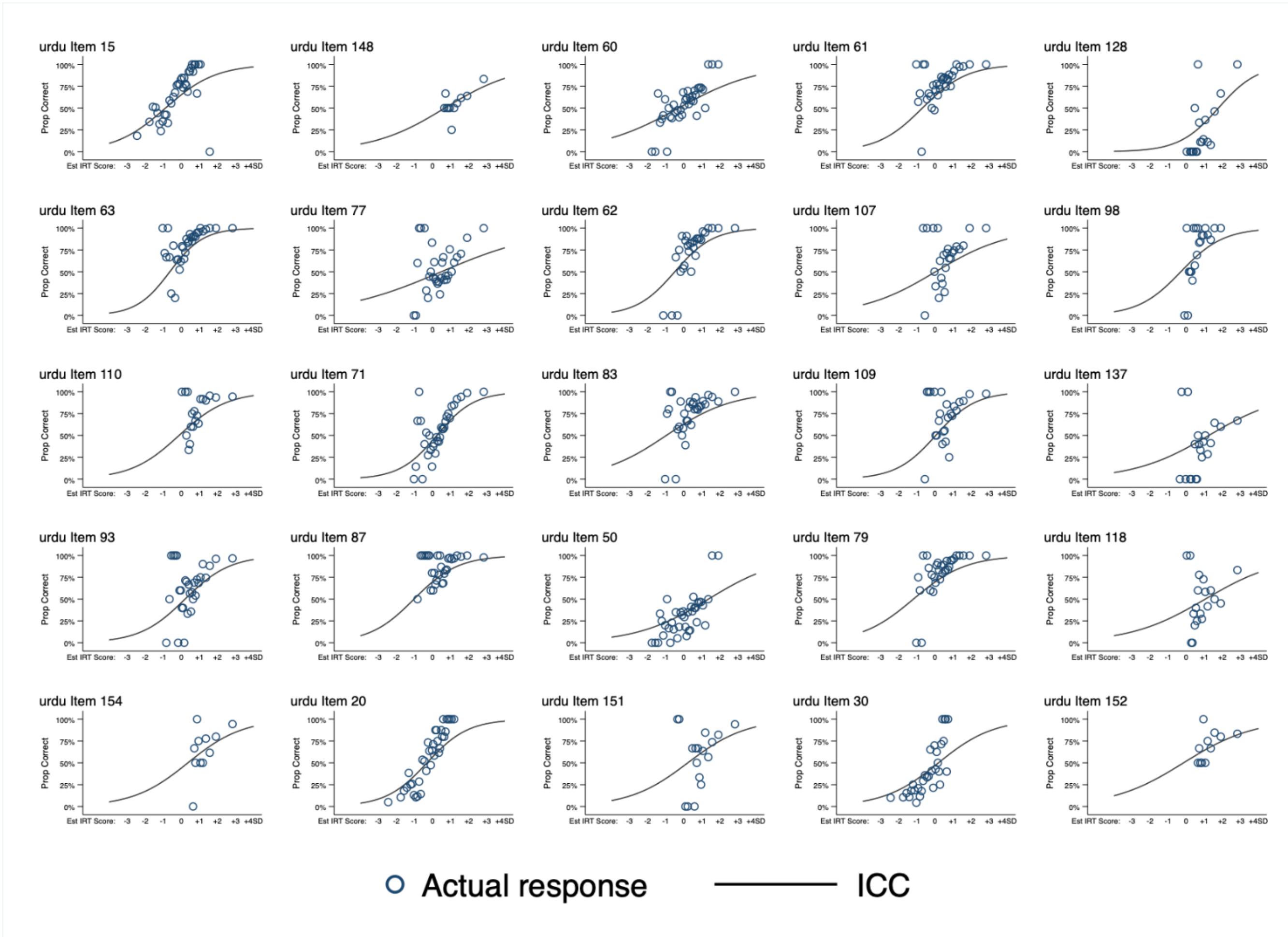
Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

Figure A5. Item Characteristics Curves and Actual Response Patterns – Paper Test, English



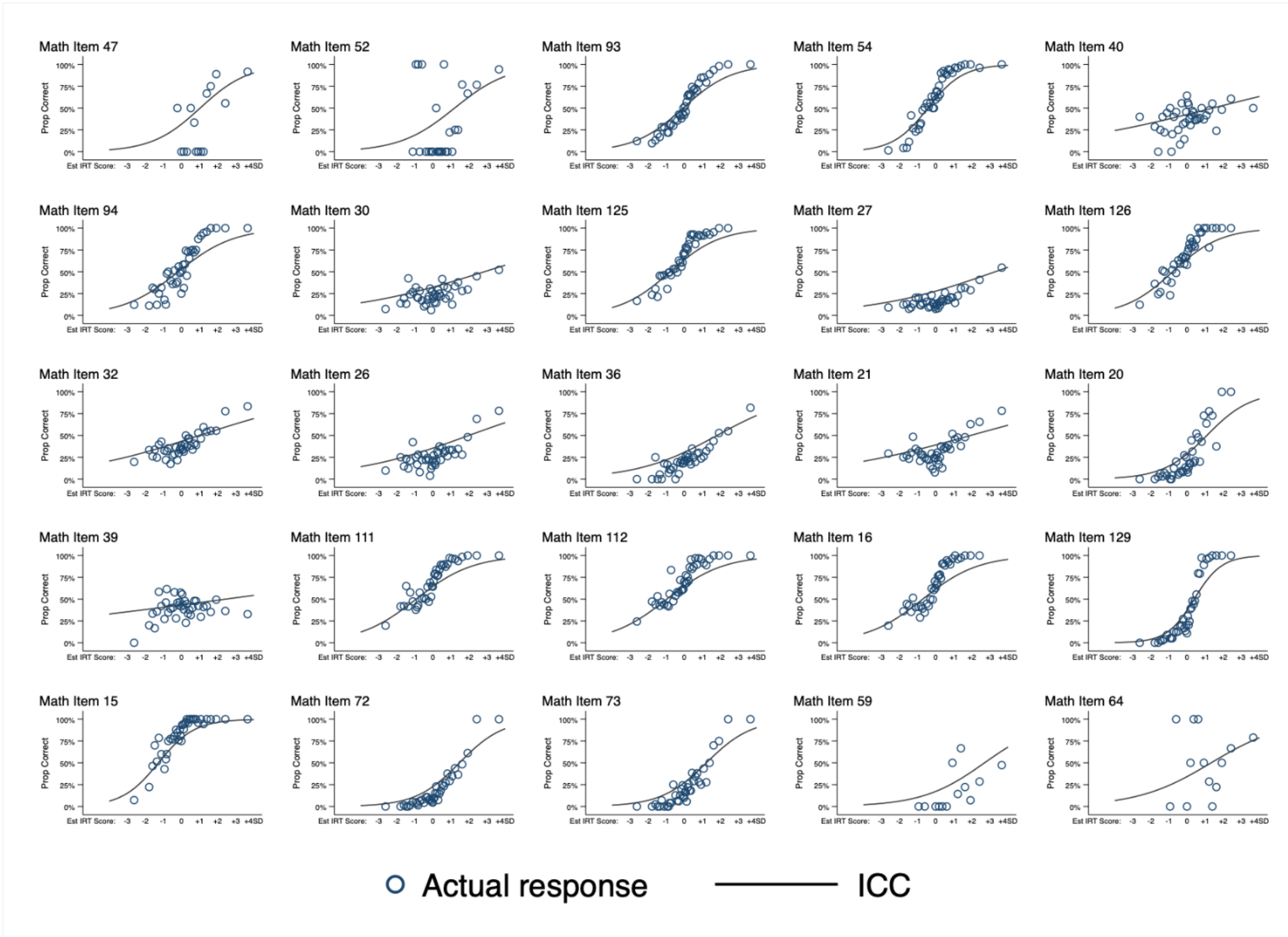
Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

Figure A6. Item Characteristics Curves and Actual Response Patterns – Adaptive Test, Urdu



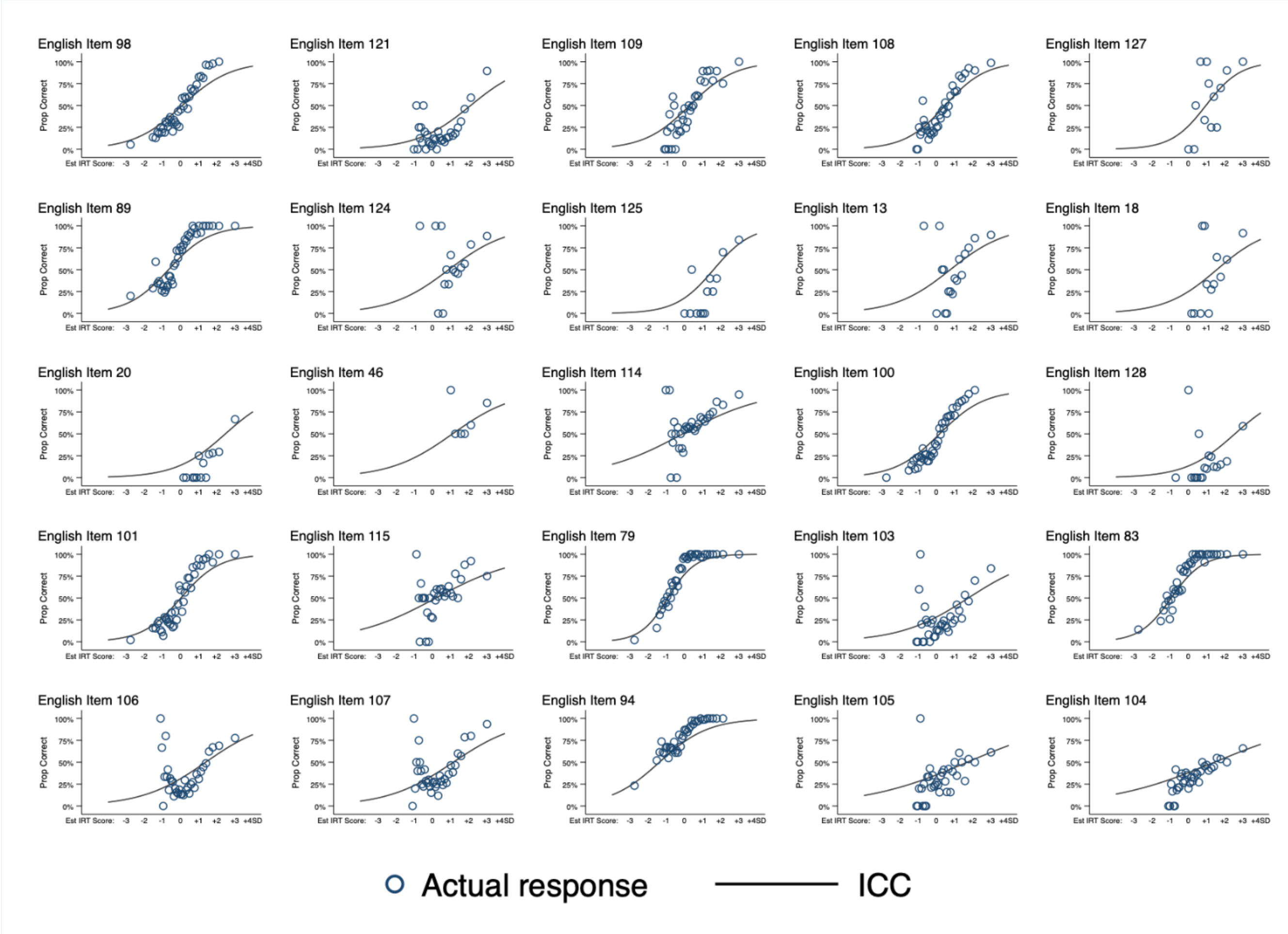
Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

Figure A7. Item Characteristics Curves and Actual Response Patterns – Adaptive Test, Mathematics



Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

<u>Figure A8</u>. Item Characteristics Curves and Actual Response Patterns – Adaptive Test, English



Notes. We use a two-parameter Item Response Theory (IRT) to model the likelihood of answering a question correctly. In this model, the probability of getting a question right is determined by the ability of the respondent, θ, and two items parameters – difficulty and discrimination. The solid line in each graph is the Item Characteristic Curve (ICC), which represents the expected patterns of responses for each θ. The actual pattern of responses against 40 quantiles of θ is then plotted against it.

APPENDICES – TABLES

Table A1. Definition of attrition

	In-person	Phone	Indirect	Died	4 Attock villages	No info	Refused/Other	Total
No Attrition	4391	0	0	0	0	0	0	4391
Skills measures missing	15	79	471	0	0	0	0	565
Full Attrition	0	0	0	43	186	285	395	909
Total	4406	79	471	43	186	285	395	5865

Notes. Respondents with "No Attrition" are respondents for whom we have the full questionnaire, cognitive skills measures (either from the test on tablet or on paper, or both), and socio-emotional skills measures. Respondents with "skills measures missing" are respondents for whom we either do not have the cognitive skills measures (both paper test and tablet test missing) or the socio-emotional skills measures, or neither. These respondents will be excluded from the regressions in which we include skills measures. For the in-person version of the survey, there are 4 respondents who did not finish the survey and 11 respondents for whom the paper test was forgotten and there was a bug with the test on tablets. Respondents with "Full Attrition" are respondents we were not able to collect any long-questionnaire information about. 43 of them died, 186 were living in four villages we could not visit in Attock because they fell under military control, 285 were part of households we could not track and 395 refused.

Table A2. Analysis of differential attrition (Panel A)

		(1) Full Data		(2) No skills measures		(3) Full Attrition	t-test *Diff* *(1)-(2)*	t-test *Diff* *(1)-(3)*	*Total Obs*
	N	*Mean [SD]*	*N*	*Mean [SD]*	*N*	*Mean [SD]*			
Location - 2018									
Original village	4391	0.67	565	0.17	395	NA	0.50***	NA	4956
		[0.47]		[0.38]			(0.02)		
Within district	4391	0.19	565	0.15	395	NA	0.04**	NA	4956
		[0.39]		[0.36]			(0.02)		
Within country	4391	0.13	565	0.29	395	NA	-0.16***	NA	4956
		[0.34]		[0.46]			(0.02)		
Outside country	4391	0.01	565	0.39	395	NA	-0.37***	NA	4956
		[0.12]		[0.49]			(0.02)		
Individual - 2018									
Age	4391	23.62	565	24.96	395	NA	-1.34***	NA	4956
		[3.64]		[3.23]			(0.15)		
Years of schooling	4391	8.56	564	8.40	395	NA	0.16	NA	4955
		[4.96]		[4.33]			(0.20)		
Ever Married	4391	0.45	565	0.48	395	NA	-0.04*	NA	4956
		[0.50]		[0.50]			(0.02)		
Age at first marriage	1974	20.89	274	21.40	395	NA	-0.51**	NA	2248
		[3.39]		[3.69]			(0.25)		
Has children	4391	0.33	565	0.35	395	NA	-0.02	NA	4956
		[0.47]		[0.48]			(0.02)		
Working	4391	0.42	565	0.64	395	NA	-0.22***	NA	4956
		[0.49]		[0.48]			(0.03)		
Main work is farming	4391	0.04	565	0.01	395	NA	0.03***	NA	4956
		[0.19]		[0.08]			(0.00)		
HH has toilets on premises	4391	0.96	13	0.85	395	NA	0.12	NA	4404
		[0.19]		[0.38]			(0.09)		
HH has access to electricity	4391	0.98	13	0.92	395	NA	0.06	NA	4404
		[0.14]		[0.28]			(0.07)		

* p<.10 ** p<.05 *** p<.01

Notes. Columns (1) to (3) display the means for the group that answered all parts of the survey, the group surveyed that did not complete the cognitive and socio-emotional assessments (either because surveyed over the phone or information collected indirectly), and the group who attrited fully. Standard deviations in brackets. Column (4) is the difference between the mean of respondents with full data and the mean of respondents who do not have skills measures. Column (5) is the difference between the mean of respondents with full data and the mean of attritors. Panel A shows differences in variables measured in this latest round of data collection, in 2018 so that we cannot report data on full attritors. Panel B reports data measured during Tagging and Tracking, in 2016. Panel C reports data measured in 2003 at the individual level. Panel D reports data measured in 2003 at the household level. Differences in means are computed by OLS regressions. All standard errors in parentheses are clustered at the village level.

Table A2. Analysis of differential attrition (Panel B)

		(1)		(2)		(3)	t-test	t-test	
		Full Data		**No skills measures**		**Full Attrition**	***Diff***	***Diff***	***Total Obs***
	N	***Mean [SD]***	***N***	***Mean [SD]***	***N***	***Mean [SD]***	***(1)-(2)***	***(1)-(3)***	
Location - 2016									
Original village	4375	0.74	562	0.42	552	0.71	0.32***	0.03	5489
		[0.44]		[0.49]		[0.45]	(0.02)	(0.03)	
Within district	4375	0.15	562	0.12	552	0.11	0.02	0.04***	5489
		[0.35]		[0.33]		[0.31]	(0.02)	(0.02)	
Within country	4375	0.09	562	0.21	552	0.13	-0.12***	-0.04**	5489
		[0.29]		[0.41]		[0.34]	(0.02)	(0.02)	
Outside country	4375	0.02	562	0.25	552	0.05	-0.22***	-0.03**	5489
		[0.16]		[0.43]		[0.22]	(0.02)	(0.01)	
Individual - 2016									
Sex (1=Female)	4337	0.49	531	0.33	339	0.54	0.16***	-0.05*	5207
		[0.50]		[0.47]		[0.50]	(0.02)	(0.03)	
Age	4337	22.36	531	23.83	339	22.73	-1.48***	-0.38**	5207
		[3.72]		[3.17]		[3.46]	(0.15)	(0.18)	
Ever married	4337	0.38	531	0.40	339	0.29	-0.03	0.09***	5207
		[0.48]		[0.49]		[0.45]	(0.02)	(0.03)	
Enrolled	4173	0.15	381	0.03	299	0.19	0.12***	-0.03	4853
		[0.36]		[0.17]		[0.39]	(0.01)	(0.02)	

* p<.10 ** p<.05 *** p<.01

Notes. Columns (1) to (3) display the means for the group that answered all parts of the survey, the group surveyed that did not complete the cognitive and socio-emotional assessments (either because surveyed over the phone or information collected indirectly), and the group who attrited fully. Standard deviations in brackets. Column (4) is the difference between the mean of respondents with full data and the mean of respondents who do not have skills measures. Column (5) is the difference between the mean of respondents with full data and the mean of attritors. Panel A shows differences in variables measured in this latest round of data collection, in 2018 so that we cannot report data on full attritors. Panel B reports data measured during Tagging and Tracking, in 2016. Panel C reports data measured in 2003 at the individual level. Panel D reports data measured in 2003 at the household level. Differences in means are computed by OLS regressions. All standard errors in parentheses are clustered at the village level.

Table A2. Analysis of differential attrition (Panel C)

	N	(1) **Full Data** *Mean [SD]*	N	(2) **No skills measures** *Mean [SD]*	N	(3) **Full Attrition** *Mean [SD]*	t-test *Diff (1)-(2)*	t-test *Diff (1)-(3)*	*Total Obs*
Individual - 2003									
Respondent's Age	4391	9.79	565	10.64	909	10.13	-0.85***	-0.34***	5865
		[2.99]		[2.78]		[2.85]	(0.13)	(0.10)	
Respondent's Sex	4391	0.49	565	0.33	909	0.50	0.16***	-0.01	5865
		[0.50]		[0.47]		[0.50]	(0.02)	(0.02)	
Respondent's highest grade completed	4267	2.10	552	2.76	879	2.22	-0.67***	-0.13	5698
		[2.22]		[2.30]		[2.30]	(0.12)	(0.10)	
Respondent suffers from disability	4236	0.01	550	0.03	863	0.05	-0.01	-0.03***	5649
		[0.12]		[0.16]		[0.22]	(0.01)	(0.01)	
How good is respondent health, max is 16	4308	15.62	557	15.63	881	15.60	-0.01	0.02	5746
		[1.26]		[1.40]		[1.47]	(0.08)	(0.07)	
How intelligent is respondent, max is 5	4279	3.19	554	3.19	884	3.10	-0.00	0.09**	5717
		[0.69]		[0.74]		[0.71]	(0.03)	(0.03)	
How hardworking is respondent, max is 5	4279	3.06	554	2.98	884	2.99	0.08**	0.07*	5717
		[0.76]		[0.81]		[0.79]	(0.04)	(0.03)	
Was tested in 2003	4391	0.15	565	0.18	909	0.15	-0.02	0.00	5865
		[0.36]		[0.38]		[0.36]	(0.02)	(0.02)	
English IRT Score (2003)	670	-0.55	99	-0.88	138	-0.44	0.33**	-0.11	907
		[1.16]		[1.31]		[1.02]	(0.16)	(0.13)	
Math IRT Score (2003)	670	-0.31	99	-0.44	138	-0.41	0.13	0.11	907
		[0.96]		[1.04]		[1.07]	(0.13)	(0.11)	
Urdu IRT Score (2003)	670	-0.49	99	-0.77	138	-0.49	0.28**	0.01	907
		[1.18]		[1.24]		[1.13]	(0.12)	(0.12)	
Mean IRT Score (2003)	670	-0.45	99	-0.69	138	-0.45	0.25**	0.00	907
		[0.93]		[1.00]		[0.94]	(0.12)	(0.10)	

* $p<.10$ ** $p<.05$ *** $p<.01$

Notes. Columns (1) to (3) display the means for the group that answered all parts of the survey, the group surveyed that did not complete the cognitive and socio-emotional assessments (either because surveyed over the phone or information collected indirectly), and the group who attrited fully. Standard deviations in brackets. Column (4) is the difference between the mean of respondents with full data and the mean of respondents who do not have skills measures. Column (5) is the difference between the mean of respondents with full data and the mean of attritors. Panel A shows differences in variables measured in this latest round of data collection, in 2018 so that we cannot report data on full attritors. Panel B reports data measured during Tagging and Tracking, in 2016. Panel C reports data measured in 2003 at the individual level. Panel D reports data measured in 2003 at the household level. Differences in means are computed by OLS regressions. All standard errors in parentheses are clustered at the village level.

Table A2. Analysis of differential attrition (Panel D)

		(1)		(2)		(3)	t-test	t-test	
		Full Data		**No skills measures**		**Full Attrition**	*Diff*	*Diff*	*Total Obs*
	N	*Mean [SD]*	*N*	*Mean [SD]*	*N*	*Mean [SD]*	*(1)-(2)*	*(1)-(3)*	
Household - 2003									
HH SES in 2003	4382	-0.12	564	-0.02	905	-0.38	-0.09	0.27*	5851
		[1.99]		[1.90]		[1.99]	(0.12)	(0.14)	
Family owns house it is living in	4391	0.95	565	0.96	909	0.90	-0.01	0.05**	5865
		[0.21]		[0.18]		[0.30]	(0.01)	(0.02)	
Number of rooms house	4391	2.50	565	2.53	909	2.45	-0.03	0.04	5865
		[1.36]		[1.36]		[1.36]	(0.08)	(0.11)	
Type of house is permanent	4391	0.68	562	0.67	909	0.59	0.01	0.09***	5862
		[0.46]		[0.47]		[0.49]	(0.02)	(0.03)	
HH has toilets on premises	4390	0.58	565	0.60	909	0.52	-0.02	0.06	5864
		[0.49]		[0.49]		[0.50]	(0.03)	(0.04)	
HH has hard roof	4391	0.55	565	0.51	909	0.53	0.03	0.01	5865
		[0.50]		[0.50]		[0.50]	(0.03)	(0.04)	
HH has access to electricity	4383	0.88	565	0.88	905	0.84	-0.01	0.04	5853
		[0.33]		[0.32]		[0.37]	(0.02)	(0.04)	
Relative HH wealth compared to rest of village, max is 4	4391	3.37	565	3.36	909	3.42	0.02	-0.04	5865
		[0.89]		[0.87]		[1.01]	(0.05)	(0.07)	
HH size	4391	8.86	565	8.57	909	8.24	0.28	0.62**	5865
		[4.07]		[3.15]		[2.77]	(0.22)	(0.28)	
HH religion is not Islam	4381	0.02	565	0.01	909	0.03	0.01**	-0.01	5855
		[0.15]		[0.09]		[0.17]	(0.01)	(0.01)	
Male interview language is not Punjabi	4383	0.26	561	0.17	906	0.21	0.10***	0.06*	5850
		[0.44]		[0.37]		[0.41]	(0.02)	(0.03)	
Father is not living in HH	4293	0.12	555	0.16	884	0.21	-0.03*	-0.08***	5732
		[0.33]		[0.36]		[0.40]	(0.02)	(0.02)	
Mother is not living in HH	4298	0.02	555	0.03	884	0.02	-0.01	-0.00	5737
		[0.15]		[0.17]		[0.16]	(0.01)	(0.01)	
Parents - 2003									
Parent years of schooling	7791	3.13	989	2.86	1538	2.84	0.27	0.29	10318
		[4.31]		[4.05]		[4.08]	(0.18)	(0.23)	
Parent can read	7944	0.38	1001	0.35	1561	0.36	0.02	0.02	10506
		[0.48]		[0.48]		[0.48]	(0.02)	(0.02)	
Parent number of children	7386	5.37	937	5.50	1421	5.29	-0.13	0.08	9744
		[1.82]		[1.82]		[1.65]	(0.11)	(0.12)	
Parent is working	7958	0.51	1006	0.50	1564	0.48	0.02	0.03*	10528

		[0.50]		[0.50]		[0.50]	(0.01)	(0.02)	
Parent main work is farming	7958	0.18	1006	0.18	1564	0.13	0.00	0.05***	10528
		[0.39]		[0.39]		[0.34]	(0.01)	(0.02)	

* p<.10 ** p<.05 *** p<.01

Notes. Columns (1) to (3) display the means for the group that answered all parts of the survey, the group surveyed that did not complete the cognitive and socio-emotional assessments (either because surveyed over the phone or information collected indirectly), and the group who attrited fully. Standard deviations in brackets. Column (4) is the difference between the mean of respondents with full data and the mean of respondents who do not have skills measures. Column (5) is the difference between the mean of respondents with full data and the mean of attritors. Panel A shows differences in variables measured in this latest round of data collection, in 2018 so that we cannot report data on full attritors. Panel B reports data measured during Tagging and Tracking, in 2016. Panel C reports data measured in 2003 at the individual level. Panel D reports data measured in 2003 at the household level. Differences in means are computed by OLS regressions. All standard errors in parentheses are clustered at the village level.

Table A3. Employment category and respondent location for men

		Respondent Location			
		In Village	**Out of Village**	**Out of District**	**Out of Country**
Employment	**Daily Wage**	30%	23%	14%	15%
	Salaried	31%	51%	76%	81%
	Self-employed or Family Business	26%	24%	9%	4%
	Agriculture	13%	2%	2%	0%
		100%	100%	100%	100%

Notes. This table shows the share of respondents in each employment category depending on where they currently live. Daily Wage refers to someone working for an employer that pays a wage daily. Salaried refers to someone working for an employer that pays a wage monthly. It can be either in the formal or the informal sector as long as the individual receives a wage. Self-employed or Family Business refers to someone working for themselves or a family member (outside Agriculture & Livestock). Agriculture refers to someone working in agriculture and livestock for themselves or their family. If the respondent is doing agriculture for someone else for a monthly wage, they are categorized as Salaried and not Agriculture. The sample includes all men currently working, including those who are simultaneously currently enrolled, that is 2,043 men.

Table A4. Instruments Overview

Index	Construct	Instrument	Mode
Cognitive	Urdu	LEAPS Test	Paper Test
Cognitive	Mathematics	LEAPS Test	Paper Test
Cognitive	English	LEAPS Test	Paper Test
Cognitive	Urdu	Adaptive Test Designed for Study	Tablet
Cognitive	Mathematics	Adaptive Test Designed for Study	Tablet
Cognitive	English	Adaptive Test Designed for Study	Tablet
Life	Literacy	Read and Interpret Real Life Electricity Bill	Administered by enumerator
Life	Numeracy	Read text messages in Urdu, roman Urdu and English	Administered by enumerator
Socio-Emotional	Grit	Grit scale	Self-reported instrument, administered by enumerator
Socio-Emotional	Big Five	Big Five scale	Self-reported instrument, administered by enumerator
Socio-Emotional	Self-Control	GoNoGo task	Task-based (Tablet)
Socio-Emotional	Risk taking behavior	Balloon Analogue Risk Task (BART)	Task-based (Tablet)

Notes. In this table, we present the instruments used to measure the three different types of skills we capture in our survey: cognitive skills, life skills, and socio-emotional skills. To measure cognitive skills, we first used a paper test with 12 items for Urdu, 13 items for Mathematics, and 10 items for English. Then, we also used an adaptive test administered on tablets designed especially for this study. We also designed 17 questions testing functional literacy and numeracy useful in the respondents' everyday life. Examples included understanding how much is due from an electricity bill or reading text messages. These items are aggregated in an index labeled "life-skills". The socio-emotional skills instruments included self-reported scales and tasks administered on tablets. We use the term "self-reported" as the respondents answered the items but an enumerator was reading the item out loud given that some respondents were illiterate. Grit is defined as "the combination of passion and perseverance for long-term goals" (Duckworth and Quinn, 2009). We used a 10-items version of the self-reported grit scale in Pakistan. The second self-reported scale measures the “Big Five”, a taxonomy of traits that encompasses five dimensions of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To measure these traits, we used the short 15-item Big Five Inventory (Lang et al., 2011), which consists of three items for each of the five personality traits. The GoNoGo task was used to measure impulse control: The participant is presented with a square on the screen for a very short period. If the square is of any color but black, the participant must touch the screen as quickly as possible. If the square is black (the “no go” stimulus), the respondent must inhibit their response. A total of 72 trials are completed (48 Go and 24 NoGo trials) and the main outcome we use is the average response time. The Balloon Analogue Risk Task (BART) measures risk-taking behavior by asking participants to maximize the amount of money they can win from the game. Respondents earned on average 322 PKR (approximately 3 USD at the time) from the game. On each trial, they were presented with a balloon, which they can pump. Each pump earned them money but increased the likelihood that the next pump would "pop" the balloon, in which case they lost the accrued money for that balloon. If they instead chose to stop pumping the balloon, they collected their accrued money and moved to the next trial. The main outcome is the average number of pumps on the balloons that did not explode.

Table A5. Learning over time in Pakistan

Subject	What is the question	% correct 2003	% correct 2011	% correct 2018	% correct 2018 , college	N
Panel A. Respondents with test scores in 2003, 2011 and 2018						
Mathematics	Tick box next number that matches the number of objects	49 %	89 %	94 %	99 %	467
Mathematics	678+923	56 %	83 %	77 %	94 %	467
Mathematics	7/3=__	3 %	16 %	7 %	18 %	467
English	Match picture: Banana	63 %	94 %	91 %	99 %	467
English	Missing letter to match picture: Flag	27 %	76 %	70 %	96 %	467
English	Use word in sentence: deep	1 %	29 %	22 %	63 %	467
Urdu	Match picture: Book	74 %	98 %	95 %	98 %	467
Urdu	Join letters and write word: m-a-l-k	36 %	76 %	64 %	83 %	467
Urdu	Fill blank in the story by selecting the correct word 3	NA	NA	NA	NA	0
Panel B. Respondents with test scores in 2011 and 2018						
Mathematics	Tick box next number that matches the number of objects	NA	82 %	85 %	100 %	1643
Mathematics	678+923	NA	71 %	62 %	93 %	1643
Mathematics	7/3=__	NA	13 %	5 %	20 %	1643
English	Match picture: Banana	NA	88 %	82 %	99 %	1643
English	Missing letter to match picture: Flag	NA	69 %	60 %	97 %	1643
English	Use word in sentence: deep	NA	21 %	18 %	61 %	1643
Urdu	Match picture: Banana	NA	90 %	86 %	100 %	1643
Urdu	Join letters and write word: m-a-l-k	NA	66 %	53 %	85 %	1643
Urdu	Fill blank in the story by selecting the correct word 3	NA	44 %	36 %	74%	1643

Notes. This table shows a sample of questions for each of the three subjects tested on paper: Mathematics, Urdu, and English. We show the same sample questions as in Table 4 (Part 1). For each of the questions, we show the sample's proportion of correct answers in 2003, 2011, and 2018, as well as the number of observations. Panel A is restricted to respondents who were tested in 2003, 2011, and 2018 and has 467 respondents, while Panel B is restricted to respondents who were tested in 2001 and 2018 and has 1,643 respondents. The number of respondents for "% correct 2018, college" is restricted to the sub-sample of those respondents who went to college: the sample is 82 respondents for Panel A and 234 respondents for Panel B.

Table A6. Correlations between skills measures and education

	Schooling	Cognitive	SEMS	Life Skills
Years of schooling	1			
Cognitive Skills Index	0.8	1		
Socio-Emotional Skills Index	0.15	0.14	1	
Life Skills Index	0.81	0.77	0.18	1

Notes. This table shows the bivariate correlations between years of schooling, the cognitive skills index, the socio-emotional skills index, and the life skills index. The Cognitive Skill Index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (leaving out items that less than 50 respondents answered, and those that less than 5% or more than 95% of respondents got the correct answer). The socio-emotional skills (SEMS) index is computed using principal factor analysis on the Big Five items, Grit items, BART, and GoNoGo scores. The dependent variable in columns (5) and (6) is a life skills index computed using principal factor analysis on 17 functional literacy and numeracy questions.

Table A7. Relationship migration, schooling and skills in Pakistan

	Men						Women		
	(1) Out of Village	(2) Out of Village	(3) Out of Village	(4) Out of Pakistan	(5) Out of Pakistan	(6) Out of Pakistan	(7) Out of Village	(8) Out of Village	(9) Out of Village
Years of schooling	0.006**	0.007**	-0.002	-0.000	0.001	0.002	-0.002	-0.002	0.003
	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)
Cognitive Skills			0.039**			-0.005			-0.031
			(0.015)			(0.005)			(0.020)
SEMS			0.046***			-0.000			-0.009
			(0.010)			(0.003)			(0.012)
Mother highest grade	0.006	0.004	0.003	0.003	0.001	0.001	0.003	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)	(0.001)	(0.004)	(0.005)	(0.005)
Father highest grade	-0.002	0.000	-0.000	-0.003*	-0.001	-0.001	-0.002	-0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
HH SES in 2003	-0.005	-0.010*	-0.010	0.006*	0.002*	0.002*	-0.014**	-0.016***	-0.015***
	(0.006)	(0.006)	(0.006)	(0.003)	(0.001)	(0.001)	(0.006)	(0.006)	(0.005)
Constant	0.007	0.003	0.106**	0.017	0.019	0.011	0.007	-0.001	-0.033
	(0.040)	(0.040)	(0.045)	(0.025)	(0.011)	(0.014)	(0.019)	(0.019)	(0.027)
Observations	2,595	2,217	2,217	2,595	2,217	2,217	2,360	2,174	2,174
Adjusted R-squared	0.026	0.023	0.035	0.025	0.006	0.005	0.128	0.129	0.130
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Has skills measures	Has skills measures	All	Has skills measures	Has skills measures	Has skills measures	Has skills measures	Has skills measures
N_Clusters	108	108	108	108	108	108	108	108	108
Mean Dependent	0.348	0.267	0.267	0.104	0.0289	0.0289	0.430	0.395	0.395

* p<.10 ** p<.05 *** p<.01

Notes. The dependent variable "Out of village" in columns (1) to (3), and (7) to (9) is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household (their natal village most of the time). The dependent variable "Out of Pakistan" in columns (4) to (6) is a dummy variable taking the value 1 if the respondent lives outside of Pakistan. Out of the 2,360 women in the sample, only 11 reported living outside of Pakistan. Therefore, we do not run these regressions for the women sample. The cognitive skills index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills index (SEMS) is computed using principal component analysis on the Big-Five items, Grit items, BART, and GoNoGo scores. The sample for columns (1) and (4) is all men surveyed. The sample for columns (2), (3), (5), and (6) is all men who answered the direct version of the questionnaire (for whom we have skills measures). The sample for column (7) is all women. The sample for columns (8) and (9) is all women who answered the direct version of the questionnaire (for whom we have skills measures). Standard errors are in parentheses (clustered at the village level). All the regressions include age and district fixed effects. The R-squared shown is the adjusted R-squared.

Table A8. Socio-emotional skills factors and Earnings for Men in Pakistan

	Conscientiousness/Grit (1)				Openness to Experience (2)				Agreeableness (3)			
	Monthly income				*Monthly income*				*Monthly income*			
	Mean		*Median*		*Mean*		*Median*		*Mean*		*Median*	
Years of Schooling (a1)	2.96**	0.86	1.36**	0.087	3.27***	1.09	1.85***	0.78	3.39***	1.27	2.26***	1.36*
	(1.16)	(1.09)	(0.67)	(0.78)	(1.15)	(1.02)	(0.65)	(0.74)	(1.16)	(1.04)	(0.70)	(0.78)
Cognitive Skills (a2)	6.63	2.83	7.41***	-0.18	5.63	0.87	7.81***	-0.29	6.26	2.05	6.40**	-1.71
	(4.71)	(5.01)	(2.70)	(3.60)	(4.79)	(4.98)	(2.99)	(2.47)	(4.78)	(4.99)	(3.03)	(3.33)
SEMS (a3)	12.6***	8.70**	15.7***	14.7***	6.69**	12.0***	5.36***	7.17***	4.56	7.85**	3.22	5.66**
	(3.74)	(3.88)	(2.13)	(2.47)	(3.23)	(3.37)	(1.75)	(2.51)	(3.17)	(3.04)	(2.45)	(2.75)
Out Village (a4)		5.88		29.3**		11.9		38.9**		17.1		41.3***
		(22.4)		(12.3)		(21.8)		(15.7)		(22.5)		(15.0)
Interaction YrsSchooling and Out Village (b1)		9.23***		5.01***		8.92***		4.30**		8.53***		4.07***
		(2.53)		(1.35)		(2.49)		(1.71)		(2.52)		(1.57)
Interaction Cog and Out Village (b2)		-5.00		13.9***		-0.36		16.2***		-2.71		16.3***
		(11.4)		(5.20)		(11.4)		(6.29)		(11.4)		(5.92)
Interaction SEMS and Out Village (b3)		-2.37		-8.43*		-22.4***		-8.95*		-20.5***		-5.37
		(7.81)		(5.11)		(7.81)		(4.98)		(6.84)		(5.76)
Constant	48.5***	37.3***	36.0	30.1	44.3***	35.8***	26.7	22.1	50.3***	47.2***	26.0	26.8
	(9.60)	(9.82)	(69.1)	(36.3)	(9.71)	(9.45)	(74.9)	(51.3)	(11.1)	(11.4)	(69.9)	(45.5)
Observations	1978	1978	1978	1978	1978	1978	1978	1978	1978	1978	1978	1978
R-squared	0.050	0.12	0.043	0.091	0.040	0.13	0.032	0.086	0.040	0.13	0.032	0.086
Median/Mean Dependent	134.71	134.71	115.38	115.38	134.71	134.71	115.38	115.38	134.71	134.71	115.38	115.38
a1+b1=0		10.09***		5.10***		10.01***		5.08***		9.80***		5.43***
		(2.49)		(1.17)		(2.45)		(1.58)		(2.48)		(1.40)
a2+b2=0		-2.16		13.67***		0.51		15.94***		-0.66		14.57***
		(10.06)		(3.92)		(10.10)		(5.86)		(10.03)		(5.01)
a3+b3=0		6.33		6.27		-10.46		-1.77		-12.61*		0.29
		(6.78)		(4.54)		(6.88)		(4.36)		(6.36)		(5.11)

(continued)

Table A8 (continued)

	Extroversion (4)				Emotional Stability (5)			
	Monthly income				*Monthly income*			
	Mean		*Median*		*Mean*		*Median*	
Years of Schooling (a1)	3.39***	1.20	2.21***	0.81	3.08***	0.90	1.89***	1.08
	(1.16)	(1.05)	(0.65)	(0.79)	(1.18)	(1.04)	(0.49)	(0.78)
Cognitive Skills (a2)	6.35	2.18	6.97**	-0.73	5.84	2.43	5.68**	-1.10
	(4.79)	(5.02)	(2.71)	(3.83)	(4.75)	(5.00)	(2.33)	(3.63)
SEMS (a3)	-0.32	4.55	0.48	4.60**	11.5***	9.98***	7.98***	5.26**
	(3.12)	(2.93)	(2.11)	(2.28)	(4.06)	(3.60)	(1.84)	(2.52)
Out Village (a4)		8.93		37.0**		7.13		36.1**
		(21.8)		(16.2)		(21.6)		(16.1)
Interaction YrsSchooling and Out Village (b1)		8.97***		4.48***		9.21***		4.28**
		(2.50)		(1.74)		(2.52)		(1.70)
Interaction Cog and Out Village (b2)		-3.70		16.0**		-5.66		14.2**
		(11.4)		(6.67)		(11.2)		(6.39)
Interaction SEMS and Out Village (b3)		-8.54		-5.11		-1.93		-1.00
		(7.28)		(4.89)		(8.28)		(5.33)
Constant	41.8***	33.7***	22.7	20.6	43.9***	34.5***	22.2	19.2
	(9.76)	(9.51)	(55.0)	(36.5)	(9.56)	(9.40)	(68.1)	(46.5)
Observations	1978	1978	1978	1978	1978	1978	1978	1978
R-squared	0.040	0.12	0.031	0.085	0.050	0.13	0.034	0.086
Median/Mean Dependent	134.71	134.71	115.38	115.38	134.71	134.71	115.38	115.38
a1+b1=0		10.17***		5.29***		10.10***		5.35***
		(2.47)		(1.58)		(2.52)		(1.55)
a2+b2=0		-1.52		15.31***		-3.24		13.06**
		(10.12)		(5.58)		(9.81)		(5.37)
a3+b3=0		-3.99		-0.51		8.04		4.25
		(6.95)		(4.33)		(8.17)		(4.72)

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, migration, and earnings for men in Pakistan. The dependent variable Monthly income is the raw monthly income for median regressions and is top coded at 100,000 PKR per month (961.5 USD) for mean regressions. The cognitive index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills index (SEMS) is computed using principal factor analysis on the Conscientiousness sub-scale of the Big-Five and Grit items for the "Conscientiousness/Grit" columns. The SEMS index is computed using principal factor analysis on the Openness to Experience, Agreeableness, Extroversion, and Emotional stability sub-scales of the Big-Five for the respective columns. The "Out of village" variable is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household (their natal village most of the time). The sample is all men who answered the direct version of the questionnaire, in-person: 1,978 men. All the regressions include age and district fixed effects. Robust standard errors are in parentheses. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.

Table A9. Socio-emotional skills factors and Earnings for Women in Pakistan

	Conscientiousness/Grit			Openness to Experience			Agreeableness		
	(1)			(2)			(3)		
	Monthly income			*Monthly income*			*Monthly income*		
	All	*Working women*		*All*	*Working women*		*All*	*Working women*	
	Mean	*Mean*	*Median*	*Mean*	*Mean*	*Median*	*Mean*	*Mean*	*Median*
Years of Schooling	1.36***	8.16*	6.10***	1.32***	7.74*	4.92*	1.39***	8.17**	5.75**
	(0.43)	(4.13)	(2.32)	(0.43)	(4.28)	(2.84)	(0.43)	(4.09)	(2.83)
Cognitive Skills	3.02	1.96	-5.65	3.14*	1.53	-0.96	3.05*	1.97	-3.51
	(1.83)	(18.4)	(9.88)	(1.79)	(18.1)	(13.1)	(1.81)	(18.1)	(11.2)
SEMS	1.25*	0.45	-9.20*	1.30	8.71	1.84	-1.92**	1.81	1.93
	(0.75)	(8.74)	(5.18)	(0.79)	(8.57)	(7.07)	(0.83)	(8.67)	(7.92)
Constant	-8.05*	-24.2	1.36	-8.35*	-19.5	2.52	-6.81	-22.7	7.62
	(4.15)	(47.6)	(48.9)	(4.50)	(49.2)	(49.6)	(4.12)	(49.7)	(48.6)
Observations	1925	111	111	1927	111	111	1927	111	111
R-squared	0.070	0.25	0.23	0.070	0.25	0.22	0.070	0.25	0.22
Median/Mean Dependent	7.84	107.43	57.69	7.84	107.43	57.69	7.84	107.43	57.69

	Extroversion			Emotional Stability		
	(4)			(5)		
	Monthly income			*Monthly income*		
	All	*Working women*		*All*	*Working women*	
	Mean	*Mean*	*Median*	*Mean*	*Mean*	*Median*
Years of Schooling	1.36***	8.19**	6.76***	1.35***	7.55*	6.44***
	(0.43)	(4.04)	(2.27)	(0.42)	(4.11)	(2.43)
Cognitive Skills	3.26*	1.90	-4.63	3.24*	4.33	-3.40
	(1.81)	(18.1)	(10.5)	(1.80)	(18.5)	(11.2)
SEMS	-0.38	0.49	-1.82	1.24	13.8	3.41
	(0.72)	(7.28)	(8.28)	(1.34)	(11.9)	(12.8)
Constant	-7.90*	-24.8	-9.61	-8.09*	-0.76	-8.19
	(4.15)	(46.0)	(48.0)	(4.72)	(46.0)	(51.3)
Observations	1927	111	111	1927	111	111
R-squared	0.070	0.25	0.23	0.070	0.26	0.22
Median/Mean Dependent	7.84	107.43	57.69	7.84	107.43	57.69

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, and earnings for women in Pakistan. The dependent variable Monthly income is the raw monthly income for median regressions and is top coded at 100,000 PKR per month (961.5 USD) for mean regressions. The cognitive index is the mean of the Urdu, Mathematics, and English scores computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the paper test and the computer adaptive test on tablet (excluding items that less than 50 respondents answered and those that less than 5% or more than 95% of respondents got it right). The socio-emotional skills index (SEMS) is computed using principal factor analysis on the Conscientiousness sub-scale of the Big-Five and Grit items for the "Conscientiousness/Grit" columns. The SEMS index is computed using principal factor analysis on the Openness to Experience, Agreeableness, Extroversion, and Emotional stability sub-scales of the Big-Five for the respective columns. The sample for the first column for each factor is all women who answered the direct version of the questionnaire, in-person: 1,925 women. The sample for the rest of the columns is all women who answered the direct version of the questionnaire, in-person and are working: 111 women. All the regressions include age and district fixed effects. Robust standard errors are in parentheses. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.

Table A10. Relationship between schooling and skills formation in Cambodia

	(1) Cognitive Skills	(2) Cognitive Skills	(3) Socio-Emotional Skills	(4) Socio-Emotional Skills
Years of schooling	0.191***	0.183***	0.060***	0.051***
	(0.012)	(0.014)	(0.011)	(0.013)
Respondent Age	-0.097***	-0.102***	0.002	0.007
	(0.009)	(0.010)	(0.008)	(0.010)
Sex of the respondent = 1, Female	-0.280***	-0.242***	-0.106***	-0.091**
	(0.030)	(0.033)	(0.034)	(0.039)
Household head can read (2011)	0.016	0.015	0.336***	0.338***
	(0.108)	(0.127)	(0.105)	(0.122)
Household head can write (2011)	0.128	0.095	-0.269**	-0.285**
	(0.109)	(0.129)	(0.107)	(0.122)
HH SES in 2008	0.038	0.022	0.006	-0.027
	(0.024)	(0.027)	(0.021)	(0.026)
Constant	0.208	0.349	-1.739***	-0.297
	(0.199)	(0.222)	(0.183)	(0.219)
Observations	3,285	3,285	3,264	3,264
Adjusted R-squared	0.175	0.222	0.014	0.038
Sample	All	All	All	All
Village FE	No	Yes	No	Yes
Province FE	Yes	No	Yes	No
N_Clusters	423	423	423	423

* p<.10 ** p<.05 *** p<.01

Notes. The dependent variable in columns (1), and (2) is the Mathematics score, computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the computer adaptive test on tablet. The dependent variable in columns (3) and (4) is a socio-emotional skills index computed using principal factor analysis on 15 Big-Five items and 8 Grit items. The sample for this long-term follow-up is 3,294 respondents. Among them, some did not complete the computer adaptive test and some answered "Don't know" to the socio-emotional skills questions. The sample is therefore 3,285 respondents for the cognitive index, and 3,264 respondents for the socio-emotional skills index. Regressions in even columns control for village fixed effects, while regressions in odd columns control for province fixed effects. Standard errors are in parentheses (clustered at the village level).

Table A11. Relationship migration, schooling and skills in Cambodia

	All		Men		Women	
	(1) Out of village	(2) Out of village	(3) Out of village	(4) Out of village	(5) Out of village	(6) Out of village
Years of schooling	-0.002	-0.003	-0.015*	-0.014	0.005	0.004
	(0.005)	(0.005)	(0.009)	(0.010)	(0.008)	(0.008)
Cognitive Skills		0.006		-0.002		0.005
		(0.007)		(0.013)		(0.009)
Socio-Emotional Skills (SEMS)		0.002		-0.006		0.004
		(0.005)		(0.010)		(0.009)
Household head can read (2011)	-0.025	-0.026	-0.075	-0.072	0.033	0.033
	(0.032)	(0.032)	(0.052)	(0.053)	(0.053)	(0.053)
Household head can write (2011)	0.002	0.003	0.070	0.068	-0.056	-0.057
	(0.030)	(0.030)	(0.052)	(0.052)	(0.053)	(0.053)
HH SES in 2008	-0.008	-0.008	0.000	-0.000	-0.014	-0.014
	(0.009)	(0.009)	(0.016)	(0.016)	(0.012)	(0.012)
Sex of the respondent = 1, Female	0.001	0.002				
	(0.012)	(0.012)				
Constant	0.124***	0.138***	0.162**	0.136	0.199***	0.211***
	(0.035)	(0.040)	(0.073)	(0.091)	(0.053)	(0.055)
Observations	2,688	2,688	1,249	1,249	1,439	1,439
Adjusted R-squared	0.066	0.066	0.084	0.083	0.055	0.054
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All	All
N_Clusters	378	378	309	309	324	324
Mean Dependent	0.0763	0.0763	0.0817	0.0817	0.0716	0.0716

* p<.10 ** p<.05 *** p<.01

Notes. The dependent variable "Out of village" is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household. The cognitive skills index is the Mathematics score, computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the computer adaptive test on tablets. The socio-emotional skills index is computed using principal factor analysis on 15 Big-Five items and 8 Grit items. The sample for this long-term follow-up is 3,294 respondents. We only have information on where the respondent currently lives for 2,706 respondents, 1259 men, and 1,447 women. Among them, some respondents did not complete the computer adaptive test on tablets or answered "Don't know" to the socio-emotional skills questions, leading to a sample of 2,688 respondents for regressions that include measures of scores, 1,249 men and 1,439 women. All regressions control for village and age fixed effects. Standard errors are in parentheses (clustered at the village level).

Table A12. Schooling, Skills, Migration, and Earnings for Men in Cambodia

	Median Regressions				**Mean Regressions**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Monthly income	Monthly income	Monthly income	Monthly income	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)
Years of Schooling (a1)	-3.74	-1.24	-1.05	-2.33	-13.0	-15.5	-13.8	-15.5
	(2.82)	(4.52)	(4.68)	(4.59)	(10.9)	(12.3)	(12.5)	(12.7)
Cognitive Skills (a2)			-0.62	1.35			-22.5	-14.3
			(5.56)	(6.24)			(17.4)	(17.1)
SEMS (a3)			4.03	2.03			34.3**	38.1**
			(4.35)	(4.69)			(16.8)	(17.7)
Out Village (a4)				-22.2				-50.1
				(63.4)				(452.3)
Interaction YrsSchooling and Out Village (b1)				1.79				16.1
				(11.0)				(67.6)
Interaction SEMS and Out Village (b2)				1.36				-50.4
				(32.6)				(56.3)
Interaction Cog and Out Village (b3)				-17.4				-87.3
				(21.1)				(71.9)
Constant	44.4	72.2	31.9	38.7	100.7	115.6	187.1	219.1*
	(670756.7)	(658360.5)	(169809.0)	(167865.0)	(87.7)	(98.6)	(127.5)	(127.6)
Observations	1451	1173	1173	1173	1451	1173	1173	1173
Pseudo R-squared	0	0	0	0	0.0090	-0.022	-0.017	-0.016
Median/Mean Dependent	160.9	160.9	160.9	160.9	296.3	292.1	292.1	292.1
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
a1+b1=0				-0.54				0.61
				(11.18)				(65.33)
a2+b2=0				-16.03				-101.61
				(19.24)				(71.30)
a3+b3=0				3.39				-12.31
				(32.22)				(52.84)

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, migration, and earnings for men in Cambodia. The dependent variable "Monthly income" in columns (1) to (4) is the raw monthly income while the dependent variable "Monthly income (TC)" in columns (5) to (8) is top coded at 2000 USD per month. The cognitive index is the Mathematics score, computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the computer adaptive test on tablets. The socio-emotional skills index is computed using principal factor analysis on 15 Big-Five items and 8 Grit items. The "Out of village" variable is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household. The sample for columns (1), and (5) are all men in the sample who are working and are not currently enrolled, that is 1,451 men. For the rest of the columns, the sample is only those who have skills and location measures, that is 1,173 men. Robust standard errors are shown in parentheses. All the regressions include age and village fixed effects. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.

<u>Table A13</u>. Schooling, Skills, Migration, and Earnings for Women in Cambodia

	Median Regressions				Mean Regressions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Monthly income	Monthly income	Monthly income	Monthly income	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)	Monthly income (TC)
Years of Schooling (a1)	2.27	5.00***	3.65*	1.99	7.57	9.21	7.39	8.35
	(1.75)	(1.72)	(2.11)	(2.43)	(9.11)	(10.3)	(10.7)	(10.0)
Cognitive Skills (a2)			2.23	2.43			6.26	8.92
			(3.34)	(3.34)			(13.3)	(13.6)
SEMS (a3)			7.44***	7.31**			18.5	19.6
			(2.52)	(2.87)			(13.2)	(13.3)
Out Village (a4)				-32.7				26.9
				(78.6)				(371.6)
Interaction YrsSchooling and Out Village (b1)				4.02				-5.46
				(12.2)				(53.4)
Interaction SEMS and Out Village (b2)				14.1				-19.8
				(14.8)				(28.2)
Interaction Cog and Out Village (b3)				13.7				-38.0
				(24.4)				(38.7)
Constant	163.7	141.1	167.1	215.1	-243.1***	-244.5***	-223.7***	-235.6***
	(602.3)	(576.4)	(472.3)	(1248.2)	(56.8)	(60.7)	(66.5)	(65.6)
Observations	1642	1354	1354	1354	1642	1354	1354	1354
Pseudo R-squared	0	0	0	0	-0.0054	-0.051	-0.050	-0.053
Median/Mean Dependent	123.1	123.8	123.8	123.8	208.7	205.0	205.0	205.0
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
a1+b1=0				6.01				2.90
				(12.10)				(52.82)
a2+b2=0				16.16				-29.09
				(25.02)				(38.10)
a3+b3=0				21.43				-0.17
				(14.49)				(29.03)

* p<.10 ** p<.05 *** p<.01

Notes. This table reports estimates of the relationships between schooling, skills, migration, and earnings for women in Cambodia. The dependent variable "Monthly income" in columns (1) to (4) is the raw monthly income while the dependent variable "Monthly income (TC)" in columns (5) to (8) is top coded at 2000 USD per month. The cognitive index is the Mathematics score, computed using Item Response Theory with a two-parameter logistic (2PL) model. The items included in the model are those from the computer adaptive test on tablets. The socio-emotional skills index is computed using principal factor analysis on 15 Big-Five items and 8 Grit items. The "Out of village" variable is a dummy variable taking the value 1 if the respondent lives outside the village where we originally surveyed their household. The sample for columns (1), and (5) are all women in the sample who are working and are not currently enrolled, that is 1,642 women. For the rest of the columns, the sample is only those who have skills and location measures, that is 1,354 women. Robust standard errors are shown in parentheses. All the regressions include age and village fixed effects. The R-squared shown is the pseudo R-squared for median regressions and the adjusted R-squared for mean regressions.