European Journal of Psychological Assessment

A Psychometric Evaluation of the Short Grit Scale: A Closer Look at its Factor Structure and Scale Functioning

Oscar Gonzalez, Jessica R. Canning, Heather Smyth, and David P. MacKinnon

Online First Publication, July 2, 2019. http://dx.doi.org/10.1027/1015-5759/a000535

Multistudy Report

A Psychometric Evaluation of the Short Grit Scale

A Closer Look at its Factor Structure and Scale Functioning

Oscar Gonzalez[1], Jessica R. Canning[2], Heather Smyth[3], and David P. MacKinnon[3]

[1]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, NC, USA
[2]Department of Psychology, University of Washington, Seattle, WA, USA
[3]Department of Psychology, Arizona State University, Tempe, AZ, USA

Abstract: Grit, the passion and perseverance for long-term goals, has received attention from personality psychologists because it predicts success and academic achievement. Grit has also been criticized as simply another measure of self-control or conscientiousness. A precise psychometric representation of grit is needed to understand how the construct is unique and how it overlaps with other constructs. Previous research suggests that the Short Grit Scale (Grit-S) has several psychometric limitations, such as uncertain factor structure within and across populations, uncertainty about reporting total or subscale scores, and different assessment precision at low and high levels on the construct. We conducted modern psychometric techniques including parallel analysis, measurement invariance, extrinsic convergent validity, and Item Response Theory models on two American samples. Our results suggest that the Grit-S is essentially unidimensional and that there is construct overlap with the self-control construct. Subscale factors were the result of an item doublet, where two items had highly correlated uniquenesses, showed similar item information, and were more likely to exhibit measurement bias. Findings replicated across samples. Finally, we discuss recommendations for the use of the Grit-S based on the theoretical interpretation of the unidimensional factor and our empirical findings.

Keywords: grit, self-control, psychometrics, conscientiousness

The construct of grit has received attention from personality psychologists because research suggests that it can predict success in education and other areas uniquely over talent or opportunity alone (Duckworth, Peterson, Matthews, & Kelly, 2007). Past research has focused on studying the relations of grit with other constructs. For example, grit is associated with positive outcomes like success in work or school (Duckworth & Quinn, 2009; Duckworth et al., 2007; O'Neal et al., 2016; Strayhorn, 2014) and goal achievement (Duckworth, Kirby, Tsukayama, Berstein, & Ericsson, 2011; Eskreis-Winkler, Shulman, Beal, & Duckworth, 2014). Additionally, research suggests that grit is related to conscientiousness, self-control, emotional stability, self-efficacy, mental toughness, and positive affect (Credé, Tynan, & Harms, 2017). Grit has also gained attention as a potential target for interventions because it may be more malleable than intelligence and other cognitive abilities (Duckworth, Grant, Loew, Oettingen, & Gollwitzer, 2011; Duckworth & Gross, 2014; Eskreis-Winkler et al., 2016; Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013). On the other hand, the construct map of grit has also been questioned. A recent meta-analysis suggests that grit researchers have fallen victim to the *jangle* fallacy because grit may be a repackaged version of conscientiousness (Credé et al., 2017). For example, previous research suggests that grit and conscientiousness are highly correlated ($r > .60$), and that grit may not have incremental prediction on academic performance after accounting for conscientiousness (Credé et al., 2017).

Jangle fallacies prevent the consolidation of many research efforts in psychology. Instead of wasting efforts, the decade-long research on the grit construct could be integrated with research on other personality constructs. One of many examples is that both grit and self-control, a conscientiousness facet, involve goal-setting and attainment (Credé et al., 2017; Duckworth & Gross, 2014; Muenks, Wigfield, Yang, & O'Neal, 2017). Empirical evidence of construct overlap between grit and self-control could help consolidate those research areas. Similar evidence could be used to study construct overlap between grit and persistence, motivation, or tenacity. A potential barrier to investigating construct overlap between grit and other constructs is that previous research has criticized the precision of grit measures, specifically the Short Grit Scale (Grit-S; Duckworth & Quinn, 2009). Research suggests that there are inconsistencies in the dimensionality of the

https://doi.org/10.1027/1015-5759/a000535

Grit-S, along with factor-structures across populations (Muenks et al., 2017). Lack of measurement precision could preclude or exacerbate observed relations with other constructs, which would compromise conclusions of construct overlap. The goal of this paper is to examine the psychometric properties of the Grit-S using modern methods and to provide some groundwork on how to study construct overlap between grit and other constructs.

The structure of the article is as follows. First, we review previous criticisms of the psychometric properties on the Grit-S. Then, we apply modern psychometric analyses to the Grit-S to identify its factor structure and item properties. Finally, we replicate our analyses in an independent dataset and demonstrate how to obtain extrinsic convergent validity evidence for construct overlap, using the overlap between grit and self-control as an example. The overall rationale of this study is to understand how the items from the Grit-S function and recommend how to best use the Grit-S score as researchers continue to obtain evidence of construct overlap.

Previous Psychometric Work on Grit

Grit is defined as "trait-level perseverance and passion for long-term goals" (Duckworth et al., 2007, p. 1087). The operational definition of grit is broken down into two factors, consistency of interest (CI) and perseverance of effort (PE; Duckworth et al., 2007). CI is described as the ability to maintain the same goal over long periods of time, and PE is described as the ability to keep working toward one's goal despite setbacks. During the development of the Grit-S, the CI sum score had an adequate reliability of α = .73–.79 and the reliability of the PE sum score was below conventional standards, α = .60–.78 across several samples (Duckworth & Quinn, 2009). The fit of the two-factor model was acceptable for some samples, but not for all (Muenks et al., 2017). Both PE and CI also had differential prediction with outcomes, yet a total Grit-S score was more reliable and a stronger predictor of outcomes (Duckworth & Quinn, 2009). Grit was also assumed to be a hierarchical construct (Duckworth & Quinn, 2009), and that might be the reason why most researchers use a Grit-S total score instead of subscale scores (Credé et al., 2017). However, a higher order model with two first-order factors and a two-factor model are mathematically equivalent, and they cannot be distinguished by model fit alone (Credé et al., 2017).

In the development of the Grit-S, items were selected based on prediction of several outcomes instead of using exploratory factor analysis. This procedure assumes that the two-factor structure from the original measure will hold after removing items. Empirically, this might not be the case. For example, Muenks and colleagues (2017) found that a two-factor model might fit the Grit-S well in a sample of high school students, but that a bifactor model could best describe the Grit-S in a sample of college students. Weston (2014) found that a one-factor solution with six Grit-S items was the best fitting model in a sample of economically disadvantaged students. Haktanir, Lenz, Can, and Watson (2016) evaluated the Turkish version of the Grit-S and found that a unidimensional factor with the full eight items did not fit the data well. However, model fit improved by removing two items and adding correlated residuals. Schmidt, Fleckenstein, Retelsdorf, Eskreis-Winkler, and Möller (2017) confirmed the two-factor structure of a German version of the Grit-S, but correlated residuals were needed between two items. Although the two-factor solution has been confirmed in other studies and cultures (Li et al., 2016), there might be situations where the two-factor solution might not be empirically supported.

Overall, research suggests some evidence that the Grit-S is multidimensional and other evidence that it is largely unidimensional. Dimensionality assessments on the Grit-S have been scarce (Christensen & Knezek, 2014; Rojas, Reser, Usher, & Toland, 2012) as researchers have focused on confirming the factor structure of the Grit-S originally proposed (Duckworth & Quinn, 2009). In this paper, we evaluate whether the Grit-S is more reasonably considered a multidimensional or unidimensional scale. The factor structure of the Grit-S also could differ across subgroups, such as education level (Muenks et al., 2017), which in turn could lead to measurement bias if groups are compared on that construct. The psychometric methods discussed below are useful for understanding the properties of the Grit-S and the construct overlap of grit with other constructs.

Psychometric Methods

Dimensionality Assessment

Parallel analysis (PA) is a widely accepted method to determine factor structure (Horn, 1965). In PA, the number of factors is determined by how many rank-ordered eigenvalues of the correlation matrix are higher than the rank-ordered eigenvalues of randomly simulated data. Exploratory and confirmatory factor analyses are then conducted based on the number of factors that PA suggests. Measures of factor strength also help determine if a measure is essentially unidimensional or multiple factors are needed to represent the construct (Rodriguez, Reise, & Haviland, 2016). Slight violations of unidimensionality might be manifested by correlated uniquenesses in the items and could have adverse effects on parameter estimation. Local dependence statistics, such as the Jackknife Slope Index (JSI), could help researchers identify items whose residual correlation biases parameter estimates (Edwards, Houts, & Cai, 2018). Convergent information from PA, measures of factor strength, and factor model fit would suggest the factor structure of the Grit-S.

Measurement Bias

Measurement invariance procedures provide information about the presence of measurement bias in a scale. A measure exhibits measurement bias when participants from subgroups of interest who have the same latent level of a construct express different observed scores on a measure. The confirmatory factor analysis framework for measurement invariance examines measurement bias by fitting a series of nested models that test if the factor structure is the same across subgroups (Millsap, 2011). The *configural* invariance model tests if items load onto the same corresponding factors across groups. The *metric* invariance model tests if the factor loadings are the same across groups. If *metric* invariance holds, then the observed differences between the item covariances are due to the latent variable and not to a faulty measure (Millsap & Olivera-Aguilar, 2012). Finally, the *scalar* invariance model tests if the item intercepts are the same across groups. If *scalar* invariance holds, then the observed differences between item means are due to differences in the latent variable and not to a faulty measure (Millsap & Olivera-Aguilar, 2012). If more restrictive models do not fit the dataset, it is a common procedure to look for items that violate measurement invariance and remove their equality constraint across groups. These models are known as *partial invariance* models. Overall, measurement invariance procedures help test for the presence of measurement bias and investigate if subpopulations have the same Grit-S factor structure.

Scale Functioning

A property of the Grit-S that has not been widely studied is its performance at assessing participants at low, average, and high levels on the grit construct (Credé et al., 2017). Item response theory (IRT) models determine the range of the grit construct in which the Grit-S is most precise at measuring grit. IRT is a family of latent variable models used to analyze item responses (Embretson & Reise, 2000). IRT models describe the probability of endorsing an item as a function of parameters related to the participant and item parameters. Specifically, the *severity* item parameters describe the range of the latent variable in which each item response is most likely; and the *slope* item parameters describe the relationship between each item and the latent construct (similar to a factor loading). These item parameters can then be transformed into item information functions that indicate the range of the grit construct where each item is most useful. For example, an item might be more relevant to measure participants below the mean on the grit construct than to those above the mean of the grit construct. Item information is additive, so one can investigate if the Grit-S can estimate precise latent scores for participants across the whole range of the grit construct.

Construct Overlap in Reference to Criteria

Two constructs should not be considered functionally interchangeable until there is enough evidence that both constructs have similar correlation profiles with outside criteria, or when extrinsic convergent validity is demonstrated (Fiske, 1971; Lubinski, 2004). A high correlation between the constructs is not enough to declare construct overlap because the two constructs could map differently to a criterion space (McCornack, 1956). For example, extrinsic convergent validity evidence for grit and self-control could be obtained by testing if the correlations of grit and self-control with other criteria of interest are the same (i.e., self-regulation, emotion regulation, and mindfulness). Evidence in favor of extrinsic convergent validity would suggest construct overlap between grit and self-control, and that the constructs are functionally interchangeable for predicting the criteria tested.

Present Study

In this study, we used the psychometric methods outlined above to investigate the factor structure and scale functioning of the Grit-S. The main contributions of this research are to help clarify psychometric inconsistencies found in the literature, and to propose a method to help consolidate evidence of construct overlap between grit and other constructs. These psychometric analyses are also important because exploratory factor analyses and scale functioning analyses of the Grit-S are scarce. Based on prior research, it is hypothesized that the Grit-S largely measures a single construct instead of two facets. Also, it is hypothesized that the factor structure will differ with education level. Our analyses were carried out in two independent samples to investigate whether the results replicate, and convergent information would suggest robust findings.

Materials and Methods

Sample 1

Publicly available data of 4,270 participants were used in this study (http://www.personality-testing.info/_rawdata/). We limited our sample to US participants who identified themselves as either male or female, and who passed a basic validity check (n = 2,047). The mean age was 25.66 years (SD = 12.23), 70% of participants were white, 68%

were female, and 36% were at least college-educated. The Grit-S has eight items, shown in Table A in the Electronic Supplementary Material (ESM 1), and are rated on a 5-point Likert scale ranging from 1 = *not at all like me* to 5 = *very much like me* (Duckworth & Quinn, 2009). The items hypothesized to reflect perseverance of effort were reverse-coded so that higher scores indicated more grit. Descriptive statistics are presented in Table B in ESM 1.

Sample 2

The dataset consisted of 522 US Mturk responders who completed the Grit-S within a large battery of self-regulation questionnaires (Eisenberg et al., 2018). The sample mean age was 33.63 years (SD = 7.87), 51% were females, 86% were white, and 44% were at least college-educated. Items hypothesized to reflect perseverance of effort were also reverse-coded so higher scores indicated more grit. Descriptive statistics are found in ESM 1, Table B. We refer to Sample 2 as the replication sample.

Psychometric Procedures

Generally, the same psychometric plan was carried out for both Sample 1 and Sample 2. All the analyses were conducted in the R statistical environment. First, the dimensionality of the Grit-S was investigated using Horn's (1965) parallel analysis using the `psych` R-package (Revelle, 2017). Given the results from PA, the factor structure of the Grit-S was determined by convergent information from measures of factor strength using the bifactor model (Rodriguez et al., 2016) and exploratory and confirmatory factor analyses using the `lavaan` R-package (Rosseel, 2012). Next, measurement bias was examined by investigating if the determined factor structure for the Grit-S varied across education. Data availability also allowed us to test measurement bias across gender and race. Model fit was compared using information from both χ^2 difference tests and alternative fit indices, such as the RMSEA (< .08), CFI (> .90), and SRMR (< .08), to suggest acceptable model fit (Lai & Green, 2016). A more restrictive model has the same fit as a more flexible model when there is a nonsignficant χ^2 difference test or when the CFI difference is less than .01 (Cheung & Rensvold, 2002). For IRT, the graded response model (GRM) was estimated because it is appropriate for Likert-type items with more than two categories (Thissen & Wainer, 2001). Item parameters for the GRM were estimated using marginal maximum likelihood (with the EM algorithm) in the `mirt` R-package (Chalmers, 2012) to investigate which items provide the most information on the grit latent variable (for more information on IRT estimation, see Thissen & Wainer, 2001). Finally, to study the construct overlap between grit and self-control, we obtained evidence for extrinsic convergent validity by testing if the correlations of the Grit-S and the BSCS (Brief Self-Control Scale; Tangney, Baumeister, & Boone, 2004) with three external criteria were the same (see ESM 1 for description of the scales that measured the outcomes). The external criteria were a measure of self-regulation (measured by the Short Self-Regulation Questionnaire, SSRQ; Carey, Neal, & Collins, 2004), cognitive appraisal and expressive suppression (both measured by the Emotion Regulation Questionnaire, ERQ; Gross & John, 2003), and mindfulness (measured by the Mindful Awareness Attention Scale, MAAS; Brown & Ryan, 2003). We used Williams' (1959) formula to test if the difference of the two dependent correlations was statistically significant. We chose self-control and these outcomes because they are all theorized to be part of the ontological network of self-regulation (Eisenberg et al., 2018), so one can study if grit could substitute self-control in that network.

Results

Psychometric Analyses in Sample 1

Parallel Analysis of the Grit-S

The eigenvalues for parallel analysis (PA) are shown in Figure A in ESM 1. PA suggests that two factors underlie the Grit-S scale, as hypothesized in the literature. However, a single factor seems to dominate the common variance in the Grit-S, with a ratio of the first to the second eigenvalue of 3.11:1 (1st eigenvalue = 3.580, 2nd eigenvalue = 1.148). The second eigenvalue is close to 1, so post hoc analyses using measures of factor strength and factor model fit could help determine if the Grit-S is essentially unidimensional.

Factor Strength

Measures of factor strength were estimated by fitting a bifactor measurement model to the Grit-S (Rodriguez et al., 2016). The bifactor measurement model is specified by fitting a general factor (grit) that explains variance common to all the items and two orthogonal group factors (CI and PE) that explain variance specific to each of the subscales theorized in the Grit-S. Here, we use the bifactor model to estimate the strength of the general factor, not as a theoretical model of the Grit-S. In this case, *omega hierarchical* (ω_H) describes if a total sum score from the Grit-S reflects reliable variance from a single dimension or multiple dimensions; and the *explained common variance* (ECV) index describes the ratio of the common variance explained by the general factor to the total common variance explained by the bifactor model, as in,

$$\omega_H = \frac{\left(\sum \lambda_{gen}\right)^2}{\left(\sum \lambda_{gen}\right)^2 + \left(\sum \lambda_{gr1}\right)^2 + \left(\sum \lambda_{gr2}\right)^2 + \sum \phi^2}; \qquad (1)$$
$$\mathrm{ECV} = \frac{\left(\sum \lambda^2_{gen}\right)}{\left(\sum \lambda^2_{gen}\right) + \left(\sum \lambda^2_{gr1}\right) + \left(\sum \lambda^2_{gr2}\right)}.$$

In this case, λ is a standardized factor loading of the item either on the general factor (λ_{gen}) or on group factor k (λ_{grk}), and ϕ^2 is the residual variance of each item. Higher values of ω_H and ECV indicate that the total score mostly reflects variance from a grit general factor. The bifactor model fit the Grit-S adequately, $\chi^2(12) = 59.418$, $p < .001$, RMSEA = .044, CFI = .990, SRMR = .018. For these analysis, ω_H = .700 and ECV = .671. For reference, the reliability coefficient α = .816 with a 95% bootstrapped confidence interval [.803, .828]; coefficient ω = .825 with 95% bootstrapped confidence interval [.813, .837]; and the total variance explained by the bifactor model was .478. The general factor accounts for 67.1% of the .478 variance explained by the bifactor model and accounts for 84.8% of the reliability coefficient ω estimate (.700/.825) for the whole scale. So, the vast majority of the reliable variance of the Grit-S is due to a single factor, and the Grit-S subscales might not provide much additional information than what is already found in the total score (Thissen & Wainer, 2001).

Factor Analysis

The estimated factor loadings for all models are presented in Table 1. The model fit of the unidimensional factor model of the Grit-S was $\chi^2(20) = 568.957$, $p < .001$, RMSEA = .116, CFI = .886, SRMR = .065. The model fit for a congeneric, two-factor model was $\chi^2(19) = 226.350$, $p < .001$, RMSEA = .073, CFI = .957, SRMR = .040, with a between-factor correlation of $r = .746$. Finally, the model fit of a noncongeneric, two-factor EFA was $\chi^2(13)$ = 51.617, $p < .001$, RMSEA = .038, TLI = .983, with a between-factor correlation of $r = .590$. The analyses suggest two factors, but upon closer inspection, there are two things to note. First, in the two-factor solutions, there is an excessively high correlation between the two factors. Second, in the noncongeneric two-factor model, the second factor is roughly defined by four items, but only two items, GS4 and GS8, load highly on the second factor and item GS7 cross-loaded onto both latent factors. Item content of GS4 and GS8 suggests that the items might be considered synonyms (see Table A in ESM 1). Similarly, modification indices for both the unidimensional model and the congeneric two-factor model suggested that a correlation between items GS4 and GS8 should be specified (similar to Schmidt et al., 2017). A unidimensional model with correlated uniqueness between GS4 and GS8 fits the dataset generally well, $\chi^2(19) = 296.306$, $p < .001$, RMSEA = .084, CFI = .942, SRMR = .050, and the model fits significantly better than the unidimensional model according to a χ^2 difference test, $\Delta\chi^2(1) = 272.651$, $p < .001$. Correlated uniquenesses violate the latent variable modeling assumption of local independence, where the items should be unrelated after accounting for the latent variable. A pair of locally dependent items is referred to as *doublet*. Violations of local independence can have adverse effects in estimating latent variables, such as biased factor loadings and their standard errors, but the JSI (see ESM 1 and Edwards et al., 2018) suggests that the presence of the *doublet* might not bias factor-loading estimates.

Measurement Bias and Measurement Invariance

Model fit information and parameter estimates from the measurement invariance models are presented in Table 2 and Table C in ESM 1, respectively. Generally, items GS4 and GS8 were flagged as items with bias across gender, race, and education level. For gender, a partial scalar invariance model fit the dataset well. Specifically, scrutiny of the residuals suggested that items GS4 and GS8 had different intercepts across males and females. At a latent grit score of zero, females rate themselves higher in items GS4 and GS8 than males. For education, a partial scalar invariance model fit the data well. Specifically, item GS4 had a different factor loading across levels of education, and items GS1, GS4, and GS8 had different intercepts across levels of education. As the latent variable increased, participants with higher education endorsed item GS4 at a lower rate than participants with lower education. Also, at a latent grit score of zero, participants with more education rated themselves higher in items GS4 and GS8 than participants with less education, and participants with less education rated themselves lower in item GS1 than participants with higher education. Finally, for race, a partial scalar invariance model fit the data well. Specifically, item GS4 had different intercepts across race groups. At a latent grit score of zero, Black participants rated themselves higher in item GS4 followed by White and then Asian participants.

Scale Information Using Item Response Theory

The slopes and threshold item parameters from the GRM are reported in Table 3 (also, see item fit information in ESM 1). Slope parameters (except one) range from 1.219 to 2.315, indicating reasonable associations between the items and the grit construct. The most discriminating items are GS6 and GS5. Consequently, these are the items that provide the most information in the top-left panel of Figure 1. The lowest slope parameter is .684 from item GS2, which is the item with the lowest factor loading in the one-factor solution. This is also the item with the flattest information curve in the top-left panel in Figure 1. All the thresholds ranged from −4.30 to 2.95. Most of the thresholds were negative or close to zero, which indicate that the

Table 1. Grit-S standardized factor loadings across four models for Studies 1 and 2

	One Factor Model	Non-congeneric Two Factor Model		Congeneric Two Factor Model		Correlated Uniqueness Model
	F1	F1	F2	F1	F2	F1
Study 1						
Factor λ						
GS1	.596	.670	−.070	.618	–	.609
GS3	.668	.740	−.040	.691	–	.682
GS5	.703	.750	−.020	.739	–	.718
GS6	.757	.670	.130	.772	–	.764
GS2	.332	.050	.350	–	.388	.320
GS4	.498	−.030	.720	–	.616	.450
GS7	.686	.400	.390	–	.739	.672
GS8	.545	.020	.710	–	.664	.503
GS4∼GS8	–	–	–	–	–	.371
Residuals ϕ						
GS1	.644	.600		.619		.630
GS3	.553	.490		.523		.536
GS5	.506	.460		.453		.485
GS6	.428	.430		.404		.416
GS2	.889	.850		.849		.898
GS4	.752	.510		.620		.798
GS7	.529	.500		.455		.548
GS8	.703	.470		.560		.747
			Study 2			
Factor λ						
GS1	.773	.820	−.010	.800	–	.787
GS3	.787	.910	−.090	.832	–	.805
GS5	.819	.800	.050	.845	–	.831
GS6	.811	.690	.170	.809	–	.815
GS2	.584	.280	.400	–	.612	.571
GS4	.566	−.050	.790	–	.691	.518
GS7	.794	.410	.500	–	.851	.776
GS8	.662	.000	.870	–	.799	.624
GS4∼GS8	–	–	–	–	–	.515
Residuals ϕ						
GS1	.402	.350		.359		.381
GS3	.381	.270		.308		.352
GS5	.330	.300		.287		.310
GS6	.342	.350		.346		.336
GS2	.658	.630		.626		.674
GS4	.680	.420		.523		.731
GS7	.370	.330		.276		.399
GS8	.562	.240		.362		.611

Note. λ = factor loading; ϕ = residual variance; ∼ = correlated uniqueness; F1 = factor 1; F2 = factor 2.

Grit-S mostly distinguishes participants at or below the mean of the grit construct. This agrees with the test information function presented in the right panel in Figure 1. Test information is also related to the reliability of the score, where 1 *minus* the inverse of the information function is the conditional reliability. The test information function suggests that participants who are within two standard deviations from the mean latent score of grit have a reliability around .75. However, Grit-S scores are less precise outside the two-standard-deviation range around the mean of the latent variable. It is important to note that items identified as potential doublets, GS4 and GS8, provide almost the

Table 2. Model comparison for measurement invariance

Model	χ^2	*df*	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	*p*-value
	Study 1							
Gender								
Configural	330.521	38	.940	.087	.048	–	–	–
Metric	343.865	45	.938	.081	.051	13.344	7	.064
Scalar	392.461	52	.930	.080	.055	48.596	7	< .001
Scalar – *i* = 8	386.803	51	.931	.080	.054	42.938	6	< .001
Scalar – *i* = 4, 8	358.674	50	.936	.078	.052	14.809	5	.011
Education								
Configural	366.760	73	.936	.089	.049	–	–	–
Metric	418.274	97	.929	.081	.062	51.514	24	< .001
Metric – *i* = 4	398.611	94	.933	.080	.057	31.851	21	.061
Scalar – *i* = 4	477.010	112	.920	.080	.062	78.399	18	< .001
Scalar – *i* = 1, 4	442.555	109	.927	.078	.060	43.944	15	< .001
Scalar – *i* = 1, 4, 8	420.421	106	.931	.076	.058	21.810	12	.040
Race								
Configural	362.330	76	.940	.086	.048	–	–	–
Metric	391.384	97	.939	.077	.054	29.054	21	.113
Scalar	432.118	118	.935	.073	.056	40.734	21	.006
Scalar – *i* = 4	429.963	115	.934	.074	.056	38.579	18	.003
	Study 2							
Gender								
Configural	203.570	38	.931	.129	.056	–	–	–
Metric	226.160	45	.924	.124	.083	22.590	7	.002
Metric – *i* = 8	222.697	44	.925	.125	.078	19.127	6	.004
Metric – *i* = 4, 8	213.567	43	.929	.123	.066	9.997	5	.075
Scalar – *i* = 4, 8	217.500	48	.929	.116	.067	3.933	5	.559
Education (high school degree and up to college degree)								
Configural	194.722	38	.930	.131	.056	–	–	–
Metric	198.335	45	.932	.119	.061	3.613	7	.823
Scalar	203.153	52	.933	.110	.061	4.818	7	.682

Notes. χ^2 = chi-square statistic; $\Delta\chi^2$ = chi-square difference; *df* = degrees of freedom; *i* = item free to vary.

same information and discriminate mainly below two standard deviations away from the mean. Thus, the information provided by these items appears to be redundant.

Psychometric Analyses in Sample 2

Parallel Analysis of the Grit-S

The distribution of eigenvalues from the Grit-S in Sample 2 is presented in the right panel of Figure A in ESM 1. PA in Sample 2 suggests that one factor underlies the Grit-S scale and that a single factor dominates the common variance in the Grit-S, with a ratio of the first to the second eigenvalue of 4.475:1 (1st eigenvalue = 4.730, 2nd eigenvalue = 1.057). Therefore, PA suggested that the Grit-S may essentially be unidimensional.

Factor Strength

Measures of factor strength were estimated by fitting the same bifactor measurement model to the Grit-S as in Sample 1 (Rodriguez et al., 2016). The bifactor model fits the Grit-S in Sample 2 adequately, $\chi^2(12) = 11.356$, $p = .499$). The measures of factor strength were $\omega_H = .807$ and ECV = .740. For reference, reliability coefficient $\alpha = .897$, 95% bootstrapped confidence interval [.883, .908], coefficient $\omega = .903$, 95% bootstrapped confidence interval [.890, .915], and the total variance explained by the bifactor model is .658. So, the general factor accounts for 74.0% of the .658 variance explained by the bifactor model and accounts for 89.4% of the reliability coefficient ω estimate (.807/.903) for the whole scale. The analysis suggests that the vast majority of the reliable variance is due to the total score and the variance explained by the bifactor model depends on a single factor.

Factor Analysis

The estimated factor loadings of all models are also presented in Table 1. The model fit of the unidimensional factor model of the Grit-S was $\chi^2(20) = 329.675$, $p < .001$, RMSEA

Table 3. IRT item parameters for the Grit-S using the Graded Response Model

Item	a1	b1	b2	b3	b4
Study 1					
GS1	1.448	−1.037	−0.037	1.325	2.956
GS3	1.741	−1.188	−0.376	0.592	1.791
GS5	1.937	−1.578	−0.655	0.284	1.706
GS6	2.315	−1.158	−0.482	0.299	1.438
GS2	0.684	−4.073	−1.360	0.573	2.554
GS4	1.219	−4.459	−2.984	−1.581	−0.253
GS7	1.867	−2.744	−1.318	−0.125	1.032
GS8	1.279	−4.305	−2.590	−0.735	0.632
Study 2					
GS1	2.495	−1.817	−0.552	0.279	1.518
GS3	2.583	−1.614	−0.467	0.273	1.394
GS5	3.077	−2.009	−0.829	0.012	1.235
GS6	3.133	−1.550	−0.762	−0.226	0.777
GS2	1.453	−2.049	−0.545	0.227	1.678
GS4	1.515	–	−2.859	−1.491	0.033
GS7	2.886	−2.445	−1.203	−0.422	0.685
GS8	1.910	–	−2.186	−1.017	0.495

Notes. a1 = slope parameters; b*k* = threshold parameter for *k*th threshold. In Sample 2, threshold b1 for GS4 and GS8 was not estimated because only one participant endorsed the lowest category in each of the items.

= .172, CFI = .870, SRMR = .077. The model fit for a congeneric, two-factor model was $\chi^2(19) = 124.674$, $p < .001$, RMSEA = .103, CFI = .956, SRMR = .050, with a between-factor correlation of $r = .778$. Finally, the model fit of a non-congeneric, two-factor EFA was $\chi^2(13) = 17.580$, $p < .174$, with a between-factor correlation of $r = .620$. Factor loading patterns similar to Sample 1 were observed. The same two items, GS4 and GS8, loaded highly on the second factor, and GS7 cross-loaded onto both latent factors. Also, the highest modification indices of both unidimensional and the congeneric two-factor model again suggested that a correlation between items GS4 and GS8 should be specified. A unidimensional model with correlated uniqueness between GS4 and GS8 fits the dataset generally well, although the RMSEA was higher than in Sample 1, $\chi^2(19) = 181.997$, $p < .001$, RMSEA = .128, CFI = .932, SRMR = .061. The unidimensional model with the correlated uniqueness fit significantly better than the unidimensional model according to a χ^2 difference test, $\Delta\chi^2(1) = 147.628$, $p < .001$. As with Sample 1, the JSI suggested that the presence of the doublet did not have adverse effects on parameter estimation.

Measurement Bias and Measurement Invariance

The covariates tested for measurement invariance in Sample 2 were limited by our sample. We did not conduct analyses on the race covariate because the replication sample was predominantly white (around 83%), so there were less than 50 cases in the other minority groups. For the education covariate, there were no participants who had less than high school education and the participants with graduate degrees were also limited. Therefore, we tested for measurement invariance across those who completed up to high school to those who completed college. Model fit information from the measurement invariance models are presented in Table 2. For education, a scalar model for the Grit-S unidimensional model with correlated uniqueness fit the data generally well, $\chi^2(52) = 203.153$, $p < .001$, RMSEA = .110, CFI = .933, SRMR = .061. In other words, items have the same factor loading and intercept across education levels. For gender, items GS4 and GS8 were flagged as items with bias, and a partial scalar invariance model fit the dataset generally well, $\chi^2(48) = 217.500$, $p < .001$, RMSEA = .116, CFI = .929, SRMR = .067. Specifically, scrutiny of the residuals suggested that items GS4 and GS8 had different factor loadings and intercepts across males and females. As the latent variable increased, males endorsed the items GS4 and GS8 at a higher rate than females. However, at a latent grit score of zero, females rated themselves higher in items GS4 and GS8 than males.

Scale Information Using Item Response Theory

The slopes and threshold item parameters from the GRM are reported in Table 3 (see ESM 1 for item fit information). Slope parameters ranged from 1.453 to 3.133, indicating reasonable associations between the items and the grit construct. Similar to Sample 1, the most discriminating items were GS6 and GS5. Consequently, these were the items that provide the most information in the bottom-left panel of Figure 1. The lowest slope parameter was 1.465 from item GS2, which was the item with the lowest factor loading in the one-factor solution. This was also the item with the flattest information curve in Figure 1. All the thresholds ranged from −2.445 to 1.678. Similar to Sample 1, most of the thresholds were negative or close to zero, which indicates that the Grit-S mostly distinguishes participants at or below the mean of the grit construct. This agrees with the test information function presented in the right panel in Figure 1. The test information function suggests that latent scores within two standard deviations from the mean of grit have a reliability around .90, however, scores outside the 2 *SD* range are more unreliable. Similar to Sample 1, GS4 and GS8 seem to provide overlapping information.

Extrinsic Convergent Validity

We investigated the hypothesis of extrinsic convergent validity for construct overlap by testing if the Grit-S had the same correlation as the BSCS with other self-regulation criteria. We used Williams' (1959) formula to test the difference of two dependent correlations. The correlation between the Grit-S and the BSCS was .728, suggesting construct overlap. The reliability estimate of the Grit-S was $\alpha = .897$, and the reliability estimate for the BSCS was

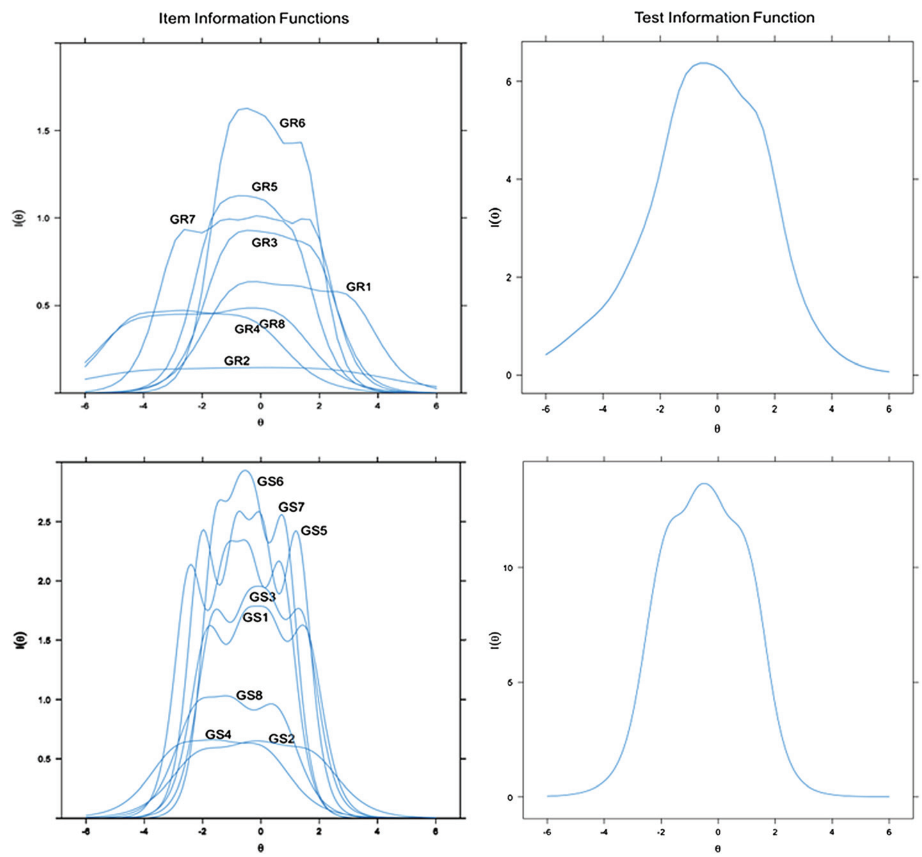

Figure 1. Item information functions and test information function for the Grit-S in Study 1 (top) and Study 2 (bottom). Note that the *Y*-axes are different across the four panels.

α = .914, so we do not expect the test of dependent correlations to be influenced by the difference in reliabilities. For the cognitive reappraisal subscale of the ERQ, the correlation with the Grit-S was r = .285, and for the BSCS was r = .299. The correlations of the Grit-S and the BSCS with cognitive reappraisal were not significantly different from each other, r_{diff} = .014, $t(519) = -.442$, p = .658. For the expressive suppression subscale of the ERQ, the correlation with the Grit-S was $r = -.146$, and for BSCS was $r = -.182$. The correlations of the Grit-S and the BSCS with expressive suppression were not significantly different from each other, r_{diff} = .036, $t(519)$ = 1.123, p = .262. For the MAAS score, the correlation with the Grit-S was r = .601, and for the BSCS was r = .587. The correlations of the Grit-S and the BSCS with the MAAS score were not significantly different from each other, r_{diff} = .014, $t(519)$ = .558, p = .577. Finally, for scores on the SSRQ, the correlation with the Grit-S was r = .763, and for the BSCS was r = .768. The correlations of the Grit-S and the BSCS with the SSRQ scores were not significantly different from each other, $r_{\mathrm{diff}} = -.006$, $t(519) = -.302$, p = .762. Therefore, extrinsic convergent validity evidence suggests that there is construct overlap between grit and self-control, and that the Grit-S and the BSCS could be functionally interchangeable measures to predict the constructs measured by the SSRQ, ERQ, and MAAS.

Discussion

The grit construct has gained attention in the areas of personality and positive psychology, but the assessment of grit has complications. In this paper, we conducted psychometric analyses of the Grit-S in two independent samples to examine the psychometric properties of the Grit-S. Our analyses suggest that the Grit-S is essentially unidimensional, and there is construct overlap between grit and self-control in the prediction of self-regulation, mindfulness, and emotion regulation. Substantively, the bifactor model and measures of factor strength suggested that a single factor accounts for most of the variance of the Grit-S items, and that subscales do not provide much additional information beyond what is already found in the total score. Therefore, evidence suggests that a total score should be reported from the Grit-S. Also, scale functioning analysis with IRT suggested that the Grit-S is more precise at assessing participants low on the latent variable as opposed to high on the latent variable.

It is important to consider the interpretation of the single factor measured by the Grit-S. Although the operational definition for grit includes the facets of perseverance of effort and consistency of interest, our empirical results of the two-factor EFA suggests that the items from the perseverance of effort did not define the construct clearly. A perseverance of effort item that measures if a person *finishes whatever they begin* also cross-loaded in the consistency of interest factor. A seemingly representative item of the perseverance of effort construct that deals with *overcoming setbacks* did not load highly on the respective factor. In the same vein, the hypothesized two-factor solution of the Grit-S is likely due to an item *doublet* (GS4 and GS8) in the perseverance of effort items, where the item pair had an excess correlation after accounting for a single latent variable. The correlated uniqueness of the item pair was observed in both samples; the items provided similar item information; and the item pair exhibited measurement bias across gender, race, and education. This excess correlation is not surprising because the word *diligent* (in GS4) is considered a synonym of *hard-working* (in GS8) according to Merriam-Webster's definition, so these items might be giving qualitatively similar information. Schmidt and colleagues (2017) also found an excess correlation among these items, and Tyumeneva, Kardanova, and Kuzmina, (2017) discuss these items as similar. As a result, the single factor of the Grit-S is empirically representing mostly consistency of interest, suggesting that the Grit-S suffers from limited content coverage. It is unlikely that the Grit-S scores could be interpreted as "passion and perseverance for long-term goals" if the items are empirically representing consistency of interest and not much on perseverance.

We ask researchers to consider the possible narrow interpretation and limited content coverage of the Grit-S scores before it is used. If the Grit-S were to be used, we suggest using a total Grit-S score and consider dropping one item of the doublet. However, it is unclear which specific item should be eliminated, because items that exhibited local dependence also exhibited measurement bias. From a scale construction perspective, these items might be redundant, so one of the items might be dropped without losing practical information. In terms of measurement bias, for the items in the doublet that violated scalar invariance, differences in intercepts are interpreted in the same metric as the item response. Most of these intercept differences were around .2, so one goal of future research is to investigate if an observed difference of .2 is meaningful in research using the scale. These decisions rely on researcher's expertise because research on measurement invariance effect sizes is limited (Millsap, 2011). Otherwise, both Grit-S items could be kept, but never administered together. Then, IRT could be used to equate the scores across two different versions of the Grit-S (Thissen & Wainer, 2001). Given previous considerations, we think that it is most sensible to drop an item based on word complexity. It is hypothesized that the word "diligent" requires a higher reading level and it is not as self-descriptive as the word "hard-working," so our preliminary recommendation is to drop the "diligent" item. Overall, understanding the dimensionality, reliability, and measurement bias in the Grit-S could prevent estimation problems, especially in studying construct overlap with other measures or if the Grit-S is used in intervention studies as a mediator (MacKinnon, 2008).

There were several limitations to our analyses. First, the publicly available sample did not have as much variability as the replication sample. The correlations among the Grit-S items in Sample 2 were higher than the correlations among items in Sample 1. Consequently, factor loadings and item information functions in Sample 2 were larger than Sample 1, although the general pattern of factor loadings and item information were the same. It is hypothesized that these differences in correlations may be due to the nature of the participants, where those in Sample 1 were volunteers that were interested in taking the survey online and those in Sample 2 were recruited and paid for their participation. Also, a unidimensional model with a correlated uniqueness for the Grit-S provides a trade-off between model simplicity and model fit, although not all the alternative fit indices suggested adequate fit to the model (Lai & Green, 2016). Finally, our dimensionality results from the Grit-S should not be generalized to other grit measures (Duckworth & Quinn, 2009). We determined that the perseverance factor in the Grit-S was poorly defined and did not follow simple structure, but factor structure might be better defined as more items from the perseverance of effort facet are added (Tyumeneva et al., 2017). Similarly, further studies might consider developing more items that measure perseverance of effort and passion, which are integral to the definition of grit, but underrepresented in the Grit-S. Also, researchers could develop items that discriminate participants high on the latent construct, especially when studying change in grit over time or when high precision is needed to measure those high on grit. Item response theory methods could help with these efforts (Thissen & Wainer, 2001).

Future directions in this research are to investigate the construct overlap between grit and other constructs. Here, we proposed a method to study construct overlap and obtained preliminary results on the overlap between grit and self-control to predict self-regulation measures. Although Duckworth and Gross (2014) suggest that, in theory, self-control is important for short-term goals and grit is important for long-term goals, the Grit-S items do not reflect long-term goals, so it is likely that there is empirical overlap between the two constructs (Muenks et al., 2017). The extrinsic convergent validity framework (Fiske, 1971) would provide necessary, but not sufficient, evidence on how constructs overlap in the same criterion space. Factor-analytic methods, testing for incremental prediction, and item content analysis could also provide supplemental evidence for construct overlap. Furthermore, more comprehensive models could be included to study (1) the overlap between grit with other constructs (e.g., tenacity, persistence), and (2) the simultaneous relationship between the overlapping constructs with external criteria. Here, we used separate analyses per outcome to clarify specific relationships (as in Muenks et al., 2017).

In summary, we found empirical results that support reporting a total score of Grit-S because it is largely unidimensional, but the Grit-S could lead to scores with limited interpretation because of its limited content coverage. These results are useful as researchers scrutinize the similarity between the grit and other personality constructs. Uncertainty on the factor structure, measurement bias, and item functioning could compromise those results.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000535

ESM 1. Tables A–C show item content, item descriptives, and parameter estimates for measurement invariance, respectively. Figure A shows scree plot of parallel analysis. Local dependence and item fit information is also discussed. Figures B–F show distributions and scatterplots (.docx).

References

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Carey, K. B., Neal, D. J., & Collins, S. E. (2004). A psychometric analysis of the Self-Regulation Questionnaire. *Addictive Behaviors, 29*, 253–260. https://doi.org/10.1016/j.addbeh.2003.08.001

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Christensen, R., & Knezek, G. (2014). Comparative measures of grit, tenacity and perseverance. *International Journal of Learning, Teaching and Educational Research, 8*, 16–30.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology, 113*, 492–511. https://doi.org/10.1037/pspp0000102

Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., & Gollwitzer, P. M. (2011). Self-regulation strategies improve self-discipline in adolescents: Benefits of mental contrasting and implementation intentions. *Educational Psychology, 31*, 17–26. https://doi.org/10.1080/01443410.2010.506003

Duckworth, A. L., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science, 23*, 319–325. https://doi.org/10.1177/0963721414541462

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science, 2*, 174–181. https://doi.org/10.1177/1948550610385872

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of Personality Assessment, 91*, 166–174. https://doi.org/10.1080/00223890802634290

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods, 23*, 138–149. https://doi.org/10.1037/met0000121

Eisenberg, I. W., Bissett, P. G., Canning, J. R., Dallery, J., Enkavi, A. Z., Gabrieli, S. W., . . . Poldrack, R. A. (2018). Applying novel technologies and methods to inform the ontology of self-regulation. *Behavior Research and Therapy, 101*, 46–57. https://doi.org/10.1016/j.brat.2017.09.014

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Eskreis-Winkler, L., Duckworth, A. L., Shulman, E. P., & Beal, S. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology, 5*, 36. https://doi.org/10.3389/fpsyg.2014.00036

Eskreis-Winkler, L., Shulman, E. P., Young, V., Tsukayama, E., Brunwasser, S. M., & Duckworth, A. L. (2016). Using wise interventions to motivate deliberate practice. *Journal of Personality and Social Psychology, 111*, 728–744.

Fiske, D. W. (1971). *Measuring the concepts of personality*. Chicago, IL: Aldine.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Haktanir, A., Lenz, A. S., Can, N., & Watson, J. C. (2016). Development and evaluation of Turkish language versions of three positive psychology assessments. *International Journal for the Advancement of Counseling, 38*, 286–297. https://doi.org/10.1007/s10447-016-9272-9

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185. https://doi.org/10.1007/BF02289447

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research, 51*, 220–239. https://doi.org/10.1080/00273171.2015.1134306

Li, J., Zhao, Y., Kong, F., Du, S., Yang, S., & Wang, S. (2016). Psychometric assessment of the Short Grit Scale among

Chinese adolescents. *Journal of Psychoeducational Assessment, 36*, 291–296. https://doi.org/10.1177/0734282916674858

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence', objectively determined and measured". *Journal of Personality and Social Psychology, 86*, 96–111. https://doi.org/10.1037/0022-3514.86.1.96

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Erlbaum.

McCornack, R. L. (1956). A criticism of studies comparing item-weighting methods. *Journal of Applied Psychology, 40*, 343–344.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). New York, NY: Gilford Press.

Muenks, K., Wigfield, A., Yang, J. S., & O'Neal, C. R. (2017). How true is grit? Assessing its relations to high school and college students' personality characteristics, self-regulation, engagement, and achievement. *Journal of Educational Psychology, 109*, 599–620. https://doi.org/10.1037/edu0000153

O'Neal, C. R., Espino, M. M., Goldthrite, A., Morin, M. F., Weston, L., Hernandez, P., & Fuhrmann, A. (2016). Grit under duress: Stress, strengths, and academic success among non-citizen and citizen Latina/o first-generation college students. *Hispanic Journal of Behavioral Sciences, 38*, 446–466. https://doi.org/10.1177/0739986316660775

Revelle, W. (2017). *psych: Procedures for personality and psychological research (Version 1.7.5.)*. Evanston, IL: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*, 137–150. https://doi.org/10.1037/met0000045

Rojas, J. P., Reser, J. A., Usher, E. L., & Toland, M. D. (2012). *Psychometric properties of the academic grit scale*. Lexington, KY: University of Kentucky.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Schmidt, F. T., Fleckenstein, J., Retelsdorf, J., Eskreis-Winkler, L., & Möller, J. (2017). Measuring grit: A German validation and domain-specific approach to Grit. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015-5759/a000407

Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Washington, DC: US Department of Education, Department of Educational Technology.

Strayhorn, T. L. (2014). What role does grit play in the academic success of black male collegians at predominantly white institutions? *Journal of African American Studies, 18*, 1–10. https://doi.org/10.1007/s12111-012-9243-0

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*, 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Thissen, D. & Wainer, H. (Eds.). (2001). *Test scoring*. New York, NY: Routledge.

Tyumeneva, Y., Kardanova, E., & Kuzmina, J. (2017). Grit: Two related but independent constructs instead of one. Evidence from Item Response Theory. *European Journal of Psychological Assessment*. Advance online publication. https://doi.org/10.1027/1015–5759/a000424

Weston, L. C. (2014). *A replication and extension of psychometric research on the grit scale* (Unpublished master's thesis). University of Maryland, College Park, MD.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B, 21*, 396–399.

History
Received April 6, 2018
Revision received January 30, 2019
Accepted February 1, 2019
Published online July 2, 2019
EJPA Section/Category Personality

Funding
This research was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1311230 and the National Institute on Drug Abuse under Grants No. R37 DA09757 and No. UH2DA041713.

Oscar Gonzalez
Department of Psychology and Neuroscience
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
USA
ogonza13@unc.edu

