



Supplementary Information for

Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills

Shuaizhang Feng, Yujie Han, James J. Heckman, and Tim Kautz

James J. Heckman

Email: [jjh@uchicago.edu](mailto:jjh@uchicago.edu)

This PDF file includes:

Tables S1 to S21

Figures S1 to S2

SI References

**Table S1.** Correlations between (younger/older) child and teacher (guardian) reports of noncognitive skills

Pair of respondents	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
Children (any) and teachers	0.32***	0.11***	0.20***	0.06***	0.23***
Children (younger) and teachers	0.30***	0.13***	0.18***	0.09***	0.22***
Children (older) and teachers	0.33***	0.10***	0.21***	0.04**	0.25***
Children (any) and guardians	0.30***	0.09***	0.16***	0.07***	0.16***
Children (younger) and guardians	0.29***	0.09***	0.15***	0.08***	0.15***
Children (older) and guardians	0.31***	0.10***	0.17***	0.05**	0.18***

Notes: The number of observations is 4,292 for the any child sample, 2,010 for the younger child sample, and 2,036 for the older child sample. 246 children are dropped when analyzing the correlations by age, because there is missing information on their age. Younger children are defined as those below age 10.5, the average age in our sample. Older children are defined as those aged 10.5 and above.

\*\*\* significant at the 1% level. \*\* significant at the 5% level.

**Table S2.** Correlations between (non-)parental guardian and child (teacher) reports of noncognitive skills

Pair of respondents	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
Children and guardians (any)	0.30 ***	0.09 ***	0.16 ***	0.07 ***	0.16 ***
Children and parental guardians	0.33***	0.12***	0.20***	0.10***	0.18***
Children and non-parental guardians	0.22***	0.03	0.11***	0.01	0.12***
Teachers and guardians (any)	0.30 ***	0.04 ***	0.15 ***	0.08 ***	0.16 ***
Teachers and parental guardians	0.34***	0.05***	0.18***	0.09***	0.18***
Teachers and non-parental guardians	0.21***	0.00	0.11***	0.06**	0.12***

Note: The number of observations is 4,292 for the whole sample, 2,990 for the parental guardian sample, and 1,227 for the non-parental guardian sample.

\*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S3.** Inter-trait correlation between noncognitive skills

	Extraversion	Agreeableness	Conscientiousness	Emotional stability
<b>Child</b>				
Agreeableness	0.39			
Conscientiousness	0.33	0.46		
Emotional stability	0.39	0.40	0.37	
Openness	0.48	0.44	0.42	0.44
<b>Teacher</b>				
Agreeableness	0.32			
Conscientiousness	0.42	0.66		
Emotional stability	0.47	0.56	0.56	
Openness	0.64	0.43	0.50	0.53
<b>Guardian</b>				
Agreeableness	0.41			
Conscientiousness	0.25	0.45		
Emotional stability	0.33	0.39	0.39	
Openness	0.41	0.47	0.42	0.41

Note: The number of observations is 4,292 for the whole sample. All coefficients are statistically significant at the 1% level.

**Table S4.** Test-retest correlations for children's skills measured in two consecutive years for the same child

Respondent	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
Child	0.55	0.46	0.45	0.40	0.49
Any teacher	0.62	0.43	0.59	0.39	0.58
Same teacher	0.68	0.54	0.65	0.50	0.64
Any guardian	0.49	0.35	0.42	0.29	0.38
Same guardian	0.58	0.50	0.59	0.39	0.50

Notes: The table shows Spearman correlation coefficients between measures of the same child collected between two different years. The number of observations is 2,231 for the whole sample, 1,692 for the same teacher respondent sample, and 1,380 for the same guardian respondent sample. All coefficients are significant at the 1% level.

**Table S5.** Cronbach's alpha of any type of guardian, parental guardian, and non-parental guardian reports of noncognitive skills

Noncognitive skill	Guardian (any)	Parental guardian	Non-parental guardian
Extraversion	0.65	<b>0.69</b>	0.56
Agreeableness	0.73	<b>0.76</b>	0.66
Conscientiousness	0.61	<b>0.67</b>	0.49
Emotional stability	0.51	<b>0.57</b>	0.39
Openness	0.67	<b>0.69</b>	0.62

Notes: The bold font indicates the respondent type with the highest Cronbach's alpha for each noncognitive skill. The number of observations is 4,846–4,870 for guardian reports, 3,352–3,369 for parental guardian reports, and 1,407–1,415 for non-parental guardian reports.

**Table S6.** Test of the equality of the Cronbach's alpha between pairs of respondent types

Noncognitive skill	Child versus teacher	<i>p</i> -value	
		Child versus guardian	Guardian versus teacher
Extraversion	<0.01	<0.01	<0.01
Agreeableness	<0.01	<0.01	<0.01
Conscientiousness	<0.01	<0.01	<0.01
Emotional stability	<0.01	<0.01	<0.01
Openness	<0.01	<0.01	<0.01

Notes: The *p*-values are based on tests of equality with the null hypothesis that two types of respondents have the same Cronbach's alpha for a given noncognitive skill. They are obtained by bootstrapping with 1,000 replications.

**Table S7.** Correlations between individual noncognitive skills and child outcomes in school measured 1 y later for various types of guardians

	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
<b>Guardian (any)</b>												
Extraversion	0.08 ***	0.09 ***	0.09 ***	0.09 ***	0.08 ***	0.03	0.09 ***	0.12 ***	0.08 ***	-0.08 ***	-0.01	0.04 **
Agreeableness	0.07 ***	0.07 ***	0.06 ***	0.05 ***	0.05 **	0.00	0.05 **	0.06 ***	0.05 ***	0.06 ***	0.05 ***	0.05 ***
Conscientiousness	0.05 **	0.09 ***	0.07 ***	0.12 ***	0.07 ***	0.07 ***	0.12 ***	0.10 ***	0.10 ***	0.13 ***	0.13 ***	0.13 ***
Emotional stability	0.05 **	0.08 ***	0.06 ***	0.05 **	0.06 ***	0.05 ***	0.06 ***	0.07 ***	0.04 **	0.03	0.03	0.07 ***
Openness	0.13 ***	0.14 ***	0.13 ***	0.09 ***	0.10 ***	0.11 ***	0.14 ***	0.12 ***	0.09 ***	-0.02	0.05 **	0.08 ***
<b>Parental guardian</b>												
Extraversion	0.10 ***	0.11 ***	0.08 ***	0.10 ***	0.09 ***	0.02	0.10 ***	0.14 ***	0.10 ***	-0.08 ***	0.01	0.04 *
Agreeableness	0.07 ***	0.07 ***	0.05 **	0.06 ***	0.05 **	0.00	0.06 ***	0.08 ***	0.06 ***	0.07 ***	0.06 ***	0.06 ***
Conscientiousness	0.06 **	0.12 ***	0.09 ***	0.15 ***	0.11 ***	0.07 ***	0.15 ***	0.13 ***	0.12 ***	0.14 ***	0.15 ***	0.15 ***
Emotional stability	0.06 **	0.09 ***	0.06 **	0.06 ***	0.06 ***	0.05 **	0.07 ***	0.09 ***	0.04 *	0.06 **	0.04 *	0.08 ***
Openness	0.13 ***	0.14 ***	0.12 ***	0.10 ***	0.12 ***	0.10 ***	0.16 ***	0.14 ***	0.08 ***	-0.02	0.05 **	0.09 ***
<b>Non-parental guardian</b>												
Extraversion	0.03	0.07 *	0.10 ***	0.07 **	0.05	0.05	0.07 **	0.07 **	0.04	-0.07 **	-0.04	0.05
Agreeableness	0.05	0.01	0.08 **	0.00	0.02	0.01	0.00	0.01	0.00	0.02	0.01	0.02
Conscientiousness	0.06 *	0.13 ***	0.11 ***	0.09 ***	0.05	0.06 *	0.12 ***	0.07 *	0.12 ***	0.10 ***	0.12 ***	0.10 ***
Emotional stability	0.04	0.09 **	0.09 **	0.04	0.10 ***	0.05	0.06	0.06 *	0.05	-0.02	0.01	0.05
Openness	0.12 ***	0.17 ***	0.18 ***	0.07 **	0.09 ***	0.13 ***	0.10 ***	0.09 **	0.10 ***	-0.03	0.04	0.05

Notes: The number of observations is 2,857 for the any guardian sample, 1,980 for the parental guardian sample, and 826 for the non-parental guardian sample. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Leader is a dummy for being elected a class or school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often."

\*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S8.** *p*-values from tests of the equality of the correlations between individual noncognitive skill reports and outcomes in school across respondent types

	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
<b>Child versus teacher</b>												
Extraversion	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.20	<0.01	<0.01
Agreeableness	0.82	0.02	0.01	0.54	0.18	0.19	<0.01	0.28	<0.01	<0.01	<0.01	<0.01
Conscientiousness	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Emotional stability	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Openness	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<b>Child versus guardian</b>												
Extraversion	0.95	<0.01	<0.01	<0.01	0.01	0.29	<0.01	0.03	0.05	<0.01	0.07	<0.01
Agreeableness	0.82	<0.01	0.11	<0.01	0.02	0.12	<0.01	<0.01	0.02	0.10	0.01	<0.01
Conscientiousness	0.22	<0.01	0.33	0.05	0.14	0.35	0.01	<0.01	0.25	0.14	1.00	0.02
Emotional stability	0.82	0.14	0.08	0.03	0.35	0.73	<0.01	0.03	<0.01	0.76	0.61	0.07
Openness	0.20	0.01	0.40	<0.01	0.50	0.35	0.23	<0.01	0.11	0.28	0.49	0.14
<b>Teacher versus guardian</b>												
Extraversion	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.07	<0.01	<0.01
Agreeableness	0.99	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Conscientiousness	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Emotional stability	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Openness	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Notes: The number of observations is 2,857 for the whole sample. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Leader is a dummy for being elected as a class or school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often." The *p*-values are based on tests of equality with the null hypothesis that the two types of respondents have the same correlation coefficient between a given noncognitive skill and a child outcome in school measured 1 y later. They are obtained by bootstrapping with 1,000 replications.

**Table S9.** Predictive power (adjusted  $R$ ) of groups of noncognitive skills for child outcomes in school measured 1 y later for various types of guardians

Respondent	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
Guardian (any)	0.14	0.16	0.14	0.14	0.13	0.15	0.17	0.15	0.11	0.18	0.14	0.12
Parental guardian	0.14	0.16	0.12	0.15	0.14	0.14	0.18	0.17	0.12	0.21	0.15	0.14
Non-parental guardian	0.14	0.21	0.19	0.14	0.13	0.16	0.17	0.11	0.13	0.15	0.14	0.10

Notes: The number of observations is 2,857 for the any guardian sample, 1,980 for the parental guardian sample, and 826 for the non-parental guardian sample. The table shows the square root of the adjusted  $R^2$  from the OLS regressions of outcomes in school on the group of Big Five noncognitive skills for each respondent type. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Leader is a dummy for being elected a class or school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often."

**Table S10.** Standardized coefficients from regressions of children's future outcomes in school on the group of noncognitive skills

	Cognitive outcomes						Behavioral outcomes					
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
<b>Child</b>												
Extraversion	0.03 (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.09*** (0.02)	0.02 (0.02)	0.06*** (0.02)	0.06*** (0.02)	0.06*** (0.02)	-0.01 (0.02)	-0.00 (0.02)	0.04* (0.02)
Agreeableness	0.01 (0.02)	0.07*** (0.02)	0.01 (0.02)	0.06*** (0.02)	0.03 (0.02)	0.01 (0.02)	0.05** (0.02)	0.05** (0.02)	0.03 (0.02)	0.12*** (0.02)	0.09*** (0.02)	0.09*** (0.02)
Conscientiousness	0.01 (0.02)	0.08*** (0.02)	0.04** (0.02)	0.09*** (0.02)	0.07*** (0.02)	0.02 (0.02)	0.12*** (0.02)	0.09*** (0.02)	0.08*** (0.02)	0.09*** (0.02)	0.11*** (0.02)	0.14*** (0.02)
Emotional stability	-0.00 (0.02)	-0.03** (0.02)	-0.00 (0.02)	-0.03 (0.02)	0.01 (0.02)	-0.00 (0.02)	0.01 (0.02)	0.02 (0.02)	0.03 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.01 (0.02)
Openness	0.08*** (0.02)	0.09*** (0.02)	0.09*** (0.02)	0.07*** (0.02)	0.05** (0.02)	0.07*** (0.02)	0.08*** (0.02)	0.10*** (0.02)	0.03 (0.02)	-0.13*** (0.02)	-0.05** (0.02)	-0.00 (0.02)
<b>Teacher</b>												
Extraversion	0.00 (0.02)	0.07*** (0.02)	0.07*** (0.02)	0.04* (0.02)	0.07*** (0.02)	0.01 (0.02)	0.02 (0.02)	0.19*** (0.02)	0.06*** (0.02)	-0.22*** (0.02)	-0.11*** (0.02)	-0.01 (0.02)
Agreeableness	-0.12*** (0.03)	-0.09*** (0.02)	-0.10*** (0.02)	-0.11*** (0.03)	-0.09*** (0.02)	-0.12*** (0.03)	-0.12*** (0.02)	-0.11*** (0.02)	0.05** (0.02)	0.04 (0.03)	0.07** (0.03)	-0.05** (0.02)
Conscientiousness	0.10*** (0.02)	0.25*** (0.02)	0.16*** (0.02)	0.26*** (0.02)	0.21*** (0.03)	0.16*** (0.03)	0.36*** (0.02)	0.26*** (0.02)	0.25*** (0.02)	0.26*** (0.03)	0.28*** (0.03)	0.38*** (0.02)
Emotional stability	0.06** (0.03)	0.05* (0.02)	0.05** (0.02)	-0.00 (0.02)	0.02 (0.02)	0.03 (0.02)	0.02 (0.02)	0.03 (0.02)	0.00 (0.02)	0.12*** (0.03)	0.02 (0.03)	0.05** (0.02)
Openness	0.22*** (0.02)	0.20*** (0.02)	0.26*** (0.02)	0.21*** (0.03)	0.17*** (0.03)	0.23*** (0.02)	0.33*** (0.02)	0.14*** (0.02)	0.13*** (0.02)	0.03 (0.03)	0.14*** (0.02)	0.18*** (0.02)
<b>Guardian</b>												
Extraversion	0.04** (0.02)	0.05*** (0.02)	0.05*** (0.02)	0.07*** (0.02)	0.04** (0.02)	0.01 (0.02)	0.06*** (0.02)	0.09*** (0.02)	0.06*** (0.02)	-0.11*** (0.02)	-0.03* (0.02)	-0.01 (0.02)
Agreeableness	0.01 (0.02)	-0.03* (0.02)	-0.00 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.07*** (0.02)	-0.06*** (0.02)	-0.04* (0.02)	-0.02 (0.02)	0.07*** (0.02)	0.00 (0.02)	-0.01 (0.02)
Conscientiousness	-0.01 (0.02)	0.04** (0.02)	0.02 (0.02)	0.09*** (0.02)	0.03* (0.02)	0.04** (0.02)	0.09*** (0.02)	0.06*** (0.02)	0.08*** (0.02)	0.15*** (0.02)	0.14*** (0.02)	0.11*** (0.02)
Emotional stability	-0.03 (0.02)	-0.02 (0.02)	-0.03 (0.02)	-0.02 (0.02)	-0.01 (0.02)	0.02 (0.02)	-0.04* (0.02)	0.00 (0.02)	-0.02 (0.02)	0.01 (0.02)	-0.02 (0.02)	0.01 (0.02)
Openness	0.12*** (0.02)	0.12*** (0.02)	0.11*** (0.02)	0.05** (0.02)	0.11*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.08*** (0.02)	0.04* (0.02)	-0.08*** (0.02)	0.01 (0.02)	0.04 (0.02)
<b>p-value from joint test of equality between respondent types</b>												
Child versus guardian	0.65	<0.01	0.12	<0.01	<0.01	0.04	<0.01	<0.01	0.06	<0.01	0.01	<0.01
Child versus teacher	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Teacher versus guardian	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01



Notes: The number of observations is 2,857 for the whole sample. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Leader is a dummy for being elected as a class or school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often." All future outcomes are standardized by subtracting the mean and then dividing the standard deviation. Standard errors in parentheses. The  $p$ -values are based on  $F$  tests with the null hypothesis that the regression coefficients from each pair of respondent types are jointly equal for each outcome and are obtained by bootstrapping with 1,000 replications.

\*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S11.** Predictive power (adjusted  $R$ ) of IQ and groups of noncognitive skills for child outcomes in school measured 1 y later

Predictor(s)	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
Guardian-reported noncognitive skills	0.14	0.16	0.14	0.14	0.13	0.15	0.17	0.15	0.11	0.18	0.14	0.12
Child-reported noncognitive skills	0.11	0.26	0.19	0.22	0.19	0.10	0.23	0.23	0.17	0.17	0.15	0.22
Teacher-reported noncognitive skills	0.31	0.48	0.45	0.39	0.35	0.33	0.58	0.43	0.42	0.35	0.39	0.50
All respondent reports of noncognitive skills	0.32	0.50	0.46	0.42	0.37	0.34	0.59	0.45	0.42	0.38	0.41	0.51
Baseline IQ	0.71	0.44	0.51	0.33	0.34	0.37	0.37	0.22	0.21	0.08	0.13	0.26
Baseline IQ + guardian-reported noncognitive skills	0.71	0.45	0.52	0.35	0.35	0.38	0.39	0.25	0.23	0.20	0.19	0.28
Baseline IQ + child-reported noncognitive skills	0.71	0.49	0.53	0.38	0.37	0.37	0.42	0.30	0.26	0.19	0.20	0.34
Baseline IQ + teacher-reported noncognitive skills	0.72	0.57	0.60	0.45	0.43	0.43	0.62	0.44	0.43	0.35	0.39	0.52
Baseline IQ + all respondent reports of noncognitive skills each entered separately	0.72	0.59	0.60	0.47	0.44	0.44	0.62	0.46	0.44	0.39	0.41	0.53

Notes: The number of observations is 2,857 for the whole sample. The table shows the square root of the adjusted  $R^2$  from the OLS regressions of the cognitive and behavioral outcomes in school on different combinations of baseline IQ and the groups of Big Five noncognitive skills. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often."

**Table S12.** Correlation between individual noncognitive skills and child outcomes in school measured 1 y later for various approaches to adjusting child reports for reference bias

	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
<b>Child (no adjustment)</b>												
Extraversion	0.08 ***	0.20 ***	0.15 ***	0.15 ***	0.14 ***	0.05 ***	0.16 ***	0.17 ***	0.12 ***	-0.01	0.03 *	0.12 ***
Agreeableness	0.06 ***	0.18 ***	0.10 ***	0.16 ***	0.10 ***	0.04 **	0.15 ***	0.15 ***	0.12 ***	0.10 ***	0.12 ***	0.16 ***
Conscientiousness	0.02	0.18 ***	0.09 ***	0.17 ***	0.11 ***	0.04 **	0.19 ***	0.17 ***	0.13 ***	0.09 ***	0.13 ***	0.19 ***
Emotional stability	0.04 **	0.11 ***	0.10 ***	0.10 ***	0.09 ***	0.04 **	0.13 ***	0.13 ***	0.12 ***	0.02	0.04 **	0.11 ***
Openness	0.09 ***	0.21 ***	0.15 ***	0.16 ***	0.12 ***	0.09 ***	0.17 ***	0.19 ***	0.12 ***	-0.05 ***	0.03	0.11 ***
<b>Child (using the lowest score)</b>												
Extraversion	0.12 ***	0.22 ***	0.18 ***	0.17 ***	0.17 ***	0.09 ***	0.20 ***	0.16 ***	0.14 ***	0.04 **	0.06 ***	0.16 ***
Agreeableness	-0.05 **	0.03	-0.02	0.08 ***	0.01	-0.01	0.02	0.05 ***	0.01	0.06 ***	0.05 ***	0.03
Conscientiousness	0.08 ***	0.21 ***	0.14 ***	0.17 ***	0.13 ***	0.08 ***	0.22 ***	0.16 ***	0.12 ***	0.08 ***	0.13 ***	0.21 ***
Emotional stability	0.07 ***	0.11 ***	0.11 ***	0.11 ***	0.12 ***	0.06 ***	0.15 ***	0.12 ***	0.13 ***	0.03	0.05 **	0.13 ***
Openness	0.12 ***	0.18 ***	0.15 ***	0.12 ***	0.08 ***	0.05 ***	0.18 ***	0.15 ***	0.11 ***	0.01	0.04 **	0.13 ***
<b>Child (using the average score)</b>												
Extraversion	0.07 ***	0.18 ***	0.14 ***	0.14 ***	0.15 ***	0.04 **	0.16 ***	0.14 ***	0.11 ***	0.02	0.04 **	0.12 ***
Agreeableness	-0.08 ***	-0.01	-0.05 **	0.06 ***	-0.01	-0.04 **	0.00	0.05 ***	0.01	0.05 ***	0.06 ***	0.02
Conscientiousness	0.04 **	0.17 ***	0.10 ***	0.15 ***	0.10 ***	0.05 ***	0.18 ***	0.15 ***	0.10 ***	0.07 ***	0.12 ***	0.18 ***
Emotional stability	-0.02	0.01	0.00	0.05 **	0.05 **	0.00	0.06 ***	0.07 ***	0.07 ***	0.02	0.01	0.06 ***
Openness	0.04 **	0.10 ***	0.08 ***	0.06 ***	0.04 **	0.01	0.11 ***	0.13 ***	0.09 ***	-0.04 **	0.02	0.07 ***
<b>Child (using the highest score)</b>												
Extraversion	-0.02	0.06 ***	0.04 *	0.04 **	0.07 ***	-0.05 ***	0.05 ***	0.08 ***	0.05 ***	-0.02	-0.01	0.03
Agreeableness	-0.11 ***	-0.05 ***	-0.08 ***	0.02	-0.03 *	-0.08 ***	-0.03	0.04 **	0.01	0.04 **	0.05 ***	0.01
Conscientiousness	-0.05 ***	0.05 ***	0.01	0.07 ***	0.04 **	-0.02	0.08 ***	0.10 ***	0.05 ***	0.03 *	0.08 ***	0.09 ***
Emotional stability	-0.14 ***	-0.12 ***	-0.14 ***	-0.07 ***	-0.07 ***	-0.09 ***	-0.10 ***	-0.03	-0.03 *	0.00	-0.04 **	-0.06 ***
Openness	-0.03	0.03 *	0.02	0.02	0.01	-0.02	0.05 ***	0.09 ***	0.07 ***	-0.06 ***	0.00	0.03

Notes: The number of observations is 2,857 for the whole sample. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Leader is a dummy for being elected a class or school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often." Child (using the lowest score) assigns the lowest possible score to the related items when children responded to the anchors inconsistently. Child (using the highest score) assigns the highest possible score to the related items when children responded to the anchors inconsistently. Child (using the average score) assigns the average of the highest possible score and lowest possible score to the related items when children responded to the anchors inconsistently.

\*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S13.** Predictive power (adjusted  $R$ ) of groups of noncognitive skills for child outcomes in school measured 1 y later for various approaches to adjusting child reports for reference bias

	Cognitive outcomes							Behavioral outcomes				
	IQ	Chinese test score	Math test score	English test score	Morality test score	Science test score	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
Child (no adjustment)	0.11	0.26	0.19	0.22	0.19	0.10	0.23	0.23	0.17	0.17	0.15	0.22
Child (using the lowest score)	0.24	0.37	0.32	0.28	0.27	0.22	0.33	0.23	0.21	0.11	0.15	0.28
Child (using the average score)	0.17	0.28	0.25	0.22	0.22	0.15	0.26	0.20	0.16	0.12	0.15	0.23
Child (using the highest score)	0.14	0.14	0.17	0.10	0.11	0.10	0.14	0.13	0.08	0.09	0.10	0.10
Child (no adjustment, controlling for whether they responded extraversion anchor inconsistently)	0.26	0.34	0.29	0.28	0.25	0.22	0.31	0.24	0.19	0.18	0.18	0.27
Child (no adjustment, controlling for whether they responded agreeableness anchor inconsistently)	0.16	0.31	0.25	0.25	0.21	0.16	0.26	0.23	0.17	0.17	0.16	0.23
Child (no adjustment, controlling for whether they responded conscientiousness anchor inconsistently)	0.14	0.27	0.21	0.23	0.20	0.11	0.24	0.23	0.17	0.17	0.15	0.22
Child (no adjustment, controlling for whether they responded emotional stability anchor inconsistently)	0.30	0.37	0.34	0.30	0.28	0.21	0.36	0.28	0.23	0.18	0.20	0.31
Child (no adjustment, controlling for whether they responded openness anchor inconsistently)	0.16	0.28	0.21	0.23	0.20	0.11	0.25	0.23	0.18	0.17	0.16	0.24
Child (no adjustment, controlling for whether they responded any of the anchors inconsistently)	0.25	0.37	0.33	0.30	0.26	0.24	0.33	0.25	0.20	0.18	0.17	0.28
Child (no adjustment, controlling for whether they responded all anchors inconsistently)	0.14	0.27	0.20	0.23	0.20	0.10	0.24	0.23	0.17	0.17	0.15	0.23

Notes: The number of observations is 2,857 for the whole sample. The table shows the square root of the adjusted  $R^2$  from the OLS regressions of the cognitive and behavioral outcomes in school on the group of Big Five noncognitive skills for each approach to adjusting the children's reports of noncognitive skills. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to "never" and five corresponding to "often." Child (using the lowest score) assigns the lowest possible score to the related items when children responded to the anchors inconsistently. Child (using the highest score) assigns the highest possible score to the related items when children responded to the anchors inconsistently. Child (using the average score) assigns the average of the highest possible score and lowest possible score to the related items when children responded to the anchors inconsistently.

**Table S14.** Cronbach's alpha of child reports of noncognitive skills for various approaches to adjusting child reports for reference bias

Noncognitive skill	Child (no adjustment)	Child (using the lowest score)	Child (using the average score)	Child (using the highest score)
Extraversion	0.57	<b>0.77</b>	0.71	0.72
Agreeableness	0.63	<b>0.80</b>	0.75	0.73
Conscientiousness	0.57	<b>0.79</b>	0.70	0.64
Emotional stability	0.62	<b>0.85</b>	0.79	0.82
Openness	0.62	<b>0.79</b>	0.75	0.75

Notes: The bold font indicates the adjustment type with the highest Cronbach's alpha for each noncognitive skill. The number of observations is 5,422–5,465. Child (using the lowest score) assigns the lowest possible score to the related items when children responded to the anchors inconsistently. Child (using the highest score) assigns the highest possible score to the related items when children responded to the anchors inconsistently. Child (using the average score) assigns the average of the highest possible score and lowest possible score to the related items when children responded to the anchors inconsistently.

**Table S15.** Correlations between child and teacher (guardian) reports of noncognitive skills for various approaches to adjusting child reports for reference bias

Pair of respondents	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
Child (no adjustment) and teacher	0.32***	0.11***	0.20***	0.06***	0.23***
Child (using the lowest score) and teacher	0.24***	0.06***	0.18***	0.05***	0.18***
Child (using the average score) and teacher	0.26***	0.06***	0.17***	0.02	0.15***
Child (using the highest score) and teacher	0.22***	0.06***	0.12***	-0.02	0.10***
Child (no adjustment) and guardian	0.30***	0.09***	0.16***	0.07***	0.16***
Child (using the lowest score) and guardian	0.21***	0.05***	0.14***	0.04***	0.12***
Child (using the average score) and guardian	0.23***	0.06***	0.14***	0.04**	0.12***
Child (using the highest score) and guardian	0.21***	0.06***	0.11***	0.02	0.10***

Note: The number of observations is 4,292 for the whole sample.

\*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S16.** Correlations between individual noncognitive skills and teacher-reported outcomes in school measured 1 y later by whether the same teacher provides the report in both years

	Same teacher respondent						Different teacher respondent					
	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
<b>Child</b>												
Extraversion	0.17 ***	0.16 ***	0.15 ***	0.01	0.03	0.13 ***	0.14 ***	0.18 ***	0.04	-0.06	0.03	0.08 **
Agreeableness	0.16 ***	0.16 ***	0.13 ***	0.12 ***	0.13 ***	0.18 ***	0.13 ***	0.10 ***	0.08 **	0.04	0.05	0.12 ***
Conscientiousness	0.20 ***	0.18 ***	0.13 ***	0.12 ***	0.15 ***	0.21 ***	0.14 ***	0.16 ***	0.12 ***	-0.01	0.06	0.13 ***
Emotional stability	0.14 ***	0.13 ***	0.13 ***	0.03	0.05 **	0.12 ***	0.11 ***	0.14 ***	0.09 **	-0.01	0.01	0.11 ***
Openness	0.19 ***	0.21 ***	0.14 ***	-0.03	0.06 ***	0.13 ***	0.13 ***	0.15 ***	0.07 **	-0.12 ***	-0.05	0.06 *
<b>Teacher</b>												
Extraversion	0.37 ***	0.37 ***	0.30 ***	-0.04 *	0.14 ***	0.28 ***	0.32 ***	0.33 ***	0.17 ***	-0.05	0.06 *	0.21 ***
Agreeableness	0.30 ***	0.16 ***	0.31 ***	0.27 ***	0.33 ***	0.32 ***	0.23 ***	0.21 ***	0.22 ***	0.17 ***	0.28 ***	0.23 ***
Conscientiousness	0.49 ***	0.34 ***	0.41 ***	0.32 ***	0.40 ***	0.48 ***	0.39 ***	0.35 ***	0.26 ***	0.18 ***	0.24 ***	0.37 ***
Emotional stability	0.34 ***	0.25 ***	0.28 ***	0.23 ***	0.28 ***	0.33 ***	0.34 ***	0.29 ***	0.18 ***	0.09 **	0.16 ***	0.29 ***
Openness	0.54 ***	0.37 ***	0.35 ***	0.12 ***	0.28 ***	0.39 ***	0.38 ***	0.35 ***	0.21 ***	0.07 *	0.23 ***	0.29 ***
<b>Guardian</b>												
Extraversion	0.07 ***	0.12 ***	0.08 ***	-0.08 ***	0.01	0.02	0.14 ***	0.11 ***	0.07 *	-0.10 ***	-0.03	0.08 **
Agreeableness	0.03	0.06 ***	0.05 **	0.06 ***	0.06 ***	0.04 *	0.11 ***	0.08 **	0.05	0.03	0.01	0.10 **
Conscientiousness	0.10 ***	0.08 ***	0.09 ***	0.13 ***	0.14 ***	0.11 ***	0.17 ***	0.14 ***	0.15 ***	0.10 **	0.07 **	0.17 ***
Emotional stability	0.05 **	0.08 ***	0.04 **	0.07 ***	0.06 ***	0.07 ***	0.08 **	0.04	0.03	-0.09 **	-0.05	0.06
Openness	0.14 ***	0.13 ***	0.09 ***	0.01	0.07 ***	0.08 ***	0.16 ***	0.10 ***	0.06 *	-0.12 ***	-0.03	0.07 *

Notes: The number of observations is 2,157 for the same teacher respondent sample and 700 for the different teacher respondent sample. Leader is a dummy for being elected a school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to “very bad” and five corresponding to “very good.” Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to “never” and five corresponding to “often.” \*\*\* significant at the 1% level. \*\* significant at the 5% level. \* significant at the 10% level.

**Table S17.** Predictive power (adjusted  $R$ ) of groups of noncognitive skills for teacher-reported outcomes in school measured 1 y later by whether the same teacher provides the report in both years

	Same teacher respondent						Different teacher respondent					
	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline	Overall academic performance	Leader	Mental health	Avoids fights	Honesty	Good discipline
Child	0.24	0.24	0.19	0.17	0.17	0.25	0.19	0.19	0.10	0.15	0.11	0.13
Teacher	0.61	0.44	0.46	0.37	0.42	0.52	0.48	0.41	0.28	0.26	0.30	0.41
Guardian	0.16	0.15	0.10	0.18	0.14	0.11	0.21	0.14	0.13	0.24	0.13	0.16

Notes: The table shows the square root of the adjusted  $R^2$  from the OLS regressions of teacher-reported outcomes in school on the group of Big Five noncognitive skills for each respondent type. The number of observations is 2,157 for the same teacher respondent sample and 700 for the different teacher respondent sample. Leader is a dummy for being elected a school leader. Overall academic performance, mental health, and good discipline are evaluated on a one- to five-point scale, with one corresponding to “very bad” and five corresponding to “very good.” Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to “never” and five corresponding to “often.”

**Table S18.** Description of variables used in the analysis

Variable	Definition	Source
Big Five	Measures of extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience that are based on Big Five Inventory-2 (BFI-2) (1). For each skill, the measure is a factor score based on factor models that were estimated separately for each group of Big Five items and respondent type. All scores are standardized within survey wave and respondent type.	Child, guardian, and teacher report
IQ	Standardized score from a 60-item version of Raven's Standard Progressive Matrices (2).	Test of children
Chinese test score	Score from a standardized test covering Hanyu Pinyin (the romanization of Mandarin), ancient poetry, reading comprehension, and writing in Mandarin. The score is standardized for each grade.	Administrative record
Math test score	Score from a standardized test covering numbers, algebra, measurement, and geometry. The score is standardized for each grade.	Administrative record
English test score	Score from a standardized test covering listening, spelling, grammar, reading comprehension, and writing in the English language. The score is standardized for each grade.	Administrative record
Morality test score	Score from a standardized test on students' morality and values that assesses whether students can analyze objectively and make reasonable judgements about different issues. Examples of topics include politeness, caring about others, confidence, perseverance, respect for others, responsibility, and integrity. The score is standardized for each grade.	Administrative record
Science test score	Score from a standardized test covering basic scientific knowledge and science process skills. Examples of topics include materials, life science, space, and energy. The score is standardized for each grade.	Administrative record
Overall academic performance	A scale that measures a child's overall academic performance at school during the survey semester. The scale ranges from one to five with one corresponding to "very bad" and five corresponding to "very good."	Teacher report
Leader	An indicator for whether a child is elected as a class or school leader.	Teacher report
Mental health	A scale that measures a child's mental health during the survey semester. The scale ranges from one to five with one corresponding to "very bad" and five corresponding to "very good."	Teacher report
Avoids fights	A scale that measures a child's frequency of avoiding fights at school during the survey semester. The scale ranges from one to five with one corresponding to "never" and five corresponding to "often."	Teacher report
Honesty	A scale that measures a child's frequency of being honest at school during the survey semester. The scale ranges from one to five with one corresponding to "never" and five corresponding to "often."	Teacher report
Good discipline	A scale that measures a child's overall disciplinary performance at school during the survey semester. The scale ranges from one to five with one corresponding to "very bad" and five corresponding to "very good."	Teacher report



**Table S19.** Summary statistics of baseline characteristics on children

<b>Baseline characteristic</b>	<b>Mean</b>	<b>Standard deviation</b>
Grade 4	0.34	0.48
Grade 5	0.34	0.47
Grade 6	0.32	0.47
Attending a boarding school	0.03	0.16
Attending an urban school	0.24	0.42
Female	0.50	0.50
No siblings	0.53	0.50
Father has urban hukou (household registration in urban area)	0.21	0.41
Mother has urban hukou (household registration in urban area)	0.20	0.40
Father has a high school education or above	0.36	0.48
Mother has a high school education or above	0.30	0.46
Currently left behind (has at least one parent that migrated without them)	0.47	0.50
<b>Sample size</b>	<b>4,586–5,573</b>	

**Table S20.** Percent of variance in each item explained by the noncognitive skill factors for each respondent type and skill

Skill	Item	Percent of variance in each item explained by the skill factor		
		Child	Guardian	Teacher
Extraversion	1. is outgoing or sociable	61.55	61.99	67.83
	2. has an assertive personality	17.25	21.22	54.91
	3. is sometimes shy, introverted	25.21	39.87	54.53
	4. is dominant, acts as a leader	12.26	14.98	39.64
Agreeableness	5. is compassionate, has a soft heart	25.89	32.24	63.17
	6. has a forgiving nature	43.50	45.29	72.88
	7. is helpful and unselfish with others	35.58	56.97	74.46
	8. is polite, courteous to others	20.70	28.87	51.21
Conscientiousness	9. leaves a mess, doesn't clean up	31.07	21.26	62.19
	10. is dependable, steady	7.48	13.97	46.99
	11. keeps things neat and tidy	61.71	68.06	74.46
	12. is efficient, gets things done	15.90	28.41	39.02
Emotional stability	13. is relaxed, handles stress well	22.78	26.47	50.66
	14. is emotionally stable, not easily upset	38.52	18.98	48.30
	15. often feels sad	15.86	8.00	21.16
	16. keeps their emotions under control	44.08	38.00	57.58
Openness	17. is inventive, finds clever ways to do things	50.42	46.85	76.17
	18. values art and beauty	13.49	18.60	27.81
	19. is curious about many different things	11.74	19.95	40.59
	20. is original, comes up with new ideas	59.96	59.17	78.48

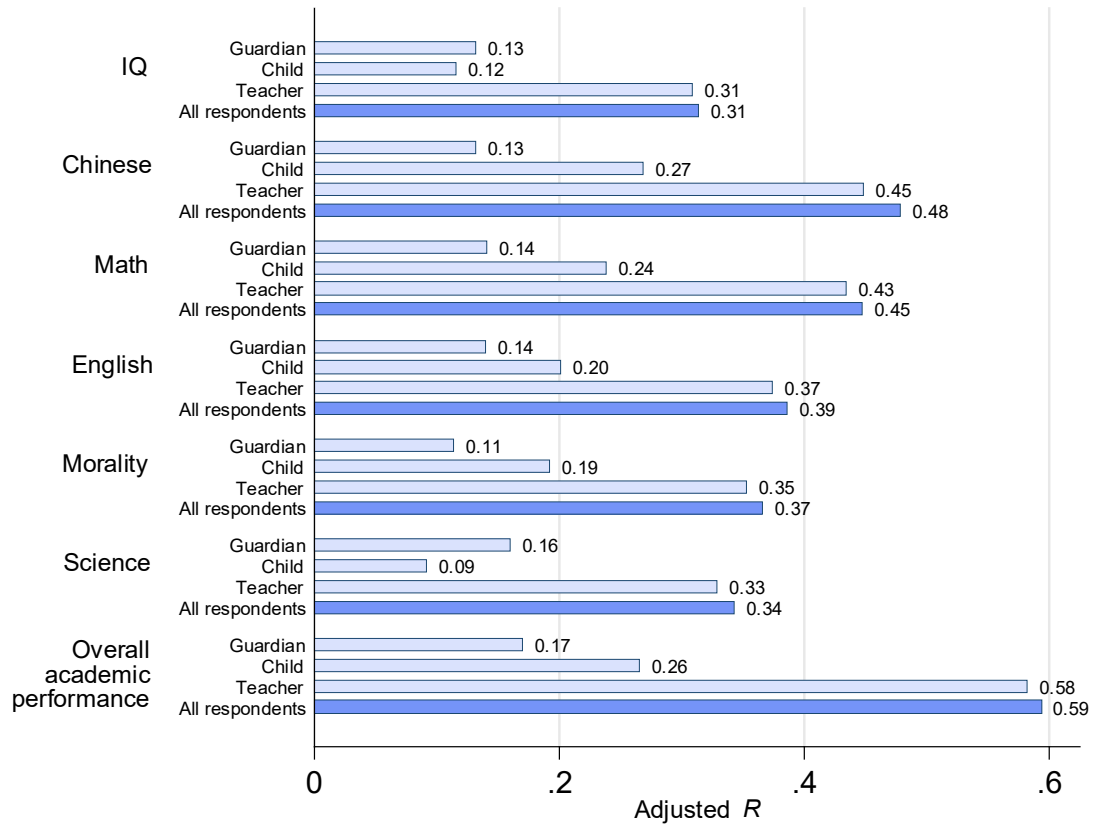
Notes: Items 3, 9, and 15 were reverse coded. The estimates are based on factor models that were estimated separately by respondent type and skill, using the four items associated with each Big Five skill. Each item is constrained to measure one skill. The estimate is the percent of the variance in each item explained by the factor for the associated skill and respondent type. The remaining variance in the item is assumed to arise from measurement error.

**Table S21.** Rules for using anchoring vignettes to adjust children’s self-reports of the Big Five

Order of ratings	Adjusted value
Original response < Low level vignette	1
Original response = Low level vignette	2
Low level vignette < Original response < Medium level vignette	3
Original response = Medium level vignette	4
Medium level vignette < Original response < High level vignette	5
Original response = High level vignette	6
High level vignette < Original response	7

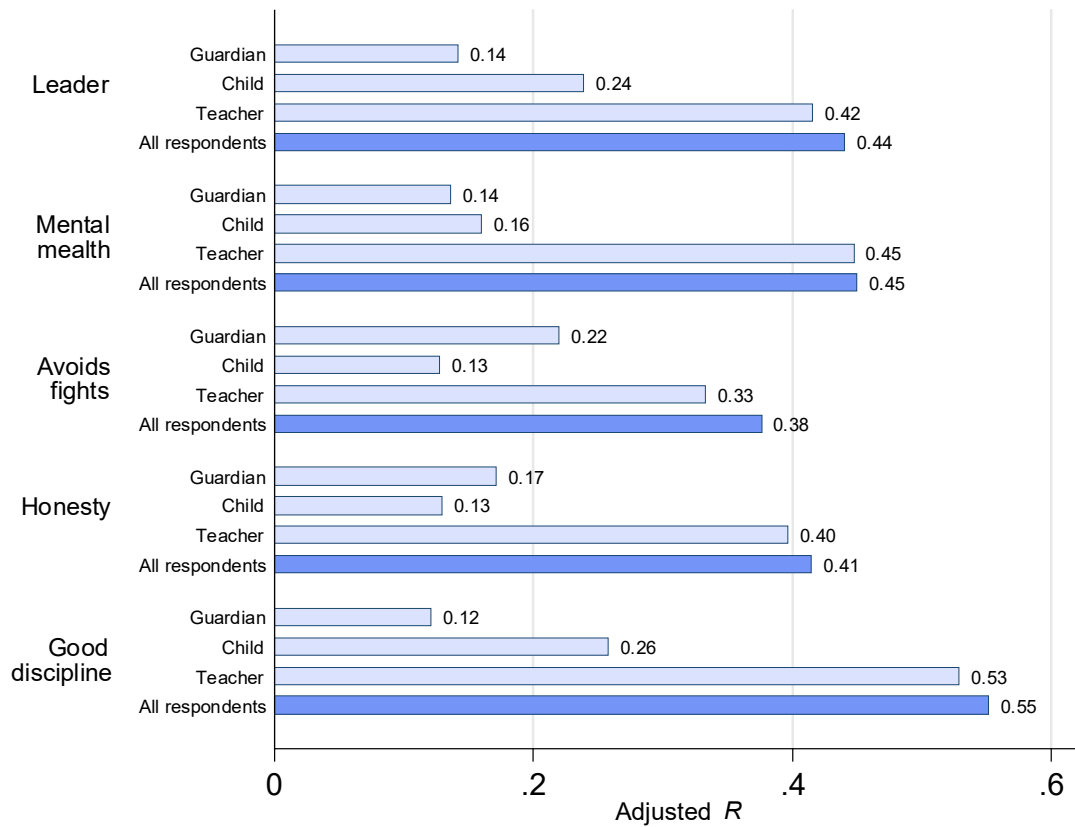
Notes: These rules summarize how responses on children’s reports of the Big Five were adjusted using the anchoring vignettes when children responded consistently on the vignettes. The “Original response” refers to children’s response on the unadjusted Big Five items, which ranges from one to five. The vignette responses refer to children’s rating of hypothetical children on the same one- to five-point scale.

**Figure S1.** Predictive power (adjusted  $R$ ) of groups of noncognitive skills for cognitive outcomes measured 1 y later for children rated by the same teachers and same guardians



Notes: The figure shows the square root of adjusted  $R^2$  from OLS regressions of cognitive outcomes on the group of Big Five noncognitive skills for each respondent type. IQ is measured using a 60-item version of Raven's Standard Progressive Matrices. Academic test scores are normalized by grade level. The morality test assesses whether students can analyze objectively and make reasonable judgements about different issues, including caring about others, confidence, perseverance, respect for others, responsibility, and integrity. Overall academic performance is evaluated on a one- to five-point scale, with one corresponding to "very bad" and five corresponding to "very good." The full sample includes 1,177 observations who are rated by the same teachers and same guardians in both years.

**Figure S2.** Predictive power (adjusted  $R$ ) of groups of noncognitive skills for behavioral outcomes in school measured 1 y later for children rated by the same teachers and same guardians



Notes: The figure shows the square root of the adjusted  $R^2$  from the OLS regressions of behavioral measures on the group of Big Five noncognitive skills for each respondent type. Leader is a dummy for being elected as a class or school leader. Mental health and good discipline are evaluated on a one- to five-point scale, with one corresponding to “very bad” and five corresponding to “very good.” Avoiding fights and honesty are evaluated on a one- to five-point scale, with one corresponding to “never” and five corresponding to “often.” The full sample includes 1,177 observations who are rated by the same teachers and same guardians in both years.

## SI References

1. C. J. Soto, O. P. John, The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J. Pers. Soc. Psychol.* 113, 117–143 (2017).
2. J. Raven, J. C. Raven, J. H. Court, “The Standard Progressive Matrices” in *Manual for Raven’s Progressive Matrices and Vocabulary Scales* (Oxford Psychologists Press, Oxford, UK, 2000).