

Statistical analysis

For statistical analysis, R version 3.6.1 was used (packages psych and eRm).¹

Factor structure

The intended factor structure of the NL-Grit scale was examined by a principal component analysis (PCA) with Varimax rotation (“structural validity”). We used a parallel analysis scree plot to determine the number of factors to retain.² Factor loadings >0.40 for individual items were considered satisfactory. We calculated internal consistency reliability (coefficient alpha, also known as Cronbach’s α), as well as item means and standard deviations of the two subscales separately.

Unidimensionality of the subscales

We assessed the unidimensionality of both subscales by examining their fit to the Rasch model. For that goal, we dichotomized the items on their mean value. The Rasch model is a probabilistic model which describes the interaction of respondents with the questionnaire items and is governed by two parameters: item difficulty and person ability. Fit to the Rasch model of individual items was examined by infit and outfit statistics.⁴ Outfit is based on a sum of squared standardized residuals between observed and expected responses, infit is an information-weighted form of outfit.⁵ Infit and outfit statistics within the range of 0.5 to 1.5 indicate that an item effectively contributes to the latent construct the instrument purports to measure.⁴ The overall fit of the items together to the unidimensional Rasch model was tested using Andersen’s likelihood ratio test.⁶ The rationale of the likelihood ratio test is that if the Rasch model holds in the population, equivalent item parameter estimates should be obtained in low and high scoring subgroups (below and above the mean NL-Grit score, respectively). A p-value >0.05 indicates that the Rasch model is accepted for an item set (“Rasch homogeneous”). Unidimensionality was further examined by a PCA of the residuals that remain after Rasch analysis. The rationale here is that there is only one dimension, therefore residuals do not

contain other significant dimensions. Reference values for the first two eigenvalues of the PCA with 100 observations are 1.8 and 1.6 and are 1.4 and 1.3 with 500 respondents, respectively.⁷

Discriminative quality of the rating scale

To investigate the discriminative quality of the item rating scale, we performed Rasch rating scale analysis using the partial credit model (PCM).⁸ We plotted item category response curves to examine whether the item category measures (“threshold difficulties”) were ordered (i.e., that at each point on the latent grit scale, a single item category score is the most probable category). In that case, threshold difficulties should increase when moving from lower to higher categories. With a disordered rating scale, on the other hand, this is not the case and item measures are reversed (i.e. lower item categories having an item measure exceeding that of an adjacent higher item category). Disordered rating scale categories and cells including <10 observations were collapsed in such a way that an ordered and logical rating category emerged (*Figure 1: A and B in the article*).

Comprehensiveness

To assess to what extent NL-Grit scale items together can reliably measure the entire range of grit levels present in the sample, we plotted the person-item map based on the amended item scoring. The person-item map displays the location of person measures and item category difficulty respectively along the same latent dimension. The person item map relates the distribution of estimated person measures (grit trait levels) to the distribution of item category measures on a common log-odds unit scale. By using logit value, person-item maps are able to show simultaneously the hierarchies of both person and items at the same scale. We verified whether there was sufficient overlap between item measures and NL-Grit person measures. Furthermore, we examined whether there were substantial gaps between the item category measures along the total range of the grit scale, indicating less discriminative capacity within that range.

References

1. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. <http://www.r-project.org/>.
2. Field A, Miles J, Field Z. Exploratory Factor Analysis. In: *Discovering Statistics Using R*. London, England: AGE Publications.; 2012:749-811.
3. Hayes AF, Coutts JJ. Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But.... *Communication Methods and Measures*. 2020;14(1):1-24.
doi:10.1080/19312458.2020.1718629
4. Linacre JM. *What Do Infit and Outfit, Mean-Square and Standardized Mean? Rasch Measurement Transactions* 16:2.; 2002.
5. Masters GN, Wright BD. The Partial Credit Model BT - Handbook of Modern Item Response Theory. In: van der Linden WJ, Hambleton RK, eds. New York, NY: Springer New York; 1997:101-121. doi:10.1007/978-1-4757-2691-6_6
6. Andersen EB. A goodness of fit test for the rasch model. *Psychometrika*. 1973;38(1):123-140.
doi:10.1007/BF02291180
7. Brentani E, Golia S. Unidimensionality in the Rasch model: how to detect and interpret. *Statistica; Vol 67, No 3 (2007)DO - 106092/issn1973-2201/3508* . September 2007.
<https://rivista-statistica.unibo.it/article/view/3508>.
8. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: MESA Press.; 1982.
<https://research.acer.edu.au/measurement/2/>.