

Chapter 8

Deconstructing Grit's Validity: The Case for Revising Grit Measures and Theory



Michael C. Tynan

Abstract Current measures of grit misrepresent its original conceptualization. This chapter aims to outline the flaws of extant grit measures, particularly in light of recent updated recommendations for maximizing construct validity during scale development. After reviewing empirical findings regarding grit's construct validity, structure, and association with success outcomes, recommendations for the future development of grit will be proposed. Grit was developed as a higher-order construct consisting of two facets: perseverance of effort and consistency of interests. However, the higher order construct of "overall grit" is not supported by item response theory, factor analytic, or structural equation modeling approaches. Grit is currently better interpreted as two separate constructs, possibly within a bifactor model in which "overall grit" does not consist of perseverance and consistency. After controlling for conscientiousness, perseverance explains little incremental variance in success outcomes. However, perseverance may be related to, but distinct from, lower-order facets of conscientiousness. Consistency is much less associated with success outcomes, and its predictive utility is unclear.

Grit researchers would benefit from revising grit measures. Positively and negatively worded items should be used to capture both constructs. An overinclusive set of new items should be administered to heterogeneous samples before the scales are refined. Further structural analysis is recommended, and similar factor analytic, item response theory, and structural equation modeling techniques used for the original grit scale should be employed for any revised measure. Meaningful inferences about grit may only be drawn after such revisions, as the first conceptual iteration of grit is psychometrically unsatisfactory.

Keywords Grit · Conscientiousness · Construct validity · Discriminant validity · Measurement

M. C. Tynan (✉)

Department of Psychology, Iowa State University, Ames, IA, USA

e-mail: mtynan@iastate.edu

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

L. E. van Zyl et al. (eds.), *Multidisciplinary Perspectives on Grit*,
https://doi.org/10.1007/978-3-030-57389-8_8

137

8.1 Introduction

8.1.1 *Deconstructing Grit's Validity: The Case for Revising Grit Measures and Theory*

Grit has received significant attention as a non-cognitive variable associated with academic success. Popularized in part outside the psychological field by a viral TED talk, a best-selling general audience book (Duckworth, 2016), multiple articles in major newspapers, and American educators' renewed interest in "character education," interest in grit seemed to grow at a faster rate than other psychological concepts. Grit's popularity has translated to school policy in some settings. Most notably, a U.S. Department of Education report recommended that grit be taught in schools (Schechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013), and the Knowledge Is Power Program (KIPP) network of 242 charter schools incorporated grit into educational materials and included the increased grit of students as one of the school's "character strength" goals (KIPP, 2019).

The initial conceptualization of grit attempted to distinguish it from related concepts like conscientiousness (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009). Outside the original authors' initial construct validation attempts, few studies have been dedicated to the validity of interpretations of grit measures. Construct validity determines the worth of real-world decisions based on psychological research (Clark & Watson, 2019). Considering the real-world implementation of interventions to increase grit and the popularity of grit research around the world, these oversights in the validity of extant grit scales are worth revisiting. This chapter will outline the validation of grit measures in light of the recent guide to best practices to maximize construct validity (Clark & Watson, 2019). This chapter follows a structure assuming Loevinger's (1957) outline of the three components of construct validity: substantive, structural, and external (Clark & Watson, 1995, 2019) as each pertains to the development and administration of existing grit scales. I will limit discussion to the original grit scale (Grit-O) and short grit scale (Grit-S), while acknowledging that many translations or adapted domain-specific versions of these scales exist. Seeing as translations and adapted versions of the primary grit scales are derivative, they are assumed to be similarly problematic in their use and interpretation, barring substantial differences in structure or scale meaning. One notable example of such deviation from the original scale stems from poor fit of the original grit concept in collectivist cultures (Datu, Valdez, & King, 2015). While not the focus of this chapter, as the Grit-O and Grit-S are by far the most commonly used grit scales and most adapted scales shared their structure, I refer interested readers to the (three-facet) Triarchic Model of Grit Scale (see Datu, Yuen, & Chen, 2017).

By outlining the construct validity considerations ideally made when introducing a new concept in psychology and comparing that ideal to the development of grit, I hope to create a case for significant revisions of grit measures and greater caution in interpreting current findings related to grit. This chapter focuses heavily on the two

foundational grit articles (Duckworth et al., 2007; Duckworth & Quinn, 2009), and at the risk of repeating information familiar to grit researchers, I hope the lens of construct validity allows readers to approach these works with a fresh perspective.

8.2 Substantive Validity

Substantive validity is the facet of construct validity concerned with the items generated and the items finalized into a scale being truly indicative of the construct being proposed (Loevinger, 1957). This section will focus on the process of the Grit-O's inception and further reductions in items to produce the Grit-S.

There is seemingly no limit to the number of possible psychological constructs to propose as possible new measures, and to illustrate this idea consider all descriptions of human characteristics in the English language in addition to all similar descriptions in the approximately 7000 existing human languages (Clark & Watson, 1995, 2019). In describing a person who is generally achievement-oriented, determined, and perseverant, one out of myriad options for that description is "grit." Grit was developed as a non-cognitive ability capturing "perseverance and passion for long-term goals" (Duckworth et al., 2007, p. 1087). This conceptualization involves generally working hard to achieve a desired outcome and, importantly, having that outcome take a considerable amount of time (e.g. Duckworth and colleagues offer "years" as a marker of "long-term," p. 1088). Within such time, distractions, discouragement, and disinterest must be overcome. Gritty individuals do not entirely avoid boredom or the desire to quit, but rather continue with their initial intention despite short-term situational hurdles. In developing grit, Duckworth and colleagues set out in attempt to write a detailed description of what a gritty individual looks like, and what behaviors are expected of that person. The conceptualization process described is largely in line with ideal construct validity practices (Clark & Watson, 1995, 2019).

8.2.1 *Literature Search and Hierarchical Structure of Constructs*

The next step in establishing a distinct construct is to search the extant literature for similar constructs, theoretical positioning against "near-neighbor" constructs, and to be able to describe the construct's level of breadth (Clark & Watson, 1995, 2019). The consideration of grit's place in both its breadth and relation to similar constructs within the nomological network seems to be lacking. One clear goal in developing grit was to establish a construct that had four characteristics: psychometric soundness, generalizability across age groups and achievement domains, low likelihood of ceiling effects for high-achievers, and relevance to the proposed grit definition

(Duckworth et al., 2007). Upon finding no scale that captured all four criteria, the initial literature search to compare grit to neighboring constructs appears to be limited to intelligence (used somewhat interchangeably with the vaguer term “talent”), need for achievement, and conscientiousness (Duckworth et al., 2007). A number of other constructs are plausible near-neighbors to grit, as later research suggests, for example: industriousness, self-control, effort regulation, engagement, and self-efficacy (Credé, Tynan, & Harms, 2017; Muenks, Wigfield, Yang, & O’Neal, 2017; Schmidt, Nagy, Fleckenstein, Möller, & Retelsdorf, 2018). In addition to testing an over-inclusive set of related constructs, Clark and Watson (1995, 2019) emphasize the need to consider the place of a new construct within a hierarchy of similar constructs. The consideration of intelligence, conscientiousness, and need for achievement appear to be assumed to have no hierarchical relation to one another, and it is unclear what level of abstraction grit is supposed to occupy. The inference that grit is at the same conceptual level as conscientiousness in particular ignores the possibility of conscientiousness’ lower order facets (competence, order, dutifulness, self-discipline, achievement-striving, deliberation) being more strongly related to grit than global conscientiousness (see Costa, McCrae, & Dye, 1991). Theoretical distinction from self-control is offered by Duckworth et al. (2007), again by emphasizing time frame: “An individual high in self-control but moderate in grit may, for example, . . . resist the urge to surf the Internet at work—yet switch careers annually” (p. 1089). This theoretical distinction is not supported by empirical findings for grit and the self-discipline facet of conscientiousness or another self-control measure. Later work suggests that grit may be fully integrated into the hierarchical structure of conscientiousness, as it is strongly related to general conscientiousness, industriousness, and self-discipline, but the original oversight of grit’s conceptual breadth delayed this theoretically germane finding (Schmidt et al., 2018).

Appropriate steps were taken, using multiple measures (SAT scores, Verbal IQ, and Whole Candidate Scores for military cadets), to explore the relation between intelligence and grit (Duckworth et al., 2007). There are sound theoretical reasons why grit is expected to be unrelated to intelligence, namely the variance in achievement between individuals known to have similar levels of intelligence and the content of the construct focusing on effort and interest rather than ability. The lack of relation between grit and intelligence found by Duckworth et al. (2007) has been confirmed by later meta-analytic estimates (Credé et al., 2017). Less establishing evidence for grit is provided in comparison to both need for achievement and conscientiousness. Duckworth et al. (2007) do not test the original grit scale against need for achievement. The evidence for this distinction appears to be primarily theoretical, in that individuals high in need for achievement are thought to require external incentives or positive feedback while gritty individuals do not (Duckworth & Quinn, 2009), but this distinction is not supported empirically.

Conscientiousness seemed to be the nearest conceptual neighbor to grit from its inception. The theoretical justification offered for why conscientiousness and grit ought to be considered separately is that grit emphasizes “long term stamina rather than short-term intensity” (Duckworth et al., 2007, p. 1089). Despite this theoretical

distinction, conscientiousness is correlated very highly with grit in both the original, $r = .77$, $r = .64$, and short scale development papers, $r = .77$ (Duckworth et al., 2007; Duckworth & Quinn, 2009). Grit is predictive of outcomes in the original paper after controlling for conscientiousness, but both grit levels and these outcomes are subject to range restriction making the conscientiousness-grit relation unclear in terms of incremental validity. Meta-analytic estimates indicate that grit does not explain additional variance in success outcomes like grades in college after controlling for conscientiousness (Credé et al., 2017).

8.2.2 *Creation of an Item Pool*

The original grit measure was derived from an initial 27-item pool (Duckworth et al., 2007). Items were written to capture “the ability to sustain effort in the face of adversity” and “the consistency of interests over time” (p. 1090). 10 items were excluded from the original 27 before an exploratory factor analysis (EFA) was conducted on the remaining 17. The available sample size was 1545, and EFA was conducted using the responses of half the sample chosen at random ($N = 772$). Retaining at least 5 items with loadings greater than .40 and examining the scree plot to determine factors, using oblique factor extraction with a promax rotation, yielded a two-factor solution with 6 items per factor.

The cardinal rule of item generation is to be over-inclusive: “*At very least, the items in the pool should be drawn from an area of content defined more broadly than the trait expected to be measured*” (Loevinger, 1957, p. 659, emphasis in original). Any initial item pool should include items assessing tangential content or information only marginally related to the theorized construct, with the intention of possible exclusion later in the scale development process (Clark & Watson, 1995, 2019). Put simply, initial scale items “*should be chosen so as to sample all possible contents which might comprise the putative trait according to all known alternative theories of the trait*” (Loevinger, 1957, p. 659, emphasis in original). Speculating upon what was *not* included in the development of the Grit-O is obviously fraught territory. The original 27 items are not included in supplemental materials or appendices to the original article. Readers should be cautioned against over-weighting inferences from unclear or incomplete information in grit’s foundational articles. However, the description of item reduction clearly prioritizes internal consistency over construct breadth, which risks construct distortion via the attenuation paradox (i.e. as the scope of a scale narrows, internal consistency increases through item redundancy, therefore reducing the amount of construct-relevant information provided by the scale; Clark & Watson, 2019; Loevinger, 1954). The first 10 items were eliminated by consulting “item-total correlations, internal reliability coefficients, redundancy, and simplicity of vocabulary” (Duckworth et al., 2007, p. 1090) *before* conducting an EFA on the remaining 17. This mixed method of item reduction muddies the conceptualization of grit as either a single construct or a multidimensional one. Item-total correlations are typically appropriate for reducing the items in a scale proposed to be

unidimensional, whereas factor analytic techniques are necessary to reduce items in a scale with multiple proposed subscales (Clark & Watson, 1995, 2019). Grit items were written so as to capture two passion and perseverance subscales, so what may have been more informative was to conduct EFA on all 27 initial items. Guidelines in conducting and interpreting exploratory factor analyses are relatively straightforward (see Russell, 2002), and I will emphasize only one additional point. The interpretation of scree plots to determine the number of factors is subjective. Parallel analysis eliminates this ambiguity by comparing the eigenvalues extracted from the scale in question to eigenvalues extracted from random data with the same parameters (sample size and number of items) and retaining the factors of the scale with eigenvalues greater than those of random data (see Hayton, Allen, & Scarpello, 2004). Eliminating ambiguity in factor retention may reduce questionable scale reduction practices, and parallel analysis is the recommended tool in all early-stage scale development.

The two grit factors retained after item-total correlation reduction and EFA are acknowledged to be confounded with positively and negatively worded items. That is, all perseverance items are positively scored (e.g. “I am a hard worker”), and all consistency items are negatively scored (e.g. “I have been obsessed with a certain idea or project for a short time but later lost interest”). The only justification offered is that the researchers were “convinced the factor structure reflected two conceptually distinct dimensions” (Duckworth et al., 2007, p. 1090). Negatively worded items are known to be able to artificially create a distinct factor within factors analyses because they are more difficult to answer (Credé, 2018; Schmitt & Stults, 1985; Swain, Weathers, & Niedrich, 2008). The implications of factors being confounded with item direction will be discussed further in the context of structural validity.

8.2.3 *Derivative Versions*

Several questions of substantive validity arise for the Grit-S, considering grit’s definition and original justification for distinction from conscientiousness. If the key proposed difference between conscientiousness and grit is the time-frame—grit refers to “long-term stamina,”—then the Grit-O items omitted from the Grit-S appear on their face to eliminate a richer measurement of time-relevant information. The following items are included in the Grit-O but not the Grit-S (emphasis added): “My interests change *from year to year*,” “I become interested in new pursuits *every few months*,” “I have achieved a goal that took *years of work*” (Duckworth & Quinn, 2009). The only remaining reference to specific long-term time frames in the Grit-S is one consistency item: “I have difficulty maintaining my focus on projects that take *more than a few months* to complete.” Losing this lack of depth in item content for only minimal gains in convenience of administration or ease of response distorts the original conceptualization of grit. Development of grit’s short version appears to have ignored this and other pitfalls of short scales. Resorting to very short measures

of individual difference constructs (e.g. four items per facet) risks attenuating effects and increasing both Type 1 and Type 2 error rates (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). The Grit-S scale not only loses face valid time-frame information but also alters the internal structure of the scale. Meta-analytic estimates indicate that the perseverance and consistency facets are much less similar for the Grit-O, $\rho = .27$, than for the Grit-S, $\rho = .66$ (Credé et al., 2017). This difference indicates that the shortening of the grit scale results in a notable alteration of the construct being measured, such that responses to the original grit scale differentiate the two facets to a greater extent than responses to the short grit scale, with apparent loss of the crucial time-frame content which supposedly separates grit from conscientiousness.

It is unclear from Duckworth and Quinn's (2009) terse introduction what the benefit of the Grit-S ought to be other than an attempt to improve the fit of the higher-order, two-factor grit model and general increased efficiency. Modest affordances in scale efficiency do not seem justified in this case. An 8-item scale would likely take most participants only a few seconds less to complete than a 12-item scale with an equal ratio of positively and negatively worded items. The modest improvement in scale administration efficiency does not appear on its face to be worth the loss in information from the original scale. As will be discussed in the structural validity section, improvements may be made in the structural model of grit, but item reduction in pursuit of these improvements produces more confusion than clarity. Grit may serve as a cautionary tale in the attenuation paradox: the Grit-S is seemingly a psychometric improvement on the Grit-O but is in fact a mischaracterization of grit's proposed definition (Clark & Watson, 2019).

Overall, improvements in substantive validity may be made by returning focus to the Grit-O. If a key component of grit is goal adherence in a long-term time frame, the Grit-O contains notably more information than the Grit-S in this regard. Revisiting the substantive validity of grit will necessitate new item generation and overinclusion of these items encompassing the theoretical spectrum of perseverance and consistency to the fullest extent possible. These items should be both positively and negatively worded, with items content encompassing the full grit concept, with particular emphasis on time frame. Scale reduction and validation against conceptual neighbors like conscientiousness should be informed by current best practices (Clark & Watson, 2019; Hayton et al., 2004; Russell, 2002).

8.3 Structural Validity

Structural validity refers to the element of construct validity concerned with the relations between scale items resembling other manifestations of the construct being proposed (Loevinger, 1957). Some issues with item selection and reduction have already been discussed with regard to substantive validity of grit scales, though they are also relevant for structural validity. This section will primarily focus on the

proposed higher-order, two-factor structure of the Grit-O and Grit-S before summarizing more recent findings related to the internal structure of grit measures.

Grit is deemed explicitly by Clark and Watson (2019) to be a structurally problematic *conglomerate* construct, in that the overall grit factor claims to be more than the sum of its parts (perseverance and consistency). Specifically, the higher-order grit factor proposed by its developers is meant to be more predictive of achievement than either of its two lower-order facets. In fact, by meta-analytic estimates, the perseverance facet is more predictive of success outcomes like GPA, $\rho = .26$, than “overall grit,” $\rho = .17$ (Credé et al., 2017). Conscientiousness is also approximately equally related to both “overall grit,” $\rho = .84$, and the perseverance facet, $\rho = .83$ (Credé et al., 2017). Another large-scale ($N = 11,750$) investigation of grit’s construct validity concludes that the higher-order model does not fit the data, and most of the predictive power of grit is explained by the perseverance facet (Fosnacht, Copridge, & Sarraf, 2018). Skepticism appears to be warranted toward studies reporting results for “overall grit,” especially when the Grit-S is the only measure used, for reasons discussed subsequently. There is growing evidence that perseverance and consistency are two separate constructs, and “overall grit” is not psychometrically meaningful (Credé et al., 2017; Disabato, Goodman, & Kashdan, 2019; Fosnacht et al., 2018; Guo, Tang, & Xu, 2019; Tyumeneva, Kardanova, & Kuzmina, 2017).

Several issues are notable in the identification of a higher order, two-facet solution for the grit scale. The Grit-S is especially problematic in this area, due to the confirmatory factor analysis performed in an attempt to replicate the Grit-O structure. A factor structure with one second-order factor and two first-order factors cannot be identified at the higher-order level (Credé et al., 2017; Kline, 2011). That is, the Grit-S developers attempted to solve the equation $a*b = y$, where a and b represent the paths from the first order facets to the second order factor and y is held constant as the correlation between perseverance and consistency (Credé, 2018). There are infinite solutions to this problem of identifying a and b (e.g. any two values between 0 and 1 that produce the correlation coefficient), and the fit of this misidentified model will be identical to a model which keeps perseverance and consistency as separate but correlated constructs with no higher-order factor (see Fig. 8.1; Credé, 2018; Credé & Harms, 2015). Duckworth et al. (2007) tested this two-factor solution and reported poor fit for the model, CFI = .83, RMSEA = .11. The higher-order solution found using confirmatory factor analysis is not meaningfully different from a two-factor solution with no higher-order factor, and therefore it does not support the idea of a higher-order model.

More generally, the proposed higher-order model equates individuals with high levels of perseverance and low levels of consistency with individuals who have low levels of perseverance and high levels of consistency, when these individuals are substantively different from one another (Credé, 2018). An alternative conceptualization of grit is that grit is composed only of high-perseverance combined with high-consistency. Duckworth et al. (2007) hint in their original conceptualization that this combination is optimal and reflective of highly successful people. Gritty individuals may need to fulfill both these requirements, and the failure to meet a

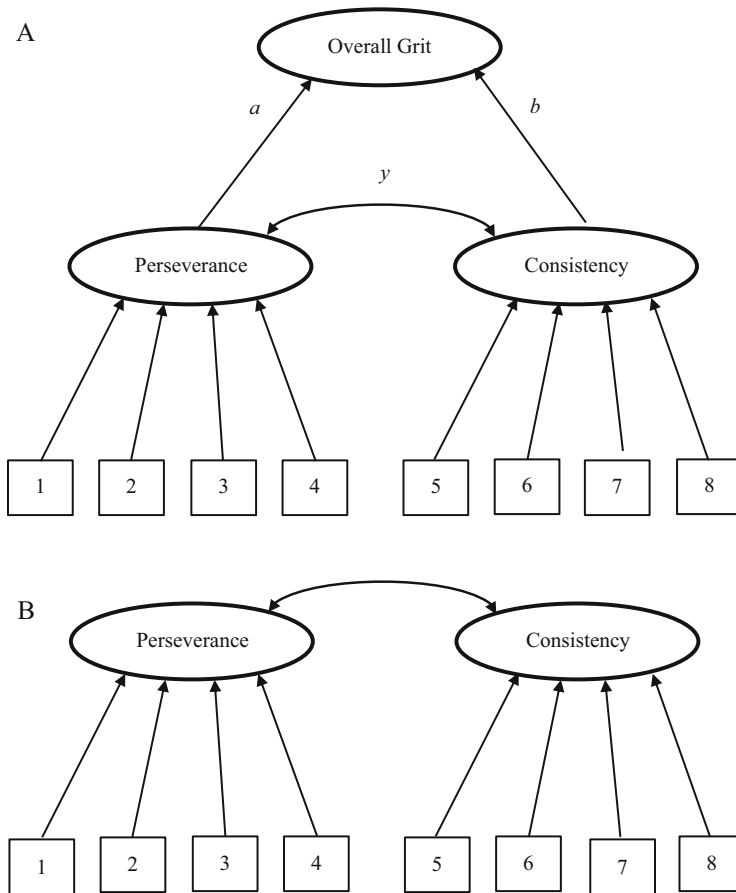


Fig. 8.1 The original proposed factor structure of the Grit-S (a) and the empirically supported alternative model (b). *Note:* a and b are paths calculated between lower order factors and the higher order factor (overall grit) with y is held constant. The fit of these two models is equal, and model B is recommended

threshold of either construct means an individual cannot be described as possessing grit. This conceptualization has not been tested empirically but may be using necessary condition analysis (see Dul, 2016) or cluster analytic approaches (Credé, 2018).

8.3.1 Item Response Theory

Item response theory (IRT) has a number of advantages in the course of investigating novel constructs. IRT can calculate probabilities of item response endorsement

based on latent trait levels, considering one or more traits at the same time, and clarify the dimensionality of a measure (Clark & Watson, 2019; Smith, 2002; Tyumeneva et al., 2017). Few studies have examined grit using IRT techniques, but findings thereof are briefly discussed here.

The Grit-O measures two distinct constructs as tested by IRT methods, further indicating that skepticism is warranted toward a higher-order, two-factor model (Tyumeneva et al., 2017). After conducting a principal components analysis of the standardized residuals of a Rating Scale Model and confirming the multidimensionality of the Grit-O through IRT model selection, Tyumeneva et al. (2017) conclude that psychometric evidence of a single construct underlying the Grit-O has not been found. Additionally, IRT based findings seem to suggest the Grit-S is more consistent with a unidimensional trait that largely overlaps with self-control (Gonzalez, Canning, Smyth, & Mackinnon, 2019). Item-level analyses point to an item doublet between the perseverance items “I am diligent” and “I am a hard worker,” wherein these items are excessively correlated after accounting for a single latent factor (Gonzalez et al., 2019; Schmidt, Fleckenstein, Retelsdorf, Eskreis-Winkler, & Möller, 2019; Tyumeneva et al., 2017). These items are therefore not providing qualitatively different information but are driving the originally hypothesized two-factor solution, when in fact the Grit-S is more aligned with a unidimensional construct more indicative of consistency than perseverance (Gonzalez et al., 2019).

Multidimensional item response theory (MIRT) analyses have been conducted on the Grit-S, allowing for examination of the interaction between the Grit-S and respondents when assumptions of a continuous response format (e.g. Likert-type scales) are violated (Wirth & Edwards, 2007). Muenks et al. (2017) found that participants were not using all five options of the response scale, and responses were not normally distributed for any Grit-S item, indicating that the scale is better examined as having categorical items. Using the same IRT-based methods, perseverance and consistency were found to overlap with neighboring constructs. One factor appears to underlie consistency, effort regulation, and behavioral disaffection; similarly, one factor underlies perseverance and self-control across both high school and college students (Muenks et al., 2017).

8.3.2 *Bifactor Models*

Since critiques of the originally proposed factor structure have surfaced, researchers have paid additional attention to the plausibility of a bifactor model of grit measures. The findings of bifactor analyses and their possible implications are summarized here.

It is possible that grit is structurally variant across age groups or education levels. Using MIRT techniques, a factor structure assuming perseverance and consistency are two separate, correlated constructs with *no* higher-order grit factor fit a high school student sample, whereas a bifactor model fit college students (Muenks et al.,

2017). However, one item, “Setbacks don’t discourage me,” having an extremely low factor loading led the bifactor model not to converge in the high school sample.¹ When this item was excluded, the bifactor model fit better than competing models in the high school sample as well. In general, if a bifactor model fits older, more educated people better than a two-factor model, the expression or understanding of grit may change such that consistency and perseverance are more easily conceptualized as separate when tasks, such as those of high-school-level work, are clearly defined. Grit may be less informed by perseverance and consistency and more an expression of one underlying general factor as tasks become more difficult or more ambiguous, as is typical in the course of university-level work.

Other researchers have found evidence for the benefit of a bifactor model over the original higher-order model (Gonzalez et al., 2019; van Zyl, Olckers, & Roll, 2020). Using both traditional factor analytic techniques and exploratory structural equation modeling, which imposes less restrictions on factor models and may provide more distinction between factors in multidimensional constructs, van Zyl et al. (2020) conclude that a bifactor structure is the only appropriate model for the Grit-O in its current form. However, the utility of a bifactor model may as yet be unclear, as one bifactor analysis has shown that the consistency and perseverance subscales contain more common variance within themselves than there is common variance in total scores explained by the general factor (Disabato et al., 2019). An alternative conceptualization may be warranted. Grit’s structural problems may be entirely due to artifacts of positively and negatively worded items. The fit of bifactor models may not be describing conceptual differences between perseverance, consistency, and a general factor, but rather positively worded items (perseverance), negatively worded items (consistency), and a unidimensional construct (grit/self-control/conscientiousness). When testing the Grit-O, this very bifactor model fit the data better than alternatives—one general trait factor with positive and negative latent method factors (see Vazsonyi et al., 2019, Appendix B).

While it should be clear to readers at this point that the original proposed higher-order model of grit is not supported, the true structure of grit remains unknown. The Grit-S muddies the waters of the original construct both substantively and structurally. Further examination of grit measures should revert to the Grit-O, and particular attention should be paid to modeling grit as two separate traits or as a bifactor conglomerate, with the possibility of bifactor structure being an artifact of item wording. The structural problems of grit may have been caused by missteps in the original scale validation. Future researchers should consider the following when revising grit measures.

¹This item in particular has seemed to confuse some respondents because of its double negative wording in conjunction with additional negative terms in the response scale, e.g. “disagree,” resulting in low factors loadings and general model improvement if excluded or rewritten (Fosnacht et al., 2018; Tyumeneva et al., 2017). It may be prudent to assume this item in particular is inappropriate for samples with less education or a lower reading level than others. Revisions of the grit scale should avoid similar double-negative wording in the interest of ease for all respondents.

8.3.3 *Sample Considerations*

Clark and Watson (1995, 2019) recommend validating new scales on several large heterogeneous samples. The original development of grit partially followed this suggestion, as the first study consisted of respondents to a public online posting. Participants were age adults of diverse ages and mostly women (73%), and the sample size was large ($N = 1545$). However, this sample was followed by smaller, much more homogenous ones: Ivy League students ($N = 138$), National Spelling Bee finalists ($N = 175$), and West Point cadets ($N = 1308$). While sample size is clearly not problematic in the West Point sample, all these samples are expected to consist of highly conscientious people. Academically talented students, precocious children, and military trainees are all expected to exhibit range restriction of conscientiousness and achievement, potentially attenuating effects found between close conceptual neighbors describing work ethic and passion. While grit items were not altered according to the responses of any of these samples, examinations of grit's place within the nomological network using these relatively homogenous samples is not advised. Future examination and validation of grit should resemble the original adult sample more than these subsequent samples.

8.3.4 *Inclusion of Comparison Scales*

As already mentioned, the original grit validation studies included conscientiousness but not need for achievement, though both were discussed as close conceptual neighbors. The step of including measure of related constructs is critical in identifying the boundaries and potential shared information given by a construct and its "near neighbors" (Clark & Watson, 2019). Informed by findings of potential overlap between grit facets and the following constructs, future examination of grit should measure revised scales against global conscientiousness (Credé et al., 2017), conscientiousness facets including industriousness and self-discipline (Schmidt et al., 2018), and self-control (Gonzalez et al., 2019; Muenks et al., 2017). The theoretical distinction from these constructs—with regards to time frame—must be included in the grit scale, which largely precludes the Grit-S from further analysis in this vein.

8.3.5 *Subscales*

Scale developers balance between two pitfalls when designing a scale with subscales: creating a measure with items too similar to one another and needlessly breaking them apart, or creating distinct measures and needlessly meshing them together (Clark & Watson, 1995). While apparently desirable—intuitively it seems the inclusion of information assessing diverse elements of a unified construct would

make the most robust scale—in the case of grit, it may be useful to remain neutral on the need for a measure with subscales. As both constructs are revised to include positively and negatively worded items, along with time-frame-relevant information, only items which load on a general factor while also loading cleanly on their respective subscales and creating factors moderately correlated with one another can be considered evidence of a grit measure with subscales (Clark & Watson, 2019). Such a measure may not be the ultimate goal of grit research, as perseverance and consistency may simply be differentially related to outcomes of interest.

8.3.6 Summary

The original two-factor higher-order structure of grit is clearly unsupported, and the true structure of grit remains unclear. In the pursuit of identifying a better fitting model, researchers should test revised scales using large heterogenous samples, remain impartial to the existence of subscales rather than separate constructs, and include comparison scales such as self-control, conscientiousness facets, and need for achievement to further validate grit or its components against existing constructs.

8.4 External Validity

While substantive and structural validity are mainly concerns surrounding the items within a scale, external validity is the extent to which total scores of a scale are reasonably derived from items and meaningful in terms of discriminative or predictive power (Loevinger, 1957). This section will deal primarily with findings of grit's relations to meaningful outcomes and its place in relation to constructs that occupy similar positions in the nomological network.

8.4.1 Criterion Validity

Is grit related to theoretically relevant outcomes? In choosing a criterion relevant to a novel construct, it is beneficial to choose a “risky test,” being deliberately difficult for the construct to pass, thereby strengthening the construct if it passes (Meehl, 1978). Relevant outcomes for grit would include performance criteria, and grit appears to have been focused on academic performance since its early development. We should expect grit to be related to high grades. Whether the relation of grit and grades constitutes a “risky test” on its own is not clear, but it may be “risky” to pit grit against well-known predictors of grades in establishing its importance: standardized intelligence tests ($\rho = .54$, Roth et al., 2015), achievement motivation ($\rho = .30$, Robbins et al., 2004), and conscientiousness ($\rho = .23$, Poropat, 2009).

Perseverance and consistency are indeed related to grades, $\rho = .20$, $\rho = .10$ respectively, and comparisons to other constructs will be discussed in the context of incremental validity (Credé et al., 2017).

8.4.2 Discriminant Validity

Is grit distinct from other well-known constructs? The primary hurdle of academic success researchers is to study variables orthogonal from intelligence. Intelligence is the best known predictor of grades in high school and college (Kuncel, Hezlett, & Ones, 2004; Roth et al., 2015). One of the original goals of grit research was to support the hypothesis that grit is distinct from intelligence (Duckworth et al., 2007). Meta-analytic estimates support this distinction ($\rho = .05$, Credé et al., 2017). The distinction between grit and intelligence allows for the possibility of grit to explain variance in academic performance over the large amount of variance explained by cognitive ability.

In the case of grit, the next challenge was to distinguish itself from close conceptual neighbors like conscientiousness. For reasons outlined in the substantive and structural validity sections, this is a notable challenge. Substantively, grit and conscientiousness cover very similar information, for example, the perseverance items “I finish whatever I begin” and “I am a hard worker” are very similar to International Personality Item Pool (IPIP) achievement-striving facet items like “I carry out my plans” and “I work hard” (Credé et al., 2017; Goldberg et al., 2006). Structurally, grit’s model is misidentified, and theoretically distinctive information is omitted by the commonly used Grit-S. Meta-analytic estimates show a strong relation between conscientiousness and “overall grit,” $\rho = .84$; perseverance, $\rho = .83$; and consistency, $\rho = .61$ (Credé et al., 2017). Grit also exhibited a strong relation with self-control, $\rho = .72$, which is considered a facet of conscientiousness (Credé et al., 2017). The strength of these relations indicates that grit and conscientiousness overlap significantly, to the point of valid concern that the two constructs are largely isomorphic. For reference, consider that the relation between grit and conscientiousness is stronger than that of two different global measures of conscientiousness, $\rho = .63$ (Credé et al., 2017; Pace & Brannick, 2010). While grit and conscientiousness may contain very similar information, grit may yet have utility if able to explain variance in outcomes of interest after controlling for conscientiousness.

8.4.3 Incremental Validity

Does grit explain variance in relevant outcomes over and above related constructs? After controlling for conscientiousness using hierarchical regression methods, “overall grit” and consistency explained no meaningful additional variance in either

high school or college grades (Credé et al., 2017). However, after controlling for conscientiousness and consistency, perseverance explained additional variance in both high school and college grades (Credé et al., 2017). This finding again points to consideration of perseverance and consistency being of more utility as separate constructs. The limitation of this analysis was the reliance on only global measures of conscientiousness as controls. Further research is needed to examine the unique variance explained by perseverance after controlling for lower order conscientiousness facets like self-control and industriousness (Schmidt et al., 2018).

8.4.4 Summary

In the course of exploring relations to academic success, grit passes initial validation tests but fails others. Grit is distinct from intelligence, but not from conscientiousness. The substantive and structural issues with extant grit scales compound difficulties in supporting their external validity. Treating grit in its current iteration as two separate constructs and ultimately revising grit scales is needed before any meaningful relations between grit and academic success may be claimed.

8.5 Conclusion

To this point, the development of grit has been unique as a lesson in the rapid pace of a construct's potential growth, given the speed of modern data collection and dissemination. The goal of this chapter was to pause this process momentarily to reassess what ideals of construct validation grit has met and which it has failed to meet. In approaching this goal, I hope the following points have been made clear. The substance of grit scales is critical for their theoretical meaning and practical utility. The Grit-O and the Grit-S are not equal to the task of reflecting the original definition of grit. Substantively and structurally, the Grit-S is particularly off the mark of the original developers' intentions. Items in current measures do not reflect a theoretically meaningful or predictive construct. "Setbacks don't discourage me" appears to be particularly damaging to meaningful responses to a grit scale by its double-negative framing. The items "I am diligent" and "I am a hard worker" also appear to be redundant. While "overall grit" made up of two subscales may have been intuitively appealing, responses to extant grit measures do not support such a concept's existence. Future work in exploring grit's structure would benefit by treating perseverance and consistency as separate or as unrelated to a general factor underlying grit measures (such as in a bifactor model). Grit's usefulness as a predictor of performance outcome, in part due to these internal inconsistencies, is unclear. Conscientiousness and intelligence, coupled with practical skills such as study habits and class attendance (Credé & Kuncel, 2008; Credé, Roch, & Kieszczynka, 2010) clearly explain most variance in academic performance, and a

unified grit construct does not add to this understanding. The future of grit development may lie in examining perseverance against lower-order facets of conscientiousness or domain-specific practical skills.

As scale development and construct validation are iterative processes, there is value in going “back to the drawing board” for grit. The following practical recommendations for revision ought to be considered by future grit researchers. An efficient revision to both the substance of these items and a structural artifact of the grit scale would be to revise items of the Grit-O to be worded both positively and negatively. Having a mix of reverse scored items within both perseverance and consistency scales would eliminate the possibility of misinterpreting factor analytic results as purely artifacts of wording. An over-inclusive set of items based on the Grit-O should be analyzed and reduced using criterion-based and IRT methods to clarify what information is captured by grit items, particularly which items are capturing crucial time-frame information theoretically distinguishing grit from conscientiousness and other near-neighboring constructs. This revised scale should be examined alongside conscientiousness facet scales and other related constructs to determine the discriminant and incremental validity of a revised grit scale.

This continuation of the construct validation process would not establish grit as having passed the scale development phase, as scale development ought to be continuously ongoing (Clark & Watson, 2019), but the next steps in grit research must acknowledge failures to support grit’s construct validity to this point. After a reasonable amount of support can be produced in reference to these construct validity questions, more complex psychometric issues such as cross-cultural validation and measurement invariance may be addressed (see van de Vijver, 2002). Until grit, perseverance, and consistency can be supported as psychometrically sound and theoretically meaningful constructs, educational and professional institutions would benefit from devoting attention and resources to more well-established correlates of performance.

References

- Clark, L. A., & Watson, D. B. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>.
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. <https://doi.org/10.1037/pas0000626>.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences, 12*(9), 887–898.
- Credé, M. (2018). What shall we do about grit? A critical review of what we know and what we don’t know. *Educational Researcher, 47*(9), 606–611. <https://doi.org/10.3102/0013189X18801322>.
- Credé, M., & Kuncel, N. R. (2008). Study habits, study skills, and study attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*, 415–453.

- Credé, M., & Harms, P. D. (2015). 25 years of higher order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior*, *36*, 845–872.
- Credé, M., Harms, P. D., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *Journal of Personality and Social Psychology*, *102*, 874–888.
- Credé, M., Roch, S. G., & Kieszczyńska, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*, 272–295.
- 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*(3), 492–511. <https://doi.org/10.1037/pspp0000102>.
- Datu, J. A. D., Valdez, J. P. M., & King, R. B. (2015). Perseverance counts but consistency does not! Validating the short grit scale in a collectivist setting. *Current Psychology*, *35*, 121–130. <https://doi.org/10.1007/s12144-015-9374-2>.
- Datu, J. A. D., Yuen, M., & Chen, G. (2017). Development and validation of the Triarchic model of grit scale (TMGS): Evidence from Filipino undergraduate students. *Personality and Individual Differences*, *114*, 198–205. <https://doi.org/10.1016/j.paid.2017.04.012>.
- Disabato, D. J., Goodman, F. R., & Kashdan, T. B. (2019). Is grit relevant to well-being and strengths? Evidence across the globe for separating perseverance of effort and consistency of interests. *Journal of Personality*, *87*, 194–211.
- Duckworth, A. L. (2016). *Grit: The power of passion and perseverance*. New York: Scribner.
- 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*(6), 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*(2), 166–174. <https://doi.org/10.1080/00223890802634290>.
- Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of “necessary but not sufficient” causality. *Organizational Research Methods*, *19*(1), 10–52. <https://doi.org/10.1177/1094428115584005>.
- Fosnacht, K., Copridge, K., & Sarraf, S. (2018). How valid is grit in the postsecondary context? A construct and concurrent validity analysis. *Research in Higher Education*. <https://doi.org/10.1007/s11162-018-9524-0>.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The international personality item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>.
- Gonzalez, O., Canning, J. R., Smyth, H., & Mackinnon, D. (2019). A psychometric evaluation of the short grit scale: A closer look at its factor structure and scale functioning. *European Journal of Psychological Assessment*, 1–12. <https://doi.org/10.1027/1015-5759/a000535>.
- Guo, J., Tang, X., & Xu, K. M. (2019). Capturing the multiplicative effect of perseverance and passion: Measurement issues of combining two grit facets. *Proceedings of the National Academy of Sciences*, *116*(10), 3938–3940.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191–205.
- KIPP. (2019). *Character strengths*. Retrieved November 22, 2019, from <http://www.kipp.org/approach/character/>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford publications.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161. <https://doi.org/10.1037/0022-3514.86.1.148>.

- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, *51*, 493–504.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
- 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*(5), 599–620.
- Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the “same” construct? A meta-analytic investigation. *Personality and Individual Differences*, *49*, 669–676.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*, 261–288.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. <https://doi.org/10.1016/j.intell.2015.09.002>.
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in personality and social psychology bulletin. *Personality and Social Psychology Bulletin*, *28*, 1629–1646.
- Schechtman, 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: U.S. Department of Education, Office of Education Technology. Retrieved from <http://pgbovine.net/OET-Draft-Grit-Report-2-17-13.pdf>
- Schmidt, F. T., Fleckenstein, J., Retelsdorf, J., Eskreis-Winkler, L., & Möller, J. (2019). Measuring grit: A German validation and domain-specific approach to grit. *European Journal of Psychological Assessment*, *35*(3), 436–447.
- Schmidt, F. T., Nagy, G., Fleckenstein, J., Möller, J., & Retelsdorf, J. (2018). Same same, but different? Relations between facets of conscientiousness and grit. *European Journal of Personality*, *32*, 705–720.
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*, 367–373.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*, 205–231.
- Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*, 116–131.
- 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*, *35*(4), 469–478. <https://doi.org/10.1027/1015-5759/a000424>.
- van de Vijver, F. J. R. (2002). Cross-cultural assessment: Value for money? *Applied Psychology: An International Review*, *51*(4), 545–566. <https://doi.org/10.1111/1464-0597.00107>.
- van Zyl, L. E., Olckers, C., & Roll, L. C. (2020). The psychometric properties of the grit-O scale within the Twente region in the Netherlands: An ICM-CFA vs. ESEM approach. *Frontiers in Psychology*, *11*, 796. <https://doi.org/10.3389/fpsyg.2020.00796>.
- Vazsonyi, A. T., Ksinan, A. J., Jiskrova, G. K., Mikuška, J., Javakhishvili, M., & Cui, G. (2019). To grit or not to grit, that is the question! *Journal of Research in Personality*, *78*, 215–226.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79.

Michael C. Tynan recently earned a PhD in psychology from Iowa State University. He earned a B.S. in psychology from Loyola University Chicago and a M.S. in psychology from Iowa State University. He has published in the *Journal of Personality and Social Psychology*, *Learning and Individual Differences*, and *The Leadership Quarterly*. His research interests include judgment and decision-making, personality trait measurement, academic success antecedents, and predictors of professional well-being