Supplementary Material

Section 1.

Statistical Analyses

Validating the New MTQ-10 Measure. Missing data were imputed by regression, considering that less than 5% of the data was missing. After screening the data, the sample was divided into two subgroups of wave 1 (W1) and wave 2 (W2) to verify the factor structure of the scale. Using AMOS 23 software, confirmatory factor analyses (CFA) were performed on the MTQ-10 initially with the W1 data to test the assumption that a single factor comprised the measure. The most plausible model from the W1 data was replicated using the W2 data to cross-validate the confirmed model. The validated model across W1 and W2 data was assessed for measurement invariance over time using multigroup CFA. Goodness of fit was evaluated through Chi-square, CFI (Comparative Fit Index), IFI (Incremental Fit Index), RMSEA (Root Mean Squared Error of Approximation) and SRMR (Standardized Root Mean Squared Residual). For these indices, literature suggests that values above .86 for CFI and IFI are marginal; values equal to or greater than .90 are deemed acceptable; and values close to 1 are indicative of good fit (Nigg, Nikolas, Miller, Burt, Klump, & von Eye, 2009). For RMSEA and SRMR, values of .08 and lower are considered optimal (Hair, Black, Babin, Anderson, & Tatham, 2009). For the initial factor structure assessment, modification indices (MI) were also examined to identify parameter misfit.

Through multigroup CFA, two models were tested: For the first model, parameters were freely estimated for each group, with only the plausibility of the factor structure tested. In the second model, factor loadings were constrained to be equal between groups. The CFI and chi-square differences between the models were considered. Finally, composite reliability (CR) was

calculated to determine the internal consistency of items in the factor structure. CR values above .60 are considered acceptable (Diamantopoulos & Siguaw, 2000). For completeness, Cronbach's alpha was reported along with test-retest reliability from W1 to W2 responses. The intraclass correlation coefficient was used to assess test-retest reliability with values greater than .70 considered acceptable (Cicchetti, 1994).

Descriptive Statistics and Covariates. Demographics and questionnaire data were examined using Descriptive Statistics in SPSS Version 24 (IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp). Kurtosis and skewness, were calculated to test for normality in the distribution (*skewness < 1.0*). All variables were approximately normally distributed. Cronbach's alphas were calculated to estimate the measures' internal consistency (see Table S1 below for details).

Previous research (e.g. Marchant et al., 2009; Lin et al., 2017; Paulhus & Williams, 2002) has shown that age and sex influence MT and narcissism. In our sample males scored significantly higher than females on MT at wave 1 [F (1, 338) = 35.34, p < ,001, η^2 = .09] and wave 2 [F (1, 338) = 21.44, p < ,001, η^2 = .06]; and on narcissism at wave 1 [F (1, 338) = 18.50, p < ,001, η^2 = .05] and wave 2 [F (1, 338) = 14.36, p < ,001, η^2 = .04]. Age showed a very weak correlation with school achievement at wave 1 (r = .11, p < .05). As such, age and sex were used as covariates in the partial correlation, multiple linear regression and mediation analyses. To explore whether subclinical narcissism and MT associate significantly with school achievement, when controlling for cognitive ability, a mean composite score of two cognitive tests (RPM and IVT-80) was used as a covariate in the aforementioned analyses.

Correlations, Multiple Linear Regression, Cross-Lagged and Mediation Analyses. Partial correlations were used to test the first hypothesis (MT will associate positively with

subclinical narcissism); multiple linear regressions were used to test the second hypothesis (MT and subclinical narcissism will associate positively with school achievement). Cross-lagged panel analysis was used to investigate the longitudinal associations between MT, narcissism and achievement across collection waves 1 and 2 (hypothesis 3). Cross-lagged analyses were run using the MPlus 7.0 software (Muthén & Muthén, 1998-2012). To test the fourth hypothesis, that individual differences in MT mediate the relationship between subclinical narcissism and school achievement, hierarchical regression analysis was performed using the PROCESS macro for SPSS (Version 2.13; Hayes, 2012). The direct effect provides an estimate of the effect of the independent variable (IV) on the dependent variable (DV). The indirect effect of the IV on the DV via a potential mediator (M) can be estimated from bias-corrected bootstrap 95% confidence intervals. The total effect provides an estimate of the combined direct and indirect effects. In the present study we used 5,000 bootstrap resamples.

Section 2.

MTQ-10 Psychometric Properties

A confirmatory factor analysis of the one-factor model for the MTQ-10 with the W1 data revealed unsatisfactory fit on all indices but SRMR, χ^2 (35, N = 343) = 147.88, p < .001, CFI = .84, IFI = .84, SRMR = .07, RMSEA = .09 (CI of .08 to .11). An inspection of modification indices (MI) revealed that allowing within-item error correlations between items 4 and 9, 2 and 8, 2 and 3, and 3 and 8 would improve model fit, χ^2 (31, N = 343) = 98.33, p < .001, CFI = .90, IFI = .90, SRMR = .05, RMSEA = .08 (CI of .06 to .09). Byrne (2016) recommends that item errors should not be correlated unless there exists appropriate justification. In this case, a rationale existed given items 4 and 9 both belonged to the *Challenge* subscale from the initial MTQ-48; and items 2, 8, and 3 were all reverse-scored items representing a degree of

commonality in item phrasing. The suitability of the one-factor model can be further supported in relation to its factor loadings, as all items loaded higher than the minimum threshold of .30 (Brown, 2015) apart from items 2 and 3 (loadings of .28 and .23 respectively). This final model including within-item correlations was specified and tested with the W2 data, and revealed acceptable fit, thus validating the one-factor structure of the model, χ^2 (31, N = 343) = 95.94, p < .001, CFI = .91, IFI = .91, SRMR = .06, RMSEA = .07 (CI of .06 to .09). As with the W1 data, all items loaded higher than the minimum threshold of .30 apart from items 2 and 3 (loadings of .22 and .18 respectively).

The multigroup CFA compared W1 with W2 data to verify the equivalence of parameters across time. The multigroup analyses indicated acceptable fit for the model with parameters freely estimated, χ^2 (64, N = 343) = 202.76, p < .001, CFI = .90, IFI = .90, SRMR = .05, RMSEA = .06 (CI of .05 to .07). The model with constrained factor loadings suggested acceptable fit, χ^2 (71, N = 343) = 215.31, p < .001, CFI = .90, IFI = .90, SRMR = .06, RMSEA = .06 (CI of .05 to .06). The difference between the CFI indices did not exceed .01. Also, a chi-square (χ2) difference test revealed a non-significant difference between models: χ^2 (7, N = 343) = 12.55, p = .08, suggesting that both the factor structure of the scale and the factor weights of the items are similar between W1 and W2. Composite reliability values for the final one-factor solution were acceptable (greater than .60) for both W1 (ρc = .77) and W2 (ρc = .73). Cronbach's alpha additionally indicated good internal consistency for W1 (α = .76) and W2 (α = .75). The test-retest intraclass correlation coefficient comparing W1 and W2 was .75 (CI of .70 to .80), indicating acceptable reliability of the measure over time. Overall, the results provide support for the suitability and stability of the one-factor solution of the MTQ-10.

Table S1. Descriptive Statistics for Mental Toughness, Subclinical Narcissism and School Grades (Wave 1 and Wave 2)

Variable	Mean	Variance	Median	Range	Kurtosis	Skewness	Cronbach's α
Mental Toughness 1	3.22 (.54)	.30	3.18	2.91	.07	-.05	.76
Mental Toughness 2	3.18 (.55)	.30	3.18	3.36	.71	.07	.75
Narcissism 1	2.86 (.53)	.29	2.77	3.56	.46	.29	.65
Narcissism 2	2.90 (.58)	.33	2.88	3.11	.35	.20	.69
Mathematics 1	6.54 (1.44)	2.07	7.00	6.00	-.68	-.05	—
Mathematics 2	6.80 (1.25)	1.57	7.00	6.00	-.34	-.001	—
Literacy 1	6.90 (.98)	.96	7.00	5.00	-.23	.23	—
Literacy 2	7.06 (.98)	.97	7.00	6.00	-.049	-.005	—
Second Language 1	6.84 (1.23)	1.51	7.00	5.00	-.63	-.03	—
Second Language 2	7.00 (1.16)	1.36	7.00	6.00	-.40	-.005	—
School Achievement 1	6.76 (.99)	.98	6.66	4.33	-.70	.15	—
School Achievement 2	6.95 (.93)	.87	7.00	5.67	-.19	.33	—

Note: N = 339. Standard deviations are given in parentheses. Numbers 1 and 2 after variable names refer to the assessment waves.

Table S2. Multiple Linear Regressions between Mental Toughness at Wave 1 and School Grades at Wave 2

N=339							
Independent Variable: Mental Toughness at Wave 1							
Dependent Variable	B	β	t	95% CI for β Lower Bound	95% CI for β Upper Bound	R^2	p-value
Mathematics (Wave 2)	.21	.08	1.45	-.075	.50	.006	.14
Literacy (Wave 2)	.27	.15	2.87	.08	.45	.02	.004
Foreign Language (Wave 2)	.09	.04	.80	-.14	.33	.002	.42

Note: The "B" and "β" refer to the unstandardized and standardized regression coefficients respectively.

Table S3. Multiple Linear Regressions between Narcissism at Wave 1 and School Grades at Wave 2

N=339							
Independent Variable: Narcissism at Wave 1							
Dependent Variable	B	β	t	95% CI for β Lower Bound	95% CI for β Upper Bound	R^2	p-value
Mathematics (Wave 2)	-.22	-.08	-1.57	-.50	.05	.007	.11
Literacy (Wave 2)	.15	.08	1.63	-.03	.33	.007	.10
Foreign Language (Wave 2)	-.04	-.018	-.35	-.27	.19	.00	.72

Note: The "B" and "β" refer to the unstandardized and standardized regression coefficients respectively.

References

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New London, UK: Guildford Press.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). New York, NY: Routledge.

Cicchetti, D. V. (1994). Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. Leicester, UK: Sage.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados*. Porto Alegre, RS: Bookman.

Hayes, A. F. (2012). PROCESS [Macro]. Retrieved from http://afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Marchant, D. C., Polman, R. C. J., Clough, P., Jackson, J. G., Levy, A. R., & Nicholls, A. R. (2009). Mental toughness: Managerial and age differences. *Journal of Managerial Psychology, 24*(5), 428-437.

Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

Nigg, J. T., Nikolas, M., Miller, T., Burt, S. A., Klump, K. L., & von Eye, A. (2009). Factor Structure of the Children's Perception of Interparental Conflict Scale for Studies of Youths with Externalizing Behavior Problems. *Psychological Assessment*, *21*, 450-456.