Table S1. Fit Indices for Confirmatory Factor Analysis With Maximum Likelihood Estimation

	Time 1		Time 2	
	Two factors	One factor	Two factors	One factor
AIC[+]	15920	16026	7324	7416
BIC[+]	16050	16150	7426	7514
CFI[+]	0.968	0.901	0.956	0.828
TLI[+]	0.953	0.861	0.929	0.743
RMSEA[+]	0.045	0.078	0.056	0.090
χ^2 difference test[++]	1284.3 (df=6), p<0.001		709,7 (df=6) p<0.001	

[+]Lower values in the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are indicative of better model fit for the two-factor solution. The comparative-fit index (CFI) compares the fit of the proposed model with that of a null model (which assumes no relationship between the observed variables) (Raykov, 2006a). CFI values "in the mid- .90s or above" (Raykov, 2006a) suggest plausible approximations to the data. As can be seen values above 0.95 are only achieved for the two factor model. The Tucker-Lewis Index (TLI) follows a similar principle to that of the CFI and models with TLI close to 1 indicate a good fit (Loehlin, 2004). As can be seen, the two factor models are consistently closer to that. The fit of the proposed model with that of a null model also underlies the logic of the root mean square error of approximation (RMSEA) fit index. An RMSEA of 0.05 or less is considered a good approximation to the data (Raykov, 2006b).

[++]This test, implemented in MplusVersion 5, compares the fit of the two to that of the one factor model. The highly significant result indicates that the two-factor model fits the data better than the one-factor solution.

Table S2. Longitudinal Phenotypic Predictions From Irritability (Excluding the *Stubborn* Item) and Headstrong/Hurtful Behaviors to Dimensional Outcomes of Psychopathology

	Depression (Time 2)	Delinquency (Time 2)
Irritability (time 1)	**.14*****	.06*
	(.09 to .20)	(.00 to .11)
Headstrong/ hurtful (time 1)	.01	**.19*****
	(-.05 to .07)	(.12 to .27)

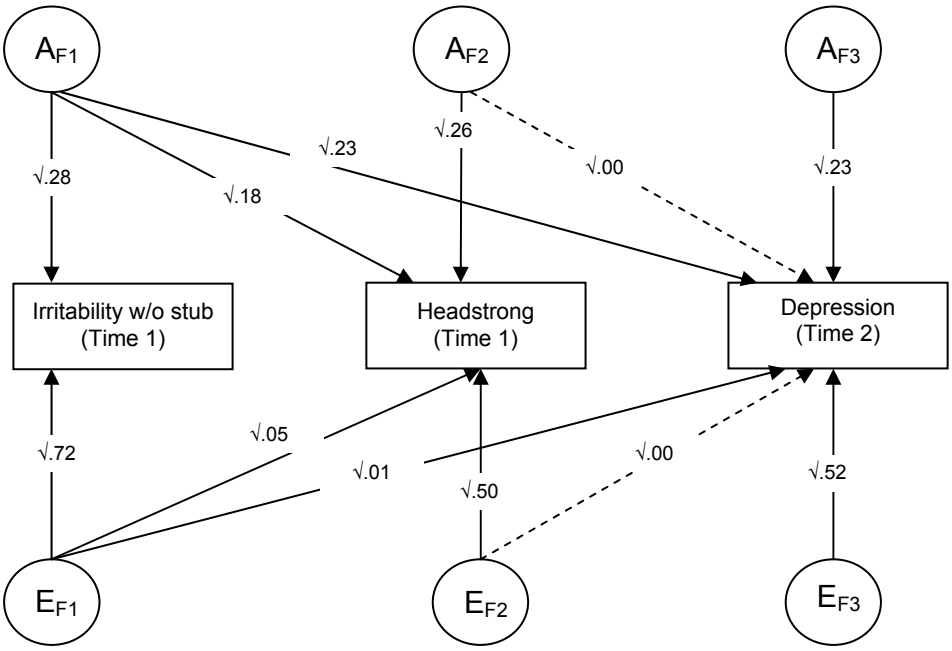
This table corresponds to the analyses presented in Table 2 of the main paper, except that the irritability scale here does not include the *stubborn* item.

Table S3. Cross-Sectional Multivariate Genetic Model Results at Time 2

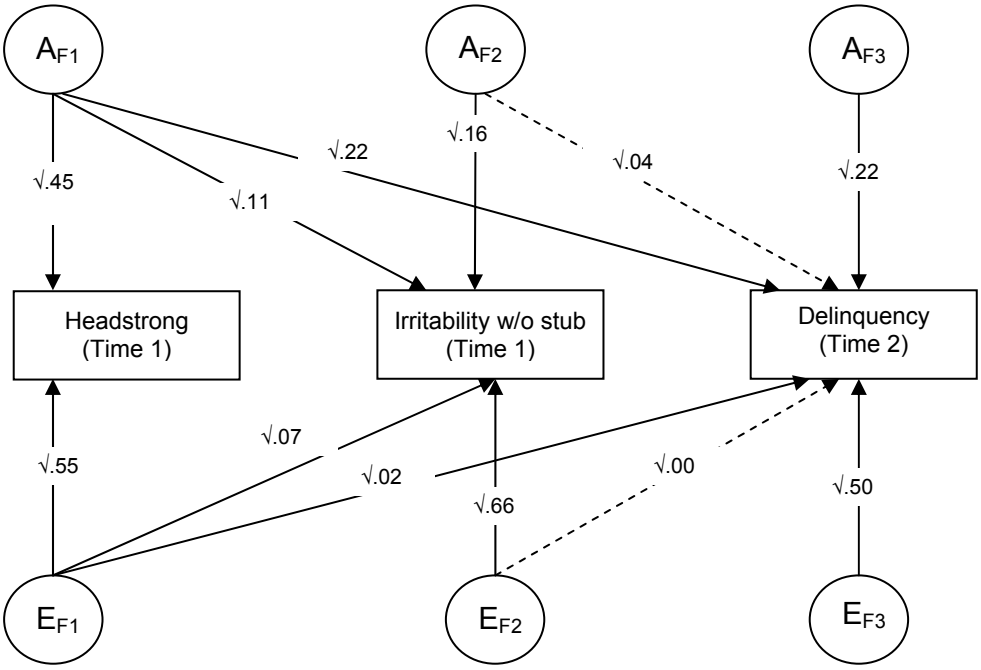
Cholesky Model	Irritability	Headstrong/Hurtful	Depressive Ratings	Delinquency
Genetic influences				
Irritability	**0.35 (0.25-0.44)**	0.66 (0.47-0.84)	0.76 (0.63-0.90)	0.68 (0.53-0.82)
Headstrong/hurtful		**0.34 (0.24-0.44)**	0.44 (0.26-0.61)	0.72 (0.58-0.85)
Depression ratings			**0.46 (0.37-0.54)**	0.58 (0.45-0.71)
Delinquency				**0.46 (0.37-0.55)**
Nonshared environmental influences				
Irritability	**0.65 (0.56-0.75)**	0.27 (0.17-0.36)	0.35 (0.25-0.44)	0.28 (0.17-0.37)
Headstrong/hurtful		**0.66 (0.56-0.76)**	0.20 (0.09-0.30)	0.41 (0.32-0.50)
Depression ratings			**0.54 (0.47-0.63)**	0.23 (0.12-0.33)
Delinquency				**0.54 (0.45-0.63)**

This table corresponds to Table 3 of the main manuscript, except that cross-sectional analyses from time 2, rather than time 1, are presented.

Figure S1a. Longitudinal Cholesky decomposition with depression as part of the model. The irritability scale does not include the *stubborn* item.



Suppl Figure 1b. Longitudinal Cholesky decomposition with delinquency as part of the model. The irritability item does not include the *stubborn* item.



These figures correspond to Figure 2 and b of the main manuscript, except that the *stubborn* item has been removed from the irritability scale.

References

Loehlin JC: Goodness of Fit Indices: Latent Variable Models, 4th ed. Mahwah, NJ, Lawrence Erlbaum Associates, 2004

Raykov T, Marcoulides GA: Parameter and Model Identification: A First Course in Structural Equation Modeling. Mahwah, NJ, Lawrence Erlbaum Associates, 2006a

Raykov T, Marcoulides GA: Confirmatory Factor Analysis: A First Course in Structural Equation Modeling. Mahwah, NJ, Lawrence Erlbaum Associates, 2006b