Intelligence, Personality, or Work Ethic? The Labour Market Returns to Cognitive and Noncognitive Skills in Latin America*

Peter Leighton†

April 5, 2018

Preliminary : Do not cite

Abstract

This paper estimates the returns to cognitive and noncognitive skills across 10 Latin American countries. Results from an exploratory factor analysis reveal that a range of noncognitive skill measures can be explained by two underlying latent factors, which I interpret as "work ethic" and "strength of personality". I estimate standard Mincerian earnings regressions, which provide strong evidence that both cognitive and noncognitive skills are important determinants of hourly earnings. A one standard deviation increase in cognitive skills is associated with a 9.6% increase in hourly earnings, whereas an increase in the two latent noncognitive skills is associated with a 6.8% and 5.5% return respectively. Specifically, antagonism (the reverse of agreeableness) is important only for women, whereas neuroticsm is important only for men. Noncognitive skills also predict significant increases in labour force participation. One of the distinguishing features of the labour market in developing countries is the size of the informal sector. I show that there is significant self selection of individuals with high cognitive skills away from the informal sector and into the formal sector. However there is no strong evidence of a difference in returns between sectors. This finding runs contrary to theories of a segmented labour market.

---

*I would like to thank Orazio Attanasio, Sarah Cattan, Pramila Krishnan, and seminar participants at the IFS and CSAE, Oxford for helpful comments. I would also like to thank Pablo Sanguinetti and the development bank of Latin America for kindly sharing their data. A version of this paper was submitted as an extended essay for the MSc in Economics for Development at Oxford university.

†Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE. Email: peter_leighton@ifs.org.uk

1 Introduction

It has long been recognised that education and cognitive skills are important determinants of wages and labour market success (Psacharopoulos, 1985; Hanushek and Woessmann, 2012). However, a growing literature has moved beyond this uni-dimensional view of human capital to focus on a wide range of noncognitive abilities; including personality traits, social skills, and motivation. Common practices such as the interviewing of job applicants are indicative of the importance of these traits, and a number of studies have found a strong empirical relationship between noncognitve skills and labour market outcomes (Bowles, Gintis and Osborne, 2001; Heckman, Stixrud and Urzua, 2006; Lindqvist and Vestman, 2011). Research in this area is especially pertinent given the rising labour market value of social skills observed in recent decades (Deming, 2017); with the rising automation of jobs, human comparative advantage is increasingly in the ability to interact.

Despite these recent advances, there is still very little known about the return to cognitive and noncognitive skills in developing countries. The small number of current studies have found large returns to cognitive skills, yet mixed evidence on the importance of noncognitive skills in determining wages (Glewwe, Huang and Park, 2017; Muller, Sarzosa and Acosta, 2015; Valerio et al., 2016). However, findings from recent interventions have been more promising. Positive labour market returns have been achieved from improving the noncognitive skills of different groups; ranging from garment workers (Adhvaryu, Kala and Nyshadham, 2016), to adolescents (Krishnan and Krutikova, 2013) and micro-enterprise owners (Glaub et al., 2014). This malleability of noncognitive skills, even into adulthood and for the most marginalised in society, offers large potential for policies targeting these skills.

In this paper, I contribute to this new literature by estimating the labour market returns to cognitive and noncognitive skills across 10 lower and middle income countries (LMIC's) in Latin America. To better understand the nature of skills and simplify the analysis, I first conduct an exploratory factor analysis. This reveals that cognitive skills can be summarised by one latent factor, and noncognitive skills by two; which I interpret as "work ethic" and "strength of personality" respectively. Estimates from Mincerian earnings regressions show that both cognitive and noncognitive skills have a large and statistically significant impact on hourly earnings. A one standard deviation increase in cognitive skills is associated with a 9.6% increase in hourly earnings, whereas an increase in the two noncognitive factors is associated with a 6.8% and 5.5% return respectively. These results are robust to including different fixed effects or correcting for sample selection. When looking instead at individual skill measures; numerical ability, anatagonism (the opposite of agreeableness), conscientiousness, and risk tolerance are particularly important. Contrary to previous research, I also find evidence that cognitive skills and certain noncognitive skills, namely the work ethic factor, are substitutes and not complements in the labour market. There are also some significant differences by gender, with antagonism being rewarded only for women, and neuroticsm only for men.

One of the main differences between the labour market in developed countries and lower income countries is the degree of informality. In Latin America, an estimated 46.8% of all jobs are in the informal sector (International Labour Organization, 2014). There exists two competing views of the informal labour market in developing countries. The first follows from the classic dualistic models of Lewis (1954) and Harris and Todaro (1970). Under this theory, the rationing of formal sector jobs, for example through minimum wages and other regulations, leads to a segmented labour market in which some individuals are excluded from the formal sector. The competing view holds that competitive markets do indeed exist, and instead workers rationally select into each sector based on comparative advantage, as in Roy (1951). In the absence of barriers to the formal sector, workers in the lower paid sector would enter the higher paid sector until wages are equalised. Therefore one implication of the segmentation hypothesis is that observably identical workers will be paid differently across sectors, something that can be tested empirically. The evidence on this has been largely mixed with early studies seemingly suggesting strong evidence of segmentation (Heckman and Hotz, 1986; Gindling, 1991), but others arguing that the competitive model is more applicable (Pratap and Quintin, 2006; Magnac, 1991). However, in estimating differences in returns, previous research has only looked at traditional measures of human capital such as years of education, and not noncognitive skills. This study therefore makes a contribution to this literature, through incorporating latent cognitive and noncognitive skills into the standard tests of segmentation.

I first estimate the employment decision of workers through a multinomial logit model. This reveals that high cognitive skill workers strongly select into the formal sector and away from the informal sector. In contrast, those with greater work ethic are significantly more likely to be in work, but there is no differential sorting by formality. Other factors such as being single and higher socio-economic status are also significant predictors of working in the formal sector. I then estimate the earnings equations separately across sectors. In order to correct for the selection of workers into sectors

I adopt a multinomial sample selection correction from Bourguignon, Fournier and Gurgand (2007). There is some evidence that the returns to cognitive skills and work ethic are greater in the formal sector, and the returns to strength of personality greater in the informal sector, however these differences are not statistically significant and largely disappear when controlling for important characteristics of the informal sector; such as the size of the firm. Therefore I cannot reject the hypothesis that observably identical workers receive the same compensation for their skills across sectors. This result combined with the selection of worker based on skills and other characteristics, supports the theory of competitiveness and not segmentation of the informal sector.

The rest of the paper proceeds as follows. Section 2 describes the data being used and how it is constructed, Section 3 estimates the total labour market return to skills, before Section 4 analyses the difference between sectors. Finally, Section 5 concludes.

2 Data

2.1 Skill measures

I use data from a survey conducted by the Corporacion Andina de Fomento - Banco de Desarrollo de América Latina (CAF) (the development Bank of Latin America), in 2015. This is an individual level survey consisting of a representative sample of 9,634 individuals aged 15-55 from the capital cities of Argentina, Bolivia, Brazil, Colombia, Ecuador, Mexico, Panama, Peru, Uruguay, and Venezuela. All results are estimated using the relevant sample weights to ensure that the estimates are representative of the populations being studied.

Cognitive Skills: The survey contains three separate measures of cognitive skills along three domains; fluid intelligence, verbal skills, and numerical skills. The first test is the widely used Ravens progressive matrices (Raven, 1936), a non verbal measure of fluid intelligence. This survey uses a short form of the test which has been used in other studies across Latin America. The second measure is a short verbal conceptualization test, designed to test a respondent's ability to abstract, generalise, and find relationships between verbal concepts. This test is made up of a selection of 6 items from the "Anologies" sub scale of the Wechsler Adult Intelligence II (Wechsler and Hsiao-pin, 2011). The final measure of cognitive ability is an index of basic numeracy skills. This includes a test of counting backwards from 20, and 3 questions involving simple mathematical problems.

Noncognitive Skills: The survey includes a wide range of measures of personality traits, preferences, and psychological well being, that have been commonly used in the previous literature. The first measure is the ten-item personality inventory (TIPI) from Gosling, Rentfrow and Swann (2003), this is a very short measure of the Big Five (or Five Factor Model) dimensions. These factors are 5 broad domains that psychologists believe define human personality; openness to experience, conscientiousness, extraversion, agreeableness (or it's inverse antagonism), and emotional stability (or it's inverse neuroticsm).[1] The second measure is the self-efficacy scale from Schwarzer and Jerusalem (1995). This is a 10 item scale designed to assess optimistic self-beliefs; the extent to which respondents believe that their actions are responsible for successful outcomes. The third measure is the 8 item Grit scale (Duckworth et al., 2007). This is a measure of the respondent's passion and perseverance for achieving long-term goals. There are also measures of risk preferences and mental health. A measure of risk tolerance is obtained through asking for a number of choices between a job with a secure payment and jobs with a higher expected value but with some degree of uncertainty. A screening test for depressive disorder, the CES-D scale (Radloff, 1977) is also collected. The simplest approach to scoring each measure would be to calculate an arithmetic mean of the scores on each question. However this may not be the most efficient way of discriminating between individuals of different latent ability. In order to address this problem, I proceed by scoring each individual skill measure using Item Response Theory (IRT).[2] All test scores are subsequently standardised to have a mean of zero and a standard deviation of one.

2.2 Reducing dimensionality

In order to better understand the structure of the data and simplify the analysis, it is useful to reduce the dimensionality of the skill measures. To achieve this, I conduct exploratory factor analysis (EFA). The relevant scales, scored by IRT where possible, are included in a seperate EFA for cognitive and noncognitive skills.[3] I do not consider either depression nor risk tolerance as a noncognitive skill, due to both their conceptual differences and low correlation with other measures.[4] The three main criteria from the

[1]See Table A1 for a description of these traits

[2]For the binary cognitive measures, a two-parameter logistic model from Birnbaum (1968) is fitted. As the noncognitive skills are all measured on a Likert scale, an IRT rating scale model for ordinal items from Andrich (1978) is fitted. Scores are then calculated as empirical Bayes means of the latent variables. The big 5 are not scored using IRT due to there only being two items per trait.

[3]I include IRT scores for each measure in the EFA rather than the full set of individual questions, largely for ease of interpretation of the factors. However results are largely unchanged when instead conducting the EFA using the full set of noncognitive skill questions

[4]See Appendix Table A2

literature are used for selecting the number of factors to extract,[5] and the estimated factor loadings are then rotated using the oblique rotation, oblimin. This method is preferred to methods that assume orthogonility, as the different latent factors are very likely to be correlated. To estimate the final factors, I create a score using the regression approach, which is then standardised to have a mean of zero and standard deviation of one.

The results of this exploratory factor analysis are reported in Table 1. The analysis suggests extracting one factor for cognitive ability, in keeping with the concept of g theory (Jensen, 1998). For noncognitive ability, the analysis instead suggests extracting two factors. The main loadings on the first factor are: conscientiousness, openness to experience, self-efficacy, and grit. I interpret this factor as "work ethic", comprising traits that determine an individuals effort on a task, such as self belief, determination, and vigilance. The second factor is composed mainly of neuroticsm, antagonism, and extraversion. I interpret this as "strength of personality", containing elements of sociability, selfishness, and self-consciousness that shape an individuals interactions with others. Note that I do not use "strength" in any normative sense, but instead to denote that an individual with such traits is likely to be highly noticeable. Strength of personality actually has a strong negative correlation with work ethic, and a small negative correlation with cognitive skills (see Appendix Table A3). Work ethic and cognitive skills, on the other hand, show a fairly strong positive correlation. As a result, whereas cognitive skills and work ethic are expected to have a labour market return, ex-ante it is not clear whether strength of personality is a trait that will be rewarded or penalised in the labour market.

Table 1: Rotated factor loadings

	Factor1	Factor2
Conscientiousness	**0.43**	-0.17
Emotional Stability	0.16	**-0.42**
Openness to experience	**0.51**	0.07
Extraversion	0.25	**0.33**
Antagonism	-0.03	**0.43**
Self efficacy (IRT)	**0.39**	-0.03
Grit (IRT)	**0.46**	-0.06

Noncognitive skills

	Factor1
Verbal (IRT)	0.41
Numerical (IRT)	0.42
Ravens (IRT)	0.29

Cognitive skills

[5]These are ; Horns Parallel Analysis (Horn, 1965), Minimum Average Partial Correlation (Velicer, 1976), and the Kaiser Criteria (Kaiser, 1961)

2.3 Other variables

The survey also asks for a range of employment outcomes. Reported earnings are converted into US$ using the same exchange rate as in CAF (2016), and divided by total hours worked to give an estimate of hourly earnings. I follow CAF (2016), by defining an individual as working in the formal sector if they report that their employer makes contributions to social security or a pension fund, and in the informal sector otherwise. The economic sector that an individual works in is also reported and defined under three categrories; manufacturing and construction, commerce, and services. Experience is calculated as the difference between the date of the interview and the year that an individual entered their first job. Cognitive and noncognitive skill requirements for each job category are based on O*NET classifications, as reported in CAF (2016, p.44)

2.4 Summary statistics

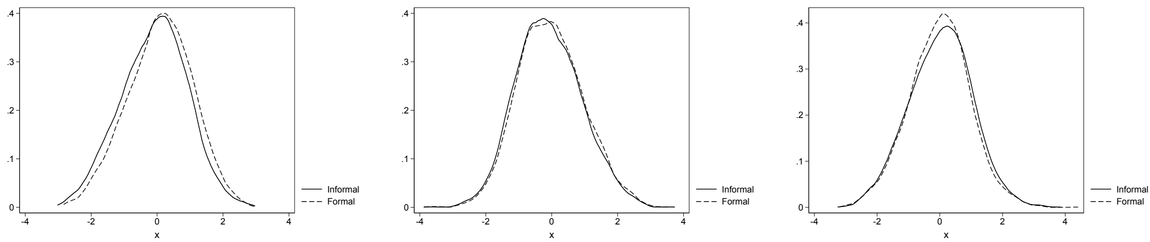
Table 2 presents means of some of the key variables in the CAF data, both for the sample as a whole and separately for those in the formal and informal sector. It can be seen that individuals in the sample are relatively wealthy, with mean monthly earnings of $592. However this masks considerable heterogeneity, men earn 40% more than women, those in the formal sector earn 42% more than those in the informal sector, and average earnings in the richest country in the sample, Venezuela, are nearly 6 times that of the poorest, Bolivia. In total 55% of the sample are in work, with half of these in the informal sector. Women are much less likely to be in the labour market, but the unemployment rate is similar between males and females at around 10%. Most people work in either commerce or services, with informal jobs particularly concentrated in commerce. Formal jobs are also different along a number of other domains; formal firms have on average three and a half times the number of employees, and are more likely to require the use of a computer or team work. Around 20% of individuals work in jobs demanding high cognitive skills and 40% work in jobs requiring high noncognitive skills, with formal jobs twice as likely to have a high cognitive skill requirement. Individuals are relatively well-educated with around 60% completing secondary and 14% also completing further education. Figures 1 and 2 plot the distribution of skills by sector and gender. This reveals that those in the informal sector have greater cognitive skills throughout the whole distribution. The differences in noncognitive skills are smaller, however the informally employed tend to have higher strength of personality and lower work ethic. Men have consistently higher cognitive skills throughout the distribution, however women seem to have slightly higher levels of both noncognitive skills.

Table 2: Means of key variables

	Full sample		By formality		
	All	N	Formal	Informal	p-value
Earnings and Employment Status					
Monthly Earnings (USD)	592.26	3572	687.99	485.60	0.01***
Hourly Earnings	3.26	3512	3.67	2.80	0.01***
Working Hours per Week	46.47	5086	46.79	46.09	0.42
Employed	0.55	9251	1.00	1.00	0.00***
Active	0.62	9251	1.00	1.00	0.00***
Self Employed	0.26	9251	0.26	0.72	0.00***
Job Characteristics					
Manufacturing and construction	0.21	4340	0.20	0.21	0.59
Commerce	0.38	4340	0.28	0.51	0.00***
Services	0.41	4340	0.52	0.28	0.00***
Number of employees	17.65	4295	28.21	7.97	0.00***
Use a computer at work	0.74	9457	0.62	0.44	0.00***
Job involves team work	0.86	9444	0.84	0.67	0.00***
High Cognitive Skill Job	0.20	4672	0.26	0.13	0.00***
High Non Cognitive Skill Job	0.40	4672	0.38	0.43	0.08*
Recieved on the job training	0.32	2664	0.36	0.18	0.00***
Time in Current Job	7.91	3409	8.09	7.68	0.64
Experience	15.39	4851	16.22	15.35	0.25
Time in Current Job	7.91	3409	8.09	7.68	0.64
Demographics					
Completed Secondary School	0.60	9476	0.75	0.56	0.00***
Complete College/University	0.14	9476	0.25	0.13	0.00***
Mother Completed Secondary School	0.35	8827	0.37	0.27	0.05**
Mother Completed University	0.09	8827	0.11	0.07	0.03**
Age	33.42	9496	36.12	35.26	0.15
Male	0.48	9496	0.60	0.56	0.11
Married	0.34	9475	0.40	0.35	0.06*
Has Children	0.59	9450	0.65	0.65	1.00
Has a Child under 5	0.39	5479	0.35	0.36	0.90
N	9497		2749	2732	

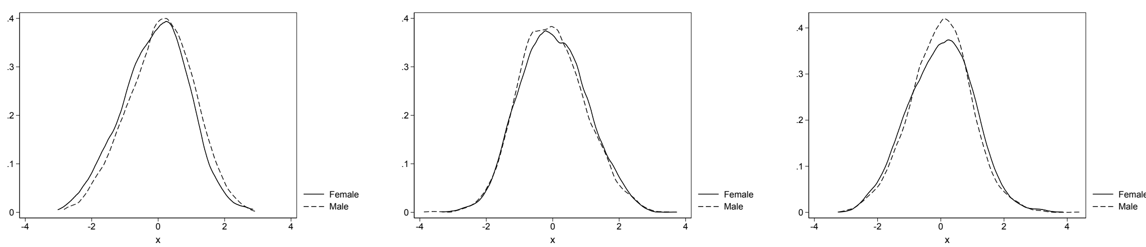
Note: P-values from a test of difference in means between columns (3) and (4), calculated following the wild bootstrap procedure of Cameron, Gelbach and Miller (2008)

Figure 1: Distribution of skills by formality



(a) Cognitive (b) Work Ethic (c) Strength of personality

Figure 2: Distribution of skills by gender



(a) Cognitive (b) Work Ethic (c) Strength of personality

In order to investigate the main determinants of cognitive and noncognitive skills, Table 3 reports OLS regressions of the relationship between each skill and a number of key variables. Firstly in columns (1), (4) and (7), each skill is regressed on just indicators for an individuals level of education. This reveals that education is indeed significantly correlated with all three factors. However, education levels alone still only explain a very small portion of the variance in skills; 9% of cognitive skills, 2% of work ethic, and essentially none of the variation in strength of personality. This confirms the importance of looking at human capital as a multi-dimensional construct, comprising of more than just education. Columns (2), (5) and (8) add in various demographic variables. This reveals that across multiple domains, there is greater inequality in cognitive skills than noncognitive skill. Socio-economic status, as proxied by parental education level, is significantly correlated with cognitive skills but not noncognitive skills. Conditional on other variables, men have significantly higher cognitive skills yet actually slightly lower levels of both noncognitive skills. Columns (3), (6) and (9) estimate the relationship between various job characteristics and skills. Both cognitive skills and work ethic are significantly greater in the formal sector and in jobs requiring greater use of the respective skill. However, strength of personality is seemingly unrelated to job characteristics and the R^2 in column (9) is still only 2%. Therefore noncognitive skills, and strength of personality in particular, are largely unexplained by observable factors.

Table 3: Correlates of skill factors

	Cognitive Factor			Work Ethic			Strength of Personality		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education									
Completed Secondary	0.4855***	0.4219***	0.2348**	0.1583***	0.1509***	0.1497**	0.0340	0.0224	0.0049
School	(0.0944)	(0.0948)	(0.0954)	(0.0425)	(0.0203)	(0.0642)	(0.0447)	(0.0222)	(0.0469)
Complete	0.3215***	0.2865***	0.1228	0.2920***	0.2645***	0.2073***	-0.0972**	-0.0736	-0.0007
College/University	(0.0619)	(0.0472)	(0.0866)	(0.0382)	(0.0408)	(0.0502)	(0.0391)	(0.0440)	(0.0688)
Demographics									
Mother Completed		0.1532***	0.1120		0.0252	-0.0440		-0.0035	0.0559
Secondary School		(0.0270)	(0.0709)		(0.0547)	(0.0581)		(0.0416)	(0.0483)
Mother Completed		0.1480*	0.1354**		0.0000	-0.0823		0.0291	0.0865
University		(0.0805)	(0.0590)		(0.0537)	(0.0702)		(0.0703)	(0.0811)
Age		-0.0086	0.0719***		0.0208	0.0322		-0.0080	-0.0313
		(0.0108)	(0.0144)		(0.0122)	(0.0187)		(0.0100)	(0.0240)
Age Squared		0.0001	-0.0010***		-0.0001	-0.0003		-0.0001	0.0003
		(0.0001)	(0.0002)		(0.0002)	(0.0002)		(0.0001)	(0.0003)
Male		0.1410**	0.0311		-0.0339	-0.0227		-0.0426	-0.0811
		(0.0434)	(0.0888)		(0.0293)	(0.0472)		(0.0470)	(0.0656)
Job Characteristics									
Received on the job			0.0129			0.1117			-0.0335
training			(0.0533)			(0.0614)			(0.0853)
Time in Current Job			0.0079**			0.0012			-0.0040
			(0.0026)			(0.0021)			(0.0031)
High Cognitive Skill			0.1258*			-0.0222			0.0710
Job			(0.0626)			(0.0752)			(0.0821)
High Non Cognitive			0.0408*			0.1026*			-0.0305
Skill Job			(0.0209)			(0.0482)			(0.0360)
Formal sector			0.2336*			0.1809**			-0.0684
			(0.1158)			(0.0624)			(0.0533)
Observations	9476	8816	1623	9279	8653	1608	9279	8653	1608
R^2	0.088	0.101	0.087	0.022	0.043	0.048	0.001	0.020	0.020

* $p < 0.05$, ** $p < 0.01$. P-values calculated using Wild bootstrap procedure with 100 replications. Standard errors clustered at the country level in parenthesis.

3 The return to skills

3.1 Empirical strategy

In the baseline specification, I estimate the returns to different skills through augmented Mincerian earnings regressions (Mincer, 1974) of the following form:

$$lnw_{i,c} = \beta_0 + C'_{i,c}\beta + N'_{i,c}\pi + \gamma x_{i,c} + \delta_1 e_{i,c} + \delta_2 e^2_{i,c} + \lambda_c + \mu_{i,c} \tag{1}$$

Where $w_{i,c}$ is the hourly earnings in US$ for individual i in country c, $C_{i,c}$ is a vector of cognitive skills, $N_{i,c}$ is a vector of noncognitive skills, $e_{i,c}$ is experience, and $x_{i,c}$ is gender. λ_c is a country fixed effect for each country, therefore controlling for country specific labour market conditions. As is common in the literature, I drop the top 1% of income earners from the analysis. To avoid the confounding effects of education, I also only include those aged 25 and older, not in full-time education, and working full-time.[6] To account for potential correlation in the error term for individuals within the same country, I cluster standard errors at the country level. However as shown by Cameron, Gelbach and Miller (2008), when there are less than 42 clusters, conventional cluster robust standard errors can perform poorly. To address this, I report p values calculated from the wild bootstrap procedure outlined in Cameron, Gelbach and Miller (2008).

The coefficients estimated from equation 1 represent the partial effect of each skill conditional on all other observed skills. However there are reasons to prefer a more parsimonious specification. Firstly, different skills could themselves be outcomes of each other, leading to a "bad control" problem (Angrist and Pischke, 2008). Secondly, these partial effects may not be the most policy relevant parameters. Interventions are more likely to impact broader latent traits than any one narrowly defined trait alone. Therefore, I also estimate equation 1 using latent factor scores instead of a vector of skill measures. As these factor scores are not observed variables but instead "generated regressors", there is some additional sample variability in their estimation. This is something that is not taken into account by conventional or even cluster robust standard errors. In order to correct for this, the standard errors are calculated through bootstrapping the whole procedure (see Gensowski (2014)).

However, there remain a number of reasons why the estimation of equation 1 may not yield consistent estimates of the true returns to skills.

Firstly, there may be many factors, such as parental characteristics, that are correlated with both the skills that individuals acquire as well as their ultimate labour market success. This would result in omitted variables bias. In additional specifications I address this through including a set of controls for maternal education level. However, this clearly does not fully exhaust all potential omitted variables. Therefore, I further explore the robustness of the results to these concerns using the procedure of Oster (2016), in Section 3.2.

Secondly, there is the related concern of reverse causality. For example, working in better paid sectors, such as services, could lead to greater skill accumulation. To address this concern, additional specifications also include sector fixed effects or a dummy for formality, therefore estimating the returns to skills within sector. This of course does not fully solve the problem; there could still be some reverse relationship between

[6]Defined as working more than 15 hours a week. Results are robust to varying definitions.

skills and earnings within sector. However the fact that noncognitive skills are not significantly related to many observables, including experience in the current job and on-the-job training (see Table 3), provides at least some evidence that skills are largely predetermined.

Thirdly, as hourly earnings are only observed for those in employment, there could be a sample selection problem. To address this, results are reported from a standard Heckman selection model (Heckman, 1979). The method of maximum likelihood is used, and marriage status and the presence of young children are used as excluded instruments.

Finally, within the psychology literature there has been well documented bias from the use of factor scores within OLS regression. Although the Bartlett method is often preferred when estimating factor scores, Devlieger, Mayer and Rosseel (2016) highlight that this leads to biased estimates of the true coefficients. Therefore, I use the regression method in order obtain unbiased estimates of the returns to latent factors.

3.1.1 Earnings returns

Table 4 reports various estimates of equation 1, with log hourly earnings as the dependent variable. Column (1) reports results from the basic OLS specification. This shows that both cognitive and noncognitive skills have a significant return; a one standard deviation increase in the cognitive factor predicts 9.18 log point higher hourly earnings, compared to 6.55 log points for work ethic factor and 5.4 log points for the strength of personality factor. Column (2) presents results from a Heckman selection model, correcting for potential problems of selection into the sample of income earners. The estimated returns are actually slightly higher, and the coefficient on the inverse Mills ratio is not statistically significant. This suggests that the initial OLS results are not driven by selection bias. Columns (3) and (4) add in fixed effects for the sector of employment and formality respectively. This only leads to relatively small changes in the estimated coefficients. The similarity of returns within sector means that the results in column (1) are unlikely to be purely driven by differential sorting or skill acquisition across sectors. Similarly as seen in column (5), results remain largely unchanged when adding in additional controls for individuals socio-economic background, as proxied by parental education levels. Finally, the last column includes quadratic terms and the interactions between different skills. This reveals that hourly earnings are linear in cognitive ability and strength of personality but convex in work ethic. The interaction term between cognitive skills and work ethic is negative and statistically significant, suggesting that certain noncognitive skills and cognitive skills could be substitutes and not complements. This finding is the opposite to what has been found in the context of developed countries (for example Cunha et al. (2006); Lindqvist and Vestman (2011)).

Table 4: Estimated effect of latent skills on log hourly earnings

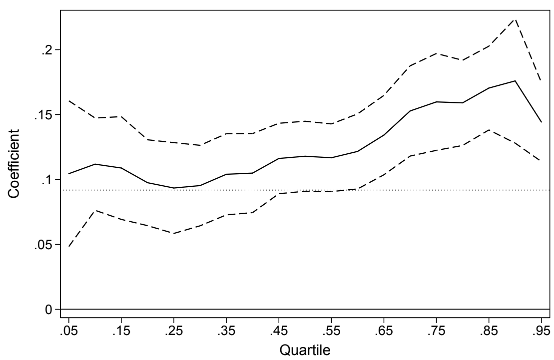
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.0918***	0.0998***	0.0872***	0.0786***	0.0801***	0.102***
	(0.0205)	(0.0211)	(0.0278)	(0.0205)	(0.0230)	(0.0205)
Work ethic	0.0655**	0.0734**	0.0615*	0.0622**	0.0523**	0.0541**
	(0.0264)	(0.0352)	(0.0331)	(0.0253)	(0.0205)	(0.0262)
Strength of personality	0.0537**	0.0581**	0.0476	0.0537**	0.0379	0.0509**
	(0.0245)	(0.0269)	(0.0428)	(0.0232)	(0.0284)	(0.0257)
Cognitive sq.						0.00558
						(0.0139)
Work ethic sq.						0.0402**
						(0.0180)
Strength of personality sq.						0.00881
						(0.0190)
Cognitive * Work ethic						-0.0604**
						(0.0297)
Cognitive * Strength of personality						0.000566
						(0.0257)
Work ethic * Strength of personality						0.0250
						(0.0288)
Selection Correction	No	Yes	No	No	No	No
Sector Fixed Effects	No	No	Yes	No	No	No
Formality Control	No	No	No	Yes	No	No
Additional Controls	No	No	No	No	Yes	No
Observations	2070	3450	1761	2070	1953	2070
R^2	0.366		0.398	0.401	0.400	0.373

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors computed from bootstrapping of whole estimation procedure using 50 replications and reported in parentheses.

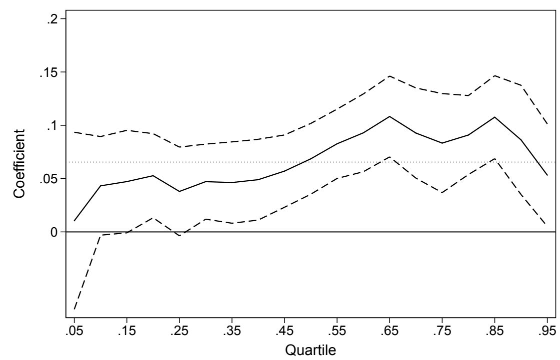
Note: All specifications include controls for gender, experience, experience squared and country fixed effects . Heckman selection model estimated using maximum likelihood using marriage status and the presence of young children (aged under 5) as excluded instruments. Additional controls include dummy variables for mothers education level.

The OLS results provide estimated returns at the mean of the hourly earnings distribution, however there could be interesting differences at different points throughout the distribution. To investigate this, I estimate conditional quantile regressions of the specification from column (1), with the results being displayed in Figure 3. It can be seen that the returns to all three factors are generally increasing over the conditional earnings distribution, with the greatest returns between the 85th and 90th percentiles for all three factors. This implies that cognitive and noncognitive skills are particularly valuable within higher earning jobs. Nevertheless, the returns remain pretty consistent throughout the earnings distributions. Returns are statistically significant at all quantiles for cognitive skills, above the 25th percentile for work ethic, and above the 40th percentile for strength of personality.

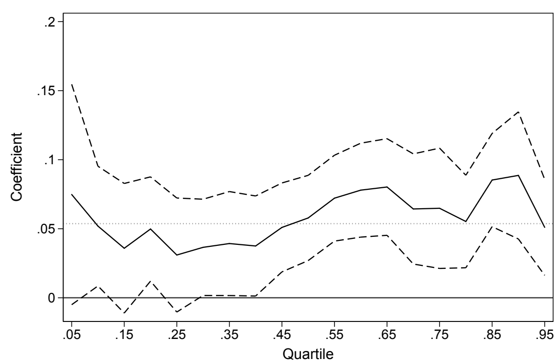
Figure 3: Quantile Regressions:



(a) Cognitive Skills



(b) Work Ethic



(c) Strength of Personality

Note: Coefficients plotted from quantile regressions between the 5th and 95th percentiles, at 5th percentile increments. Specification include all three latent skill factors, and controls for gender, experience, experience squared, and country fixed effects. 95% confidence intervals shown in dashed lines. Dotted line represents OLS coefficient from column (1) of Table 4.

Table 5 reports results for the analogous specifications to Table 4, but including all individual skill measures instead of latent factor scores. This shows that the most highly valued cognitive skill is numerical ability, which predicts an increase in hourly earnings of between 10.6 and 13 log points. For noncognitive skills the most important skills are antagonism, conscientiousness, and risk tolerance. This perhaps surprising relationship between antagonism and earnings has been previously found in a variety of contexts (Nyhus and Pons, 2005; Mueller and Plug, 2006; Valerio et al., 2016)). There are a number of potential explanations for the economic return to strength of personality, and antagonism in particular. There could be benefits to a "Machievellian" personality; comprising the ability to manipulate others for one's own gain. This trait has been shown to be highly related to anatagonism (Paulhus and Williams, 2002), and also socio-economic outcomes (Turner and Martinez, 1977). Additionally, more agreeable people may be willing to accept lower wages in any wage bargainning process.

Table 5: Estimated effect of individual skills on log hourly earnings

	(1)	(2)	(3)
Cognitive Skills			
Raven	0.0159	0.0039	0.0185
	(0.0354)	(0.0325)	(0.0290)
Verbal	-0.0011	-0.0000	0.0008
	(0.0120)	(0.0157)	(0.0105)
Numerical	0.1273***	0.1295***	0.1064***
	(0.0287)	(0.0261)	(0.0271)
The Big Five			
Antagonism	0.0335*	0.0349***	0.0266*
	(0.0148)	(0.0088)	(0.0140)
Conscientousness	0.0410	0.0375	0.0374
	(0.0228)	(0.0255)	(0.0209)
Emotional Stability	-0.0259	-0.0186	-0.0283
	(0.0256)	(0.0144)	(0.0260)
Openness to Experience	0.0266	0.0278	0.0379
	(0.0263)	(0.0253)	(0.0283)
Extraversion	-0.0117	-0.0181	-0.0109
	(0.0153)	(0.0162)	(0.0142)
Other Traits			
Self efficacy	0.0171	0.0163	0.0110
	(0.0237)	(0.0216)	(0.0188)
Depression	0.0158	0.0257	0.0162
	(0.0249)	(0.0246)	(0.0257)
Grit	-0.0059	-0.0016	-0.0171
	(0.0238)	(0.0150)	(0.0180)
Risk Tolerance	0.0710***	0.0787***	0.0727***
	(0.0179)	(0.0212)	(0.0185)
Sector Fixed Effects	No	Yes	No
Formality Control	No	No	Yes
Observations	2055	1749	2055
R^2	0.383	0.417	0.417

* $p < 0.05$, ** $p < 0.01$. P-values calculated using Wild bootstrap procedure with 100 replications. Standard errors clustered at the country level in parenthesis.

Note: Also controls for gender, experience, experience squared and country fixed effects .

3.1.2 Wage Returns by gender

Table 6 estimates the returns to skills by gender. Although the returns to cognitive skills and work ethic are very similar for men and women, the returns to strength of personality are twice as high for men. However due to large standard errors none of these differences are statistically significant at conventional levels. Table 7 estimates the earnings equations when instead including all individual skills. This reveals some interesting differences by gender. From comparison of columns (1) and (4) it can be seen that antagonism is strongly rewarded for women and not for men, with the difference being statistically significant. This striking difference remains when including different fixed effects, however due to increased standard errors the differences are no longer statistically significant. The opposite result is found with respect to emotional stability; there is a small reward for emotional stability for women, however there is a strong reward for it's inverse, neuroticsm, for men, and this difference is statistically significant across all specifications. The negative effect of agreeableness amongst women has been shown in previous studies (for example Nyhus and Pons (2005)). However, Mueller and Plug (2006) find the opposite effects of neuroticsm in Germany, with a reward for emotional stability for men and not women.

Table 6: Estimated effect of latent skills on log hourly earnings: by gender

	Female			Male		
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	0.0867**	0.102**	0.0775**	0.0971***	0.0798***	0.0815***
	(0.0346)	(0.0428)	(0.0351)	(0.0255)	(0.0294)	(0.0260)
Work ethic	0.0712	0.0750	0.0572	0.0626	0.0523	0.0655*
	(0.0762)	(0.0681)	(0.0785)	(0.0393)	(0.0359)	(0.0392)
Strength of personality	0.0313	0.0266	0.0278	0.0654	0.0584	0.0689
	(0.0714)	(0.0719)	(0.0751)	(0.0481)	(0.0496)	(0.0509)
Sector Fixed Effects	No	Yes	No	No	Yes	No
Formality control	No	No	Yes	No	No	Yes
Observations	793	676	793	1277	1085	1277
R^2	0.358	0.399	0.406	0.350	0.382	0.378

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors computed from bootstrapping of whole estimation procedure using 50 replications and reported in parentheses. † means that coefficients are statistically significantly different between males and females at the 10% level

Note: Also controls for gender, experience, experience squared and country fixed effects .

Table 7: Estimated effect of individual skills on log hourly earnings : by gender

	Female			Male		
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive Skills						
Raven	-0.0038	-0.0066	0.0051	0.0269	0.0097	0.0268
	(0.0446)	(0.0358)	(0.0347)	(0.0354)	(0.0382)	(0.0311)
Verbal	0.0236	0.0397	0.0264	-0.0139	-0.0200	-0.0138
	(0.0308)	(0.0390)	(0.0283)	(0.0163)	(0.0236)	(0.0165)
Numerical	0.0927**	0.0917*	0.0743*	0.1503**	0.1541**	0.1296*
	(0.0347)	(0.0424)	(0.0349)	(0.0577)	(0.0496)	(0.0576)
The Big Five						
Antagonism	0.0665***†	0.0900**	0.0468***	0.0069	-0.0041	0.0073
	(0.0180)	(0.0353)	(0.0142)	(0.0209)	(0.0180)	(0.0202)
Conscientousness	0.0508**	0.0306	0.0401	0.0433	0.0529**	0.0430*
	(0.0195)	(0.0395)	(0.0238)	(0.0279)	(0.0203)	(0.0229)
Emotional Stability	0.0187†	0.0540**†	0.0046†	-0.0594*	-0.0700**	-0.0573
	(0.0311)	(0.0219)	(0.0240)	(0.0322)	(0.0253)	(0.0349)
Openness to Experience	0.0426*	0.0481*†	0.0570**	0.0118	0.0110	0.0218
	(0.0215)	(0.0235)	(0.0233)	(0.0342)	(0.0275)	(0.0364)
Extraversion	-0.0208	-0.0330	-0.0271	-0.0041	-0.0066	-0.0005
	(0.0260)	(0.0295)	(0.0305)	(0.0095)	(0.0136)	(0.0078)
Other Traits						
Self efficacy	0.0187	0.0187	0.0097	0.0211	0.0208	0.0166
	(0.0208)	(0.0167)	(0.0184)	(0.0287)	(0.0278)	(0.0278)
Depression	-0.0102	-0.0105†	-0.0076	0.0312	0.0479*	0.0318
	(0.0225)	(0.0258)	(0.0176)	(0.0287)	(0.0242)	(0.0309)
Grit	-0.0111	0.0140	-0.0281	0.0000	-0.0086	-0.0080
	(0.0314)	(0.0244)	(0.0223)	(0.0297)	(0.0220)	(0.0269)
Risk Tolerance	0.0650*	0.0694	0.0717*	0.0727***	0.0781***	0.0719***
	(0.0301)	(0.0415)	(0.0355)	(0.0202)	(0.0176)	(0.0194)
Sector Fixed Effects	No	Yes	No	No	Yes	No
Formality Control	No	No	Yes	No	No	Yes
Observations	790	673	790	1265	1076	1265
R^2	0.378	0.425	0.425	0.376	0.416	0.400

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. P-values calculated using Wild bootstrap procedure with 100 replications. Standard errors clustered at the country level in parenthesis. † means that coefficients are statistically significantly different between malea and females at the 10% level

Note: Same as Table 6

3.2 Robustness checks

There are still likely to be many omitted variables that are correlated with both an individuals skill level and their ultimate labour market outcomes, leading to inconsistency in the estimates. To investigate the likely extent of these concerns, I follow the procedure of Oster (2016) to calculate the degree of selection on unobservables that would be required to eliminate the estimated returns. The results of this exercise are reported in Table 8. I find that the selection on unobservables would have to be between roughly 1 and 5 times larger than the selection on observables to eliminate the observed effect on cognitive skills, between 1.5 and 7 times larger to eliminate the effect on work ethic, and between 3 and 13 times larger to eliminate the effect on strength of personality. Given that the observables include a country fixed effect, it seems unlikely that unobservables would have as a great an effect as observables, let alone many orders of magnitude greater. Therefore although omitted variables are still a concern, it seems unlikely that they would fully explain the size of the returns found.

Table 8: Robustness to omitted variables bias

	Max R-squared					
Variable	0.5	0.6	0.7	0.8	0.9	1
Cognitive	4.87	2.89	2.05	1.59	1.30	1.10
Work ethic	-6.67	-3.87	-2.73	-2.11	-1.71	-1.45
Strength of personality	-12.58	-7.27	-5.12	-3.95	-3.21	-2.71

Note: Results from Oster (2016) procedure, for a variety of assumptions on the maximal R^2. Coefficients are δ estimates; the degree of selection on unobservables (relative to the observed degree of selection on observables) that would be required in order for the true coefficients from the estimation of equation 1 to be zero, given the parameter estimates reported in Table 4.

I also conducted a number of additional robustness checks (available on request). Firstly, I estimated equation 1 using total monthly earnings instead of hourly earnings. This leads to coefficients on the two noncognitive skills that are slightly higher and a coefficient on cognitive skills that is slightly lower. This suggests that those with higher noncognitive skills work more hours, whereas those with higher cognitive skills work slightly shorter hours. Secondly, I also estimate all results using a simple index, the average of all standardised measures for each skill. This gives estimated returns that are very similar, but the use of factor scores is preferred due to the greater ease in interpretation. Thirdly, using simple aggregation instead of IRT leads to similar estimates. However, the greater precision in estimates means that the IRT specification is preferred. Fourthly, returns to the whole sample of income earners, not excluding those in education, below 25, and not in full-time work, are actually substantially higher. Therefore the results are not sensitive to these assumptions on who to exclude. Finally, there are also some concerns about the existence of missing data on earnings for those in employment. To address this, I impute missing values for hourly earnings using the joint multivariate normal multiple imputation method (MVNI) (see Schafer, 1997).[7] The results are robust to this procedure, with the regression coefficients actually being slightly higher for all three factors.

4 Skills and the informal sector

4.1 Empirical strategy

Having established that both cognitive and noncognitive skills are important determinants of labour market success, I now turn to estimating how the returns to these skills differ across the informal and formal sectors. To fix ideas, I consider a simple Roy model (Roy, 1951).

An individual has three possible choices; to not be in employment, to be in formal employment, or to be in informal employment. The earnings process in each working sector $k \in \{F, I\}$ is given by:

$$lnW_i^k = \alpha_0^k + \alpha_1^k \theta_i^c + \alpha_2^k \theta_i^n + X'\alpha_2^k + f^k(e_i) + \mu_i^k \tag{2}$$

Where W_i^k represents the hourly earnings for individual i in sector k, θ_i^c and θ_i^n are latent cognitive and noncognitive skills respectively, X is a vector of other characteristics, $f^k(e_i)$ is some function of experience, e_i, and μ_i^k is a mean zero i.i.d error term. Individuals have utility defined over expected hourly earnings $E(W_i^k)$, some other utility from each sector $\eta^k(z)$ which depends on a set of covariates z, and an additional random utility component ϵ_i^k. Hence utility in the formal and informal sector is given by:

$$U^k = U(E(W_i^k), \eta_i^k(z), \epsilon_i^k) \tag{3}$$

And utility when not in employment (and hence when there are no earnings) is given by:

$$U^N = U(\eta_i^k(z), \epsilon_i^k) \tag{4}$$

[7]Estimation uses 10 imputations. Variables used for imputation are education level, mothers education level, country, and all three skill factors

This is a simple static model, whereby at the beginning of each period workers choose the option with the highest utility. Hence their employment decision in any one period is:

$$J_i = \begin{cases} F & \text{if} \quad U(E(W_i^F), \eta_i^F(z), \epsilon_i^F) > max\{U(E(W_i^I), \eta_i^I(z), \epsilon_i^I), U(\eta_i^N(z), \epsilon_i^N)\}, \\ I & \text{if} \quad U(E(W_i^I), \eta_i^I(z), \epsilon_i^I) > max\{U(E(W_i^F), \eta_i^F(z), \epsilon_i^F), U(\eta_i^N(z), \epsilon_i^N)\}, \\ N & \text{if} \quad U(\eta_i^N(z), \epsilon_i^N) \geq max\{U(E(W_i^F), \eta_i^F(z), \epsilon_i^F), U(E(W_i^I), \eta_i^I(z), \epsilon_i^I)\}. \end{cases}$$

I assume that utility functions $U(.)$ and $\eta_i(.)$ have a simple additive structure, and that ϵ_i^k is i.i.d with a Type 1 Extreme Value Distribution. This means that an individuals employment decision J_i can be estimated using a multinomial logit model.[8]

However, the above model highlights a concern about the estimation of equation 2 through standard OLS. Wages are only observed for an individual in the sector in which they choose to work. This mean that what is actually observed in the formal sector for example is given by:

$$\begin{aligned} &E(W^F|U^F{>}max\{U^N, U^I\}) = \\ &\quad \alpha_0^F + \alpha_1^F\theta^c + \alpha_2^F\theta^n + X'\alpha_2^F + f^F(e) + E(\mu^F|U^F{>}max\{U^N, U^I\}) \end{aligned} \tag{5}$$

This leads to selection bias arising from the term $E(\mu^k|U^F{>}max\{U^N, U^I\})$. If μ_i and ϵ_i are not independent then OLS estimates of the coefficients of interest; $\underline{\alpha^k} = (\alpha_1^k, \alpha_2^k)$ will be inconsistent. However, in a similar fashion to Heckman (1979), consistent estimates can still be obtained through estimating this selection bias term and then including it in estimates of equation 2. To achieve this, I use the method proposed by Bourguignon, Fournier and Gurgand (2007). This requires a vector of variables, that determine whether an individual is selected into the sample or not, J_i, but not their value of the outcome of interest; W_i. It can be seen that the vector z satisfies this condition; it shifts the utility of different employment options without directly affecting hourly earnings. I take z to consist of marriage status and whether an individual has young children. All these factors will likely shift preferences for different jobs, and in the absence of discrimination, I assume that they do not directly enter into the earnings equation.

[8]The results are largely unchanged when estimated using a multinomial probit model. Hence estimates are not particularly sensitive to these assumptions.

4.2 Results

4.2.1 Employment decision

Table 9 reports results from a multinomial logit model of individuals employment choices. A one standard deviation increase in cognitive skills predicts a 5 percentage point increase in the probability of being employed in the formal sector, and a 3 percentage point decrease in the probability of being employed in the informal sector. Work ethic is particularly important for labour force particpation; a one standard deviation increase in work ethic is associated with a reduction in the probability of being not in work of around 5 percentage points. Greater work ethic also predicts greater employment in both sectors, but has a larger effect on the probability of being employed in the informal sector. There is no significant correlation between strength of personality and an individuals employment choice. These results are largely similar to what has been found in previous studies. The association between cognitive skills and participation in the formal sector has been displayed in Colombia (Muller, Sarzosa and Acosta, 2015), and the importance of noncognitive skills in reducing unemployment has been shown in previous studies in both developed and developing countries (Lindqvist and Vestman, 2011; Muller, Sarzosa and Acosta, 2015). The instruments that are used for the selection model are also significantly related to employment decisions, for example being married predicts a 7 percentage point reduction in the probability of being employed in the informal sector. These results suggest that there is significant sorting into the formal sector based primarily on cognitive ability, and to a lesser extent noncognitive skills. Multinomial logit estimates including all measures are reported in Appendix Table A4. As previously, most of the effect of cognitive skills is coming from numerical skills. For noncognitive skills, the strongest drivers of formal sector participation are antagonism and grit, with extraversion and agreeableness being most important for informal sector participation.

Table 9: Multinomial logit estimates of the employment decision

	(1) Not in work	(2) Informal sector	(3) Formal sector
Cognitive	-0.0175**	-0.0365***	0.0540***
	(0.00755)	(0.00808)	(0.00801)
Work ethic	-0.0565***	0.0253*	0.0311**
	(0.0128)	(0.0141)	(0.0136)
Strength of personality	-0.0121	0.00610	0.00603
	(0.0133)	(0.0139)	(0.0128)
Selection instruments			
Married	0.0602***	-0.0659***	0.00568
	(0.0160)	(0.0188)	(0.0210)
Has a Child under 5	0.0558***	-0.0331*	-0.0227
	(0.0181)	(0.0183)	(0.0184)
Observations	4395	4395	4395

Note: Mean marginal effects. Also controls for gender, experience, experience squared, and country fixed effects. Unemployed defined as not being in any form of paid work.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors computed from bootstrapping of whole estimation procedure using 50 replications and reported in parentheses.

4.2.2 Returns by sector

Panel A of Table 10 estimates the skill returns in both the formal and informal sectors, by OLS. Columns (1) and (2) estimate the same specifications as equation 1, whereas columns (3) and (4) add in additional controls for some of the main differences between the formal and informal sector such as the number of employees. There are no significant differences in the returns to skills between the two sectors, although the returns to work ethic are over twice as large in the first specification, and the returns to cognitive skills over twice as large once including the larger set of controls. In addition, a Chow test can not reject the null hypothesis of the equality of the earnings equation between sectors. Panel B displays results correcting for potential sample selection bias. Similarly, there are some differences in the point estimates; the returns to cognitive skills and work ethic are quite a bit higher in the formal sector, whereas the returns to strength of personality are higher in the informal sector. This suggests that the type of skills valued in the informal sector could be different, however none of these differences are statistically significant. In addition, once including the larger set of controls, the point estimates for cognitive skills and work ethic are extremely similar, although the returns to strength of personality are substantially higher in the informal sector. Therefore any differences by formality might be explained by the type of work, and not anything specific about the informal sector. Although the differences aren't statistically significant, across all specifications the returns to experience are greater in the formal sector, and the additional earnings from being male are greater in the informal sector. Overall there is no strong evidence that people of the same latent ability receive different hourly earnings in the formal and informal sector; running contrary to theories of a segmented labour market. However, the lack of precision in estimates combined with some interesting differences in the coefficients means that more research in this area is required.

Table 10: Estimates of the earnings equation by sector

	(1) Informal	(2) Formal	(3) Informal	(4) Formal
Panel A: OLS				
Cognitive	0.0729**	0.0889***	0.0375	0.0870**
	(0.0332)	(0.0284)	(0.0329)	(0.0423)
Work ethic	0.0350	0.0816	0.0424	0.0326
	(0.219)	(0.0638)	(0.136)	(0.0782)
Strength of personality	0.0588	0.0528	0.0475	0.0710
	(0.211)	(0.0643)	(0.124)	(0.0886)
Male	0.293***	0.164***	0.267***	0.160**
	(0.0528)	(0.0418)	(0.0554)	(0.0670)
Experience	0.0147	0.0248***	0.0127	0.0343***
	(0.0135)	(0.00763)	(0.0110)	(0.0108)
Panel B: Selection model				
Cognitive	0.0177	0.0660	0.0573	0.0594
	(0.165)	(0.0916)	(0.146)	(0.0987)
Work ethic	0.0285	0.0728	0.0340	0.0446
	(0.0835)	(0.0577)	(0.0619)	(0.0468)
Strength of personality	0.0736	0.0424	0.0517	-0.00124
	(0.0714)	(0.0475)	(0.0734)	(0.0527)
Male	0.338*	0.127	0.311***	0.218***
	(0.190)	(0.113)	(0.114)	(0.0701)
Experience	0.0107	0.0243	0.00384	0.0174
	(0.0445)	(0.0233)	(0.0332)	(0.0183)
Additional Controls	No	No	Yes	Yes
Observations	864	1206	654	781

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors computed from bootstrapping of whole estimation procedure using 50 replications and reported in parentheses. † means that the difference in coefficients between the formal and informal sectors is statistically significant at the 10% level.

Note: Dependent variable : log hourly earnings ($USD). Multinomial selection correction following Bourguignon, Fournier and Gurgand (2007), excluded instruments are: marriage status and the presence of young children (aged under 5). All specifications also include controls for experience squared, and country fixed effects. Additional controls include: number of employees, sector fixed effects, use a computer at work, job involves team work.

5 Conclusion

This paper has shown that both cognitive and noncognitive skills are important determinants of labour market outcomes in 10 Latin American countries, both in the formal and informal sectors. Due to the malleability of noncognitive skills found in other studies, these findings have important policy implications; improving noncognitive skills can increase earnings and help individuals into employment. The suggestive evidence of substitutability between cognitive and noncognitive skills also implies large potential for policy to tackle the inequality in cognitive skills arising from highly divergent schooling outcomes, through a focus on noncognitive skills. There are however a number of limitations to these results. Firstly, issues of omitted variables bias and reverse causality could still lead to inconsistency in the estimates. Secondly, all individuals in this sample are relatively rich and from major urban cities, hence it is unclear how much these results apply to poorer, rural contexts. The increasing recognition of noncognitive skills in economics has broadened our understanding of human capital and the determinants of labour market success. However, this new literature still lacks a consistent definition or unified framework for understanding noncognitive skills. The two factor structure proposed in this paper presents one potential way for making progress in this area.

References

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2016. "Soft Skills to Pay the Bills: Evidence from Female Garment Workers." *mimeo, University of Michigan.*

American Psychological Association. 2007. *APA Dictionary of Psychology.* Vol. xvi, Washington, DC, US:American Psychological Association.

Andrich, David. 1978. "Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers." *Applied Psychological Measurement*, 2(4): 581–594.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Birnbaum, Allan. 1968. "Some latent trait models and their use in inferring an examinee's ability." *Statistical theories of mental test scores.*

Bourguignon, Franois, Martin Fournier, and Marc Gurgand. 2007. "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons." *Journal of Economic Surveys*, 21(1): 174–205.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "Incentive-Enhancing Preferences: Personality, Behavior, and Earnings." *American Economic Review*, 91(2): 155–158.

CAF, Development Bank of Latin America. 2016. "More skills for work and life: the contributions of the family, school, environment, and labor market."

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the evidence on life cycle skill formation." *Handbook of the Economics of Education*, 1: 697–812.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics*, 132(4): 1593–1640.

Devlieger, Ines, Axel Mayer, and Yves Rosseel. 2016. "Hypothesis testing using factor score regression: A comparison of four methods." *Educational and Psychological Measurement*, 76(5): 741–770.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology*, 92(6): 1087.

Gensowski, Miriam. 2014. "Personality, IQ, and lifetime earnings."

Gindling, T. H. 1991. "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal, and Informal Sectors in San Jos, Costa Rica." *Economic Development and Cultural Change*, 39(3): 585–605.

Glaub, Matthias E., Michael Frese, Sebastian Fischer, and Maria Hoppe. 2014. "Increasing Personal Initiative in Small Business Managers or Owners Leads to Entrepreneurial Success: A Theory-Based Controlled Randomized Field Intervention for Evidence-BasedManagement." *Academy of Management Learning & Education*, 13(3): 354–379.

Glewwe, Paul, Qiuqiong Huang, and Albert Park. 2017. "Cognitive skills, noncognitive skills, and school-to-work transitions in rural China." *Journal of Economic Behavior & Organization*, 134: 141–164.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. "A very brief measure of the Big-Five personality domains." *Journal of Research in personality*, 37(6): 504–528.

Hanushek, Eric A., and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth*, 17(4): 267–321.

Harris, John R., and Michael P. Todaro. 1970. "Migration, unemployment and development: a two-sector analysis." *The American economic review*, 60(1): 126–142.

Heckman, James, and V. Joseph Hotz. 1986. "An Investigation of the Labor Market Earnings of Panamanian Males Evaluating the Sources of Inequality." *Journal of Human Resources*, 21(4): 507–542.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1): 153–161.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411–482.

Horn, John L. 1965. "A rationale and test for the number of factors in factor analysis." *Psychometrika*, 30(2): 179–185.

International Labour Organization, (ILO). 2014. "Thematic Labour Overview 1: Transition to Formality in Latin America and the Caribbean (Revised version)."

Jensen, Arthur Robert. 1998. "The g factor: The science of mental ability."

Kaiser, Henry F. 1961. "A Note on Guttman's Lower Bound for the Number of Common Factors1." *British Journal of Statistical Psychology*, 14(1): 1–2.

Krishnan, Pramila, and Sofya Krutikova. 2013. "Non-cognitive skill formation in poor neighbourhoods of urban India." *Labour Economics*, 24: 68–85.

Lewis, W. Arthur. 1954. "Economic development with unlimited supplies of labour." *The manchester school*, 22(2): 139–191.

Lindqvist, Erik, and Roine Vestman. 2011. "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment." *American Economic Journal: Applied Economics*, 3(1): 101–128.

Magnac, Th. 1991. "Segmented or Competitive Labor Markets." *Econometrica*, 59(1): 165–187.

Mincer, Jacob A. 1974. "Schooling, Experience, and Earnings." *NBER*.

Mueller, Gerrit, and Erik Plug. 2006. "Estimating the Effect of Personality on Male and Female Earnings." *Industrial and Labor Relations Review*, 60(1): 3–22.

Muller, Noel, Miguel Alonso Sarzosa, and Pablo Ariel Acosta. 2015. "Beyond qualifications : returns to cognitive and socio-emotional skills in Colombia." *Policy Research Working Paper;No. 7430. World Bank, Washington, DC.*, 1–49.

Nyhus, Ellen K., and Empar Pons. 2005. "The effects of personality on earnings." *Journal of Economic Psychology*, 26(3): 363–384.

Oster, Emily. 2016. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 0–0.

Paulhus, Delroy L, and Kevin M Williams. 2002. "The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy." *Journal of Research in Personality*, 36(6): 556–563.

Pratap, Sangeeta, and Erwan Quintin. 2006. "Are labor markets segmented in developing countries? A semiparametric approach." *European Economic Review*, 50(7): 1817–1841.

Psacharopoulos, George. 1985. "Returns to Education: A Further International Update and Implications." *The Journal of Human Resources*, 20(4): 583–604.

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

Raven, James C. 1936. "Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive." *Unpublished masters thesis, University of London.*

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146.

Schafer, Joseph L. 1997. *Analysis of incomplete multivariate data.* CRC press.

Schwarzer, Ralf, and Matthias Jerusalem. 1995. "Optimistic self-beliefs as a resource factor in coping with stress." In *Extreme stress and communities: Impact and intervention.* 159–177. Springer.

Turner, Charles F., and Daniel C. Martinez. 1977. "Socioeconomic Achievement and the Machiavellian Personality." *Sociometry*, 40(4): 325–336.

Valerio, Alexandria, Maria Laura Sanchez Puerta, Namrata Tognatta, and Sebastian Monroy-Taborda. 2016. "Are there skills payoffs in low-and middle-income countries? empirical evidence using STEP data." *Policy Research working paper; no. WPS 7879; WDR 2018 background paper. Washington, D.C. : World Bank Group.*

Velicer, Wayne F. 1976. "Determining the number of components from the matrix of partial correlations." *Psychometrika*, 41(3): 321–327.

Wechsler, David, and Chou Hsiao-pin. 2011. *WASI-II: Wechsler abbreviated scale of intelligence.* Pearson.

Appendix

Appendix A: Skill measures

Table A1: Definitions from American Psychological Association (2007)

Trait	**Definition**
Openness to Experience	The tendency to be open to new aesthetic,cultural, or intellectual experiences.
Conscientiousness	The tendency to be organized, responsible, and hardworking.
Extraversion	An orientation of one's own interest and energies towards the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.
Agreeabless	The tendency to act in a cooperative, unselfish manner.
Emotional Stability*	Predictability and consistency in emotional reactions, with absence of rapid mood changes.

Note: * The inverse of emotional stability is referred to as neuroticsm, and is used in some presentations of the big 5.

Table A2: Correlation Coefficients

	Ra.	Nu.	Ve.	Co.	Em.	Op.	Ex.	An.	Se.	Gr.	De.	Ri.
Raven	1.00											
Numerical	0.17	1.00										
Verbal	0.14	0.25	1.00									
Conscientousness	-0.00	0.06	0.06	1.00								
Emotional Stability	0.02	0.06	0.05	0.25	1.00							
Openness to Experience	0.07	0.13	0.12	0.25	0.14	1.00						
Extraversion	0.04	0.05	0.03	-0.03	-0.11	0.13	1.00					
Antagonism	0.05	0.06	0.03	-0.20	-0.30	-0.08	0.12	1.00				
Self efficacy	-0.01	0.08	0.06	0.22	0.14	0.23	0.06	-0.09	1.00			
Grit	0.11	0.12	0.12	0.19	0.14	0.18	0.01	-0.05	0.02	1.00		
Depression	-0.08	-0.21	-0.14	-0.15	-0.20	-0.14	-0.04	0.09	-0.12	-0.22	1.00	
Risk Tolerance	0.00	-0.00	-0.05	-0.03	-0.03	-0.00	0.01	0.04	0.03	-0.05	0.04	1.00

Table A3: Correlation of latent skill factors

	Cognitive	Work ethic	Strength of personality
Cognitive	1		
Work ethic	0.170***	1	
Strength of personality	-0.0358***	-0.639***	1

5.1 Appendix B: Additional tables

Table A4: Multinomial logit estimates of the employment decision

	(1) Not in work	(2) Formal sector	(3) Informal sector
Cognitive Skills			
Raven	-0.0128*	0.0212***	-0.00837
	(0.00692)	(0.00784)	(0.00672)
Verbal	0.00185	0.0175**	-0.0193*
	(0.00856)	(0.00729)	(0.0104)
Numerical	-0.0350***	0.0541***	-0.0191*
	(0.0113)	(0.0116)	(0.00989)
The Big Five			
Antagonism	0.00123	0.0161**	-0.0173***
	(0.00732)	(0.00628)	(0.00614)
Conscientousness	-0.0130	0.0151	-0.00216
	(0.00885)	(0.00928)	(0.00640)
Emotional Stability	0.00381	0.00162	-0.00543
	(0.00562)	(0.00650)	(0.00687)
Openness to Experience	-0.0170**	0.00595	0.0110
	(0.00670)	(0.00767)	(0.00775)
Extraversion	-0.0170***	0.00534	0.0116***
	(0.00398)	(0.00585)	(0.00419)
Other Traits			
Self efficacy	-0.00408	-0.000907	0.00499
	(0.0135)	(0.00729)	(0.00986)
Depression	0.0120*	-0.0225***	0.0105*
	(0.00644)	(0.00778)	(0.00570)
Grit	-0.0141	0.0143**	-0.000184
	(0.00863)	(0.00687)	(0.00726)
Risk Tolerance	0.0114	-0.0104	-0.000997
	(0.00795)	(0.00801)	(0.00837)
Observations	5613	5613	5613

Note: Mean marginal effects. Also controls for gender, experience, experience squared, and country fixed effects

* $p < 0.05$, ** $p < 0.01$. P-values calculated using Wild bootstrap procedure with 100 replications. Standard errors clustered at the country level in parenthesis.