# Supplementary Material

For the following paper:

Kevenaar\*, van Bergen\*, Oldehinkel, Boomsma & Dolan (2024). The relationship of school performance with self-control and grit is strongly genetic and weakly causal. *npj Science of Learning*

## Supplementary Methods

Abbreviations

SP School performance, a phenotypic observed variable, conveyed in a square (Fig 1)

SC Self-control, a phenotypic observed variable, conveyed in a square (Fig 1)

Grit Grit, a phenotypic observed variable, conveyed in a square (Fig 1)

A Additive genetic variable, a latent variable, conveyed in a circle (Fig 1)

D Dominance variable, a latent variable, conveyed in a circle in the path diagram (Fig 1)

E Unshared environmental variable, a latent variable, conveyed in a circle in the path diagram

(Fig 1)

This supplement provides an account of the bias in the regression model originating in additive genetic confounding. The model is depicted in Figure 1. This model includes the phenotypic regression of SP on SC and Grit. The additive genetic confounding is represented by the dashed double headed arrows. The confounding represents a source of association between SP and SC and between SP and Grit, which is independent of the direct, phenotypic regression relations (coefficients bSP,SC and bSP,Grit).

**Supplementary Figure 1**



Given the model in Figure 1, we determine how the A confounding affects the phenotypic regression relationship given the following regression model (Figure 2):

 **Supplementary Figure 2**



The model in Figure 1 can be fitted given twin data; the model in Figure 2 can be fitted given data obtained in unrelated individuals, as this is the straightforward regression of SP on SC and Grit. The results of fitting the modelling in Figure 2 are informative from a predictive point of view, but the relations cannot be interpreted causally. Any confounding will affect the explained variance of SP (R2) in the phenotypic regression model. This R2 could be completely due to confounding, in the most extreme case.

To determine how the A confounding (as shown in Fig 1) affects the results of the phenotypic regression analysis, we first determine the expected phenotypic covariance matrix associated with the “true” model (Figure 1),  This 3x3 covariance matrix can be expressed as follows:

**** = (**I**-****)-1(**A** + **D** + **E**) (**I**-****)-1t, where

**A** = a11 **a21 a31** = 2Ares ASP,SC ASP,Grit

 **a21** a22 a32 ASP,SC 2ASC ASC,Grit

 **a31** a32 a33 ASP,Grit ASC,Grit 2AGrit

where a21 and a31 are the source of confounding, i.e., a21 = SP,SC and a31 = SP,Grit.

**D** = d11 0 0 = 2Dres 

 0 d22 d32 2DSC DSC,Grit

 0 d32 d33 DSC,Grit 2DGrit

**E**  = e11 0 0 = 2Eres 

 0 a22 e32 2ESC ESC,Grit

 0 e32 e33 ESC,Grit 2EGrit

(**I**-****) = 1 -bs -bg

 0 1 0

 0 0 1

(**I**-****)-1 = 1 bs bg

 0 1 0

 0 0 1,

where bs and bg are the regression coefficients (i.e., bSP,SC and bSP,Grit, in Figure 2).

We represent the 3x3 covariance matrix  = **A**+**D**+**E** as follows:

*  * = a11+d11+e11 a21 a31

*  *  a21 a22+d22+e22 a32+d32+e32

*  *  a31 a32+d32+e32 a33+d33+e33

We introduce the Greek notation merely to avoid clutter. The expected covariance matrix is  = (I-)-1 (I-)-1t =

|  |  |  |
| --- | --- | --- |
| ** + bs\*** +bg\***+ bs\*(**+ bs\*** +bg\***) + bg\*(**+ bs\*** +bg\***) | ** + bs\*** + bg\*** | ** + bs\*** + bg\*** |
| ** + bs\*** + bg\*** | ** | ** |
| ** + bs\*** + bg\*** | ** | ** |

In fitting the linear regression model (Figure 2), we estimate the regression coefficients ***b***(2x1) as follows:

***b***= **X**-1**XY**, where 

**X=***  =* a22+d22+e22 a32+d32+e32

* * a32+d32+e32 a33+d33+e33

and

**XY** = **+ bs\*** + bg\***

** + bs\**E* + bg\***

The inverse**X**-1 equals

**/(** \** - 2*) *-*/(** \** - 2*)

*-*/(** \** - 2*) **/(** \** - 2*)*,*

where (* \* – 2*) is the determinant of **X**. Let **X**-1 equal

** /(** \**- 2*) *-*/(** \** - 2*) *=* x11 x21

*-* /(** \**- 2*) ** /(** \** - 2*) x21 x22

So, we can express the regression coefficients in the phenotypic regression model (Fig 2) as follows:

***b*** =

x11\*(**+ bs \*** +bg\***) + x21\*(**+ bs\*** + bg\***)

x21\*(**+ bs \*** +bg\***) + x22\*(**+ bs\*** + bg\***)

***b*** =

** /(** \*** -** 2)\* (** + bs\*** + bg\***) + (-** /(** \*** -** 2))\*( ** + bs\*** + bg\***)

(-** /(** \*** -** 2))\*(** + bs\*** + bg\***)+ **/(** \*** -** 2)\*(** + bs\*** + bg\***)

We know that the A confounding is due to ** =a21 and ** =a31. So, making this substitution, we have (conveying a21 and a31 in green to high these parameters):

***b*** =

x11\*(a21+bs\*** + bg\***) + x21\*(a31+bs\*** + bg\***)

x21\*(a21+bs\*** + bg\***) + x11\*(a31+bs\*** + bg\***)

The a21 and a31, being positive (see main article), result in an upwards bias in the regression estimates. Given a21=a31=0, we have

x11\*(bs\*** + bg\***) + x21\*(bs\*** + bg\***)

x21\*(bs\*** + bg\***) + x11\*(bs\*** + bg\***)

So, the regression coefficients in the phenotypic regression analysis (Figure 2) are overestimated by

x11\*a21 + x21\*a31 = ** /(** \*** -** 2)\*a21 + (-** /(** \*** -** 2))\*a31

and

x21\*a21 + x11\*a31 = (-** /(** \*** -** 2))\*a21 + **/(** \*** -** 2)\*a31

The overestimation results in upward biased estimate of the proportion of explained variance in the phenotypic regression model.

**Numerical results**

We illustrate this numerically using the results as obtained. The values of bs and bg in the causal + A confounding model (Figure 1):

> print(c(bs, bg))

[1] 0.151 0.047

These values maximum likelihood estimates taken from the OpenMx output. We calculate these values based on the above, and obtain about the same results:

> print(c(f1\_,f2\_))

[1] 0.15100000 0.05110704

So ~.15 (self-control) and ~.05 (grit) are the values of the regression coefficients in the true model (Figure 1)

The values as calculated in the phenotypic regression model (Figure 2) are shown below (based on OpenMx output):

> print(c(f1,f2))

[1] 0.1906227 0.3307272

The values based on the above are about the same:

> print(c(g1,g2))

[1] 0.2067383 0.3123943

The bias in the regression coefficients due to confounding.

**> print(c(bias1, bias2))**

**[1] 0.05573831 0.26128727**

This corresponds to

x11\*a21 + x21\*a31 = ** /(** \*** -** 2)\*a21 + (-** /(** \*** -** 2))\*a31

and

x21\*a21 + x11\*a31 = (-** /(** \*** -** 2))\*a21 + **/(** \*** -** 2)\*a31

The confounding is the source of the bias. In the true model (Figure 1), the regression coefficients are about .151 (SC) and .05 (Grit). In the regression model (Figure 2), we obtain about .20 (SC) and .32 (Grit). The differences about .151 vs .20 and .05 vs .32 is due to confounding (i.e., parameters a21 and a32, or, as conveyed in Figure 1, the covariances ASP,SC andASP,Grit).