On-line appendix (not for publication)

Data appendix: Construction of non-cognitive skills measures

The objective of this section is to provide an overview of the construction of the primary non-cognitive indices employed in the main analysis. More specifically, there are three objectives: we seek to describe the internalizing and externalizing indices in the context of the broader psychological literature; describe the item response theory methodology used to construct the indices; and present some evidence of psychometric validity. Some of this discussion draws on Glewwe et al. (2017), a paper that analyzes the same non-cognitive measures in the Gansu survey data.

First, as noted in the paper, the objective of the internalizing index is to capture challenges that are primarily intrapersonal: anxiety, depression and withdrawal. The objective of the externalizing index is to capture challenges that are primarily extrapersonal: destructive behavior, impulsivity, aggression, or hyper-activity (Achenbach and Edelbrock, 1978). Broadly speaking, both measures capture dimensions of emotional health and instability. The psychology literature argues that internalizing challenges reflect environmental factors that limit a child's sense of control around her environment. Externalizing challenges, by contrast, reflect environments or shocks that limit self-regulatory behavior, or adult role models that show limited self-regulation (Achenbach and Edelbrock, 2006; Chorpita and Barlow, 1998; Evans, 2004). As Glewwe et al. (2017) note in their discussion of these measures in the GSCF context, both the internalizing and externalizing scales have high internal consistency (Cronbach alpha $> .8$), and have been validated across a range of contexts (Ivanova et al., 2005), as well as in China itself (Chen et al., 2002).

To measure internalizing and externalizing behaviors, children were asked to respond to a series of statements; in each case, they stated whether they fully agreed, agreed, disagreed, or fully disagreed with the statement. Thus the non-cognitive indices are based entirely on direct reports from the sampled children, rather than indirect reports from parents or other adults. (The full list of statements used in the surveys is provided at the conclusion of this subsection.) Using these scores, internalizing and externalizing item response theory (IRT) scores were calculated using the rating scale model (Andrich, 1978).

While item response theory is a commonly used tool to analyze cognitive or achievement tests, it is also increasingly widely used in psychological and psychosocial assessments. Compared to traditional methods of psychological measurement that seek to characterize a respondent's average level of some latent characteristics, IRT seeks to estimate the probability that an individual's response to an item will fall into a particular category. Relative to more traditional methods, IRT methods also have advantages in enabling the use of shorter scales, and enabling different instruments to be equated (Embretson and Reise, 2000).[1] Other useful overviews of the relevance of IRT in psychological research are provided in Reise et al. (2005) and Edelen and Reeve (2007); the latter also includes an example focusing on adolescent psychological health.

In order to implement the IRT analysis, we estimate a rating scale model (Andrich, 1978), a sub-case of the partial credit model.[2] This model can be described as follows. Given i items to

[1] A detailed overview of applications of IRT in personality and attitude assessment can be found in Chapter 12 in this volume. The authors emphasize that IRT models can be applied to assess the psychometric properties of personality scales, and can also used to address some substantive questions. These include the effect of cultural differences in item functioning, and the assessment of examinee person-fit: i.e., to what degree a subject's item response pattern for a particular instrument is consistent with the pattern predicted by a normative model of response.

[2] More specifically, we use the command "glamm" in Stata, designed to estimate parameters of dichotomous and ordinal item response models.

construct the index and ordered response categories $0, 1...m_i$ for item i, the partial credit model specifies the probability of responding in the jth category of item i for person n as a function of the person's underlying latent characteristic θ_n and step parameters δ_{ij} $(j > 0)$. (In this case, the underlying latent characteristics are internalizing and externalizing challenges.) The probability function can be written as follows.

$$Pr(x_{im} = j|\theta_n) = \frac{exp \sum_{l=0}^{j}(\theta_n - \delta_{il})}{\sum_{k=0}^{m_i} exp \sum_{l=0}^{k}(\theta_n - \delta_{il})} \tag{1}$$

where $\sum(\theta_n - \delta_{il}) = 0$. The parameter δ_{ij} is interpreted as the step difficulty associated with category j of item i.

Intuitively, the probability of a response that reflects an underlying characteristic (e.g., internalizing challenges) increases with the respondent's intensity of that characteristic, and decreases with the item's difficulty. (In this context, difficulty can be interpreted as follows: a more difficult question will only elicit a report of internalizing challenges from a respondent who experiences a high intensity of those challenges.)

The rating scale model is a sub-case in which it is assumed that the respondent categories have the same meanings for all items, and that the differences in the step difficulty are the same for all items. (In this analysis, given that the categories are relatively simple subcategories of strong agreement or disagreement, these seem to be plausible assumptions.) More details about the relevant functional forms are provided in Zheng and Rabe-Hesketh (2007).

We can also provide some evidence of the psychometric validity of the indices constructed. We report in Table 2 of the main paper correlations between non-cognitive characteristics and subjects' other characteristics (gender, age, sibling parity, height-for-age and cognitive skills). To summarize this evidence, there is no evidence of any significant correlation between the internalizing index and any child demographic characteristic. However, for the externalizing index we observe that non-cognitive measures are lower, suggestive of more behavioral challenges, for second-born children, younger children, boys and children enrolled in lower grades in school. With respect to human capital characteristics, there are weakly significant and negative correlations between the Chinese and mathematics test scores and the internalizing index, and a weakly significant and positive correlation between height-for-age and the externalizing index.

We also report in Table A1 in this appendix additional evidence around how these correlations vary with maternal education. In general, there is little evidence that these correlations significantly differ for children of more or less educated mothers. The only exceptions are the positive interaction terms between maternal education and age and maternal education and parity, significant at the ten percent level in the multivariate specification employing the internalizing index as a dependent variable (Column (9) of Panel A). We regard this evidence as consistent with the psychometric validity of the indices employed.

Statements employed In the GSCF survey, the following statements are used to measure internalizing challenges.

1. I don't want others to meddle in my own business.
2. I can't concentrate on what I am doing.
3. I have many strange/weird ideas (often daydream).

4. I easily get flushed. (I am easily frustrated or anxious.)
5. I can't do things well when my parents are not present. (I usually need help from adults to do something well.)
6. I am very indifferent to others.
7. I am very shy.
8. I am often teased by classmates.
9. It bothers me if others do things better than I do.
10. My temper changes quickly and easily.
11. I feel inferior to others.
12. I often am suspicious of others.
13. I prefer to be alone.
14. I often feel nervous.
15. I am often bored.
16. I stay quiet when I am with my classmates or friends.
17. There is always something to worry about.

The following statements are used to measure externalizing challenges

1. I quarrel with others.
2. I lose my temper.
3. I like to brag.
4. I like to show off my strengths in front of others.
5. I steal things from others or my home.
6. I always want to be the center of attention.
7. I break things on purpose.
8. I do not observe school discipline.
9. I do not feel guilty, even if I have done something wrong.
10. Even if I know I am wrong, I am reluctant to listen to others.
11. I act impulsively.
12. I often am suspicious of others.

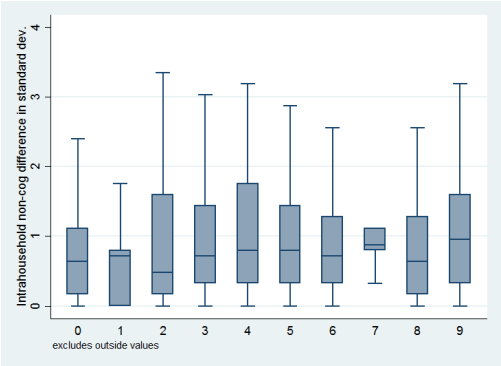
13. I often say obscenities.
14. I often make fun of others.
15. I sometimes tell lies.
16. I am easily angered.
17. I often disregard other people's ideas.
18. I sometimes menace and even hurt others.

Figures and Tables

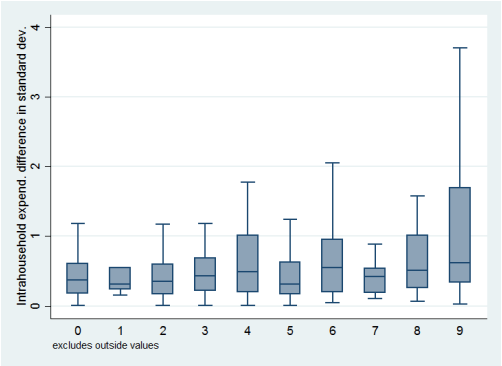
Table of Contents

Figure A1: Intra-household differences in non-cognitive characteristics and expenditure by maternal education

(a) Non-cognitive characteristics

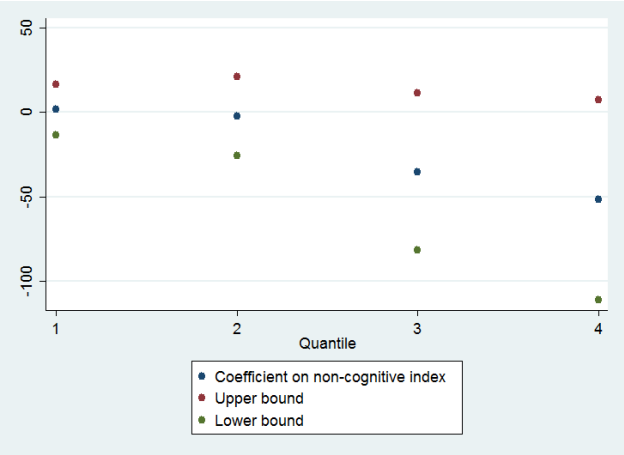


(b) Expenditure



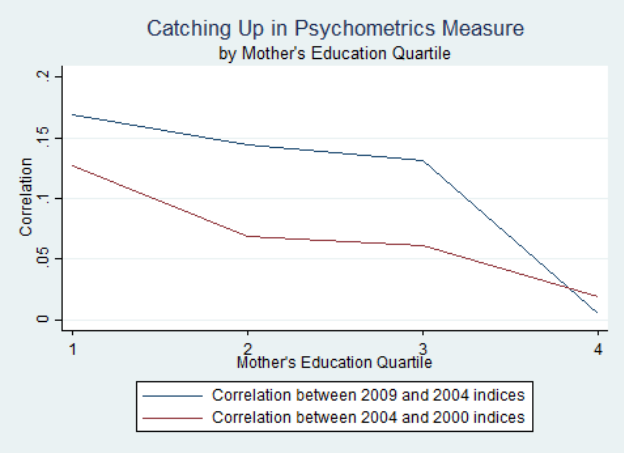
Notes: For each level of maternal education in years, the bottom bar corresponds to the minimum value of between-sibling absolute differences in non-cognitive characteristics (Figure A1a) or between-sibling absolute differences in normalized educational expenditure residuals (Figure A1b), while the top bar corresponds to the maximum value. The rectangle corresponds to the interquartile range, with the median value represented by the bold line bisecting the rectangle. Normalized expenditure residuals are calculated by regressing expenditure on child characteristics (gender, birth parity, cognitive skills, and height-for-age), generating the residuals and standardizing them to have mean zero and standard deviation one.

Figure A2: Parental response to non-cognitive characteristics by quantile of maternal education



Notes: This graph shows the estimated coefficients in regressions of total educational expenditure on the non-cognitive indices of interest, including the full set of child-level controls included in the main specification as well as household fixed effects, estimated by quantile of maternal education. The upper and lower bounds correspond to the 95% confidence interval for the estimated coefficients.

Figure A3: Correlation in non-cognitive characteristics over time by quantile of maternal education



Notes: This graph shows the estimated correlation between the summary non-cognitive indices as measured in 2000 and 2004, and the non-cognitive index in 2004 and self-esteem as measured in 2009, by quantile of maternal education.

Table A1: Non-cognitive characteristics and child characteristics

	Internal (1)	External (2)	Internal (3)	External (4)	Internal (5)	External (6)	Internal (7)	External (8)	Internal (9)	External (10)
Panel A: Child characteristics										
Sibling parity	-.089	-.170							-.097	-.137
	(.095)	(.122)							(.234)	(.300)
Parity int.	.020	.011							.062*	.070
	(.015)	(.020)							(.037)	(.043)
Age			.019	.037					-.070	-.125
			(.028)	(.033)					(.086)	(.085)
Age int.			-.001	.002					.026*	.019
			(.005)	(.005)					(.016)	(.016)
Female					-.047	.231			-.087	.185
					(.127)	(.148)			(.133)	(.148)
Female int.					.018	.023			.031	.032
					(.026)	(.025)			(.028)	(.026)
Grade level							.042	.091**	.089	.170***
							(.034)	(.037)	(.059)	(.062)
Grade int.							-.005	-.004	-.016	-.008
							(.005)	(.005)	(.011)	(.012)
Panel B: Cognitive skills and health										
Height-for-age	.055	.091					.035	.087		
	(.053)	(.057)					(.054)	(.060)		
Height int.	-.003	-.004					.001	-.002		
	(.008)	(.009)					(.009)	(.009)		
Chinese test score			-.012	-.0006			-.013*	.002		
			(.008)	(.008)			(.007)	(.008)		
Chinese int.			.00008	-.001			.0009	-.001		
			(.002)	(.001)			(.002)	(.001)		
Math test score					-.002	-.002	.003	-.003		
					(.007)	(.007)	(.007)	(.007)		
Math int.					-.001	-.0009	-.002	-.0006		
					(.001)	(.0009)	(.001)	(.0009)		
Obs.	816	816	816	816	816	816	816	816	816	816

Notes: The dependent variables are the internalizing and externalizing indices. A higher internalizing or externalizing index is indicative of fewer non-cognitive challenges. The independent variable is the specified child characteristic and the child characteristic interacted with maternal education, all measured in the second wave; all specifications include household fixed effects and standard errors clustered at the village level. Asterisks indicate significance at the ten, five, and one percent level.

Table A2: Within-household and cross-household variation

	Internal		External		Height-for-age		Achievement score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fixed effects	Household	Year	Household	Year	Household	Year	Household	Year
Obs.	816	816	816	816	816	816	816	816
R-squared	.59	.021	.591	.019	.611	.06	.532	.053

Notes: Each column reports the R-squared for the regression of the specified child characteristic on the specified set of fixed effects.

Table A3: Heterogeneity with respect to maternal education: Adding control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Discretionary								
Internalizing index	15.754	15.464*	15.333	20.752**	19.312**	21.122**	21.552**	20.499**
	(10.601)	(9.317)	(9.434)	(9.368)	(9.326)	(9.320)	(9.268)	(8.885)
Index x mother educ.	-7.512**	-8.003***	-7.968***	-8.650***	-7.505***	-8.165***	-8.348***	-7.983***
	(3.276)	(2.687)	(2.716)	(2.694)	(2.758)	(2.860)	(2.878)	(2.782)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	816	816	816	816	816	816	816	816
Panel B: Discretionary								
Externalizing index	18.142**	15.835**	16.529**	20.164***	17.888**	23.837***	23.539***	22.318***
	(7.163)	(7.300)	(7.881)	(7.591)	(7.951)	(8.815)	(8.513)	(7.962)
Index x mother educ.	-4.117	-6.600**	-6.721**	-6.703**	-6.130**	-7.805***	-8.022***	-7.199**
	(2.649)	(2.628)	(2.729)	(2.608)	(2.559)	(2.931)	(3.004)	(2.899)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	816	816	816	816	816	816	816	816
Panel C: Tuition								
Internalizing index	9.470	2.530	2.678	8.563	5.325	6.194	5.940	5.574
	(8.590)	(7.501)	(7.545)	(8.020)	(7.197)	(7.410)	(7.375)	(7.506)
Index x mother educ.	-2.586	-1.416	-1.435	-2.181	-1.029	-1.345	-1.237	-1.111
	(1.880)	(1.593)	(1.600)	(1.633)	(1.652)	(1.722)	(1.710)	(1.736)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	816	816	816	816	816	816	816	816
Panel D: Tuition								
Externalizing index	17.766**	11.034	10.696	15.128*	12.514	15.491*	15.663*	15.258*
	(8.965)	(7.284)	(7.577)	(8.554)	(8.578)	(9.206)	(9.224)	(9.058)
Index x mother educ.	-2.130	-2.874	-2.754	-2.859	-2.681	-3.519	-3.394	-3.121
	(3.034)	(2.636)	(2.528)	(2.817)	(3.054)	(3.217)	(3.173)	(2.903)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	816	816	816	816	816	816	816	816

Notes: The dependent variables are educational expenditure per semester per child in the specified category. The independent variable is the specified index of internalizing or externalizing behavior, as well as the index interacted with maternal education in years. A higher internalizing or externalizing index is indicative of fewer non-cognitive challenges. The control variables are sequentially added as specified in each column, and each column's specification also includes the control variables listed in the columns to the left; all specifications include household fixed effects. Standard errors are clustered at the level of the village. Asterisks indicate significance at the ten, five, and one percent level.

Table A4: Heterogeneity with respect to paternal education: Adding control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Discretionary								
Internalizing index	-8.283	-22.975	-20.842	-20.858	-21.439	-25.704	-23.646	-23.348
	(20.342)	(18.797)	(18.771)	(20.658)	(20.488)	(20.518)	(20.509)	(20.507)
Index x father educ.	-1.034	.611	.325	.728	1.260	1.791	1.530	1.475
	(2.446)	(2.378)	(2.403)	(2.541)	(2.456)	(2.461)	(2.496)	(2.480)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	806	806	806	806	806	806	806	806
Panel B: Discretionary								
Externalizing index	-8.437	-25.738	-22.804	-28.685*	-27.277*	-27.221*	-25.543	-25.294
	(14.392)	(15.817)	(16.085)	(16.521)	(16.098)	(16.381)	(16.245)	(16.313)
Index x father educ.	1.545	2.439	2.131	3.437*	3.548*	3.592*	3.327*	3.362*
	(1.692)	(1.872)	(1.959)	(1.953)	(1.856)	(1.931)	(1.951)	(1.945)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	806	806	806	806	806	806	806	806
Panel C: Tuition								
Internalizing index	11.622	3.835	2.647	6.494	2.796	-.0001	-1.987	-2.194
	(11.420)	(8.006)	(8.227)	(9.987)	(9.549)	(10.027)	(9.951)	(10.120)
Index x father educ.	-1.812	-1.106	-.932	-1.023	-.461	-.113	.140	.177
	(1.660)	(1.315)	(1.331)	(1.439)	(1.433)	(1.518)	(1.506)	(1.530)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	806	806	806	806	806	806	806	806
Panel D: Tuition								
Externalizing index	4.254	-3.191	-4.577	-7.028	-6.875	-6.833	-8.626	-8.733
	(16.844)	(15.825)	(15.073)	(17.555)	(16.463)	(16.423)	(16.506)	(16.499)
Index x father educ.	.850	.564	.853	1.734	1.649	1.683	1.967	1.952
	(2.056)	(1.801)	(1.816)	(2.035)	(1.845)	(1.838)	(1.841)	(1.851)
Control variable added		Age FE	Sibling-gender FE	Height-age + int.	Cog. + int.	Age int.	Parity int.	Gender int.
Obs.	806	806	806	806	806	806	806	806

Notes: The dependent variables are educational expenditure per semester per child in the specified category. The independent variable is the specified index of internalizing or externalizing behavior, as well as the index interacted with paternal education in years. A higher internalizing or externalizing index is indicative of fewer non-cognitive challenges. The control variables are sequentially added as specified in each column, and each column's specification also includes the control variables listed in the columns to the left; all specifications include household fixed effects. Standard errors are clustered at the level of the village. Asterisks indicate significance at the ten, five, and one percent level.

Table A5: Heterogeneous effects with respect to parental education: Expenditure categories

	Supplies (1)	Transportation (2)	Food (3)	Tutoring (4)	Other (5)
Panel A: Heterogenous effects with respect to maternal education					
Index	-.447	4.266**	13.267**	1.449*	3.795*
	(1.568)	(2.073)	(6.723)	(.874)	(2.137)
Index x mother educ.	.422	-1.422**	-6.984***	-.556*	-.437
	(.387)	(.649)	(2.393)	(.292)	(.441)
Obs.	816	816	816	816	816
Panel B: Heterogenous effects with respect to paternal education					
Index	-3.097*	2.452	-19.942	-3.143*	-5.596**
	(1.835)	(3.513)	(15.788)	(1.740)	(2.444)
Index x father educ.	.534*	-.500	.827	.258	1.041**
	(.324)	(.481)	(1.809)	(.195)	(.493)
Obs.	806	806	806	806	806

Notes: The dependent variables are educational expenditure per semester per child in the specified category. The independent variable is the specified index of internalizing or externalizing behavior, as well as the index interacted with maternal or paternal education in years. A higher internalizing or externalizing index is indicative of fewer non-cognitive challenges. All specifications include controls for sibling parity, height-for-age and cognitive skills measured contemporaneously with non-cognitive characteristics, a dummy for middle school, and household, year of birth, and gender-sibling gender fixed effects. Standard errors are clustered at the village level. Asterisks indicate significance at the ten, five, and one percent level.

Table A6: Additional robustness checks

	Discretionary				Tuition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Maternal favoritism and saturated specification								
Index	12.329	19.592	5.554	-15.266	8.958	11.297	7.551	-.179
	(10.897)	(15.534)	(16.994)	(10.269)	(11.411)	(11.169)	(11.564)	(5.072)
Index x mother educ.	-8.890**	-9.310***	-9.928***		-3.832	-.475	-1.504	
	(3.516)	(3.453)	(3.353)		(3.026)	(1.761)	(2.307)	
Index x father educ.			3.033				.005	
			(2.200)				(1.518)	
Index x dif. educ.				-5.497**				-.795
				(2.169)				(1.579)
Sample	Excluding weaker favorites	Excluding harsh parenting	Full		Excluding weaker favorites	Excluding harsh parenting	Full	
Obs.	446	380	806	806	446	380	806	806
Panel B: Additional robustness checks for socioeconomic status								
Index	19.196**	18.685*	18.511*	22.387***	8.996	8.616	7.572	8.503
	(9.246)	(11.317)	(10.775)	(8.676)	(8.543)	(8.827)	(8.778)	(8.405)
Index x maternal educ.	-8.039***	-7.891**	-7.479**	-9.188***	-2.102	-1.989	-1.589	-1.930
	(3.109)	(3.157)	(2.913)	(3.224)	(2.276)	(2.361)	(2.214)	(2.261)
Index x per capita income	-9.256				1.433			
	(6.086)				(3.647)			
Index x capital		-31.702				.918		
		(34.333)				(11.086)		
Index x expenditure			-22.342				-5.489	
			(17.903)				(10.022)	
Income x house value				9.080				-1.166
				(5.585)				(2.560)
Obs.	816	816	816	816	816	816	816	816

Notes: The dependent variables are educational expenditure per semester per child in the specified category; discretionary expenditure is the sum of all categories of expenditure excluding tuition. The independent variable is the specified index of internalizing or externalizing behavior, as well as the index interacted with the specified measure of parental education. A higher non-cognitive index is indicative of fewer non-cognitive challenges. All specifications include controls for sibling parity, height-for-age and cognitive skills measured contemporaneously with non-cognitive characteristics, a dummy for middle school, and household, year of birth, and gender-sibling gender fixed effects. Standard errors are clustered at the village level.

In Panel A, specifications in Columns (1) and (4) are restricted to exclude households in which the mother denotes a favorite ex ante in wave one, and that child shows evidence of weaker non-cognitive skills in wave two. Specifications in Columns (2) and (6) are restricted to exclude households that are above the median in the use of harsh parenting methods. The other specifications use the full sample, but include interaction terms between the non-cognitive index and both maternal and paternal education (in Columns (3) and (7)), or the interaction term between the non-cognitive index and the difference between maternal and paternal education (in Columns (4) and (8)).

In Panel B, the main specification is re-estimated including interaction terms between the non-cognitive index and various measures of household socioeconomic status: per capita income, household stock of working capital, annual expenditure on consumption goods, and estimated home value.

Table A7: Absolute difference in child characteristics and parental education

	Height		Achievement		Internalizing		Externalizing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother educ.	-.014		.378*		-.001		-.014	
	(.014)		(.210)		(.014)		(.013)	
Father educ.		-.003		.226		.010		.001
		(.012)		(.198)		(.011)		(.011)
Obs.	566	562	566	562	566	562	566	562

Notes: The dependent variables are the differences between the specified covariates (height-for-age, achievement test scores, internalizing index scores, and externalizing index scores) for the first-born and second-born children. The independent variables are maternal and parental education. Asterisks indicate significance at the 10, 5 and 1 percent level.

References

Achenbach, Thomas and Craig Edelbrock, "The classification of child psychopathology: a review and analysis of empirical efforts," *Psychological Bulletin*, 1978, *85*, 1275–1301.

_ **and** _ , "Within–child associations between family income and externalizing and internalizing problems," *Psychological Bulletin*, 2006, *42* (2), 237–252.

Andrich, David, "An extension of the rasch model for ratings providing both location and dispersion parameters," *Psychometrika*, 1978, *47* (1), 105–113.

Chen, Xinyin, Mowei Liu, Kenneth Rubin, Dan Li, Zhenyun Li, Guozhen Cen, and Boshu Li, "Parental reports of externalizing and internalizing behaviors in Chinese children: relevancy to social, emotional and school adjustment," *Journal of Psychology in Chinese Societies*, 2002, *3*, 233–260.

Chorpita, Bruce and David Barlow, "The development of anxiety: the role of control in the early environment," *Psychological Bulletin*, 1998, *2*, 3–21.

Edelen, Maria Orlando and Bryce B. Reeve, "Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement," *Quality of Life Research*, 2007, *16*, 5–18.

Embretson, Susan E. and Steven P. Reise, *Item Response Theory for Psychologists*, New York: Lawrence Erlbaum Associates, 2000.

Evans, Gary, "The environment of childhood poverty," *American Psychology*, 2004, *59* (2), 77–92.

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

Ivanova, Masha, Thomas M. Achenbach, Leslie A. Rescorla, Levent Dumenci, Fredrik Almqvist, Niels Bilenber, Hector Bird, Anders G. Broberg, Anca Dobrean, Manfred Dopfner, Nese Erol, Maria Forns, Helga Hannesdottir, Yasuko Kanbayashi, Michael C. Lambert, Patrick Leung, Asghar Minaei, Mesfin Mulatu, Torunn Novik, Kyung Ja Oh, Alexandra Roussos, Michael Sawyer, Zeynep Simsek, Hans-Christophe Steinhausen, Sheila Weintraub, Christa Metzke Winkler, Tomasz Wolanczyk, Nelly Zilber, Rita Zukauskiene, and Frank C. Verhulst, "The generalizability of the Youth Self-Report syndrome structure in 23 societies," *Journal of Consulting and Clinical Psychology*, 2005, *75*, 729–738.

Reise, Steven P., Andrew T. Ainsworth, and Mark G. Haviland, "Item response theory: Fundamentals, applications and promise in psychological research," *Current Directions in Psychological Science*, 2005, *14*, 95–101.

Zheng, Xiaohui and Sophia Rabe-Hesketh, "Estimating parameters of dichotomous and ordinal item response models with gllamm," *Stata Journal*, 2007, *7* (3), 313–333.