

ARTICLE OPEN



The relationship of school performance with self-control and grit is strongly genetic and weakly causal

Sofieke T. Kevenaar^{1,2,3,6}, Elsje van Bergen^{1,2,3,6}✉, Albertine J. Oldehinkel⁴, Dorret I. Boomsma^{1,3,5} and Conor V. Dolan¹

The non-cognitive skills self-control and grit are often considered predictors of school performance, but whether this relationship is causal remains unclear. We investigated the causality of this association using a twin design. Specifically, we evaluated the direct impact of self-control and grit on school performance, while controlling for genetic or environmental influences common to all three traits (i.e., confounding). Teachers of 4891 Dutch 12-year-old twin pairs (of which 3837 were complete pairs) completed a survey about school performance (school grades), self-control (ASEBA self-control scale), and the perseverance aspect of grit. Our analysis aimed to determine the direct impact of self-control and grit on school performance, while simultaneously controlling for genetic or environmental confounding. Establishing the regression relationship corrected for confounding supports the interpretation of the regression relationship as causal. In all analyses, we corrected for sex, rater bias of the teachers, and parental socioeconomic status. Initially, in the standard regression, self-control, and grit explained 28.4% of the school performance variance. However, allowing for genetic confounding (due to genetic pleiotropy) revealed that most of this association could be attributed to genetic influences that the three traits share. In the presence of genetic pleiotropy, the phenotypic regression of school performance on self-control and grit accounted for only 4.4% (i.e., the effect size association with the causal hypothesis). In conclusion, self-control and grit predict school performance primarily due to genetic pleiotropy, with a much smaller causal effect ($R^2 = 4.4\%$). This suggests that interventions targeting self-control and grit alone may yield limited improvements in school performance.

npj Science of Learning (2023)8:53; <https://doi.org/10.1038/s41539-023-00198-3>

INTRODUCTION

Much research has focused on self-control and grit as predictors of school and academic performance. Grit comprises consistency of interest and perseverance, and self-control is the ability to self-regulate conflicting impulses (Duckworth et al., 2016). Perseverance and self-control have generally been found to be associated with school and academic performance^{1–7}. For instance, Lam & Zhou⁵ reported an average correlation of 0.17 between grit and school performance in school children (based on 56 correlations). The average correlation was 0.14 in students in higher education (based on 60 effect sizes; see also Fernández Martín et al.⁸). In a recent study of Czech school children, Vazsonyi, et al.⁹ found that self-control predicted school performance (both teacher-rated and grades) while controlling for motivation and intelligence. In a twin study, Kevenaar et al.¹⁰ found that self-control and grit together explained 28.4% of the variance in school performance. The decomposition of the phenotypic regression relationship into genetic and environmental components revealed that the phenotypic associations were mainly due to genetic influences common to self-control, grit, and school performance.

While it is well established that self-control and grit predict academic outcomes, most studies tend not to address causality^{2,11,12}. One of the few studies that provide a basis for a causal interpretation regarding self-control was conducted by Duckworth et al.¹³, who showed that within-individual changes in self-control over time predict changes in academic achievement, but not vice versa, which suggests a causal effect only from self-control to achievement. A small ($N = 53$) intervention study on self-

regulation indicated that self-regulation training affected math performance, which is also consistent with a causal effect of self-regulation¹⁴. Regarding grit, Jiang et al.¹⁵ followed 193 children longitudinally, and found reciprocal effects between grit and academic achievement, consistent with a reciprocal causal relation. Achievement was found to be related to the perseverance, but not to the consistency facet of grit. Postigo et al.¹⁶ studied a large sample of children ($N = 5371$) longitudinally from age 10 to 14. They reported an effect of grit on school performance (grades) in a two-occasion panel model, which is consistent with a causal model. Hence, as far as it has been studied, most research suggests a causal effect of self-control and grit on school performance, rather than the other way around.

The interpretation of the effects of self-control and grit on school performance as causal is appealing, as it is plausible that these non-cognitive factors facilitate school or academic performance. However, more research, employing different designs, is needed to establish causal pathways, and to rule out possible—correlational, non-causal—sources of association. Such non-causal sources may be both genetic and environmental. For instance, the association may be due to common genetic influences (pleiotropy: the same genes affect multiple traits; MacKay¹⁷), or due to a rearing environment that is conducive to both cognitive and non-cognitive influences on school performance. It is also important to take note of the challenging possibility that causal and non-causal accounts of the associations are not mutually exclusive.

In contrast to our previous work, which addressed prediction (Kevenaar et al.¹⁰) the current study addressed causation.

¹Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. ²Research Institute LEARN!, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ³Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. ⁴University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁵Amsterdam Reproduction and Development Research Institute, Amsterdam, the Netherlands. ⁶These authors contributed equally: Sofieke T. Kevenaar, Elsje van Bergen. ✉email: e.van.bergen@vu.nl

Kevenaar et al.¹⁰ examined the phenotypic regression relationship of self-control and grit with school performance, and decomposed these into genetic and environmental components. Thus, Kevenaar et al.¹⁰ focussed on the genetic and environmental contributions to the prediction of school performance, but did not address causality. In contrast, the present focus is on establishing the direct phenotypic effects of self-control and grit on school performance, corrected for genetic or environmental confounding. Demonstrating the direct phenotypic regression relationships corrected for confounding supports the interpretation of these relationships as causal. In the current article, we analyze the same cohort data, but fitted causal models.

The classical twin design provides a means to estimate the putatively causal phenotypic relationships between the non-cognitive factors and school performance while accounting for the traits' background genetic and environmental correlations. Kohler et al.¹⁸ and McAdams et al.¹⁹ provided detailed discussions of causal modelling based on the classical twin design. Pingault et al.²⁰ discuss more generally the role of genetic data in causal inference using observational data (including twin data). For a recent application of this approach in the study of alcohol use and anxiety, we refer to Torvik et al.²¹ For applications in educational sciences, we refer to van Bergen et al.^{22,23} and Malanchini et al.²⁴ who studied the causal effects between literacy skills and literacy enjoyment. Using causal modelling in the classical twin design, we investigated in the current work the causal relationship between teacher-rated self-control, grit, and school performance, in the presence of possible genetic and environmental confounding, that is, non-causal associations that are attributable to genetic and environmental influences common to the phenotypes.

The outline of this article is as follows. First, we introduce the statistical model based on the classical twin design, and provide a summary of the results of Kevenaar et al.¹⁰. Next, we present a causal twin model, which we apply to explore the putative causal influence of self-control and grit and school performance. Then, we present the results of fitting the causal twin model including genetic confounding (a.k.a. genetic pleiotropy), and environmental confounding.

The twin design is a genetically informative design, which is applied to decompose phenotypic variance and covariance into genetic and environmental components. With respect to the environmental components, we distinguish shared (C) and unique environmental (E) variance. The latter (E) is unique to the individual twins, not shared, and, as such, contributes to the phenotypic variance, but not the phenotypic covariance (resemblance) of the twins. Shared environmental variance originates in environmental influences that twins share and contributes to the phenotypic covariance of the twins. With respect to the genetic components, we can distinguish additive genetic (A) variance and dominance (D) variance, where the former is due to the additive (linear) effects of alleles, and the latter is due to non-additive effects of alleles within relevant genetic loci on the phenotype²⁵. Because monozygotic (MZ) twins are genetically (nearly) identical, both A and D contribute 100% to the MZ phenotypic covariance. Dizygotic (DZ) twins, like full sibs, on average share 50% of their alleles, as inherited from their biological parents. Based on allele sharing, we expect 50% of the additive genetic variance to contribute to the DZ phenotypic covariance. The dominance variance attributable to a given locus contributes to the phenotypic resemblance, only if the twins are genetically identical by descent at the locus. Considering, as an example, a diallelic locus with alleles B and b, 25% of the DZ twins are genetically identical (i.e., both BB, Bb, or bb). Therefore, we expect 25% of the dominance variance to contribute to the DZ phenotypic covariance. When fitting the classical twin model to data from MZ and DZ twin pairs to identify the variance components, we need to limit the number of components to three, that is, an ADE or an ACE model. The choice is usually based on the following rule

of thumb concerning the phenotypic twin correlation, r_{mz} and r_{dz} :

$$r_{mz} > 2 * r_{dz} \quad (1)$$

$$r_{mz} < 2 * r_{dz} \quad (2)$$

If Eq. (1) holds, it suggests an ADE model, and if Eq. (2) holds, it suggests an ACE model^{26,27}. Based on our earlier work on the same data¹⁰, we fitted an ADE model to all three phenotypes (i.e., self-control, grit, and school performance) and decomposed the 3×3 covariance matrix Σ_{ph} as follows:

$$\Sigma_{ph} = \Sigma_A + \Sigma_D + \Sigma_E \quad (3)$$

This decomposition (Eq. 3) is achieved by modelling the 6×6 MZ and DZ twin covariance matrices. The matrices are 6×6 , because of the three phenotypes for both twin 1 and twin 2 (the first and second born, for example), as in Table 1. The MZ and DZ twin covariance matrices $\Sigma_{ph|MZ}$ (4) and $\Sigma_{ph|DZ}$ (5) are as follows:

$$\Sigma_{ph|MZ} = \begin{matrix} \text{MZtwin1} \\ \text{MZtwin2} \end{matrix} \begin{bmatrix} \Sigma_A + \Sigma_D + \Sigma_E & \Sigma_A + \Sigma_D \\ \Sigma_A + \Sigma_D & \Sigma_A + \Sigma_D + \Sigma_E \end{bmatrix} \quad (4)$$

And

$$\Sigma_{ph|DZ} = \begin{matrix} \text{DZtwin1} \\ \text{DZtwin2} \end{matrix} \begin{bmatrix} \Sigma_A + \Sigma_D + \Sigma_E & \frac{1}{2}\Sigma_A + \frac{1}{4}\Sigma_D \\ \frac{1}{2}\Sigma_A + \frac{1}{4}\Sigma_D & \Sigma_A + \Sigma_D + \Sigma_E \end{bmatrix} \quad (5)$$

The covariance matrices Σ_A , Σ_D , and Σ_E may be subject to various parameterizations, depending on computational or substantive considerations (see below).

As reported in Kevenaar et al.¹⁰, we previously analyzed teacher ratings of self-control, grit, and school performance in MZ and DZ twins using the same data that were analyzed for the present article. The results were obtained with a correction for the main effects of sex and SES, and a correction for the rater (i.e., the teacher of the twins). Given the ceiling effect in the distribution of the data (see below), we fitted ADE models using maximum likelihood estimation with a correction for right-censoring (see also de Zeeuw et al.²⁸). First, school performance was regressed on self-control and grit using linear regression analysis of these observed variables. Second, the regression analyses were conducted at the broad-sense genetic level ($\Sigma_A + \Sigma_D$), and at the unshared environmental level (Σ_E). Because self-control and grit are correlated (about 0.65 in the present data), the decomposition of school performance variance (conditional on the covariates) comprised four variance components: a component due to self-control, a component due to grit, a component involving the covariance between self-control and grit, and the residual variance component. At the level of the phenotypic regression model (see Fig. 1 top), self-control and grit explained 28.4% of the school performance variance, with the following decomposition: 4.4% due to self-control, 13.0% due to grit, and 10.9% involving the covariance of self-control and grit). Considering the unique contributions of self-control and grit, grit emerged as the stronger predictor (13% vs 4.4%).

Subsequently, the ADE model was fitted to the twin data, and the regression analyses were conducted twice: once at the level of $\Sigma_A + \Sigma_D$ (the broad-sense genetic covariance matrix) and once at the level of Σ_E (the unshared environmental covariance matrix) (see Fig. 1, bottom two panels). The results showed that the phenotypic decomposition of school performance variance was largely attributable to broad-sense genetic factors. Thus, the phenotypic regression relationship between the predictors self-control and grit and school performance was largely a reflection of common genetic influences.

Table 1. MZ and DZ correlation matrices of school performance (SP), self-control (SC), and Grit, conditional on fixed sex and SES effects, and random rater effects, and corrected for censoring.

MZ	SP1	SC1	GRIT1	SP2	SC2	GRIT2
SP1	1.000					
SC1	0.458	1.000				
GRIT1	0.523	0.721	1.000			
SP2	0.809	0.395	0.458	1.000		
SC2	0.428	0.715	0.587	0.480	1.000	
GRIT2	0.499	0.566	0.751	0.510	0.710	1.000

DZ	SP1	SC1	GRIT1	SP2	SC2	GRIT2
SP1	1.000					
SC1	0.478	1.000				
GRIT1	0.511	0.717	1.000			
SP2	0.419	0.214	0.181	1.000		
SC2	0.226	0.276	0.184	0.474	1.000	
GRIT2	0.248	0.166	0.176	0.506	0.723	1.000

The correlations shown in dark blue represent the within-person correlations between traits, which are expected to be similar in MZ and DZ. The correlations shown in bold represent the within-trait twin correlations. These are higher in MZ than DZ, suggesting genetic influences on the traits. The correlations shown in light blue represent the cross-trait, cross-twin correlations (e.g., the correlation between school performance of one twin and self-control of the cotwin). These are higher in MZ than DZ, suggesting genetic correlations between the traits. 1 = Twin 1; 2 = Twin 2.

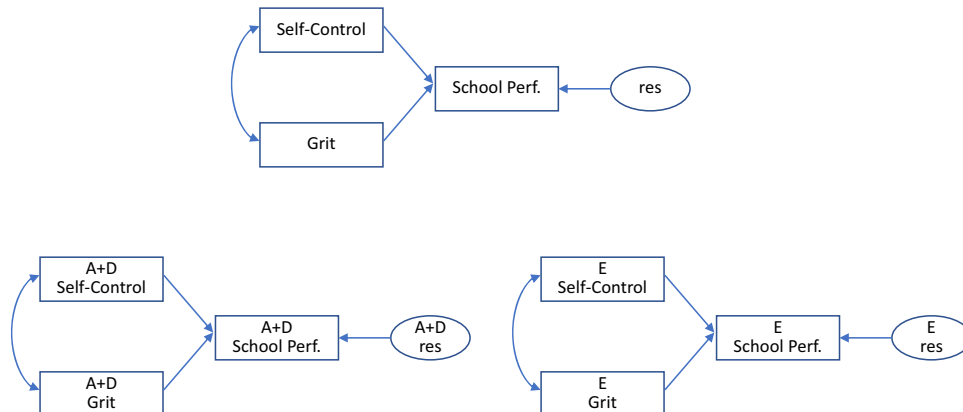


Fig. 1 The phenotypic regression model and the decomposed regression models. Top panel: the phenotypic regression model. The regression residual is denoted *res*. Bottom panel: the A + D regression model and the E regression model. These E and A + D models decompose the phenotypic regression results into A + D (based on $\Sigma_A + \Sigma_D$) and unshared environmental E regression results (based on Σ_E). The covariates (SES, sex, and rater) are not depicted.

The results of the regression analyses, both at the phenotypic level and at the genetic and environmental level, are consistent with a causal model, but do not prove causality. This is because correlation—established at the phenotypic level or the genetic and environmental level—does not imply causation. Below we present a causal twin model that addresses causality by fitting the phenotypic regression model, while accounting for the possibility of genetic or environmental background correlation (i.e., genetic or environmental confounding²⁹).

In the current paper, we fit a causal twin model. The causal twin model is depicted in Fig. 2. This model allows us to assess the putative causal regression relationships, while taking into account A, D, or E background correlation, i.e., A, D, or E confounding^{18,20,30–33}.

In Fig. 2, the background A, D, and E correlations are represented by the dashed double-headed arrows. In this approach, the strongest support for the causal hypothesis would be the finding that the phenotypic regression coefficients (denoted $b_{SP,SC}$ and $b_{SP,Grit}$) are significant, while the background correlations are all zero. This would support causality in that the results then demonstrate the phenotypic regression relations between the predictors and school performance are not due to background (A, D, or E) confounding, but to the direct phenotypic, putatively causal, relations. The causal model is refuted if the regression coefficients are zero in the presence of A, D, and/or E background correlations, as this means that the associations between self-control and grit and the dependent variable school performance are not due to direct,

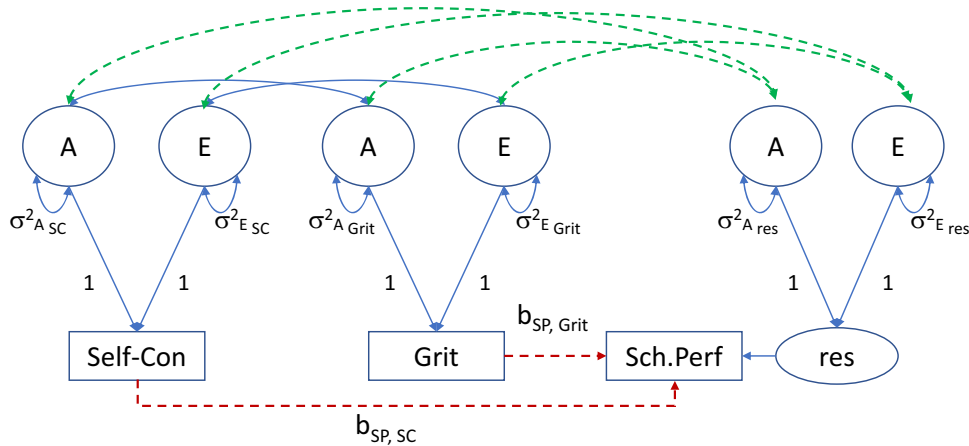


Fig. 2 The causal model. The causal model with background A, D (not shown), and E correlations is represented by dashed double-headed arrows, shown in green. SP school performance, SC self-control. The parameters $b_{SP,SC}$ and $b_{SP,Grit}$, shown in red, are the causal regression coefficients. The residual in the regression of SP on SC and Grit is denoted *res*. To avoid clutter, the Ds (and for each phenotype) and the covariates (SES, sex, and rater) are not depicted.

causal relations. Rather, they are attributable to environmental or genetic influences common to the three phenotypes (see Kohler et al.¹⁸ for a detailed treatment of this and related twin models). Note that the finding that the parameters $b_{SP,SC}$ and $b_{SP,Grit}$ (Fig. 2) differ significantly from zero does not rule out A, D, or E confounding. As mentioned above, the direct (phenotypic) causal effects and confounding are not mutually exclusive. In this causal twin model, we explore this possibility by fitting the phenotypic regression model, while allowing for A, D, or E confounding. We do this by including the dashed double-headed arrows in the model (Fig. 2). We modelled data of 8728 twin children from the Netherlands Twin Register. The twins' grit, self-control, and school performance were rated by their teacher when they were about 12 years old.

RESULTS

Descriptives

Histograms of the raw data are given in Fig. 3. The right censoring (ceiling effects) is evident in all three phenotypes. In the MZ and DZ twin 1 members, the skewnesses equal -0.63 (school performance), -1.63 (self-control), and -0.31 (grit); in the MZ and DZ twin 2 members, these equal -0.57 , -1.93 , and -0.48 , respectively. The estimates of the MZ and DZ correlation matrices, based on the saturated model, are given in Table 1 (these are conditional on the covariates sex, SES, and rater, and corrected for censoring).

The MZ twin correlations in Table 1 for school performance, self-control, and grit equal 0.809 (95% CIs: 0.788 – 0.928), 0.715 (95% CIs: 0.685 – 0.716), and 0.751 (95% CIs: 0.722 – 0.778), respectively. The DZ twin correlations equal 0.419 (95% CIs: 0.376 – 0.460), 0.276 (95% CIs: 0.250 – 0.289), and 0.176 (95% CIs: 0.114 – 0.234), respectively. The twin correlations of self-control and grit suggest an ADE model ($r_{MZ} > 2 * r_{DZ}$). The twin correlations of school performance suggest an AE model, but the 95% CIs do not rule out the possibility of an ADE model (see Keller & Coventry²⁶). The correlations between self-control and grit are about 0.72 . The correlations between school performance on the one hand and self-control or grit on the other hand range from 0.458 to 0.523 .

Phenotypic regression model (Model 1)

We fitted the phenotypic regression model in OpenMx, correcting for the family clustering of the data (i.e., MZ and DZ twins in pairs). The aim of this is to obtain regression results of regressing school performance of self-control and grit at the population level. We decomposed the proportion of explained variance of school

performance (i.e., R^2 statistic), into the part due to self-control, the part due to grit, and the part that involves the covariance of self-control and grit. Because the third part involves covariance, it cannot unambiguously be attributed to either self-control or grit. As reported by Kevenaar et al., (2023), we found that self-control and grit explained 28.4% of the variance in school performance ($R^2 = 0.284$). The unique contributions of self-control and grit equalled 4.4% (95% CIs: 2.07% – 7.91%) and 13.0% (95% CIs: 8.03% – 19.48%), respectively. The remaining 10.9% was a function of the covariance of the predictors (95% CIs: 9.08% – 12.43%). The regression coefficients equalled $b_{SP,SC} = 0.191$ (95% CIs: 0.132 – 0.251) and $b_{SP,Grit} = 0.331$ (95% CIs: 0.252 – 0.412). From these results, grit emerges as the stronger predictor. Based on the results in Table 2, we found the predictors (self-control and grit) are correlated 0.71 on average. We checked whether this correlation resulted in multicollinearity by calculating the variance inflation factor (VIF³⁴). The VIF associated with the predictors was about 2.06 . The rule of thumb concerning the interpretation varies considerably, with $VIF > 2.5$ to $VIF > 10$ signalling multicollinearity. As the present value is about 2.06 , we conclude that multicollinearity is not an issue here.

ADE twin model (Model 2)

The 3×3 covariance matrices Σ_A , Σ_D , and Σ_E were parameterized using lower triangle matrices (i.e., the Cholesky decomposition³⁰):

$$\Sigma_A = \Delta_A \Delta_A^t \quad (6)$$

$$\Sigma_D = \Delta_D \Delta_D^t \quad (7)$$

$$\Sigma_E = \Delta_E \Delta_E^t \quad (8)$$

In Eqs. (6), (7), and (8), Δ_A , Δ_D , and Δ_E are 3×3 lower triangular matrices. The results of fitting the ADE twin model are given in Table 2.

The right columns of Table 2 show the proportions of the phenotypic variances and covariances attributable to A, D, and E factors. These proportions provide an interpretable decomposition of phenotypic (co)variance. For instance, the standardized variance of grit, conditional on the covariates (SES, sex, and rater), is expressed in proportions as follows 0.179 (A), 0.556 (D), and 0.265 (E). So, we know that about 73% of the phenotypic variance is due to genetic effects ($17.9\% + 55.6\%$). The phenotypic correlation between school performance and grit is 0.502 . This correlation is expressed as proportions 0.736 (A), 0.184 (D), and

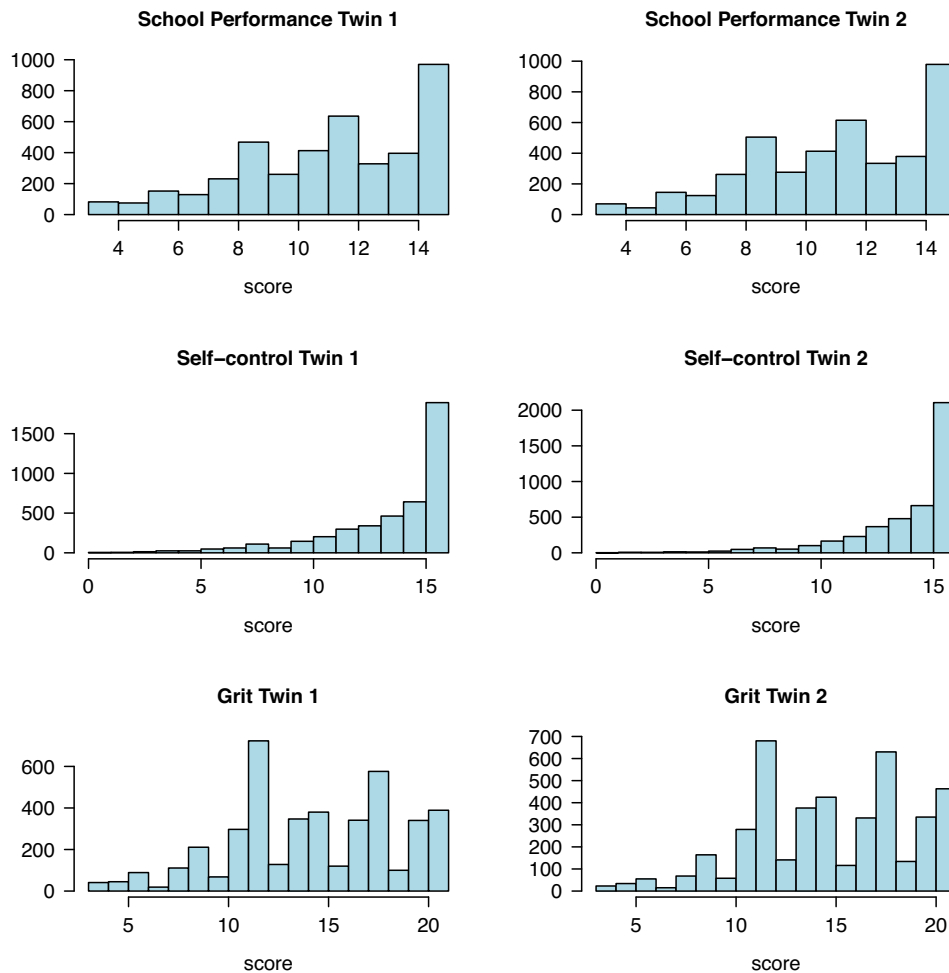


Fig. 3 Histograms. Histograms of the raw data of the first twin members and the second members (MZ and DZ twins pooled).

0.080 (E). So, 8% of the phenotypic correlation is attributable to E, unshared environmental factors, and 92% (73.6% + 18.4%) is attributable to genetic factors). The correlation matrices are displayed in the middle columns. So, for example, the additive genetic correlation between self-control and grit is 0.715. The results in Table 2 are conditional on the covariates sex, SES, and rater (teacher). As mentioned above, the rater effect was modelled as a random effect, i.e., part of the covariance structure. Table 3 contains the standardized variance components including the proportion attributable to the rater effect. Table 3 includes the estimates of the covariance matrices Σ_A , Σ_D , and Σ_E , the phenotypic covariance matrix Σ_{Ph} (i.e., $\Sigma_A + \Sigma_D + \Sigma_E$), and the associated correlation matrices (i.e., Σ_A , Σ_D , Σ_E and Σ_{Ph} standardized).

The standardized variance of grit, conditional on the covariates (SES, sex, and rater), is expressed as proportions as follows: 0.139 (A), 0.437 (D), 0.209 (E), and 0.212 (rater). We note that the rater (teacher) effects, in terms of standardized variance are quite variable, ranging from 21.2% (grit) to 0.2% (self-control).

Causal regression model without confounding (Model 3)

The causal regression model is depicted in Fig. 2. In this model, the background correlations (associated with the dashed double-headed arrows in Fig. 2) are fixed to zero, meaning that there is no background correlation due to common A, D, or E influences (i.e., no confounding). As such, this model is consistent with the causal hypothesis that self-control and grit are causes of school performance. The LRT of this model relative to the ADE model

equals $LRT = 155.8$, $df = 4$ ($p < 0.008$). Table 4 includes an overview of the LRTs and Akaike information criterion (AIC). The AIC is a goodness of fit measure that balances model complexity and model fit to achieve a parsimony-related fit statistic. In theory, the model with the lowest AIC is the model of choice.

The ADE model includes six parameters to model the phenotypic covariance between self-control and school performance and grit and school performance (two A covariances, two D covariances and two E covariances). But the causal model includes two parameters to model these covariances (i.e., the regression coefficients $b_{SP, SC}$ and $b_{SP, Grit}$). The difference in the number of parameters, which equals the degrees of freedom, is four. The LRT (155.8, $df = 4$, $p < 0.008$) clearly indicates that the causal model, without confounding, does not fit well, relative to the ADE model. The AIC value of this model (i.e., 105599.7) is also the largest in Table 4. This suggests at least that the effects of the predictors self-control and grit on the outcome school performance are not purely causal.

The causal model with confounding (Models 4, 5, 6)

We added A, D, and E confounding to the model by including the relevant background A, D, and E correlations (dashed double-headed arrows in Fig. 2). We considered A, D, and E confounding consecutively. We did not consider more than one source of confounding, as this, in combination with the phenotypic regression coefficients, renders the model equivalent to the ADE model in terms of the number of parameters used to model the associations. The LRT statistics, based on the comparison of the

Table 2. Covariance (cov) matrices, correlation (cor) matrices; and proportions, based on the ADE twin model.

Phenotypic cov matrix				Phenotypic cor matrix			
Σ_{Ph}	SP	SC	GRIT	SP	SC	GRIT	GRIT
SP	11.725	6.207	6.615	1.000	0.474	0.502	0.502
SC	6.207	14.616	10.565	0.474	1.000	0.718	0.718
GRIT	6.615	10.565	14.824	0.502	0.718	1.000	1.000
A cov matrix							
proportions Σ_A/Σ_{Ph}							
Σ_A	SP	SC	GRIT	SP	SC	GRIT	GRIT
SP	9.009	4.806	4.868	1.000	0.653	0.996	0.736
SC	4.806	6.006	2.852	0.653	1.000	0.715	0.411
GRIT	4.868	2.852	2.649	0.996	0.715	1.000	0.270
D cov matrix							
proportions Σ_D/Σ_{Ph}							
Σ_D	SP	SC	GRIT	SP	SC	GRIT	GRIT
SP	0.288	0.587	1.220	1.000	0.543	0.792	0.184
SC	0.587	4.054	5.449	0.543	1.000	0.943	0.516
GRIT	1.220	5.449	8.239	0.792	0.943	1.000	0.556
E cov matrix							
proportions Σ_E/Σ_{Ph}							
Σ_E	SP	SC	GRIT	SP	SC	GRIT	GRIT
SP	2.428	0.813	0.527	1.000	0.244	0.170	0.080
SC	0.813	4.557	2.264	0.244	1.000	0.535	0.214
GRIT	0.527	2.264	3.936	0.170	0.535	1.000	0.265

The results are conditional on the sex and SES effects, and the random rater (teacher) effects. The values shown in bold are the standardized variance components. For instance for school performance, 76.8% is attributable to additive genetic factors (i.e., the narrow-sense heritability is 0.768), 2.5% is attributable to genetic dominance factors (i.e., the broad-sense heritability is 0.793), and 20.7% is attributable to non-shared environmental influences and measurement error.

Table 3. ADE standardized variance components (corrected for sex and SES), including the variance attributable to rater (95% CIs in parentheses).

	A	D	E	Rater (Teacher)
School Performance	0.712 (0.609–0.757)	0.022 (0.003–0.105)	0.192 (0.183–0.221)	0.073 (0.044 –0.101)
Self-control	0.410 (0.261–0.538)	0.276 (0.207–0.374)	0.311 (0.286–0.315)	0.002 (0.000–0.113)
Grit	0.139 (0.093–0.226)	0.437 (0.371–0.512)	0.209 (0.194–0.231)	0.212 (0.183–0.250)

The four variance components are standardized, so add up to 1.

The standardized A component gives narrow-sense heritability and the standardized A component + the standardized D component gives the broad-sense heritability.

Table 4. Model fit comparison tests, with the preferred model printed in bold.

Model	Δdf	ΔLL	AIC	p value
ADE			105451.9	
Causal ADE	4	155.8	105599.7	<0.001
ADE			105451.9	
Causal ADE + A confounding	2	0.407	105448.3	0.815
ADE			105451.9	
Causal ADE + E confounding	2	24.8	105472.7	<0.001
ADE			105451.9	
Causal ADE + D confounding	2	38.4	105486.3	<0.001

Δdf is the difference of degrees of freedom of the models, ΔLL is the difference in the minus 2 log-likelihood of the models, AIC is the Akaike information criterion. The preferred model is printed in bold.

ADE model with the causal model with confounding, are $LRT = 0.407$, $df = 2$, $p = 0.815$, $LRT = 24.8$, $df = 2$, $p < 0.0001$, and $LRT = 38.4$, $df = 2$, $p < 0.001$, given A, E, and D confounding, respectively. The tests have two degrees of freedom, because the ADE model includes six parameters to model the phenotypic covariance between self-control and school performance and grit and school performance (two A covariances, two D covariances and two E covariances). The causal regression model with confounding does this with four parameters: the regression coefficients and two A, D, or E covariances. The LRTs suggest that the causal model with A confounding fits the data well, relative to the ADE model ($\chi^2_{\Delta}(2) = 0.407$, $p = 0.81$), but the other models clearly do not. The AIC also identifies the causal model with A confounding as the best fitting model, as this model has the lowest AIC value (i.e., 105448.3). The model of choice with direct causal effects and A confounding is shown in Fig. 4.

In the causal model with A confounding, the estimates of the causal regression coefficients $b_{SP,SC}$ and $b_{SP,Grit}$ are 0.151 (s.e. 0.027) and 0.047 (s.e. 0.033), respectively. The LRT statistics of the tests of $b_{SP,SC} = 0$ and $b_{SP,Grit} = 0$ are 2.44, $df = 1$, $p = 0.118$ ($b_{SP,SC}$) and 29.18, $df = 1$, $p < 0.001$ ($b_{SP,Grit}$). While the test of $b_{SP,SC}$ is not statistically significant, we retained this parameter in the model, and in the calculation of components of variance of school performance. The decomposition of the variance of school performance in raw and standardized variance components is given in Table 5.

The total explained variance of school performance is 16.8%, with by far the largest part (12.4%) due to genetic confounding. The causal effects account for 4.4% (i.e., see Table 4: 2.84% + 0.27% + 1.26%) of the school performance variance. The decomposition of the 4.4% reveals that self-control (2.84% of the 4.4%) is a stronger predictor than grit (0.27% of the 4.4%).

The results based on the causal regression model with A confounding differ appreciably from the phenotypic regression

results both in terms of explained variance and in terms of the relative contributions of self-control and grit. In the phenotypic regression analyses, we found that self-control and grit accounted for 28.4% of the school performance variance, and we found that grit was the stronger predictor in terms of unique contributions (grit contributed 13%, self-control contributed 4.4% to the total of 28.4%). In the causal regression model with A confounding, we found that the total explained variance is lower at 16.8%: 4.4% due to the causal effects of self-control and grit, and 12.4% due to additive genetic confounding. In contrast to the phenotypic regression, the stronger predictor here is self-control. Due to the remarkably high genetic correlation between grit and school performance (0.996, as shown in Table 2), grit emerges as the more influential factor in the phenotypic regression model. This strong association primarily arises from the presence of A confounding, which is not accounted for in the phenotypic regression model. Consequently, grit emerges as the stronger predictor in this model. Once we account for A confounding, the predictive value of grit is greatly reduced, and self-control emerges as the stronger predictor. The difference in total explained variance (28.4% in Model 1 vs 16.8% in Model 4) is a consequence of the influence of A confounding on the regression coefficients. The regression coefficients in the phenotypic regression model (Model 1) are $b_{SP,SC} = 0.191$ and $b_{SP,Grit} = 0.331$, compared to $b_{SP,SC} = 0.151$ and $b_{SP,Grit} = 0.047$ in the causal regression model with A confounding (Model 4). The bias in the regression model is due to confounding, as explained in detail in the Supplementary Material.

DISCUSSION

We investigated the association between two non-cognitive skills (self-control and grit) and school performance. Demonstrating the direct regression relations, while taking into account genetic or environmental confounding, lends credence to the causal interpretation of the regression relations (for reviews relating specifically to the twin design, see Kohler et al.¹⁸; McAdams et al.¹⁹). The present results support the hypothesis that school performance is causally dependent on self-control and grit. However, we also found that additive genetic confounding made a relatively large contribution to the associations. This confounding is due to pleiotropy: genes that are common to all three variables, independent of the direct regression relationships. With respect to effect sizes, we found that in total 16.8% of the school performance variance was explained. Of this 16.8%, genetic pleiotropy accounted for 12.4% and the causal effects accounted for 4.4%.

An educational implication is that children who lack self-control and grit may face a double disadvantage. Firstly, their genetic predisposition for low self-control and grit coexists with a genetic predisposition for low school performance. The genetic pleiotropy that we showed suggests that certain genetic variants associated with lower self-control and grit also contribute to lower academic achievement, regardless of the direct relationship. Secondly, their low self-control and grit directly impact their school performance,

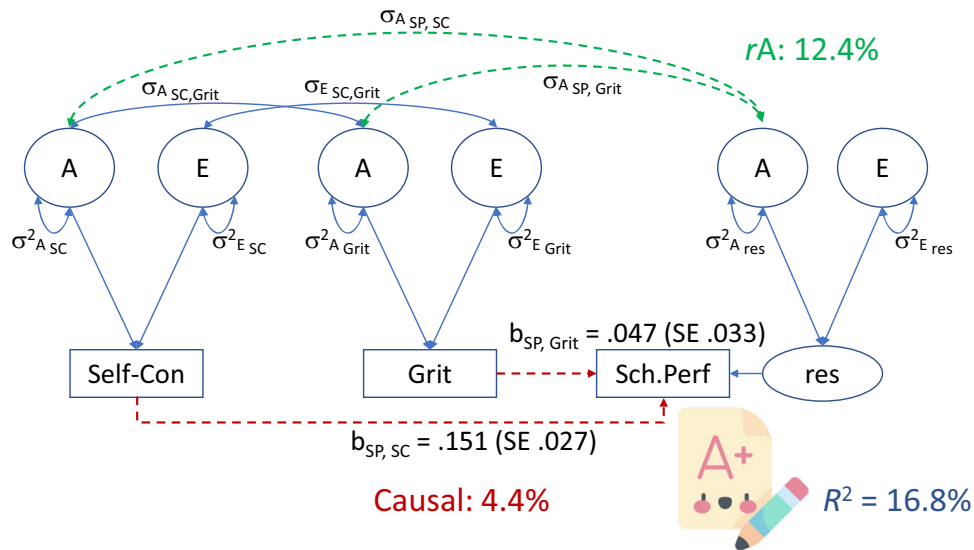


Fig. 4 The model of choice. The model of choice shows direct A confounding (in green; dashed double-headed arrows with covariances $\sigma_{A, SP, Grit}$, $\sigma_{A, SC, Grit}$, $\sigma_{E, SC, Grit}$) and causal effects (in red; parameters $b_{SP, SC}$ and $b_{SP, Grit}$). 16.8% of the variance in school performance is explained, for a large part due to genetic confounding (12.4%) and a small part due to the causal effects (4.4%), driven by the effect of self-control. To avoid clutter, the Ds (i.e., dominance genetic variance) and covariates are not depicted. A additive genetic variance, E non-shared environmental variance and measurement error, SC self-control, SP school performance, res residual.

Variance components of school performance		Raw estimate	Proportion of variance (95% CIs)
Causal due to self-control (SC)	$b_{SP, SC}^2 * (\sigma^2_{A, SC} + \sigma^2_{D, SC} + \sigma^2_{E, SC})$	0.333	0.0284 (0.019–0.051)
Causal due to grit	$b_{SP, Grit}^2 * (\sigma^2_{A, Grit} + \sigma^2_{D, Grit} + \sigma^2_{E, Grit})$	0.032	0.0027 (0.001–0.011)
Causal due to covariance SC-grit	$2 * b_{SP, SC} * b_{SP, Grit} * (\sigma_{A, SC, Grit} + \sigma_{D, SC, Grit} + \sigma_{E, SC, Grit})$	0.148	0.0126 (0.007–0.208)
Confounding due to A	$2 * b_{SP, SC} * \sigma_{A, SC, SP} + 2 * b_{SP, Grit} * \sigma_{A, Grit, SP}$	1.457	0.124 (0.101–0.134)
Residual (res) variance	$\sigma^2_{A, res} + \sigma^2_{D, res} + \sigma^2_{E, res}$	9.769	0.832 (0.816–0.868)
Total	σ^2_{SP}	11.739	1

as indicated by the direct regression relationship. Therefore, these children face challenges both due to the genetic factors influencing multiple traits and their own limited self-control and grit, which together contribute to their lower academic performance. However, bearing in mind that 83.2% (i.e., 100%–16.8%) of the school performance variance was unexplained, we emphasize that there are many other factors that may offset both the direct relationship and the genetic confounding.

Interventions designed to enhance self-control and grit (with the objective of improving school performance) are based on the assumption that self-control and grit are causally related to school performance. The present results demonstrate the truism that correlation does not imply causation. Notably, the straightforward regression analysis (Model 1, above) showed that self-control and grit accounted for 28% of the school-performance variance, which corresponds to a multiple correlation of 0.53 (i.e., $\sqrt{.28}$). However, taking into account confounding, the explained variance due to the direct, putatively causal relations is 4.4%, that is, a multiple correlation of 0.21 ($\sqrt{.044}$). In addition, straightforward regression analysis (Model 1) may identify a predictor as important, but its importance may be due to confounding, rather than its causal influence. Specifically, in the straightforward regression analysis (Model 1), grit emerged as the stronger predictor. However, when correcting for genetic confounding (Model 4), we found that self-control was a stronger causal predictor. This difference is due to

the high genetic correlation between grit and school performance (see Table 2).

Taken the present results, the causal effect size (4.4% of variance, a multiple correlation of 0.21) suggests that an intervention is likely to have a relatively small effect in the population studied here (i.e., 11.5–12.5-year-old children in the Netherlands). However, it is important to be aware that, on the one hand, intervention studies, and on the other hand causal observational studies (including the present study), answer distinctive questions. As a result, they may yield contrasting findings, despite both being concerned with causality. Observational studies that use causal-inference techniques provide insights into causal relations within natural settings, addressing the question of “what is”. Intervention studies essentially alter the natural setting to examine the effectiveness of specific interventions, addressing the question of “what could be” in the light of the interventions^{29,35}. As these questions are distinct, so could be the answers. Viewed from the “what is” perspective of the present study, one would expect an increase in self-control of one standard deviation to result in an increase in school performance of about 0.21 standard deviation units (ignoring the role of grit, to ease presentation). The relative size of this effect is hard to judge in isolation, as it should be evaluated relative to the effect sizes of other interventions targeting non-cognitive skills. More importantly, the effect of an intervention, designed ultimately to improve school performance, is unlikely to focus exclusively on

self-control and grit, and unlikely to result in an intervention effect on school performance that is only attributable to self-control and grit. A comprehensive understanding of what works in educational programmes requires well-designed intervention research, including, where possible, randomized controlled trials. Note that randomized controlled trials, theoretically, do not suffer from genetic or environmental confounding, as the random assignment to intervention and control groups controls for all forms of confounding. In practice, however, controlling for confounding in randomized controlled trials is partially mitigated through volunteer bias and selective dropout.

The present study has several limitations. It is based on the classical twin design, which is a design that relies on underlying assumptions. Testing these assumptions is beyond the scope of this paper. We therefore emphasize that the reliable inference of causal relations and the associated effect sizes requires the triangulation of results from different designs and datasets. Other causal-inference methods, which may be used, include instrumental variable regression^{20,36} and random-intercept cross-lagged panel modelling³⁷ (Mulder & Hamaker, 2021), as well as (quasi-) experimental designs (for application to non-cognitive abilities and school performance, see Yeager et al.³⁸; Diamond et al.³⁹).

The present results are based on the hypothesis that self-control and grit influence school performance. We have not ruled out the possibility of reverse causation or bidirectional causation. The decision to adopt the present unidirectional causal model is based on the prevailing literature (as discussed in the introduction), which predominantly suggests causal effect of self-control and grit on school performance.

We acknowledge that our phenotypic measures are not optimal. The measures of self-control, grit, and school performance were based on teacher ratings. With respect to school performance, we note that, in a subsample of the present sample, the teacher ratings correlate 0.70 with scores on the nationally standardized test of educational achievement (i.e., the CITO test)¹⁰. Nonetheless, ideally one would include teacher ratings as well as objective measures⁴⁰. With respect to the non-cognitive skills, the teacher ratings may be biased by the teacher's awareness of students' school performance. We recognize that such awareness may introduce a bias, and result in an overestimation of the correlation between non-cognitive skills and academic achievement. However, the inclusion of the random rater effect accounts for any source of bias originating in the teachers as raters (including halo effects).

The grit measure used here mainly captures the perseverance of effort aspect of grit, not the consistency of interest. However, the literature consistently demonstrates that the perseverance of effort aspect holds greater importance for academic outcomes^{2,15,41,42}.

In summary, our study sheds light on the relationship between non-cognitive traits and school performance. We found that self-control and grit are causally related to school performance (accounting for 4.4% of the school performance variance), with self-control emerging as the stronger causal predictor. Equally importantly, we found that additive genetic confounding contributed greatly to the associations between self-control and grit and school performance. This finding demonstrates the importance of taking into account confounding (regardless of its source) in interpreting regression relationships as causal.

METHODS

Participants

The sample consisted of children registered in the Netherlands Twin Register (NTR). The NTR collects data from twins, their parents, and their siblings. The data of the children include self-ratings and parental and teacher ratings^{43–45}. The data for this

study are teacher ratings of the grit, self-control and school performance in 11.5–12.5-year-old twins. First, the parents of these twins were asked for permission to contact the teachers. Twins could be either in the same class and share a teacher or be different classes and be rated by different teacher. The sample included 3837 complete pairs and 1054 incomplete pairs (i.e., data missing on one member). Data was available on 8728 individuals. The sample consisted of 1957 monozygotic and 2934 dizygotic twin pairs. To ascertain the zygosity of the same-sex twin pairs, a DNA or blood test was conducted for 32.2% of the pairs, while for the remainder, parents completed a questionnaire that contained items related to the twins' resemblance. Based on this questionnaire, zygosity is correctly determined in more than 96% of cases⁴⁵.

The data collection procedure was approved by the ethical committee (called 'Vaste Commissie Wetenschap en Ethiek') at the Vrije Universiteit Amsterdam (VCWE-2021-111). Parents provided written informed consent.

Materials

Self-control. The measure of self-control was based on the teacher ratings. The teachers completed the 8 items of the Achenbach Self-Control Scale (ASCS⁴⁶ in the ASEBA-TRF⁴⁷). The response options of each item are 0 (*not true*), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*). If more than three items were missing the sum scores was coded as missing. If three or fewer items were missing, the missing items were imputed using the individual (twin) level mean of the available items⁴⁶. Of the self-control sum scores, 34.5% were constructed with one to three mean imputed items, due to changes in the content of the ASEBA-TRF over the years. The scores were reverse-coded, so the total score ranged from 0–16, with higher scores indicating better self-control. The Cronbach's alpha in this sample is 0.87.

Grit. The measure of grit was based on the following three items relating to the perseverance aspect of grit: *Compared to typical pupils of the same age, (1) how hard does he/she work; (2) how appropriately does he/she behave, and (3) how task-oriented is he/she*. The teachers rated the twins with respect to these items using on a 7-point Likert scale. The item scores were summed to create grit sum scores. If a single item was missing, the individual (twin) level mean of the two observed items scores was used in calculating the individual grit sum score. If more than one item was missing, the grit sum score was coded as missing. In our sample, 55.2% had a missing score on item 3, due to changes in the content of the surveys over the years. The grit sum scores ranged from 1–21, with higher scores indicating more grit. Cronbach's alpha of the grit measure is 0.90.

From the point of view of face validity, the second grit item appears to be quite general. However, Kevenaar et al.¹⁰ demonstrated that this item highly correlates with the other two items (items 1 and 2 correlate 0.70, items 2 and 3 correlate 0.71, and items 1 and 3 correlate 0.83). Moreover, they showed that the predictive value for school performance is not driven by this one item.

School performance. The measure of school performance was the sum score of teacher ratings, on a five-point scale, of the twins' math, reading, and literacy performance^{24,48}. School performance scores ranged from 3 to 15, with higher scores indicating better school performance. If a single rating was missing the individual (twin) level mean of the other two ratings was used for the missing value. The school performance sum score was coded as missing if more than one rating was missing, which was the case for 22.9% in our sample. Regarding reliability, the items correlations are 0.73 for reading-literacy, 0.51 for reading-math, and 0.67 for literacy-math. The Cronbach's alpha of the school-

performance measure was 0.84. Regarding validity, the school-performance measure correlated 0.70 with the nationally standardised test of academic achievement (i.e., the CITO test, see Kevenaar et al.¹⁰).

Sex and socioeconomic status (SES). The covariate sex was coded 0 for boys and 1 for girls. The covariate parental socio-economic status (SES) was determined by a combination of their parents' occupation and education, as described in de Zeeuw et al.²⁸. The SES variable was coded on a scale of 1 to 4, with 1 indicating low SES and 4 indicating high SES. Sex and SES were included in the twin model as fixed covariates. That is, we fitted the twin model, while simultaneously regressing self-control, grit, and school performance on these covariates.

Same/different teacher. Twins in the same class were rated by the same teacher, while twins in different classes were rated by different teachers. To account for systematic rater bias, we coded rater sharing as 1 (twins in the same class, and rated by the same teacher), or 0 (twin members in different classes, and so rated by different teachers). We included rater as a random effect, i.e., as a source of variance common to the variables, due to systematic rater bias.

Statistical modelling

We modelled the data in R using the OpenMx library⁴⁹. As shown below, the distributions of all three phenotypes display negative skewness, as a consequence of ceiling effects. This is most notable in the distributions of self-control and grit. We fitted the models using full information maximum likelihood estimation, assuming that the data follow a right-censored multivariate normal distribution (as in de Zeeuw et al.²⁸). We took the censoring into account explicitly to avoid bias stemming from the apparent ceiling effects.

We first fitted the saturated model, which serves to obtain estimates of the 6×6 MZ and DZ covariance (and correlation) matrices, corrected for censoring, and corrected for the covariates. We subsequently fitted the following six models: (1) the standard phenotypic regression model (taking into account the clustering of twins in families); (2) the trivariate ADE model to estimate the 3×3 covariance matrices Σ_A , Σ_D , and Σ_E ; (3) the causal regression model as depicted in Fig. 2, without A, D, or E confounding (i.e., the model with the correlations associated with the dashed double-headed arrows fixed to zero); and models 4, 5, and 6, i.e., the causal regression model with A confounding (Model 4), D confounding (Model 5), or E confounding (Model 6). Model 1 produces results based on the regression of school performance of self-control and grit, as one would obtain them in a sample of unrelated children. Model 2 is a standard trivariate ADE twin model. This model does not include any regression analyses; it provides estimates of the 3×3 covariance matrices Σ_A , Σ_D , and Σ_E , and serves as a baseline model to evaluate the fit of Model 3, and Models 4, 5, and 6, as these models are nested under Model 2. If there is no confounding and if the regression relations are causal, we expect Model 2 to produce regression results comparable to those of Model 3, and we expect Model 3 to fit well (compare to Model 2). In case of A confounding, for instance, we expect Model 3 to fit poorly (compared to Model 2), and we expect Model 4 (causal regression with A confounding) to fit well compared to Model 2).

We conducted a total of six likelihood-ratio tests: the comparison of the causal regression model without confounding with the ADE model (one test with 4 degrees of freedom (*df*)); the comparison of the causal regression model with A, D, or E confounding with the ADE model (three tests, each with 2 *df*); and the test of the causal regression coefficients (two tests, each 1 *df*) in the ultimate model of choice. As we conducted 6 likelihood

ratio tests (LRTs), we corrected our family-wise alpha level of 0.05 using the Bonferroni correction⁵⁰, resulting in an alpha of $0.05/6 = \sim 0.008$ for each LRT. As mentioned, in all models, we included sex and SES as fixed covariates. We included rater (teacher) as a random covariate, to accommodate covariance among the measures rising from rater bias⁵¹.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Requests for access to NTR data: please see the procedures outlined on the following website: <https://ntr-data-request.psy.vu.nl/>.

CODE AVAILABILITY

Part of the code is available in the Supplementary Materials. Additional code is available upon request.

Received: 27 March 2023; Accepted: 17 October 2023;

Published online: 04 December 2023

REFERENCES

- Christopoulou, M., Lakioti, A., Pezirkianidis, C., Karakasidou, E. & Stalikas, A. The role of grit in education: a systematic review. *J. Psychol.* **9**, 2951–2971 (2018).
- Credé, M., Tynan, M. C. & Harms, P. D. Much ado about grit: a meta-analytic synthesis of the grit literature. *J. Pers. Soc. Psychol.* **113**, 492 (2017).
- de Ridder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M. & Baumeister, R. F. Taking stock of self-control: a meta-analysis of how trait self-control relates to a wide range of behaviors. *Pers. Soc. Psychol. Rev.* **16**, 76–99 (2012).
- Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M. & Gross, J. J. Self-control and academic achievement. *Annu. Rev. Psychol.* **70**, 373–399 (2019).
- Lam, K. K. L. & Zhou, M. Examining the relationship between grit and academic achievement within K-12 and higher education: a systematic review. *Psychol. Sch.* **56**, 1654–1686 (2019).
- Oriol, X., Miranda, R., Oyanedel, J. C. & Torres, J. The role of self-control and grit in domains of school success in students of primary and secondary school. *Front. Psychol.* **8**, 1716 (2017).
- Wolters, C. A. & Hussain, M. Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacogn. Learn.* **10**, 293–311 (2015).
- Fernández, F. D., Arco, J. L. & Hervás, M. Grit as a predictor and outcome of educational, professional, and personal success: a systematic review. *Psicol. Educ.* **26**, 163–173 (2020).
- Vazsonyi, A. T., Javakhishvili, M. & Blatny, M. Does self-control outdo IQ in predicting academic performance? *J. Youth Adolesc.* **51**, 499–508 (2022).
- Kevenaar, S. T., Dolan, C. V., Boomsma, D. I. & van Bergen, E. Self-control and grit are associated with school performance mainly because of shared genetic effects. *JCPP Adv.* **3**, e12159 (2023).
- Eskreis-Winkler, L., Shulman, E. P., Beal, S. A. & Duckworth, A. L. The grit effect: predicting retention in the military, the workplace, school and marriage. *Front. Psychol.* **5**, 36 (2014).
- Tangney, J. P., Baumeister, R. F. & Boone, A. L. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* **72**, 271–324 (2004).
- Duckworth, A. L., Tsukayama, E. & May, H. Establishing causality using longitudinal hierarchical linear modeling: an illustration predicting achievement from self-control. *Soc. Psychol. Pers. Sci.* **1**, 311–317 (2010).
- Perels, F., Dignath, C. & Schmitz, B. Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *Eur. J. Psychol. Educ.* **24**, 17–31 (2009).
- Jiang, W. et al. Reciprocal relations between grit and academic achievement: a longitudinal study. *Learn. Individ. Differ.* **71**, 13–22 (2019).
- Postigo Gutiérrez, Á. & Cuesta Izquierdo, M. Fernández Alonso, R., García Cueto, E., & Muñoz, J. Temporal stability of grit and school performance in adolescents: a longitudinal perspective. *Psicol. Educ.* **27**, 77–84 (2021).
- Mackay, T. F. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nat. Rev. Genet.* **15**, 22–33 (2014).

18. Kohler, H. P., Behrman, J. R. & Schnittker, J. Social science methods for twins data: Integrating causality, endowments, and heritability. *Biodemography Soc. Biol.* **57**, 88–141 (2011).
19. McAdams, T. A., Rijdsdijk, F. V., Zavos, H. M. S., & Pingault, J.-P. Twins and causal inference: leveraging nature's experiment. *Cold Spring Harb. Perspect. Med.* **11**, a03955 (2021).
20. Pingault, J. B. et al. Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).
21. Torvik, F. A. et al. Explaining the association between anxiety disorders and alcohol use disorder: a twin study. *Depress. Anxiety* **36**, 522–532 (2019).
22. van Bergen, E. et al. Why do children read more? The influence of reading ability on voluntary reading practices. *J. Child. Psychol. Psychiatry* **59**, 1205–1214 (2018).
23. van Bergen, E. et al. Literacy skills seem to fuel literacy enjoyment, rather than vice versa. *Dev. Sci.* **26**, e13325 (2023).
24. Malanchini, M. et al. Reading self-perceived ability, enjoyment and achievement: a genetically informative study of their reciprocal links over time. *Dev. Psychol.* **53**, 698–712 (2017).
25. Falconer, D. S., & Mackay, T. F. Quantitative genetics. London, UK: Longman (1983).
26. Keller, M. & Coventry, W. Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Res. Hum. Genet.* **8**, 201–213 (2005).
27. Posthuma, D. et al. Theory and practice in quantitative genetics. *Twin Res.* **6**, 361–376 (2003).
28. de Zeeuw, E. L. et al. The moderating role of SES on genetic differences in educational achievement in the Netherlands. *NPJ Sci. Learn.* **4**, 13 (2019).
29. Hart, S. A., Little, C. & van Bergen, E. Nurture might be nature: cautionary tales and proposed solutions. *NPJ Sci. Learn.* **6**, 2 (2021).
30. Bruins S., Franić S., Dolan C. V., Borsboom, D., & Boomsma, D. I. Structural equation modeling in genetics. In: R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, 2nd edn. (pp. 646–663). The Guilford Press, New York (2023).
31. Duffy, D. L. & Martin, N. G. Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. *Genet. Epidemiol.* **11**, 483–502 (1994).
32. Heath, A. C. et al. Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet.* **23**, 29–50 (1993).
33. Verhulst, B. & Estabrook, R. Using genetic information to test causal relationships in cross-sectional data. *J. Theor. Polit.* **24**, 328–344 (2012).
34. Sheather. *Simon A modern approach to regression with R*. New York, NY: Springer (2009).
35. Plomin, R., Shakeshaft, N. G., McMillan, A. & Trzaskowski, M. Nature, nurture, and expertise. *Intelligence* **45**, 46–59 (2014).
36. Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
37. Mulder, J. D. & Hamaker, E. L. Three extensions of the random intercept cross-lagged panel model. *Strict. Equ. Model.* **28**, 638–648 (2021).
38. Yeager, D. S. et al. Using design thinking to improve psychological interventions: the case of the growth mindset during the transition to high school. *J. Educ. Psychol.* **108**, 374 (2016).
39. Diamond, A., Lee, C., Senften, P., Lam, A. & Abbott, D. Randomized control trial of Tools of the Mind: marked benefits to kindergarten children and their teachers. *PLoS One* **14**, e0222447 (2019).
40. Rimfeld, K. et al. Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *J. Child. Psychol. Psychiatry* **60**, 1278–1288 (2019).
41. Muenks, K., Wigfield, A., Yang, J. S. & O'Neal, C. R. How true is grit? Assessing its relations to high school and college students' personality characteristics, self-regulation, engagement, and achievement. *J. Educ. Psychol.* **109**, 599 (2017).
42. Rimfeld, K., Kovas, Y., Dale, P. S. & Plomin, R. True grit and genetics: predicting academic achievement from personality. *J. Pers. Soc. Psychol.* **111**, 780 (2016).
43. Boomsma, D. I. et al. Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.* **9**, 849–857 (2006).
44. van Beijsterveldt, C. E. et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Res. Hum. Genet.* **16**, 252–267 (2013).
45. Ligthart, L. et al. The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Res. Hum. Genet.* **22**, 623–636 (2019).
46. Willems, Y. E. et al. Genetic and environmental influences on self-control: assessing self-control with the ASEBA self-control scale. *Behav. Genet.* **48**, 135–146 (2018).
47. Achenbach, T. M., & Rescorla, L. A. Manual for the ASEBA preschool forms and profiles (Vol. 30). Burlington, VT: University of Vermont, Research center for children, youth, & families (2000).
48. de Zeeuw, E. L. et al. Do children perform and behave better at school when taught by same-gender teachers? *Learn. Individ. Differ.* **36**, 152–156 (2014).
49. Neale, M. C. et al. OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* **81**, 535–549 (2016).
50. Bland, J. M. & Altman, D. G. Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170 (1995).
51. Hoyt, W. T. Rater bias in psychological research: when is it a problem and what can we do about it? *Psychol. Methods* **5**, 64–86 (2000).

ACKNOWLEDGEMENTS

We warmly thank all participating families and teachers in the Netherlands Twin Register who supplied data. NTR gratefully acknowledges funding by CID (Consortium on Individual Development) through the Gravitation Program of the 'Dutch Ministry of Education, Culture, and Science', the 'Dutch Research Council' (NWO; grant number 024-001-003); 'Decoding the gene-environment interplay of reading ability' (EvB's VENI - NWO: 451-15-017); 'Netherlands Twin Registry Repository: researching the interplay between genome and environment' (NWO: 480-15-001/674); 'Twin-family study of individual differences in school achievement' (NWO: 056-32-010); 'Longitudinal data collection from teachers of Dutch twins and their siblings' (NWO: 481-08-011) and the European Union (EvB's ERC, InterGen, 101076726). Elsje van Bergen is a Jacobs Foundation Research Fellow.

AUTHOR CONTRIBUTIONS

S.K.: data management, analyses, figures, writing, revision; E.v.B.: funding, supervision, figures, writing, revision; A.J.O.: writing, revision; D.I.B.: funding, supervision, study design, revision; C.V.D.: supervision, analysis plan, data analyses, figures, writing, revision. S.K. and EvB are co-first authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41539-023-00198-3>.

Correspondence and requests for materials should be addressed to Elsje van Bergen.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023