Supporting Information

Duckworth et al. 10.1073/pnas.1018601108

SI Materials and Methods

Study 1. ***Coded variables. Age.*** Participant age was frequently reported as a range of grade levels or years, so we treated age as a categorical variable where 1 = 5 y old or younger, 2 = 6–11 y, 3 = 12–18 y, and 4 = 19 y or older.

Article type. We recorded type of article (dissertation vs. published article).

Intelligence quotient (IQ) score. We recorded control and experimental group intelligence-test means and SD. Where available, baseline scores were also recorded.

IQ measure. Measures of intelligence included the Lorge–Thorndike Intelligence Test (LTIT), the McCarthy Scales of Children's Abilities (MSCA), the Otis–Lennon Mental Ability Test (OLMAT), the Otis Self-Administering Test (OSAT), the Peabody Picture Vocabulary Test (PPVT), Raven's Progressive Matrices (RPM), the Stanford–Binet Intelligence Scale, Third Revision (SBIS-III), the Wechsler Adult Intelligence Scale (WAIS), the Wechsler Intelligence Scale for Children (WISC), the Wechsler Intelligence Scale for Children–Revised (WISC-R), and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). Although most measures of intelligence were administered individually, the OSAT, LTIT, and OLMAT were administered in group settings.

Incentive size. To take into account the effect of inflation on the value of incentives, we converted incentives to 2007 dollar equivalents. We treated the value of incentives participants received for their intelligence-test performance as a categorical variable where 1 = less than $1, 2 = between $1 and $9, and 3 = $10 or more.

Study design. We recorded whether the study design included baseline and posttest IQ measures or only a posttest measure.

Effect-size analyses. Where reported, means and SD were used to calculate g (i.e., the bias-corrected standardized mean intelligence score difference between control and incentive groups). When these values were not available, we calculated effect sizes from t scores or from raw data. When pretest and posttest scores were available, we calculated g as the difference between control and incentive groups on mean change scores. When only posttest scores were reported, we calculated g as the difference between control and incentive group means at posttest. Of the 25 articles included in the meta-analysis, 14 presented results for more than one sample. We treated individual samples as the basic unit of our analyses, yielding a k of 46 samples and an aggregated sample size of n = 2,008. When individual samples were tested on multiple measures of IQ, we computed a mean effect across measures within sample for use in our analyses.

Moderator analyses. The mixed-effects model relies on the same assumptions as the random-effects model within each level of the moderator. That is, at each level of the moderator, there is random variation in the distribution of effect sizes across studies. However, in comparing samples across levels of the moderator, the mixed-effects model assumes that the moderator is associated with systematic differences among effect sizes (1). Unfortunately, we were unable to test for moderation by type of IQ measure because we lacked a sufficient number of studies using each type of test. We were also unable to test for moderation by administration type (individual vs. group) because we lacked a sufficient number of group-administered tests.

Publication bias. Three of four tests indicated no publication bias. First, we followed Borenstein et al. (1) in testing the relationship between sample size and effect size. Studies of smaller sample size, and therefore larger SE, should have more effect-size variability among them. However, if studies of larger SE and smaller effect size are not published (i.e., are deposited in the file drawer), there will be a weaker relationship between large SE and variability. Egger's (2) regression intercept was not significant (0.01, P = 0.99), suggesting a lack of publication bias. Second, we conducted two fail-safe N analyses. Rosenthal's (3) fail-safe N indicated that an additional 1,885 samples of average effect size g = 0 would be required to eliminate the significance of the mean effect. Similarly, according to Orwin's (4) fail-safe N, an additional 101 samples of effect size g = 0 would be needed to reduce the medium-to-large effect we found to a small effect of g = 0.2. Third, we conducted Duval and Tweedie's (5) trim and fill. The trim and fill, which adjusts the distribution of effect sizes to account for bias, found that no adjustment was necessary.

Only one of four analyses suggested a bias in the included samples. Specifically, there was evidence that article type (published article vs. dissertation) moderated the effect of incentives on IQ scores. In a mixed-effects analysis, the mean effect size in 33 published samples, g = 0.76 [95% confidence interval (CI) = 0.46, 1.05], was significantly larger than that in 13 dissertation samples, g = 0.21 (95% CI = −0.07, 0.49), $Q_{\text{between}}(1) = 7.02$, P = 0.01. However, because we were unable to control for the simultaneous effects of other moderators, this test was of limited utility. In particular, it is possible that the lower effect size in dissertation samples can be explained by the higher proportion of high-IQ samples acquired from dissertations (69%) than from published articles (42%).

Study 2. ***Participants and procedure.*** Participants were members of the middle sample of the Pittsburgh Youth Study, a longitudinal study of boys randomly selected from public schools in 1987–1988 in Pittsburgh, PA. Details on the initial recruitment and screening of the middle sample when children (all male) were aged 10 y are given in Loeber et al. (6). Briefly, of the families contacted, 85% of the boys and their parents agreed to participate. Approximately 50% of the sample (n = 508) was identified as at-risk based on prior evidence of disruptive behavior problems. Of the total sample, 54% was black, 43% was white, and the remaining 3% were Hispanic, Asian, or of mixed ethnicity.

The men of the Pittsburgh Youth Study who participated in follow-up interviews during young adulthood (n = 251) did not differ in IQ from those who did not participate: $t(427) = 1.73$, $P = 0.09$, $d = 0.17$. At follow-up, participants did not differ from nonparticipants on years of education [$t(295) = 0.05$, $P = 0.96$, $d = 0.01$], current employment [$\chi^2(1) = 1.79$, $P = 0.18$], or recent history of unemployment [$OR = 0.65$, $P = 0.16$]. However, participants were rated higher in test motivation; $t(418) = 3.03$, $P = 0.003$, $d = 0.30$. Participants also performed better in school during adolescence [$t(505) = 3.03$, $P = 0.003$, $d = 0.27$], had fewer criminal convictions by age 26 y ($IRR = 0.60$, $P < 0.001$), were higher on the Hollingshead (7) two-factor socioeconomic status (SES) index at age 12.5 y [$t(481) = 2.91$, $P = 0.003$, $d = 0.27$], and were more likely to be Caucasian [$\chi^2(1) = 17.33$, $P < 0.001$] and from two-parent homes [$\chi^2(1) = 19.01$, $P < 0.001$].

Measures. IQ. At average age 12.5 y, participants completed a short form of the WISC-R (8) in which all 12 subtests were administered, but individual subtests were shortened by administering every other item. This procedure follows those used by Hobby (9) and Yudin (10) who reported correlations between the short form and full intelligence test of r = 0.97 for full-scale IQ scores. The reliability of the WISC-R short form full-scale IQ score has been estimated at 0.905 (11). Trained testers administered the following set of instructions to each participant individually: "I'll be asking you to try a lot of different questions and puzzles. Some are like school work, most are not. Each task will only last about 3 minutes, so if you don't enjoy a task, don't worry, we will be switching to a different task soon. Each task asks you to do something different, because everybody has things they do well and things they don't do so well, and we want everyone to have a chance to succeed and have fun. Each task starts out easy and gets harder, and the questions go all the way up to college level for kids much older than you, so don't be surprised when you get some wrong. It is important to do your very best, the very best you can."

Test motivation. Three different raters coded 15 min of videotaped behavior during the intelligence test (12). Raters were blind to both the boys' risk status and the hypotheses of the study. The raters were trained to consensus (20 h) to identify behaviors that indicated low motivation. Raters gave each boy a single rating by using a standardized coding system where 3 = severe, 2 = moderate, 1 = mild, and 0 = absent. Scores were standardized within rater and then averaged across all three raters. Intraclass correlations for each set of raters ranged from 0.85 to 0.89. We reverse-scored test motivation so that higher scores indicated more motivation and used a natural log transformation to correct for right skew before completing analyses.

Demographics. Four demographic variables were included as covariates in all analyses: race (1 = black vs. 0 = white or other ethnicity), family structure (0 = two-parent vs. 1 = not), family SES, and age at the follow-up interview. Family SES was assessed by using Hollingshead's (7) two-factor index. If a boy had both a male and female parent or caretaker, scores were averaged; if he only had one caretaker, that score was used.

School performance in adolescence. Every fall and spring from age 10–13 y, teachers of participants completed the Teacher Report Form, the teacher's version of the Child Behavior Checklist (13). Four items on the Teacher Report Form inquired about the boy's performance in reading, writing, spelling, and math using a 5-point scale where 1 = far below his grade level and 5 = far above his grade level. At each assessment point, a summary score was computed as the average of these four items. The reliability of each summary score was high, α values ranged from 0.93 to 0.96. The average correlation among these summary scores over time was $r = 0.62$, suggesting reasonable cross-time stability. For each participant, a composite school performance score was computed as the mean of summary scores from age 10–13 y.

Years of education. At follow-up interviews in young adulthood, participants reported the highest grade level of education completed.

Employment. Participants reported whether they were working for pay at the time of the follow-up interviews in young adulthood.

Lifetime criminal convictions. Records of criminal offenses through age 26 y were obtained from official sources (i.e., local, state, and federal criminal history records). As an objective measure of criminal behavior, we recorded the total number of lifetime convictions.

Data analytics strategy. We analyzed the Pittsburgh Youth Study by using structural equation modeling in the software program MPlus Version 5 (14). We created reliability-corrected latent intelligence and nonintellective trait factors by regressing observed IQ scores and test motivation onto the respective factors and specifying that the error variances for IQ and test motivation equaled 1 minus their reliabilities. The variances of latent intelligence and nonintellective traits were set to 1 to standardize the factors and facilitate model identification. In all models, we controlled for demographics by regressing all outcomes onto race, family structure, family SES, and age at the early-adult interview and permitting demographics to covary with the intelligence and nonintellective factors. Additionally, we permitted the residual variances of all early-adult outcomes to freely covary and permitted the residual variance of adolescent academic performance to covary with the residual variances of early-adult cumulative years of education and current employment. Because current employment in early adulthood was a binary variable, we used a robust weighted least-squares estimator (WLSMV) (14). Model comparisons for nested models used χ^2 difference tests adjusted for use with the WLSMV estimator, and we additionally examined the comparative fit index (CFI) and root mean square error of approximation (RMSEA) to assess goodness-of-fit (15).

1. Borenstein M, Hedges L, Higgins J, Rothstein H (2009) *Introduction to Meta-Analysis* (Wiley, West Sussex, UK).
2. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315:629–634.
3. Rosenthal R (1979) The "file drawer problem" and tolerance for null results. *Psychol Bull* 86:638–641.
4. Orwin RG (1983) A fail-safe N for effect size in meta-analysis. *J Educ Stat* 8:157–159.
5. Duval S, Tweedie R (2000) Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56:455–463.
6. Loeber R, Farrington DP, Stouthamer-Loeber M, Van Kammen WB (1998) *Antisocial Behavior and Mental Health Problems: Explanatory Factors in Childhood and Adolescence* (Lawrence Erlbaum Associates, Mahwah, NJ).
7. Hollingshead AB (1975) *Four Factor Index of Social Status* (Yale University, New Haven, CT).
8. Wechsler D (1974) *Wechsler Intelligence Scale for Children–Revised* (The Psychological Corporation, San Antonio, TX).
9. Hobby KL (1980) *WISC-R Split-Half Short Form Manual* (Western Psychological Services, Los Angeles).
10. Yudin LW (1966) An abbreviated form of the WISC for use with emotionally disturbed children. *J Consult Psychol* 30:272–275.
11. Silverstein AB (1990) Notes on the reliability of Wechsler short forms *J Clin Psychol* 46: 194–196.
12. Lynam D, Moffitt TE, Stouthamer-Loeber M (1993) Explaining the relation between IQ and delinquency: Class, race, test motivation, school failure, or self-control? *J Abnorm Psychol* 102:187–196.
13. Achenbach TM, Edelbrock CR (1983) *Manual for the Child Behavior Profile and Child Behavior Checklist* (Queen City Printers, Burlington, VT).
14. Muthén LK, Muthén BO (2007) *MPlus User's Guide* (Muthén & Muthén, Los Angeles), 5th Ed.
15. Kline RB (2005) *Principles and Practice of Structural Equation Modeling* (Guilford Press, New York).

Table S1. Between-subjects studies of the effect of incentives on IQ scores in Study 1

Study	Ref.	Study type	Sample	N	Intelligence test	Raw effect size, *g* (SE)	Baseline IQ	Age, y	Incentive value	Study design
Benton (1936)	1	Article	—	50	OSAT	−0.09 (0.28)	Low	12–18	—[a]	Baseline and posttest
Bergan et al. (1971)	2	Article	Male	16	WISC	0.19 (0.47)	High	6–11	<$1	Baseline and posttest
			Female	16	WISC	−0.06 (0.48)	High	6–11	<$1	Baseline and posttest
Blanding et al. (1994)	3	Article	Study 1	29	MSCA	0.82 (0.38)	High	0–5	$1–9	Baseline and posttest
			Study 2 low SES	23	MSCA	1.33 (0.45)	Low	0–5	$1–9	Posttest only
			Study 2 high SES	20	MSCA	0.44 (0.43)	High	0–5	$1–9	Posttest only
Bradley-Johnson et al. (1984)	4	Article	Study 1	22	WISC-R	0.17 (0.41)	Low	6–11	$1–9	Baseline and posttest
			Study 2	22	WISC-R	0.67 (0.42)	High	6–11	$1–9	Baseline and posttest
Bradley-Johnson et al. (1986)	5	Article	Early elementary	20	WISC-R	0.85 (0.45)	High	6–11	$10+	Baseline and posttest
			Late elementary	20	WISC-R	0.80 (0.45)	High	6–11	$10+	Baseline and posttest
Breuning and Zella (1978)	6	Article	Group 1	147	LTIT	2.12 (0.21)	Low	12–18	$10+	Baseline and posttest
			Group 2	129	OLMAT	2.38 (0.23)	Low	12–18	$10+	Baseline and posttest
			Group 3	209	WISC-R	1.94 (0.17)	Low	12–18	$10+	Baseline and posttest
Clingman and Fowler (1976)	7	Article	High IQ	16	PPVT	−0.06 (0.47)	High	6–11	$1–9	Baseline and posttest
			Medium IQ	16	PPVT	−0.40 (0.48)	High	6–11	$1–9	Baseline and posttest
			Low IQ	16	PPVT	1.42 (0.54)	Low	6–11	$1–9	Baseline and posttest
Devers and Bradley-Johnson (1994)	8	Article	—	25	WISC-R	0.91 (0.41)	Low	12–18	$10+	Baseline and posttest
Dickstein and Ayers (1973)	9	Article	—	32	RPM, WAIS	0.49 (0.35)	High	19+	$1–9	Posttest only
Edlund (1972)	10	Article	—	22	SBIS-III	0.89 (0.43)	Low	6–11	$1–9	Baseline and posttest
Ferguson (1937)	11	Article	—	156	OSAT	0.03 (0.16)	High	12–18	<$1	Baseline and posttest
Galbraith et al. (1986)	12	Article	—	30	WISC-R	0.73 (0.37)	Low	6–11	$1–9	Baseline and posttest
Gerwell (1981)	13	Dissertation	—	64	WAIS	0.58 (0.25)	High	19+	$1–9	Posttest only
Graham (1971)	14	Dissertation	—	128	WPPSI	−0.14 (0.18)	Low	0–5	$1–9	Baseline and posttest
Holt and Hobbs (1979)	15	Article	—	40	WISC	1.03 (0.33)	Low	12–18	—[a]	Posttest only
Kapenis (1979)	16	Dissertation	Low SES	28	PPVT	0.51 (0.37)	High	6–11	<$1	Posttest only
			Middle SES	28	PPVT	−0.17 (0.37)	High	6–11	<$1	Posttest only
			High SES	28	PPVT	0.09 (0.37)	High	6–11	<$1	Posttest only
Kieffer and Goh (1981)	17	Article	Low SES	32	WISC-R	0.66 (0.35)	Low	6–11	$1–9	Posttest only
			Middle SES	32	WISC-R	−0.26 (0.35)	High	6–11	$1–9	Posttest only
Lloyd and Zylla (1988)	18	Article	High IQ	16	WPPSI	1.04 (0.51)	High	0–5	$1–9	Baseline and posttest
			Low IQ	16	WPPSI	0.62 (0.49)	Low	0–5	$1–9	Baseline and posttest
Saigh and Antoun (1983)	19	Article	—	34	WISC-R	0.93 (0.36)	Low	12–18	—[a]	Posttest only
Steinweg (1979)	20	Dissertation	Group 1	10	SBIS-III	0.17 (0.57)	High	6–11	$1–9	Posttest only
			Group 2	10	WISC-R	0.20 (0.57)	High	6–11	$1–9	Posttest only
Sweet and Ringness (1971)	21	Article	Low SES black	36	WISC	0.39 (0.33)	Low	6–11	$1–9	Posttest only
			Low SES white	48	WISC	1.17 (0.31)	Low	6–11	$1–9	Posttest only
			Middle SES white	72	WISC	0.13 (0.23)	High	6–11	$1–9	Posttest only
Terrell et al. (1980)	22	Article	Black examiner	30	WISC-R	1.18 (0.39)	Low	6–11	<$1	Posttest only
			White examiner	30	WISC-R	1.40 (0.40)	Low	6–11	<$1	Posttest only
Tiber (1963)	23	Dissertation	Low SES black	80	SBIS-III	0.11 (0.22)	Low	6–11	<$1	Posttest only
			Low SES white	80	SBIS-III	−0.34 (0.22)	Low	6–11	<$1	Posttest only
			Middle SES white	80	SBIS-III	0.00 (0.22)	High	6–11	<$1	Posttest only
Weiss (1981)	24	Dissertation	High IQ	10	PPVT	0.57 (0.59)	High	0–5	$1–9	Baseline and posttest
			Medium IQ	10	PPVT	1.25 (0.64)	High	0–5	$1–9	Baseline and posttest
			Low IQ	10	PPVT	3.64 (1.00)	Low	0–5	$1–9	Baseline and posttest
Willis and Shibata (1978)	25	Article	—	20	WPPSI	0.59 (0.44)	Low	0–5	$1–9	Baseline and posttest

[a]There was insufficient information presented to determine the value of the incentive for these samples.

1. Benton AL (1936) Influence of incentives upon intelligence test scores of school children. *J Genet Psychol* 49:494–497.
2. Bergan A, McManis DL, Melchert PA (1971) Effects of social and token reinforcement on WISC block design performance. *Percept Mot Skills* 32:871–880.
3. Blanding KM, Richards J, Bradley-Johnson S, Johnson CM (1994) The effect of token reinforcement on McCarthy Scale performance for White preschoolers of low and high social position. *J Behav Educ* 4:33–39.
4. Bradley-Johnson S, Johnson CM, Shanahan RH, Rickert VI, Tardona DR (1984) Effects of token reinforcement on WISC-R performance of Black and White, low socioeconomic second graders. *Behav Assess* 6:365–373.
5. Bradley-Johnson S, Graham DP, Johnson CM (1986) Token reinforcement on WISC–R performance for White, low-socioeconomic, upper and lower elementary-school-age students. *J Sch Psychol* 24:73–79.
6. Breuning SE, Zella WF (1978) Effects of individualized incentives on norm-referenced IQ test performance of high school students in special education classes. *J Sch Psychol* 16:220–226.
7. Clingman J, Fowler RL (1976) The effects of primary reward on the I.Q. performance of grade-school children as a function of initial I.Q. level. *J Appl Behav Anal* 9:19–23.
8. Devers R, Bradley-Johnson S (1994) The effect of token reinforcement on WISC-R performance for fifth- through ninth-grade American Indians. *Psychol Rec* 44:441–449.
9. Dickstein LS, Ayers J (1973) Effect of an incentive upon intelligence test performance. *Psychol Rep* 33:127–130.
10. Edlund CV (1972) The effect on the test behavior of children, as reflected in the I.Q. scores, when reinforced after each correct response. *J Appl Behav Anal* 5:317–319.
11. Ferguson HH (1937) Incentives and an intelligence test. *Australasian J Psychol Philos* 15:39–53.
12. Galbraith G, Ott J, Johnson CM (1986) The effects of token reinforcement on WISC-R performance of low-socioeconomic Hispanic second-graders. *Behav Assess* 8:191–194.
13. Gerwell EL (1981) Tangible and verbal reinforcement effects on fluid and crystallized intelligence in the aged. PhD dissertation (Hofstra University, Hempstead, NY).
14. Graham GA (1971) The effects of material and social incentives on the performance on intelligence test tasks by lower class and middle class Negro preschool children. PhD dissertation (George Washington University, Washington, DC).
15. Holt MM, Hobbs TR (1979) The effects of token reinforcement, feedback and response cost on standardized test performance. *Behav Res Ther* 17:81–83.
16. Kapenis JT (1979) The differential effects of various reinforcements and socioeconomic status upon Peabody Picture Vocabulary Test performance. PhD dissertation (University of South Dakota, Vermillion, SD).
17. Kieffer DA, Goh DS (1981) The effect of individually contracted incentives on intelligence test performance of middle- and low-SES children. *J Clin Psychol* 37:175–179.
18. Lloyd ME, Zylla TM (1988) Effect of incentives delivered for correctly answered items on the measured IQs of children of low and high IQ. *Psychol Rep* 63:555–561.
19. Saigh PA, Antoun FT (1983) WISC-R incentives and the academic achievement of conduct disordered adolescent females: A validity study. *J Clin Psychol* 39:771–774.
20. Steinweg SB (1979) A comparison of the effects of reinforcement on intelligence test performance of normal and retarded children. PhD dissertation (University of North Carolina, Chapel Hill, NC).
21. Sweet RC, Ringness TA (1971) Variations in the intelligence test performance of referred boys of differing racial and socioeconomic backgrounds as a function of feedback or monetary reinforcement. *J Sch Psychol* 9:399–409.
22. Terrell F, Terrell SL, Taylor J (1980) Effects of race of examiner and type of reinforcement on the intelligence test performance of lower-class Black children. *Psychol Sch* 17:270–272.
23. Tiber N (1963) The effects of incentives on intelligence test performance. PhD dissertation (Florida State University, Tallahassee, FL).
24. Weiss RH (1981) Effects of reinforcement on the IQ scores of preschool children as a function of initial IQ. PhD dissertation (Utah State University, Logan, UT).
25. Willis J, Shibata B (1978) A comparison of tangible reinforcement and feedback effects on the WPPSI I.Q. scores of nursery school children. *Educ Treat Child* 1:31–45.

Table S2. Moderation of effect of incentives by intelligence level in Study 1

Intelligence level	*g*	95% CI of *g*	*k*	*n*	I^2	$Q_{between}$(df)
Low	0.94**	0.54, 1.35	23	1257	90.28**	
High	0.26**	0.10, 0.41	23	751	9.71	
Between-groups comparison						9.76* (1)

*$P < 0.01$; **$P < 0.001$.

Table S3. Summary statistics and zero-order correlations in Study 2

Variable	M	SD	2	3	4	5	6	7	8	9	10[b]
i) Test motivation[a]	0.09	0.84	0.28***	−0.01	−0.04	−0.08	0.15*	0.31***	0.24***	0.21**	−0.18**
ii) IQ	101.80	15.77	—	−0.47***	−0.34***	−0.21**	0.37***	0.70***	0.47***	0.21***	−0.42***
iii) Age at follow-up interview	24.02	0.91		—	0.19**	0.17**	−0.27***	−0.29***	−0.24***	−0.10	0.30***
iv) Black	44%				—	0.35***	−0.17**	−0.21***	−0.13	−0.14	0.23***
v) Single parent home	56%					—	−0.22***	−0.30***	−0.23***	−0.14*	0.31***
vi) Family SES	38.13	11.67					—	0.25***	0.28***	0.02	−0.25***
vii) Academic performance	2.76	0.81						—	0.52***	0.21***	−0.26***
viii) Years of education	12.36	2.02							—	0.28***	−0.53***
ix) Employed at follow-up	72%									—	−0.40***
x) Lifetime convictions	3.33	6.08									—

n values range from 223 to 251. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
[a]Correlations are with natural log-transformed test motivation.
[b]Spearman's ρ correlation coefficients.

Table S4. Summary of final model predicting life outcomes in Study 2

Predictor	Academic performance in adolescence	Total years of education	Employment in adulthood	Lifetime convictions, age 26 y
Race	0.12*	0.10	−0.09	0.05
Single-parent home	−0.20***	−0.15*	−0.12	0.10
Family SES	−0.07	0.08	−0.13	0.02
Age at follow-up	0.09	0.02	−0.04	0.18**
Intelligence	0.74***	0.43***	0.18*	−0.18*
Nonintellective traits	0.34***	0.24***	0.29***	−0.24***
Proportion of variance explained (R^2)	0.60	0.29	0.16	0.20

Values are standardized regression coefficients. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Other Supporting Information Files

Dataset S1 (XLS)