2015

THE SHORT GRIT SCALE: A DIMENSIONALITY ANALYSIS

Caihong Li
University of Kentucky, caihong.li@uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Caihong Li, Student

Dr. Michael D. Toland, Major Professor

Dr. Kenneth Tyler, Director of Graduate Studies

THE SHORT GRIT SCALE: A DIMENSIONALITY ANALYSIS

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Education in the
College of Education
at the University of Kentucky

By
Caihong Rosina Li

Lexington, Kentucky

Chair: Dr. Michael D. Toland, Associate Professor of Educational Psychology

Lexington, Kentucky

2015

ABSTRACT OF THESIS

THE SHORT GRIT SCALE: A DIMENSIONALITY ANALYSIS

This study aimed to examine the internal structure, score reliability, scoring, and interpretation of the Short Grit Scale (Grit-S; Duckworth & Quinn, 2009) using a sample of engineering students (N = 610) from one large southeastern university located in the United States. Confirmatory factor analysis was used to compare four competing theoretical models: (a) a unidimensional model, (b) a two-factor model, (c) a second-order model, and (d) a bi-factor model. Given that researchers have used Grit-S as a single factor, a unidimensional model was examined. Two-factor and second-order models were considered based upon the work done by Duckworth, Peterson, Matthew, and Kelly (2007), and Duckworth and Quinn (2009). Finally, Reise, Morizot, and Hays (2007) have suggested a bi-factor model be considered when dealing with multidimensional scales given its ability to aid researches about the dimensionality and scoring of instruments consisting of heterogeneous item content. Findings from this study show that Grit-S was best represented by a bi-factor solution. Results indicate that the general grit factor possesses satisfactory score reliability and information, however, the results are not entirely clear or supportive of subscale scoring for either consistency of effort subscale or interest. The implications of these findings and future research are discussed.

KEYWORDS: grit, confirmatory factor analysis, two-factor model, bi-factor model, engineering

Caihong Li

Student's Signature

April 9th, 2015

Date

THE SHORT GRIT SCALE: A DIMENSIONALITY ANALYSIS

By

Caihong Rosina Li

Dr. Michael D. Toland
Director of Thesis

Dr. Kenneth Tyler
Director of Graduate Studies

April 9th, 2015
Date

For Zhaoshuai and my family

Acknowledgements

First, I would like to thank all the members of my committee who have supported me throughout my time in graduate school and as I was working on my thesis. I would like to thank Dr. Toland, who led me into the world of methodology, and built my confidence as a researcher in this area. I would like to thank Dr. Usher, who opened the door of the P20 Motivation and Learning Lab and embraced me despite my inexperience, and who put effort for my access to the data for this thesis. I will always be grateful for your mentorship and generous help.

Second, funding for this project was supplied by the National Science Foundation under award number EEC-1240327 and -1240328. I would like to thank the engineering motivation team, Drs. Usher, Mamaril, Kennedy, and Economy. Thank you very much for offering part of the dataset from your NSF funded project. It has always been an honor for me to work with you all.

I would also like to thank my family and friends for supporting me throughout my time in graduate school. Thank you to Zhaoshuai for understanding when I had to focus on my work and had no time to take care of our family. Thanks to my friend Wenjin for being sympathetic and listening when I had a bad mood and was full of complaints. Thanks to my beloved parents, Zhuoxiang Li and Xiuying Zhang. Your words gave me the strength to persist whenever I met any setbacks. I would not be where I am today without the love and support of all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Efforts have been devoted to study intelligence or IQ (Gottfredson, 1997; Hartigan & Wigdor, 1989) as a primary indictor of achievement. However, grit, defined as the "perseverance and passion for long-term goals" (Duckworth, Peterson, Matthew, & Kelly, 2007), has been shown to be a stronger predictor of achievement than intelligence alone in samples of high achievers under super-challenging settings (Duckworth et al., 2007). Despite failures, gritty people are likely to show effort and interest in moving toward their specific goals for years, and less gritty people are likely to interpret failure as the message to give up or to change their goals.

In order to quantitatively measure grit, Duckworth et al. (2007) wrote a 27-item scale, composed of items conceptually based on a review of extant literature. A classical item analysis was conducted after responses from a sample of 1,545 adults (M_{age} = 45) were collected. After reviewing the item quality, the scale was reduced to 17 items. This 17-item scale was then examined using exploratory factor analysis (EFA) with half of the sample (n = 772), which showed 5 items should not be retained from the 17-item scale because of low loadings. EFA results indicated that two factors could be retained from the remaining 12 items. Conceptually, the two factors were named Consistency of Interest (6 negatively phrased items) and Perseverance of Effort (6 positively phrased items). Next, Duckworth and her colleagues fit a two-factor model to the rest of the sample (n = 773) using confirmatory factor analysis (CFA). They interpreted the results as supporting the two-factor solution, with comparative fit index (CFI) = .83 and root-mean-square error of approximation (RMSEA) = .11. Based on these findings the 12-item grit scale (Grit-O; see Appendix A) was suggested as a measure of grit.

In a subsequent study, Duckworth and Quinn (2009) stated that Grit-O could be improved further. The 8-item Short Grit Scale (Grit-S; see Appendix B) was developed from Grit-O by deleting four items (two per factor) showing the poorest item-level correlations with four criteria in four different samples. Two models, a unidimensional model and a two-factor model were fit to the data collected from a sample of 1,554 adults (M_{age} = 45.64 years, *SD* = 11.27; 81% female) using CFA with maximum likelihood (ML) estimation. Results suggested the two-factor solution fit the data better than the unidimensional solution for Grit-S. Fit indices of two-factor solutions for Grit-S and Grit-O were also compared using the data from the above sample. In addition, results indicated that Grit-S had better fit compared to the Grit-O.

Since these two publications by Duckworth and colleagues (2007, 2009), Grit-S has been broadly used in social science research as a measure of the latent construct grit. Grit has been shown to be predictive of academic performance in college students, retention in United States Military Academy cadets (Duckworth et al., 2007; Duckworth & Quinn, 2009), teacher effectiveness (Duckworth et al., 2009), physician satisfaction (Reed, Schmitz, Baker, Nukui, & Epperly, 2012), and resident well-being (Salles, Cohen, & Mueller, 2014). People with more grit were found to try harder (Silvia, Eddington, Beaty, Nusbaum, & Kwapil, 2013) and work longer (Duckworth et al., 2007) compared with people who possessed less grit.

However, the implications drawn from these empirical studies using the Grit-S are limited in several ways. Firstly, no external research group outside of Duckworth and her colleagues have gathered evidence on the internal structure and score reliability of the Grit-S. Secondly, Duckworth and Quinn (2009) initially referred to the model being tested as a two-factor model and then went on to discuss a second-order solution, but no rationale was given as to why a second-order model was considered. The use of this latter model is confusing given how regression analyses were conducted later using total scores. Thirdly, item wording is a potential confounding variable within the two-factor or second-order model. Specifically, all Consistency of Interest items are negatively phrased, and all Perseverance of Effort items are positively phrased. Although Duckworth and Quinn (2009) indicated that item wording could be a problem to the internal structure of the scale, they argued that the two-factor structure could be interpreted substantively. However, they did not provide empirical evidence to confirm that these two factors were indeed two dimensions rather than an artifact due to item phrasing. Fourthly, Duckworth and Quinn (2009) used coefficient alpha as an estimate of the internal consistency of score reliability. However, Cronbach's coefficient alpha has continuously been criticized for its over- or underestimate of reliability (Peters, 2014). Thus, more and more researchers have suggested abandoning the use of coefficient alpha and adopting better reliability coefficient estimates, such as coefficient omega (Peters, 2014; Shevlina, Milesb, Daviesc, & Walker, 2000; Sijtsma, 2009; Starkweather, 2012). Furthermore, previous studies have not reported confidence intervals for reliability estimates, thus no reflection of sampling variability of reliability could be obtained (Guttman, 1945; Revelle & Zinbarg, 2009; Sijtsma, 2009).

To address these issues, Grit-S was studied in a sample of 610 college students in the current study. The purpose of this study was to examine the internal structure of Grit-S using CFA, report reliability evidence for the scores generated from this scale, and determine the scoring and interpretation of the total score and subscale scores generated from this scale. The present study contributes to the broader literature on the psychometric properties of the Grit-S.

Chapter 2: Literature Review

This section provides a specific literature review of the Grit-S. Specifically, initial development and dimensionality of Grit-S, correlational evidence, score reliability evidence, population studied, common methodological problems, and statement of purpose are provided.

Initial Development and Dimensionality of Grit-S

Duckworth et al. (2007) developed the long form grit scale (Grit-O) from a sample of 1,545 adults (M_{age} = 45; 73% female) in order to measure the latent construct grit quantitatively. Initially, a pool of 27 items tapping the construct of grit was written and rated using a 5-point Likert-type response scale. This was further reduced to 17 items based on removing items with poor item-total correlations, items not contributing to the score reliability coefficient, having redundancy with other items, or having complex vocabulary. Next, an EFA was conducted to examine the internal structure of the scale in a random half of the original sample (n = 772). Five items were discarded further because of low factor loadings. A two-factor correlated model (r = .45) was retained and factors were labeled Consistency of Interest and Perseverance of Effort. Consistency of Interest was denoted as interest and Perseverance of Effort was denoted as effort in the rest of the thesis. Each factor consisted of six items. All items in the interest factor are negatively phrased, whereas all items in the effort factor are positively phrased. Next, a CFA ML estimation was conducted with the remaining 773 adults, CFI = .83, and RMSEA = .11. The type of method to estimate parameters in this CFA model was not provided in their manuscript. Additionally, Duckworth et al. suggested using a total score to measure grit because it had a higher prediction of outcomes compared to both factors alone, but the specific results to support this claim were not provided.

Duckworth and Quinn (2009) developed a short version scale – Grit-S – to measure grit on the basis of Grit-O. By examining the performance of the original 12 items on the Grit-O in four different samples, including adults and adolescents, two items from each subscale were removed due to their negative or low item-level correlation with the latent construct grit. In the four samples, score reliability (α) for the total, interest, and effort scale scores, ranged from .73 to .83, .73 to .79, and .60 to .78, respectively. A two-factor model was fit separately for each sample using CFA with ML estimation. Across the sample, CFI ranged from .86 to .95, and RMSEA ranged from .061 to .101. The authors suggested that Grit-S was a second-order structure, where consistency of interest and perseverance of effort are the first level factors and grit is the second level factor, $\chi^2(38, N = 1, 554) = 22.13$, $p < .001$. However, they did not provide the rational for the consideration of a second-order solution. In the following studies of the same manuscript, the authors interchangeably used two-factor solution and second-order solution in CFA using ML estimation and measurement invariance tests. Specifically, in a larger sample ($N = 1{,}554$, $M_{age} = 45.64$, 81% female), the authors showed that the two-factor model, $\chi^2(19) = 188.52$, $p < .01$, RMSEA = .076, 90% CI [.066, .086], CFI = .96, fit better than a unidimensional model, where $\chi^2(20) = 380.45$, $p < .01$, $\Delta\chi^2(1) = 191.93$, $p < .01$. Next, Duckworth and Quinn (2009) fit the second-order model to examine whether the internal structure of Grit-S differed by gender. They found that the second-order structure of Grit-S did not differ between men and women.

Correlational Evidence

Copious efforts have been devoted to the study of grit as a personality trait related to goals and success. Grit has been shown to be related to personality traits including hardiness and traits within the Big Five model (Duckworth et al., 2007), academic variables including academic performance, retention, and final ranking (Duckworth et al., 2007) and life outcomes including life satisfaction (Reed et al., 2012), well-being (Salles et al., 2014), and happiness (Von Culin, Tsukayama, & Duckworth, 2014).

Grit and other personality traits. Grit and traits within the Big Five model are all theoretically framed as characteristics related to success. In a sample of 1,554 adults, Duckworth and Quinn (2009) showed that Grit-S was positively correlated with conscientiousness (r = .77), agreeableness (r = .24), and extraversion (r = .20), whereas it was negatively correlated with neuroticism (r = -.40). No linear correlation between Grit-S and openness to experience has been found. Reed, Pritschet, and Cutton (2013) found a strong positive relationship between Grit-S and conscientiousness (r = .72) in a study examining the prediction of grit and conscientiousness on behavior change among 1,171 adults.

In addition to the traits in the Big Five model, moderate positive relationships have also been evidenced between Grit-S and hardiness (Maddi et al., 2012, 2013). Maddi and his colleagues (2012, 2013) found Grit-S and hardiness were positively correlated in a sample of 1,285 military cadets (r = .46) and in another sample of 425 undergraduates at a public university (r = .31).

Grit and educational variables. Researchers have shown that grit predicts various educational outcomes. Duckworth et al. (2007) showed adults (aged 25 and above) with more grit were more likely to have higher educational attainment than adults with less grit. In a sample of 139 undergraduate students, Duckworth et al. found that grit was positively correlated with SAT scores (r = .34) and college GPA (r = .25). They also found grit was a strong predictor of retention rate (β = .48) among a sample of 1,218 freshmen cadets. Among higher achievers like the finalists in the National Spelling Bee (N = 190), grit has been found to be predictive of the final rankings (r = .16), indicating those who were grittier were more likely to have a top ranking in the final competition (Duckworth & Quinn, 2009). Grit has been shown to be positively related to self-control (r = .63) in a sample of 1,218 freshman cadets (Duckworth et al., 2007).

Grit and life outcomes. Grit has been found to be predictive of life outcomes including life satisfaction (A. J. Reed et al., 2012), well-being (Salles et al., 2014) and happiness (Von Culin et al., 2014). In a study of information acquisition, Haran, Bitov, and Barbara (2013) found that Grit-S was positively correlated with need for cognition, the inclination to devote oneself to and enjoy cognitive accomplishments requiring effort (Cacioppo, Petty, & Kao, 1984). In a study of positive predictors of teacher effectiveness, Duckworth, Quinn, and Seligman (2009) tested the relationship among grit, optimistic explanatory style, and life satisfaction. Results showed Grit-S has a positive relationship with optimistic explanatory style (r = .17) and life satisfaction (r = .32).

Grit in Various Settings and Populations

Grit has been studied within diverse samples under various contexts. The examination of grit has typically been constrained to competitive settings including the military (West Point cadets), Spelling Bee competitions, and universities in the Ivy League (Duckworth et al., 2007; Duckworth et al., 2009; Maddi et al., 2013; Maddi et al., 2012). Other populations in extreme stressful working environment were also studied, including novice teachers (Duckworth et al., 2009), physicians (A. J. Reed et al., 2012), medical residents (Salles et al., 2014), and minority college students at predominantly White institutions (Strayhorn, 2013). In addition, researchers have studied grit in several non-competitive contexts. Eskreis-Winkler et al. (2014) recently published a paper about the influence of grit on retention in four different samples: soldiers, high school juniors, sales representatives, and adults who once married and now are single or keeping the married status. They found that the soldiers with high grit scores are more likely to complete the military program; high school juniors with high grit scores were likely to graduate from high school; sales representatives with high grit scores tend to keep their sales jobs after three months; grittier men are more likely to keep the marital status compared to less gritty men. Maddi et al. (2013) studied a sample of 425 undergraduate students in California and found that gritty students are less likely to be addictive to Internet and engage in compulsive buying and gambling. Concluding from the above studies, studies on grit are conducted in various types of contexts and diverse populations, which greatly enrich the understanding of its influence and its prediction of success in different areas.

Score Reliability Evidence

Coefficient alpha has been used to measure score reliability of Grit-O and Grit-S. Duckworth et al. (2007) demonstrated that the reliability for total grit scores, interest scores, and effort scores generated from Grit-O were .85, .84, and .78, separately, in a sample of 773 adults. Duckworth et al. then examined the score reliability of Grit-O in other five different samples (adults, Ivy League undergraduates, West Point cadets in class of 2008, West Point cadets in Class of 2010, and National Spelling Bee finalists), which showed that score reliability for total grit scores ranged from .77 to .85. Reliability for interest scores and effort scores were not reported. In a study by Duckworth and Quinn (2009) using Grit-S, total grit scores, interest scores, and effort scores had reliability estimates of 82, .77, and .70, respectively. Subsequent studies using Grit-S show coefficient alpha estimates ranged from .77 to .90 for total grit scores (Eskreis-Winkler et al., 2014; Strayhorn, 2013; Von Culin et al., 2014), .68 to .83 for interest scores (Silvia et al., 2013; Von Culin et al., 2014), and .52 to .84 for effort scores (Silvia et al., 2013; Von Culin et al., 2014).

Common Methodological Problems

Although Grit-S has been adopted by many researchers as a measure of the latent construct grit, no subsequent studies have been conducted since 2009 to test its internal structure. However, several problems related to its internal structure do exist. In this section, the common methodological problems related to previous studies about Grit-S were discussed.

The first problem related to Grit-S is that the two-factor solution or the second-order solution might be an artifact of negative item wording. All the items in the consistency of interest subscale are negatively phrased, that is, the higher scores indicate low grit, whereas all the items in the perseverance of effort subscale are positively phrased, that is, higher scores indicate high grit. Duckworth et al. (2007) mentioned that "[w]e considered the possibility that these two factors were an artifact of positively and negatively scored items but were convinced that the factor structure reflected two conceptually distinct dimensions " (p. 1090). However, they did not provide any empirical evidence that the two-factor solution was not due to the artifact of negative item wording. Many papers have verified that item wording leads to an artifact effect of the scale internal structure (Greenberger, Chen, Dmitrieva, & Farruggia, 2003; Schriesheim & Eisenbach, 1995). For example, Greenberger et al. (2003) rewrote all the Rosenberg Self-Esteem Scale into two alternative scales, one with all items positively phrased, and one with all items negatively phrased. They used the original scale including five positive items and five negative items, and the two alternative scales in a sample of 741 undergraduates from various majors with diverse ethnical background, and found that both re-worded scales fit a unidimensional model and the original scale fit a two-factor model. Similarly, the study done by Schriesheim and Eisenbach (1995) also found a clear wording effect on the scale structure. Thus, in order to get an accurate estimate of the dimensionality of Grit-S, researchers should either explore the scale structure using both the original scale and alternatively worded scales, or use psychometric techniques to model the wording effect when examining dimensionality.

Second, all of the previous studies have used coefficient alpha as an estimate of the reliability for Grit-S scores. However, researchers have criticized the use of coefficient alpha and have suggested abandoning its use because research often violates the underlying assumptions of coefficient alpha in empirical studies before using it to measure reliability (Peters, 2014; Shevlina et al., 2000; Sijtsma, 2009; Starkweather, 2012). Coefficient alpha is based on classical test theory (CTT; Novick, 1966), which assumes each observed score is the sum of true score and measurement error. Or, for a sample, coefficient alpha is the ratio of true score variance over observed score variance and every score is assumed to measure one variable. Another assumption about coefficient alpha is that it assumes equal item variances and covariances between items (Revelle & Zinbarg, 2009). When both of these two assumptions are tenable, coefficient alpha is an accurate estimate for gauging scale score reliability. However, the above assumptions are likely violated in empirical research (Yang & Green, 2011). Sijtsma (2009) has shown that, if any of the assumptions are not tenable, it is impossible that coefficient alpha equals the reliability of the test scores. Dunn, Baguley, and Brunsden (2014) summarize known deficiencies of Cronbach's coefficient alpha as an estimate of reliability. Dunn and colleagues argue that because coefficient alpha is a point estimate, where only one single quantity is obtained, it does not represent the best estimation of a population parameter. With a comprehensive consideration of the above flaws, researchers recommend estimating score reliability using other reliability coefficients that are more robust to assumption violations (Zinbarg, Revelle, Yovel, & Li, 2005), such as coefficient omega (McDonald, 1999).

Finally, Duckworth and Quinn (2009) have used the total score generated from Grit-S to represent the latent construct of grit. However, Duckworth and Quinn did not provide empirical evidence to support this scoring approach in the presence of multidimensionality. Other researchers have calculated two subscale scores, and interpreted them separately as persistence of effort and consistency of interest (Silvia et al., 2013). So far, no research has justified the creation and scoring of two subscales in Grit-S. Furthermore, no research has examined whether the interest and effort subscale scores represent precise and meaningful information that is unique from the general grit factor. As previous CFAs did not provide adequate guidance to practical research, the necessity of creating the subscales (interest and effort) and reporting the subscale scores should be examined. The interpretability of the total scores and subscale scores of Grit-S should also be explored before interpreting them as indicators of the latent constructs grit, interest, and effort.

Statement of Purpose

Given the use, interpretation, and scoring of Grit-S varies by researchers' perceived structure of Grit-S, studying the internal structure of Grit-S is meaningful to the development of grit in academia and its application as a personality trait in different research fields. The purpose of this study was to examine the dimensionality and score reliability of Grit-S in a sample of engineering students in one southeastern university. Based on the research literature, three research questions were addressed. First, what is the internal structure of Grit-S? Second, how reliable are scores generated from Grit-S? Third, should subscale scores and total scores be reported and interpreted as representing meaningful information?

The current study provides an evaluation of Grit-S that is independent of the work done by Duckworth and colleagues. Findings are informative to researchers who will use Grit-S to measure grit and predict educational and psychological outcomes based on Grit-S scores.

Chapter 3: Method

Participants

Data were collected as a part of a larger study. The project was designed to develop instruments to measure student motivation in engineering courses (P20 Motivation and Learning Lab, 2014). Participants were recruited within engineering specific courses from one southeastern university (*N* = 610) in the United States. Eighty percent of the sample identified as men, and 20% of the sample identified as women. Self-reports indicated that 80.5% of the sample were White students, 6.2% Asian American, 4.1% African American, 3.0% Middle Eastern, 2.1% Hispanic, 0.2% American Indian, 2.6% multiracial, and 0.3% from other ethnic groups. Seven participants preferred not to report their ethnicity. Participants were enrolled in different engineering majors, including chemical engineering (*n* = 126), mechanical engineering (*n* = 115), computer science (*n* = 76), mining engineering (*n* = 65), biosystems engineering (*n* = 47), computer engineering (*n* = 59), material engineering (*n* = 25), electrical engineering (*n* = 38), and other engineering majors (*n* = 39). One participant didn’t report his or her major.

Short Grit Scale (Grit-S)

This study used the 8-item Grit-S developed by Duckworth and Quinn (2009) to measure the perseverance and passion to pursue long time goals, but had a minor modification to the response option system used (see Appendix C). First, six response options were used instead of the original five response options in order to create a balanced response option system. Second, response options were presented horizontally by filling in a circle immediately below the column headings that displayed response options instead of vertically below each item as presented in the original Grit-S form. The response options ranged from 1 (*not at all like me*) to 6 (*very much like me*). All item responses generated from the interest subscale were reversed coded for scoring and analysis purposes. Higher scores indicate higher level of stamina for long term goals.

Procedure

Following the approval from the Institutional Review Board, paper surveys consisting of demographic questions, Grit-S items, and other scales measuring persistence in engineering, engineering self-efficacy, sources of engineering self-efficacy, achievement goals, task value, and implicit opinion were group administered in engineering classes in the fall 2013. Demographic questions were asked at the beginning of the survey. Grit-S was completed as the sixth instrument and items were arranged following the order of items in the original Grit-S (Duckworth & Quinn, 2009). Before the beginning of the survey, consent forms and verbal instructions were given to participants by trained researchers. Participants were encouraged to ask for clarifications if any word or item was not understandable. Then, they were asked to complete the instruments individually and independently. The survey took participants about 30 minutes to complete. Anonymity was ensured and teachers were not present during the data collection process.

Data Analyses

Prior to the primary data analyses, items in the interest subscale (items 1, 3, 5, and 6) were reverse coded. Data were examined by checking the item response frequencies. Two data collapsing strategies were considered for response categories with low frequency. The first data collapsing strategy was recommended by Beamish (2004), that is, if the Likert-type scale items are ordinal or categorical in nature and if data collapsing strategies are considered, response categories could be reduced into dichotomous categories to minimize respondent ambiguity over too many response categories and have scores that represent binary ends of the continuum. The other data collapsing strategy was an empirical data collapsing method, which is, collapsing the response categories with few responses with the adjacent response category. In the current study, response categories with low frequency were reduced into dichotomous categories substantively. Categories with low responses were combined with the adjacent category empirically.

Dimensionality analyses. In order to answer the first research question, CFAs were conducted. In particular, four different models were compared: a unidimensional model, a two-factor model, a second-order model, and a bi-factor model. Given the ordinal and categorical nature of the data, a polychoric correlation matrix based on the mean- and variance-adjusted weighted least square (WLSMV) estimator was used for the analyses.

Although Duckworth and Quinn (2009) compared a unidimensional model (see Figure 1) with a two-factor model (see Figure 2) and found the two-factor model was a better fit, indicated by a significant chi-square difference, $\Delta\chi 2(1) = 191.93$, $p < .001$, they reported an estimate of coefficient alpha to measure the total scale score reliability. Since an underlying assumption of coefficient alpha is unidimensionality, it is necessary to confirm whether the unidimensional model fits the observed data. Moreover, the unidimensional model served as the background model by which more complex models can be evaluated.

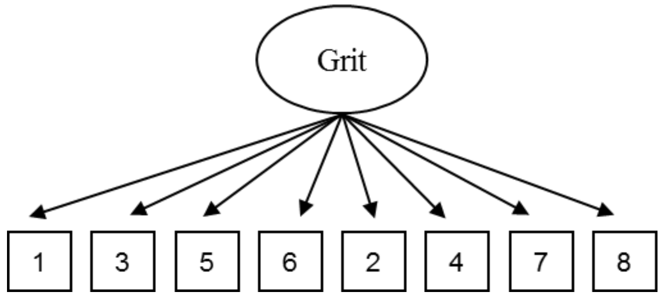

Figure 1. Unidimensional model of 8-item Grit-S.

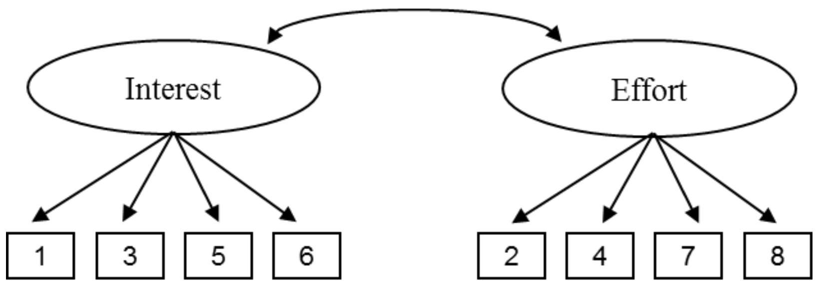

Figure 2. Two-factor model of 8-item Grit-S scale.

A two-factor model was considered in this study based on the conclusions drawn by Duckworth and her colleagues on the structures of both Grit-O and Grit-S. Duckworth et al. (2007) explored the internal structure of Grit-O using EFA and stated that it was a two-factor oblique model, where all 12 items loaded in the general factor over .40. CFAs were conducted in different samples. Findings showed that the two-factor model was the best fit of data generated from Grit-O (Duckworth et al., 2007). Duckworth and Quinn (2009) confirmed the two-factor structure of Grit-S using CFAs. According to Duckworth and Quinn, the two-factor model indicated that responses to items 1, 3, 5, and 6 can be titled Consistency of Interest and responses to items 2, 4, 7, and 8 can be titled Perseverance of Effort. The two factors were correlated with each other at r = .45. The two-factor model was referred to as a non-hierarchical correlated multidimensional model.

A second-order model (see Figure 3) was considered in this study based on the conclusion Duckworth and Quinn (2009) made about the structure of Grit-S. A second-order model contains a general factor and several first order factors. Items directly depend on the respective specific first order factors, and all the first order factors load on the general dimension, also known as the second-order factor. In Grit-S, the interest and effort serve as the first-order factors and grit serves as the second-order factor. In a second-order model, if the first order contains two factors, the second-order model is statistically the same as a two-factor model. However, they are different models conceptually. The difference between a second-order model and a two-factor model is that the second-order model is a hierarchical model, and the two-factor model is a non-hierarchical model. If the internal structure of Grit-S is indeed second-order, all items would load onto the two factors, and the common variance of the two factors composes the general latent construct grit. The assumption about the second-order model is that the first-order factors are conditionally orthogonal. In other words, the relationship between the two factors is explained by the general factor (Rijmen, 2010).

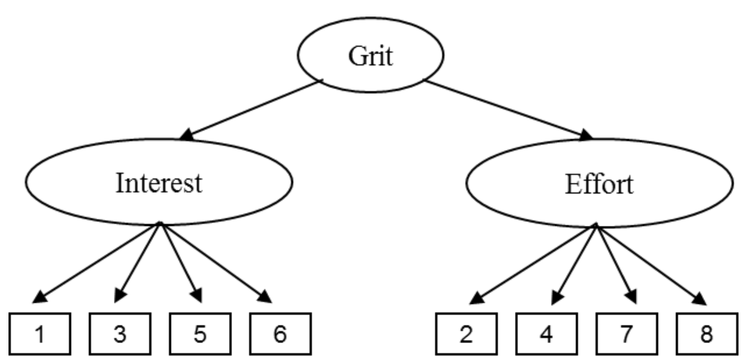

Figure 3. Second-order model of 8-item Grit-S scale.

In addition to the unidimensional, two-factor, and second-order models, a bi-factor model (see Figure 4) was considered. According to Reise, Morizot, and Hays (2007), a bi-factor model is "a useful complement to traditional (uni)dimensionality analysis" (p. 22), which provides another option for exploring the dimensionality of scales with multiple dimensions. Chen, West, and Sousa (2006) suggested that a bi-factor model is potentially applicable when a researcher is interested in a scale that has more than two factors, where a general factor runs through all the items, and the specific factors explain the uniqueness of the variance after extracting the influence of the general factor. In a bi-factor model, the item covariance has two sources: the general factor and the respective group specific factors. There is a general factor that explains the communality among the items, but there are also unique factors that explain the intercorrelations among the items which are independent from the general factor and each other. In other words, for Grit-S, grit is the general factor influencing the item covariance. Meanwhile, consistency of interest and perseverance of effort also influences the item covariance independently from the effect of the general grit factor.

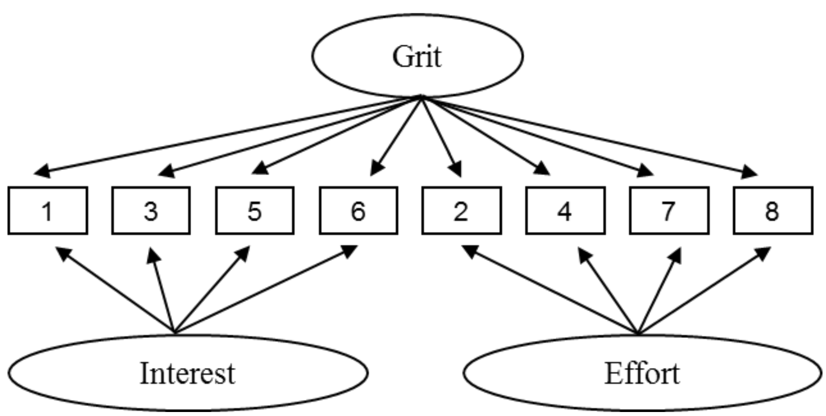

Figure 4. Bi-factor model of 8-item Grit-S.

In order to address the first research question, CFAs were conducted by analyzing a polychoric correlation matrix using the mean- and variance-adjusted weighted least square (WLSMV) estimator in M*plus* 7.11 (Muthén & Muthén, 1998-2010). The polychoric correlation matrix was used because item response categories were ordinal in nature (Brown, 2006). Four different models were fit to the data: a unidimensional model, a two-factor model, a second-order model, and a bi-factor model. The chi-square statistic, CFI, the standardized root mean square residual (SRMR), and RMSEA were used to assess the goodness of fit of each model. Conventional benchmarks suggested by Brown (2006), Hu and Bentler (1999), and Satorra and Bentler (1994) were used: RMSEA less than or close to .08, CFI and TLI greater than or close to .90, and WRMR less than or close to 1. A chi-square difference test, as implemented in M*plus*, was used to compare nested models. All analyses were done at the 5% significant level.

Score reliability. Cronbach's coefficient alpha has been criticized recently given its over- or underestimate of reliability (Peters, 2014; Shevlina et al., 2000; Sijtsma, 2009; Starkweather, 2012). In order to offer a more robust estimation of the score reliability for Grit-S, coefficient omega (McDonald, 1999) was used. Coefficient omega was estimated using the following formula

$$\text{omega} = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \sum_{i=1}^{n} (1-h_i)^2}\,, \qquad (1)$$

where n is the number of items in the factor, λ_i is the factor loading of item i, $(1- h_i)^2$ is the unique variance of item i, and assuming a standardized latent construct (i.e., variance fixed at 1). Coefficient omega for the Grit-S scale scores was denoted as omega_G, coefficient omega for the interest subscale scores was denoted as omega_I, and coefficient omega for the effort subscale was donated as omega_E. Bootstrap confidence intervals for omega were also estimated using 1,000 bootstrap samples as implemented in M*plus*. Values greater than .60 are generally considered acceptable (Bagozzi & Yi, 1988).

Scoring and interpretation. The scoring process for Grit-S depends on the internal structure. If it is unidimensional, then a total score would be computed to measure grit. If it is multidimensional, then researchers need to examine whether the total score is an adequate indicator of the observed total true scores compared to the subscale scores (Reise et al., 2007).

The first question related to scoring and interpretation of Grit-S is whether a total score is a sufficient indicator of the latent construct grit. This problem was addressed by fitting the multidimensional data to a bi-factor model and then calculating coefficients omega within the bi-factor structure to measure the percentage of observed score variance that is due to the single latent construct (Reise et al., 2010). In order to determine whether Grit-S should be scored as a univocal measure, the procedures suggested by Reise et al. (2010) were followed in this study. The percentage of explained variance due to grit, interest, and effort and the percentage of explained variance due to a single common factor (omegaH) were compared in three steps. First, the percentage of explained variance due to all common factors (omega) for Grit-S was estimated as

$$\text{omega} = \frac{\left(\sum_{i=1}^{8}\lambda_{i\_grit}\right)^2+\left(\sum_{i=1}^{4}\lambda_{i\_interest}\right)^2+\left(\sum_{i=5}^{8}\lambda_{i\_effort}\right)^2}{\left(\sum_{i=1}^{8}\lambda_{i\_grit}\right)^2+\left(\sum_{i=1}^{4}\lambda_{i\_interest}\right)^2+\left(\sum_{i=5}^{8}\lambda_{i\_effort}\right)^2+\sum_{i=1}^{n}(1-h_i)^2}, \qquad (2)$$

where λ_i is the factor loading of item *i*, and $(1- h_i)^2$ is the unique variance of item *i*. Note, Equation 2 is mathematically equivalent to Equation 1.

Second, the proportion of explained variance due to each unique factor (omegaH) was estimated. In this study, three omegaH coefficients were estimated: omegaH_G was used for general grit, omegaH_I was used for the interest factor, and omegaH_E was used for the effort factor. The percentage of explained variance that is uniquely due to the general factor (omegaH_G) was estimated as

$$\text{omegaH\_G} = \frac{\left(\sum_{i=1}^{8}\lambda_{i\_grit}\right)^2}{\left(\sum_{i=1}^{8}\lambda_{i\_grit}\right)^2+\left(\sum_{i=1}^{4}\lambda_{i\_interest}\right)^2+\left(\sum_{i=5}^{8}\lambda_{i\_effort}\right)^2+\sum_{i=1}^{n}(1-h_i)^2}. \qquad (3)$$

Similarly, the percentage of explained variance that is uniquely due to the interest factor (omegaH_I) and the effort factor (omegaH_E) were estimated by replacing the numerator with the variance explained by each respective group factor. If omegaH_G is relatively high compared to omegaH_I and omegaH_E, then researchers can be confident in concluding that the total score is an adequate indicator of the single construct – grit – underlying Grit-S. Alternatively, if omegaH_G is relatively low compared to omegaH_I and omegaH_E, then a total score is an inadequate indicator of the single construct.

Finally, the percentage of reliable variance in Grit-S scores that is due to the general grit factor was estimated as

$$p = \frac{omegaH\_G}{omega} \times 100\%. \tag{4}$$

Although no hard-and-fast guideline exists for what is considered an adequate percentage of reliable variance that is due to the general factor, in general, a higher percentage means more reliable variance. In the current study, $p > .50$ was used as the cutoff value. If p is greater than 50%, then over half of the reliable variance in Grit-S scores is due to the general factor and total scores of Grit-S can be reported.

The Haberman procedure (Haberman, 2008; Reise et al., 2013) was used to determine whether the total score generated from Grit-S was a better estimator of subscale true scores compared to the subscale scores, in other words, should subscales be created, reported, and interpreted. Two scores were computed: the proportional reduction in mean square error based on the score for the interest subscale (*PRMSE_I*) and the proportional reduction in mean square error based on the score for the effort subscale (*PRMSE_E*). Since this procedure is based on CTT, (a) coefficients alpha estimates based on the total scores (α_g) and subscale scores (α_i for the interest subscale scores and α_e for the effort subscale scores), (b) standard deviation of the total scores (*SD_G*) and subscale scores (*SD_I* for the interest subscale scores and *SD_E* for the effort subscale scores), and (c) the correlation between the interest subscale and effort subscale scores (r) were used to compute PRMSE_I and PRMSE_E in four steps. First, the true score variances for general grit (VAR_{true}_G), interest (VAR_{true}_I), and effort (VAR_{true}_E) were computed as

$$\mathrm{VAR_{true}} = \mathrm{VAR_{observed}} \times \text{coefficient alpha estimate.} \quad (5)$$

For instance, the true score variance for Grit-S ($\mathrm{VAR_{true}}$_G) was the product of observed total score variance (square of SD_G) and the coefficient alpha estimate of the total scores. Second, the covariance matrix among true subscale scores was computed. For Grit-S, this was a 2×2 covariance matrix. Values on the diagonal were the true subscale score variances ($\mathrm{VAR_{true}}$_I and $\mathrm{VAR_{true}}$_E) and values off the diagonal were the covariance of the observed subscale scores. This 2×2 matrix is represented as

$$\begin{matrix} \mathrm{VAR_{true}\_I} & r \times \mathrm{SD\_I} \times \mathrm{SD\_E} \\ r \times \mathrm{SD\_I} \times \mathrm{SD\_E} & \mathrm{VAR_{true}\_E} \end{matrix}$$

.

Third, the covariance between total true scores and the interest subscale true scores [COV (I, G)] and the covariance between total true scores and the effort subscale true scores [COV (E, G)] were computed using

$$\mathrm{COV\ (I, G)} = \mathrm{VAR_{true}\_I} + r \times \mathrm{SD\_I} \times \mathrm{SD\_E}, \quad (6)$$

$$\mathrm{COV\ (E, G)} = \mathrm{VAR_{true}\_E} + r \times \mathrm{SD\_I} \times \mathrm{SD\_E}. \quad (7)$$

Correlations squared for the interest subscale scores (ρ_i^2) and the effort subscale scores (ρ_e^2) were then computed using

$$\rho_{\mathrm{i}}^{2} = \frac{[\mathrm{COV\ (I,G)}]^{2}}{\mathrm{VARtrue\_I} \times \mathrm{VARtrue\_G}}, \tag{8}$$

$$\rho_{\mathrm{e}}^{2} = \frac{[\mathrm{COV\ (E,G)}]^{2}}{\mathrm{VARtrue\_E} \times \mathrm{VARtrue\_G}}. \tag{9}$$

Finally, PRMSE_I and PRMSE_E were computed using

$$\mathrm{PRMSE\_I} = \rho_{\mathrm{i}}^{2} \times \alpha_{g}, \tag{10}$$

$$\mathrm{PRMSE\_E} = \rho_{\mathrm{e}}^{2} \times \alpha_{g}. \tag{11}$$

PRMSE_I and PRMSE_E were then compared with the estimated coefficient alphas for both subscale scores, which was denoted as α_i for the interest subscale scores and α_e for the effort subscale scores. For instance, if PRMSE_I is greater than α_i, then the total scores is a better indicator of the interest subscale true scores and the subscale scores is a redundancy of the total scores, which means interpreting the subscale scores as a separate and unique factor can be misleading. If, however, α_i is greater than PRMSE_I, then the interest subscale scores is a better indicator of the subscale true scores. Similar logic can be applied to the effort subscale scores.

Reise et al. (2010) suggested that a bi-factor structure can also be applied to the multidimensionality structure to examine whether subscale scores represent information that is unique from the general factor. Two types of reliability coefficients are needed to determine the interpretability of a subscale: the subscale score reliability (omega_I or omega_E) and the estimate of the subscale reliability after controlling the effect of general factor (omegaS_I for the interest factor and omegaS_E for the effort factor). Omega_I and omega_E were obtained using Equation 1. OmegaS_I and omegaS_E could be computed using

$$\text{OmegaS\_I} = \frac{(\sum_{i=1}^{4} \lambda_{i\_interest})^2}{(\sum_{i=1}^{4} \lambda_{i\_grit})^2 + (\sum_{i=1}^{4} \lambda_{i\_interest})^2 + \sum_{i=1}^{n} (1-h_i)^2}, \quad (12)$$

$$\text{OmegaS\_E} = \frac{(\sum_{i=5}^{8} \lambda_{i\_effort})^2}{(\sum_{i=5}^{8} \lambda_{i\_grit})^2 + (\sum_{i=5}^{8} \lambda_{i\_effort})^2 + \sum_{i=5}^{8} (1-h_i)^2}. \quad (13)$$

A large omegaS value indicates a large amount of variance is possessed by the subscale factor that is unique from the general factor. A small omegaS value indicates little reliable variance is contained by the subscale scores which is independent from the influence of the general grit factor.

Chapter 4: Results

Preliminary Data Inspection

Table 1 displays item numbers and response distributions for the 8-item 6-point Grit-S scale. Item level response frequencies show fewer participants chose the lowest two response categories (*not at all like me* and *not much like me*) for items 4, 7, and 8. Response distributions are also displayed by gender in Table 2, with the left half of Table 2 summarizing the response frequencies for men (n = 485) and the right half of the Table 2 summarizing the response frequencies by women (n = 125). An inspection of Table 2 further emphasizes that fewer participants selected the lower two response categories for items 4, 7, and 8. In particular, none of the female participants selected the lowest two categories for items 4, 7, and 8. Fewer female students selected the third point category (*pretty much not like me*) for item 4 (n = 2), item 7 (n = 7), and item 8 (n = 2), indicating the 6-point response category system was not behaving as was expected or in other words, participants did not differentiate among the bottom response categories. As such, data collapsing strategies were considered.

Based on the initial item response frequencies, two reduced response category systems were considered: A 4-point response category system and a binary response category system. Specifically, the 8-item 6-point Grit-S was reduced empirically into an 8-item 4-point Grit-S by combining the lowest three response categories (*not at all like me*, *not much like me*, and *pretty much not like me*) across all items. However, this response category system was not balanced. Beamish (2004) recommended that if the Likert scale items are ordinal or categorical in nature and if data collapsing strategies are considered, response categories could be reduced into dichotomous categories to capture trends in the data. So, in order to have a balanced response scale that was substantively meaningful, a 8-item 2-point Grit-S was also created by combining the lower three response categories (*not at all like me*, *not much like me*, and *pretty much not like me*) to reflect choices less like the respondent and the higher three response categories (*pretty much like me*, *mostly like me*, and *very much like me*) were collapsed to represent choices more like the respondent.

Table 1

Response Frequencies for the Eight Items in the Short Grit Scale (Grit-S; N = 610)

Item	Response Frequency					
	Not At All Like Me	Not Much Like Me	Pretty Much Not Like Me	Pretty Much Like Me	Mostly Like Me	Very Much Like Me
1	49	93	227	139	81	21
3	53	83	181	170	98	25
5	37	72	133	218	111	39
6	46	75	123	179	140	47
2	26	61	119	183	146	75
4	4	4	25	116	176	285
7	4	21	62	179	183	161
8	4	10	37	158	218	183

Table 2

Response Frequencies for the Eight Items in the Short Grit Scale (Grit-S) by Men (n = 485) and Women (n = 125)

	Response Categories											
	Male						Female					
Item	Not At All Like Me	Not Much Like Me	Pretty Much Not Like Me	Pretty Much Like Me	Mostly Like Me	Very Much Like Me	Not At All Like Me	Not Much Like Me	Pretty Much Not Like Me	Pretty Much Like Me	Mostly Like Me	Very Much Like Me
1	41	75	178	108	64	19	8	18	49	31	17	2
3	46	71	142	128	80	18	7	12	39	42	18	7
5	33	63	108	166	89	26	4	9	25	52	22	13
6	39	66	97	136	112	35	7	9	26	43	28	12
2	20	44	91	144	120	66	6	17	28	39	26	9
4	4	4	23	99	142	213	0	0	2	17	34	72
7	4	21	55	146	134	125	0	0	7	33	49	36
8	4	10	35	128	177	131	0	0	2	30	41	52

Evidence of Internal Structure

Since all negative items were reverse coded before preliminary analyses, positive correlations among all items were expected. Table 3 shows the polychoric correlations among all 8 items in Grit-S using the 6-point, 4-point, and 2-point response category system. For 8-item 6-point Grit-S, all items excluding Item 2 were positively correlated with each other (ranging from .07 to .71). Item 2 (“Setbacks don’t discourage me”) was negatively correlated with item 1 (r = -.15), item 3 (r = -.16) , and item 5 (r = -.03), and positively correlated with items 4, 6, 7, and 8, indicating responses to Item 2 contradicted the latent construct effort (Duckworth & Quinn, 2009) defined by the consensus of items 2, 4, 7, and 8. Similar results could be found for Item 2 for 8-item 4-point Grit-S. Interestingly, for 8-item 2-point Grit-S, Item 2 positively correlated with all eight items. Item 2 is a double negative item. Thus, empirically, for some respondents, Item 2 might increase their cognitive loading because of the logical complexity of a double negative. Thus, responses from Item 2 were not scored as expected based on the item being misinterpreted and leading to misunderstanding in how to properly use the response category system. This confusion gives rise to inconsistency in how the response scale is used by those low or high on the latent continuum, but this inconsistency or noise is minimized when the response scale was dichotomized. Based on this initial analysis, Item 2 was flagged and a 7-item 6-point Grit-S was reduced from the 8-item 6-point Grit-S by excluding Item 2. In correspondence with the results generated from the preliminary analyses, a 7-item 4-point Grit-S and a 7-item 2-point Grit-S were also evaluated. In order to fully examine the internal structure and score reliability of the 8-

item Grit-S and the performance of Item 2 in CFAs, 8-item 6-point Grit-S, 8-item 4-point Grit-S and 8-item 2-point Grit-S were also examined.

Table 3

*Polychoric Correlations Among all Items in 8-Item 6-Point Grit-S, 8-Item 4-Point Grit-S, and 8-Item 2-Point Grit-S (*N = *610)*

Item	1	3	5	6	2	4	7
1							
3	.53 (.54) .28						
5	.35 (.37) .29	.48 (.51) .38					
6	.48 (.52) .31	.51 (.52) .37	.54(.53) .46				
2	-.15 (-.02) .25	-.16 (-.07) .29	-.03(.06) .40	.04(.11) .41			
4	.15 (.20) .38	.13 (.17) .46	.18(.26) .58	.27(.35) .58	.23(.26) .63		
7	.15 (.23) .35	.17 (.23) .43	.28(.36) .54	.32(.38) .55	.24(.25) .59	.57(.56) .83	
8	.07 (.11) .36	.13 (.18) .44	.24(.32) .57	.30(.36) .56	.26(.29) .62	.24(.72) .89	.71(.69) .82

Note. Polychoric correlations for 8-item 6-point Grit-S are reported without parentheses or an underline. Polychoric correlations for 8-item 4-point Grit-S are reported in parentheses. Polychoric correlations for 8-item 2-point Grit-S are reported with an underline.

6-point response category Grit-S. Table 4 summarizes the standardized factor loadings and fit indices for the 8-item and 7-item 6-point Grit-S for each of the four competing models. For the 8-item 6-point Grit-S, all factor loadings were statistically significant at the .01 level except item 2 in the bi-factor model. Specifically, the loading of item 2 on the general grit factor was negative, $\lambda = -.085$, $p = .104$, indicating that item 2 did not contribute to the common variance (grit) as the other items did in the bi-factor model. Moreover, although the loading of item 2 was significant in the other three solutions (i.e., the unidimensional, two-factor, and second-order models), it was the lowest in magnitude relative to the standardized loadings of the other 7 items.

Table 4

*Standardized Unidimensional (Uni), Two-Factor (2-factor), Second-Order (2nd-order), and Bi-factor Solutions of the 8-Item 6-Point Grit-S and the 7-Item (Excluding Item 2) 6-Point Grit-S (*N *= 610)*

		2-factor		2nd-order		Bi-factor		
Item	λ_{Uni}	λ_{F1}	λ_{F2}	λ_{F1}	λ_{F2}	λ_{Grit}	λ_{F1}	λ_{F2}
1	.468 (.480)	.613 (.616)		.613 (.616)		.346 (.285)	.563 (.603)	
3	.545 (.557)	.697 (.700)		.697 (.700)		.388 (.328)	.700 (.719)	
5	.567 (.572)	.674 (.674)		.674 (.674)		.587 (.547)	.326 (.385)	
6	.659 (.661)	.800 (.795)		.800 (.795)		.729 (.685)	.352 (.426)	
2	.127		.234		.234	**-.085**		.384
4	.716 (.711)		.776 (.773)		.776 (.773)	.349 (.383)		.682 (.654)
7	.714 (.706)		.787 (.783)		.787 (.783)	.442 (.491)		.632 (.586)
8	.800 (.794)		.908 (.904)		.908 (.904)	.390 (.421)		.844 (.850)
				1st-order λ				
				.707 (.707)	.480 (.512)			
r		.339 (.362)						
χ^2	879.794 (770.301)	217.642 (107.425)		217.642 (107.424)		112.327 (28.626)		
df	20 (14)	19 (13)		19 (13)		12 (7)		
RMSEA	.265 (.298)	.131 (.109)		.131 (.109)		.117 (.071)		
90% CI	[.251, .281] (.280, .316)	[.116, .147] (.091, .129)		[.116, .147] (.091, .129)		[.098, .137] (.045, .099)		
CFI	.722 (.751)	.936 (.969)		.936 (.969)		.968 (.993)		
TLI	.611 (.627)	.905 (.950)		.905 (.950)		.924 (.979)		
WRMR	3.721 (3.774)	1.495 (1.085)		1.495 (1.085)		.830 (.396)		

Note. Values in ( ) represent CFA results for 7-item Grit-S. λ = standardized factor loading; r = factor correlation; Uni = unidimensional; F1 = consistency of interest; F2 = perseverance of effort; CI = confidence interval; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; WRMR = weighted-root-mean-square residual. Threshold values for the confirmatory factor models are not provided, but can be provided upon request from the first author. All Chi-square tests were statistically significant at $p < .01$. Loading in bold was not significant at the .05 significance level.

Examination of Chi-square test results and fit indices showed the unidimensional solution did not have adequate fit for the data generated from the 8-item 6-point Grit-S, $\chi^2(20) = 879.794$, $p < .01$, RMSEA = .265, 90% CI [.251, .281], CFI = .722, TLI = .611, and WRMR = 3.721.

The two-factor solution had acceptable fit to the data, $\chi^2(19) = 217.642$, $p < .01$, RMSEA = .131, 90% CI [.116, .147], CFI = .936, TLI = .905, and WRMR = 1.495. Moreover, a Chi-square difference test showed the two-factor solution had improved fit to the data over the unidimensional solution, $\chi^2_{DIFF}(1) = 662.152$, $p < .01$. The two factors were moderately correlated with each other, $r = .339$.

A second-order solution was also fit to the data, which was statistically equivalent with the two-factor solution since there were only two factors loading on the second-order factor. Thus, all fit indices and the Chi-square difference test indicated the second-order solution was a better fit to the data compared to the unidimensional solution. The interest factor had a loading of .707 on the general grit factor, and the effort factor had a loading of .480 on the general grit factor.

Finally, a bi-factor solution provided adequate fit to the data, $\chi^2(12) = 112.327$, $p < .01$, RMSEA = .117, 90% CI [.098, .137], CFI = .968, TLI = .924, and WRMR = .830. The bi-factor solution had improved fit compared to the two-factor solution, $\chi^2_{DIFF}(7) = 105.315$, $p < .01$. This means Grit-S was best represented by a bi-factor model. Reise et al. (2013) suggested that, if the loadings for the general factor are greater than those for the subfactors, a unidimensional solution is recommended for the multidimensional scoring system. However, if loadings for the general factor are equal to or smaller than those for the subfactors and the loadings on the group factors are substantive, then subscales should be considered. In this case, six item loadings associated with the subfactors (interest and effort) were greater than those associated with the general grit factor. All of the loadings on the general factor are reasonable in magnitude. These results suggest the 8-item 6-point grit data is best represented by the bi-factor model.

Next, the same four models were fit to the data excluding item 2. All the loadings for the 7-item 6-point Grit-S were statistically significant. Similar to the results from 8-item 6-point Grit-S, a unidimensional solution did not have acceptable fit to the data, $\chi^2(14) = 770.301$, $p < .01$, RMSEA = .298, 90% CI [.280, .316], CFI = .751, TLI = .627, and WRMR = 3.774. Comparatively, a two-factor solution did fit the data better than the unidimensional solution. Chi-square difference test suggested a good fit for the two-factor model, $\chi^2_{\mathrm{DIFF}}(1) = 662.88$, $p < .01$. Also, the two-factor solution had reasonable fit to the data, $\chi^2(13) = 107.425$, $p < .01$, RMSEA = .109, 90% CI [.091, .129], CFI = .969, TLI = .950, and WRMR = 1.085. The interest subscale and the effort subscale scores are positively correlated with each other, $r = .362$. In the second-order solution, the interest factor had a loading of .707 on the general grit factor, and the effort factor had a loading of .512 on the general grit factor. Finally, a bi-factor model fit the data adequately, $\chi^2(7) = 28.626$, $p < .01$, RMSEA = .07, 90% CI [.045, .099], CFI = .993, TLI = .979, and WRMR = .396. The bi-factor solution also fit the data better than the two-factor solution, $\chi^2_{\mathrm{DIFF}}(6) = 78.799$, $p < .01$. After excluding item 2, the majority of the loadings associated with the subfactors were slightly stronger than those associated with the general factor. All loadings on the general grit factor and the subscale factors are reasonable in magnitude. Thus, bi-factor solution was the best representation of the data according to the suggestions by Reise et al. (2013).

4-point response category Grit-S. Four competing CFA models were estimated and compared in the data with four response categories. Table 5 summarized the results from these CFAs. Of the 8-item 4-point Grit-S, similar to what has been found in Table 4, all factor loadings were significant except the loading of item 2 on the general factor in the bi-factor solution. Loadings of item 2 were still the lowest in magnitude compared to other loadings, which further suggested that item 2 did not behave as what had been expected and should be excluded from the analyses. Chi-square statistics and fit indices showed that unidimensional solution was not adequate fit to the data, $\chi^2(20) = 482.263$, $p < .01$, RMSEA = .195, 90% CI [.180, .210], CFI = .825, TLI = .755, and WRMR = 2.924. Two-factor model had adequate fit to the data, $\chi^2(19) = 106.419$, $p < .01$, RMSEA = .087, 90% CI [.071, .103], CFI = .967, TLI = .951, and WRMR = 1.169; two-factor mode also fit the data better than the unidimensional model, $\chi^2_{DIFF}(1) = 375.844$, $p < .01$. The two factors had a moderately positive correlation, r =. 463. In the second-order model, the interest factor had a moderate loading on the general grit factor, $\lambda = .707$; the effort factor had a slight weaker loading on the general grit factor, $\lambda = .655$. Again, according to the fit indices and the chi-square difference test, the bi-factor solution was a better fit to the data than the two-factor solution, $\chi^2(12) = 38.526$, $p < .01$, RMSEA = .060, 90% CI [.040, .082], CFI = .990, TLI = .977, and WRMR = .563, $\chi^2_{DIFF}(7) = 67.899$, $p < .01$. The general factor pattern loadings of items 1, 2, 3, 4, and 8 were smaller than the corresponding group-specific pattern loadings. The general factor pattern loadings of items 5, 6, and 7 were greater than those of the subfactors. For the 7-item 4-point Grit-S, unidimensional solution was not adequate fit. Two-factor and second-order had acceptable fit, $\chi^2(13) = 82.191$, $p < .01$, RMSEA = .093, 90% CI [.075, .113], CFI = .973, TLI = .956, and WRMR = 1.101, $\chi^2_{\mathrm{DIFF}}(1) = 347.299$, $p < .01$, compared to the unidimensional solution. In the two-factor solution, the internal correlation between the two factors was .476; in the second-order solution, the first-order factor loadings were .707 (interest) and .673 (effort), separately. A bi-factor solution had adequate fit to the data, $\chi^2(7) = 22.633$, $p < .01$, RMSEA = .061, 90% CI [.034, .089], CFI = .994, TLI = .982, and WRMR = .425. Chi-square difference test also showed that bi-factor model was the best solution, $\chi^2_{\mathrm{DIFF}}(6) = 59.558$, $p < .01$. The general factor loadings of items 4, 5, 6, and 8 were greater than the corresponding group-specific pattern loadings. Given the fact that more than half of the loadings on the general grit factor are greater than those associated with the subfactors, all loadings associated with subfactors are moderate in size, and there is a discreprency between the unidimensional factor solution loadings and the general factor of the bi-factor solution, the bi-factor solution was deemed the best representation of the data.

Table 5

Standardized Unidimensional (Uni), Two-Factor (2-factor), Second-Order (2nd-order), and Bi-factor Solutions of the 8-Item 4-Point Grit-S and the 7-Item (Excluding Item 2) 4-Point Grit-S (N = 610)

		2-factor		2nd-order		Bi-factor		
Item	λ_{Uni}	λ_{F1}	λ_{F2}	λ_{F1}	λ_{F2}	λ_{Grit}	λ_{F1}	λ_{F2}
1	.488 (.499)	.622 (.625)		.622 (.625)		.322 (.306)	.605 (.615)	
3	.552 (.567)	.684 (.690)		.684 (.690)		.342 (.331)	.721 (.715)	
5	.591 (.596)	.698 (.697)		.698 (.697)		.561 (.536)	.393 (.420)	
6	.670 (.674)	.815 (.810)		.815 (.810)		.654 (.620)	.447 (.481)	
2	.226		.287		.287	**.096**		.350
4	.738 (.733)		.785 (.783)		.785 (.783)	.506 (.541)		.576 (.519)
7	.735 (.729)		.780 (.778)		.780 (.778)	.619 (.657)		.461(.400)
8	.827 (.822)		.893 (.890)		.893 (.890)	.546 (.567)		.760 (.792)
				1st-order λs				
				.707 (.707)	.655 (.673)			
r		.463 (.476)						
χ^2	482.263 (429.490)	106.419 (82.191)		106.419 (82.191)		38.526 (22.633)		
df	20 (14)	19 (13)		19 (13)		12 (7)		
RMSEA	.195 (.221)	.087 (.093)		.087 (.093)		.060 (.061)		
90% CI	[.180, .210]	[.071, .103]		[.071, .103]		[.040, .082]		
	(.203,.239)	(.075, .113)		(.075, .113)		(.034, .089)		
CFI	.825 (.838)	.967 (.973)		.967 (.973)		.990 (.994)		
TLI	.755 (.757)	.951 (.956)		.951 (.956)		.977 (.982)		
WRMR	2.924 (3.083)	1.169 (1.101)		1.169 (1.101)		.563 (.425)		

Note. Values in ( ) represent CFA results for 7-item Grit-S. λ = standardized factor loading; *r* = factor correlation; Uni = unidimensional; F1 = consistency of interest; F2 = perseverance of effort; CI = confidence interval; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; WRMR = weighted-root-mean-square residual. Threshold values for the confirmatory factor models are not provided, but can be provided upon request from the first author. All Chi-square tests were statistically significant at $p < .01$. Loading in bold was not significant at the .05 significance level.

2-point response category Grit-S. Eight-item Grit-S and 7-item Grit-S with two balanced response categories were created substantively for a balanced response system. The four competing models were also fit to the two dataset. Table 6 summarized the CFA results for 8-item Grit-S and 7-item Grit-S with binary response categories. Item 2 behaved even poorer in the 8-item Grit-S with binary responses: Almost all loadings of item 2 were not significant in the four solutions, and thus, should be excluded from the analyses. Table 6 showed that unidimensional solution was not adequate fit to the data. Two-factor, second-order, and bi-factor solutions all fit the data adequately well. Specifically, the two-factor model had adequate fit, $\chi^2(19) = 58.849$, $p < .01$, RMSEA = .059, 90% CI [.042, .076], CFI = .943, TLI = .916 and WRMR = 1.142 for 8-item 2-point Grit-S; $\chi^2(13) = 38.454$, $p < .01$, RMSEA = .057, 90% CI [.036, .078], CFI = .963, TLI = .941 and WRMR = 1.014 for 7-item 2-point Grit-S. A Chi-square difference test showed the two-factor model had improved fit to the data over the unidimensional solution. The estimated latent factor intercorrelation between interest and effort was .351 (8-item Grit-S) and .362 (7-item Grit-S), respectively. Finally, the bi-factor model was shown to have better fit to the data compared to the two-factor model, $\chi^2(12) = 34.156$, $p < .01$, RMSEA = .055, 90% CI [.034, .077], CFI = .968, TLI = .926, WRMR = 0.769, and $\chi^2_{DIFF}(7) = 24.673$, $p < .01$ for 8-item 2-point Grit-S; $\chi^2(7) = 11.210$, $p = .130$, non-significant, RMSEA = .031, 90% CI [.000, .064], CFI = .994, TLI = .982, WRMR = 470, and $\chi^2_{DIFF}(6) = 27.244$, $p < .01$ for 7-item 2-point Grit-S.

Table 6

*Standardized Unidimensional (Uni), Two-Factor (2-factor), Second-Order (2nd-order), and Bi-factor Solutions of the 8-Item 2-Point Grit-S and the 7-Item (Excluding Item 2) 2-Point Grit-S (*N *= 610)*

		2-factor		2nd-order		Bi-factor		
Item	λ_{Uni}	λ_{F1}	λ_{F2}	λ_{F1}	λ_{F2}	λ_{Grit}	λ_{F1}	λ_{F2}
1	.551 (.552)	.601 (.602)		.601 (.602)		.394 (.386)	.412 (.424)	
3	.657 (.659)	.714 (.718)		.714 (.718)		**.341** (**.332**)	.963 (.951)	
5	.682 (.682)	.719 (.718)		.719 (.718)		.566 (.563)	.348 (.358)	
6	.782 (.782)	.820 (.817)		.820 (.817)		.934 (.921)	**.225** (**.243**)	
2	**.025**		**.127**		**.127**	**.004**		.197
4	.594 (.595)		.758 (.770)		.758 (.770)	.310 (.310)		.672 (.688)
7	.614 (.612)		.878 (.862)		.878 (.861)	.465 (.475)		.650 (.623)
8	.581(.579)		.798 (.798)		.798 (.798)	.240 (.246)		.896 (.904)
				1st-order λs				
				.707 (.707)	.496 (.512)			
r		.351 (.362)						
χ^2	161.929 (142.216)	58.849 (38.454)		58.849 (38.454)		34.156 (**11.210**)		
df	20 (14)	19 (13)		19 (13)		12 (7)		
RMSEA	.108 (.123)	.059 (.057)		.059 (.057)		.055 (.031)		
90% CI	[.093, .124]	[.042, .076]		[.042, .076]		[.034, .077]		
	(.105,.141)	(.036, .078)		(.036, .078)		(.000, .064)		
CFI	.796 (.815)	.943 (.963)		.943 (.963)		.968 (.994)		
TLI	.714 (.722)	.916 (.941)		.916 (.941)		.926 (.982)		
WRMR	2.183 (2.305)	1.142 (1.014)		1.142 (1.014)		.769 (.470)		

Note. Values in ( ) represent CFA results for 7-item Grit-S. λ = standardized factor loading; *r* = factor correlation; Uni = unidimensional; F1 = consistency of interest; F2 = perseverance of effort; CI = confidence interval; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; WRMR = weighted-root-mean-square residual. Threshold values for the confirmatory factor models are not provided, but can be provided upon request from the first author. All Chi-square tests were statistically significant at $p < .01$. Loading in bold was not significant at the .05 significance level.

Summary of CFA results. In conclusion, a unidimensional model was not deemed an adequate solution to the data generated from the 8-item Grit-S and 7-item Grit-S with six, four, or two response categories. The two-factor and second-order solutions fit the data better compared to the unidimensional solution. However, a bi-factor solution, which included a general factor (grit) and two subfactors (interest and effort), fit the data better compared to the two-factor solution and second-order solution. Comparison between the factor loadings on the general grit factor and those on the subfactors also illustrated that the bi-factor model fit the data well. All of the following analyses including reliability, scoring, and interpretation were based on the bi-factor solution to the data.

Evidence of Reliability

Table 7 summarizes the estimates of coefficient omega along with 95% bootstrap confidence intervals for the 8-item Grit-S and 7-item Grit-S with six, four, and two response categories. Current results indicate that the general grit factor of the Grit-S possesses satisfactory reliability (omega_G ranged from .846 to .925). Similarly, the reliabilities for the two grouping factors were high, omega_I ranging from .816 to .937 and omega_E ranging from .803 to .874.

Table 7

Evidence of Reliability for the General Grit Factor, the Interest Factor, and the Effort Factor in the Final Bi-factor Solution of the 8-item Grit-S and 7-item Grit-S with Six, Four, and Two Response Categories (N = 610)

Reliability	Six Response Categories	Four Response Categories	Two Response Categories
Omega_G	.846 (.870)	.868 (.887)	.918 (.925)
95% CI	[.803, .874] (.831, .899)	[.838, .888] (.857, .906)	[.834, .991] (.865, 1.000)
Omega_I	.816 (.815)	.822 (.819)	.937 (.921)
95% CI	[.762, .862] (.763, .859)	[.771, .872] (.773, .861)	[.794, 1.000] (.815, 1.000)
Omega_E	.803 (.866)	.823 (.874)	.811 (.872)
95% CI	[.765, .832] (.836, .889)	[.791, .850] (.842, .998)	[.706, .806] (.767, .922)

Note. Values in ( ) represent results from the 7-item Grit-S. Omega_G = coefficient omega for scores generated from the general grit factor; Omega_I = coefficient omega for scores generated from the interest factor; Omega_E = coefficient omega for scores generated from the effort factor; CI = confidence interval.

Evidence of Scoring and Interpretation

Table 8 summarizes the application of the Reise et al. (2010) procedure to determine the scoring of total scores of the 8-item Grit-S and 7-item Grit-S with six, four, and two response categories based on the bi-factor solution. For 8-item 6-point Grit-S, omega was.846, indicating 84.6% of the total score variance can be attributed to the common factors. OmegaH_G was .423, indicating the general factor contributed to 42.3% of the variability in the scores. In other words, 42.3% of the total scores could be interpretable as indicators of the latent construct grit. OmegaH_G was relevantly higher compared to omegaH_I (.169) and omegaH_E (.253), and the comparison of omega to omegaH indicated that around half (50.06%) of the reliable variance was due to the general grit factor, revealing the general factor accounted for substantially similar portions of common and total variance relative to the specific group factors. Reise et al. (2010) advised that if the omegaH of the general factor was relatively high, total scores can be used as adequate indicator of the underlying latent construct regardless of the multidimensionality. Gustafsson and Aberg-Bengtsson (2010) also suggested the use of a total score in large adaptive testing instruments despite multidimensionality. Thus, the total score can be used as an indicator of the latent construct grit in Grit-S.

Table 8

Application of Reise et al. (2010) Procedure to Determine Scoring of the Total Scores From the 8-item Grit-S and 7-item Grit-S with Six, Four, and Two Response Categories (N = 610)

	Six Response Categories	Four Response Categories	Two Response Categories
OmegaH_G	.423 (.461)	.516 (.551)	.507 (.511)
OmegaH_I	.169 (.225)	.168 (.200)	.241 (.249)
OmegaH_E	.253 (.184)	.185 (.135)	.169 (.166)
Omega	.846 (.870)	.868 (.887)	.918 (.925)
p	50.06 (52.99)	59.38 (62.19)	55.29 (55.18)

Note. Values in ( ) represent results from the 7-item Grit-S. OmegaH_G = the percentage of the explained variance that is only due to the general factor; OmegaH_I = the percentage of the explained variance that is due to the interest factor; OmegaH_E = the percentage of the explained variance that is only due to the effort subfactor; Omega = the percentage of total score variance that is due to the general grit factor, the interest factor, and the effort factor; p = the ratio of omegaH_G over omega $\times$100%.

Table 9 summarizes the results following the Haberman's (2008) procedure to determine whether subscale scores should be created and reported. Results show that PRMSE_I were smaller than α_i, and PRMSE_E were smaller than α_e, indicating in the current study, subscale scores provided a relatively better indicator of subscale true scores, and thus, can be reported. In other words, both the interest subscale and effort subscale could be created and reported to indicate the subscale true scores.

Table 9

Application of the Haberman (2008) Procedure to the 8-item and 7-item Grit-S with Six, Four, and Two Response Categories

	Six Response Categories	Four Response Categories	Two Response Categories
SD_G	5.47 (5.25)	4.52 (4.22)	1.70 (1.63)
SD_I	3.90 (3.90)	2.70 (2.70)	1.39 (1.39)
SD_E	3.23 (2.67)	2.88 (2.43)	0.83 (0.64)
r	.339 (.362)	.463 (.476)	.351 (.362)
α_g	.752 (.787)	.791 (.815)	.741 (.785)
α_i	.787 (.787)	.799 (.799)	.804 (.804)
α_e	.770 (.856)	.776 (.852)	.710 (.848)
PRMSE_I	.734 (.749)	.745 (.772)	.856 (.672)
PRMSE_E	.629 (.578)	.764 (.623)	.569 (.446)

Note. Values in ( ) represent results from the 7-item Grit-S. SD_G = standard deviation of the total scores from Grit-S; SD_I = standard deviation of the interest subscale scores; SD_E = standard deviation of the effort subscale scores; r = correlation between subscales; α_g = coefficient alpha for the total scores; α_i = coefficient alpha for the interest subscale scores; α_e = coefficient alpha for the effort subscale scores; PRMSE_I = Haberman's proportional reduction in mean square error (i.e., reliability) based on total scores rather than subscales for the interest subscale; PRMSE_E = Haberman's proportional reduction in mean square error (i. e., reliability) based on total scores rather than subscales for the effort subscale.

Table 10 summarizes the application of the Reise et al. (2010) procedure to determine the interpretation of subscale scores of the 8-item Grit-S and 7-item Grit-S with six, four, and two response categories based on the bi-factor solution. The omega estimates for the interest subscale scores (omega_I) and effort subscale scores (omega_E) of the 8-item 6-point Grit-S were .816 and .803, separately. For the same dataset, the omegaS estimates for the interest subscale (OmegaS_I) and the effort subscale (OmegaS_E) were .384 and .680, separately, indicating both the interest subscale scores and the effort subscale scores contain a small to moderate amount of variance after the general factor is controlled. Plus, the majority of reliable variance (.680/.803 = 84.68%) in the effort subscale scores was independent of the general factor. Almost half of the reliable variance (.384/.816 = 47.06%) in the interest subscale scores was due to the interest latent variable. Similar results are found for 8-item 4-point Grit-S, 8-item 2-point Grit-S, and 7-item Grit-S with six, four, and two response categories (see Table 10). These results suggest that both interest and effort subscale scores contained information that is independent from the general grit factor. However, the small to moderate amount of variance unique to each group factor, does not clearly support the reporting and interpretation of interest and effort subscale scores.

Table 10

Application of Reise et al. (2010) Procedure to Determine Interpretation of Subscale Scores From the 8-item Grit-S and 7-item Grit-S with Six, Four, and Two Response Categories (N = 610)

Model	Six Response Categories	Four Response Categories	Two Response Categories
Omega_I	.816 (.815)	.822 (.819)	.937 (.921)
OmegaS_I	.384 (.463)	.464 (.493)	.428 (.435)
Omega_E	.803 (.866)	.823 (.874)	.811 (.872)
OmegaS_E	.680 (.625)	.486 (.425)	.684 (.719)

Note. Values in ( ) represent results from the 7-item Grit-S. Omega_I = coefficient omega for scores generated from the interest subscale under the bi-factor structure; Omega_E = coefficient omega for scores generated from the effort subscale under the bi-factor structure; OmegaS_I = the estimate of reliability for the interest subscale after controlling the general grit factor under the bi-factor structure; OmegaS_E = the estimate of reliability for the effort subscale after controlling the general grit factor under the bi-factor structure.

Chapter 5: Discussion

The purpose of this study was to examine the internal structure and score reliability generated from Grit-S using a sample of engineering students from one large southeastern university located in the United States. A great deal of research exists in support of Grit-S as a measure of the latent construct grit. However, not much research has been conducted to examine the internal structure and score reliability of Grit-S. The first goal of this study was to examine the internal structure of Grit-S using CFA models. However, before fitting the data, preliminary analyses showed that some response categories were not used by respondents.

When developing Grit-O, Duckworth and her colleagues (2007) indicated that Grit-O was a general measure of grit within the adolescents and adults population in a variety of domains, including work and school. Grit-S is the short form of Grit-O, which also carried on this domain-free property. Thus, the responses to Grit-S in the current study were expected to be scattered. However, preliminary analyses from current study indicated that fewer engineering students selected the lower end two response categories of 75% of the items (items 4, 7, and 8) in the effort subscale. In particular, none of the female engineering students (n = 195) selected the lowest two categories for items 4, 7, and 8, indicating female students possessed more perseverance in achieving their long term goals. Several explanations could be given for this phenomenon. First, the current study used a 6-point response scale instead of a 5-point response scale proposed by Duckworth and Quinn (2009). The 6-point response scale might not work as expected among these engineering students. Second, the timing of the study might contribute to this lack-of-low-response problem. The current study was conducted at the end of November, 2013, at which time students who were not persistent in their long term goals might have withdrawn from the engineering courses. This might lead to the possibility that the sample in this study was not as representative of the engineering population on this latent variable (i.e., grit) and only those who were persistent in their long term goals were captured in this study. Third, literature has shown that female engineering students are more effortful in their study compared to their male counterparts (Correll, 1997; Vogt, Hocevar, & Hagedorn, 2007). Thus, it is reasonable to see female engineering students scored higher in the effort subscale compared to their male counterparts. Because of this lack-of-response problem, the 8-item 6-point scale was reduced into the 8-item 4-point Grit-S empirically by combining the lowest three response categories. A 2-point Grit-S was also created by combining the lower end three response categories into one category and the upper three response categories into another category. After combination, the low response category reflected choices less like the respondent, and the high response category reflected choices more like the respondent.

Polychoric correlations among the eight items showed Item 2 was negatively correlated with three of the 8 items in Grit-S, even after all negatively phrased items had been reversed coded before analyses. Three 7-item Grit-S scales were reduced from the 8-item Grit-S with six, four, and two response categories. Four competing models were fit to the six dataset. Further examination on loadings showed that Item 2 has the lowest loading compared to the other items in all four solutions to the 8-item Grit-S with six, four, and two response categories. For all the bi-factor solution, Item 2 had non-significant and the lowest loading on the latent construct (see Tables 4, 5, and 6). Moreover, an evaluation of the wording of Item 2 shows that it is a double negative item. Thus, empirically, for some respondents, Item 2 might increase their cognitive load because of the logical complexity of a double negative. Thus, responses from Item 2 were not scored as expected. Results from this study suggested that Item 2 should be discarded or revised before Grit-S is used as a measure of grit.

CFAs showed that of the four competing models the unidimensional model was not adequate solution to the data generated from the 8-item Grit-S and 7-item Grit-S with six, four, or two response categories. The two-factor and second-order solutions fit the data better compared to the unidimensional solution. The Bi-factor solution, which included a general factor (grit) and two subfactors (interest and effort), fit the data better compared to a two-factor solution and a second-order solution. The Bi-factor solution was determined to be the best fit to the data of 8-item Grit-S and 7-item Grit-S with six, four, and two response categories, which was not originally tested by Duckworth and Quinn (2009).

Another purpose of the study was to examine the score reliability generated from Grit-S. Because coefficient alpha has been criticized for its poor estimate of reliability of scale scores, coefficient omega was used in the current study to estimate the reliability of scores generated from the Grit-S. Sample results indicate that the general grit factor, the interest factor, and the effort factor of the Grit-S possess satisfactory reliability: Omega_G ranged from .846 to .925, omega_I from .816 to .937, and omega_E from .803 to .874.

The third purpose of the current study was to examine the scoring and interpretability of total score and subscale scores from Grit-S. The bi-factor structure was fit to the data as suggested by Reise et al. (2010). First, the interpretability of total score was examined. The percentage of the total scores that can be attributed to the common factors (omega) and the percentage of the total score that can be attributed to a single factor (omegaH) were compared. For the 8-item 6-point Grit-S, omega was .846, omegaH for the general grit factor, the interest factor, and the effort factor were .423, .169, and .253, respectively, indicating compared with the contribution of the subfactors, the contribution of the general factor was relatively high to both the total score variability and the total score reliable variability. A conclusion could be made that total scores from Grit-S are an adequate representation of the latent construct grit, and thus, should be reported.

The Haberman (2008) procedure was used to determine whether the subscale scores were a better indicator of subscale true scores compared with the total scores generated from Grit-S. Results showed subscale scores generated from the interest subscale and the effort subscale provide a better representation of the subscale true scores, compared to the total scores. Thus, subscale scores can be created for Grit-S. Finally, the interpretability of subscale scores was studied by calculating the reliabilities of the subscale scores (Omega_I and omega_E) and the unique reliability for the subscale scores after controlling for the general grit factor (omegaS_I and omegaS_E). Results showed that both the interest subscale and the effort subscale contained a relatively low to moderate amount of variance that was unique from the contribution of the general grit factor to the subscale true scores. In particular, the effort subscale scores contained around 80% of the reliable variance in the observed effort subscale scores, and the interest subscale scores contained around half of the reliable variance in the observed interest subscale scores. Thus, effort subscale score might be treated as an adequate indicator of the subscale true score, but the same is not tenable for interest. Interestingly, all items on the interest factor are negatively worded and had poorer evidence for creating interest subscale scores, while while all items on the effort factor were positive and lead to marginal evidence for creating effort subscale scores.

Implications

The results from the current study have several implications for research in grit, and the scoring and interpretation of Grit-S. First, this study added new psychometric information to the research of Grit-S. By comparing four different models, this study provides more thorough model fit results. Findings from this study also suggested that the bi-factor solution was the best solution to the internal structure of Grit-S out of the four competing models. Second, the study provided information regarding the response categories and item quality. Item 2 should be revised or removed before Grit-S is used as a measure of grit in future studies. Third, the findings from this study were informative of the scoring and interpretation of Grit-S. Results from this study show that the total scores generated from Grit-S were adequate indicators of the underlying latent construct grit. Thus, total scores can be used as reliable representations of the grit variable. Results from this study also showed that the subscale scores were better measures of subscale true scores compared to the general factor score; however, results showed the subscale scores contributed a relatively small to moderate portion of variance to the subscale score variability. Thus, the creation and interpretation of subscale scores is unclear and not recommended. Fourth, the current study highlights the utility of the bi-factor model in determining the internal structure of multidimensional instruments, and its advantage in determining the scoring and interpretability of total scores and subscale scores in the presence of multidimensionality. Finally, this was the first study of grit conducted with engineering undergraduate students. Findings from this study could also be useful for those who are interested in measuring grit among engineering students. According to all the findings in this study, researchers should be aware of the bi-factor structure of Grit-S when using it to measure grit and report the total scores based on their research need. Moreover, researchers need to revise or discard Item 2 before using this instrument to measure grit.

Limitations and Future Research

Although the current study added knowledge to the research of grit and Grit-S, it also had several limitations. One limitation is that the current study used a 6-point response system whereas Duckworth and Quinn (2009) used a 5-point response system in their development and validation of Grit-S. Response frequency distribution also showed that the six-category response system did not work as expected. Response categories were reduced into four-category response system and two response system before conducting the CFAs. Although doing so improved the accuracy and stability in item parameter estimates, it would be prudent to compare the two different category systems to get rid of the influence of different categories on item responses. Otherwise, the findings from this study would be sample-specific. Future studies could also focus on examining the appropriate response scale that can be used for Grit-S. Second, the format of Grit-S is different from the original Grit-S used by Duckworth and Quinn (2009). To get rid of the noise caused by the format difference, conducting the future research on psychometrics of Grit-S, a consistent format should be adopted. Third, as there might be an artifact of two-factor models due to the wording, another study could be done comparing the scale including the original 8 items and the alternative scales including the rephrased all positive or negative items. Fourth, since this study is a confirmation of the Grit-S dimensionality and it will use many new techniques in researching grit, it is necessary to conduct more studies to examine the psychometrics of grit for further validation. Finally, although the current study had shown that both total scores and subscale scores can be scored and interpreted, preceding studies need to be conducted to justify the above finding is not unique to the present study and accumulate more evidence to decide if total and subscale scores should be reported in future studies.

Appendix A
12- Item Grit Scale

Directions for taking the Grit Scale: Here are a number of statements that may or may not apply to you. For the most accurate score, when responding, think of how you compare to most people -- not just the people you know well, but most people in the world. There are no right or wrong answers, so just answer honestly!

1. I have overcome setbacks to conquer an important challenge.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

2. New ideas and projects sometimes distract me from previous ones.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

3. My interests change from year to year.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

4. Setbacks don't discourage me.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

5. I have been obsessed with a certain idea or project for a short time but later lost interest.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

6. I am a hard worker.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me

☐ Not much like me
☐ Not like me at all

7. I often set a goal but later choose to pursue a different one.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

8. I have difficulty maintaining my focus on projects that take more than a few months to complete.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

9. I finish whatever I begin.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

10. I have achieved a goal that took years of work.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

11. I become interested in new pursuits every few months.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

12. I am diligent.
☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

Appendix B

Short Grit Scale

Directions for taking the Grit Scale: Here are a number of statements that may or may not apply to you. For the most accurate score, when responding, think of how you compare to most people -- not just the people you know well, but most people in the world. There are no right or wrong answers, so just answer honestly!

1. New ideas and projects sometimes distract me from previous ones.

- ☐ Very much like me
- ☐ Mostly like me
- ☐ Somewhat like me
- ☐ Not much like me
- ☐ Not like me at all

2. Setbacks don’t discourage me.

- ☐ Very much like me
- ☐ Mostly like me
- ☐ Somewhat like me
- ☐ Not much like me
- ☐ Not like me at all

3. I have been obsessed with a certain idea or project for a short time but later lost interest.

- ☐ Very much like me
- ☐ Mostly like me
- ☐ Somewhat like me
- ☐ Not much like me
- ☐ Not like me at all

4. I am a hard worker.

- ☐ Very much like me
- ☐ Mostly like me
- ☐ Somewhat like me
- ☐ Not much like me
- ☐ Not like me at all

5. I often set a goal but later choose to pursue a different one.

- ☐ Very much like me
- ☐ Mostly like me
- ☐ Somewhat like me
- ☐ Not much like me
- ☐ Not like me at all

6. I have difficulty maintaining my focus on projects that take more than a few months to complete.

☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

7. I finish whatever I begin.

☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

8. I am diligent.

☐ Very much like me
☐ Mostly like me
☐ Somewhat like me
☐ Not much like me
☐ Not like me at all

Appendix C

The Short Grit Scale Used in Current Study

1. How much do you think the following statements apply to you?

	1 Not At All Like Me	2 Not Much Like Me	3 Pretty Much Not Like Me	4 Pretty Much Like Me	5 Mostly Like Me	6 Very Much Like Me
1. New ideas and new projects sometimes distract me from previous ones.	○	○	○	○	○	○
2. Setbacks don't discourage me.	○	○	○	○	○	○
3. I have been obsessed with a certain idea or project for a short time but later lost interest.	○	○	○	○	○	○
4. I am a hard worker.	○	○	○	○	○	○
5. I often set a goal but later choose to pursue a different one.	○	○	○	○	○	○
6. I have difficulty maintaining focus on projects that take more than a few months to complete.	○	○	○	○	○	○
7. I finish whatever I begin.	○	○	○	○	○	○
8. I am diligent.	○	○	○	○	○	○

References

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94.

Beamish. W. (2004). *Consensus about program quality: An Australian study in early childhood special education.* Unpublished doctoral dissertation, Griffith University, Queensland, Australia.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guildford.

Byrne, B. M. (2001). *Structural equation modeling with AMOS*. London: Lawrence Erlbaum Associates.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The effcient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307. doi:10.1207/s15327752jpa4803_13

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bi-factor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. doi:10.1207/s15327906mbr4102_5

Correll, S. J. (1997). Talking about leaving: Why undergraduates leave the sciences - Seymour,E, Hewitt,NM. *Contemporary Sociology-A Journal of Reviews*, *26*(5), 644-645.

doi:10.2307/2655673

Duckworth, A. L. (2013). Angela Lee Duckworth: The key to success? Grit Retrieved from http://www.ted.com/talks/angela_lee_duckworth_the_key_to_success_grit#

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. doi: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. doi:10.1080/00223890802634290

Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, *4*(6), 540-547. doi:10.1080/17439760903157232

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British of Journal of Psychology*, *105*(3), 399-412. doi:10.1111/bjop.12046

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

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79-132. doi:10.1016/s0160-2896(97)90014-3

Greenberger, E., Chen, C. S., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, *35*(6), 1241-1254. doi:10.1016/s0191-8869(02)00331-8

Gustafsson, J.-E., & Åberg-Bengtsson, L. (2010). *Unidimensionality and interpretability of psychological instruments*. Washington: American Psychological Association.

Guttman, L. (1945). A basis for analyzing test-restest reliability. *Psychometrika, 10*(4), 255-282.

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics, 33*(2), 204-229. doi:10.3102/1076998607302636

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*(3), 188-201.

Hartigan, J., & Wigdor, A. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery*. Washington, DC: National Academy Press.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling-A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

P20 Motivation and Learning Lab. (2014). Motivation in engineering, from http://sites.education.uky.edu/motivation/research/current-projects/research-on-motivation-in-engineering/

Maddi, S. R., Erwin, L. M., Carmody, C. L., Villarreal, B. J., White, M., & Gundersen, K. K. (2013). Relationship of hardiness, grit, and emotional intelligence to internet addiction, excessive consumer spending, and gambling. *Journal of Positive Psychology, 8*(2), 128-134. doi:10.1080/17439760.2012.758306

Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B., & White, M. (2012). The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets. *Military Psychology, 24*(1), 19-28. doi:10.1080/08995605.2012.639672

McDonald, R. P. (1999). *Test theory: A unified treatment.* Hillsdale: Erlbaum.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide (sixth ed.).* Los Angeles, CA.

Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*,1-18.

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *The European Health Psychology, 16*(2), 56-69.

Reed, A. J., Schmitz, D., Baker, E., Nukui, A., & Epperly, T. (2012). Association of "grit" and satisfaction in rural and nonrural doctors. *Journal of the American Board of Family Medicine, 25*(6), 832-839. doi:10.3122/jabfm.2012.06.110044

Reed, J., Pritschet, B. L., & Cutton, D. M. (2013). Grit, conscientiousness, and the transtheoretical model of change for exercise behavior. *Journal of Health Psychology, 18*(5), 612-619. doi:10.1177/1359105312451866

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140. doi:10.1080/00223891.2012.725437

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bi-factor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19-31. doi:10.1007/s11136-007-9183-7

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145-154. doi:10.1007/s11336-008-9102-z

Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, *47*(3), 361-372.

Salles, A., Cohen, G. L., & Mueller, C. M. (2014). The relationship between grit and resident well-being. *American Journal of Surgery, 207*(2), 251-254. doi:10.1016/j.amjsurg.2013.09.006

Satorra, A., & Bentler, P. M. (1994). *Corrections to Test Statistics and Standard Errors in Covariance Structure Analsis. In Alexander von Eye and Clifford C. Clogg (Eds.)*. Thousand Oaks, CA: SAGE Publications, Inc.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An exploratory and confirmatory factor analytic investigation of item wording effects on the obtained factor structures of survey questionaire measures. *Journal of Management, 21*(6), 1177-1193. doi:10.1016/0149-2063(95)90028-4

Shevlina, M., Milesb, J. N. V., Daviesc, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*, 229-237.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Silvia, P. J., Eddington, K. M., Beaty, R. E., Nusbaum, E. C., & Kwapil, T. R. (2013). Gritty people try harder: Grit and effort-related cardiac autonomic activity during an active coping challenge. *International Journal of Psychophysiology, 88*(2), 200-205. doi:10.1016/j.ijpsycho.2013.04.007

Starkweather, J. (2012). Step out of the past: Stop using coefficient alpha; There are better ways to calculate reliability. Retrieved from http://it.unt.edu/benchmarks/issues/2012/06/rss-matters

Strayhorn, T. L. (2013). What role does grit play in the academic success of Black male collegians at predominantly White institutions? *Journal of African American Studies*, *18*(1), 1-10. doi:10.1007/s12111-012-9243-0

Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A social cognitive construct validation: Determining women's and men's success in engineering programs. *Journal of Higher Education, 78*(3), 337-364. doi:10.1353/jhe.2007.0019

Von Culin, K. R., Tsukayama, E., & Duckworth, A. L. (2014). Unpacking grit: Motivational correlates of perseverance and passion for long-term goals. *Journal of Positive Psychology*, *9*(4), 306-312. doi:10.1080/17439760.2014.898320

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377-392.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133. doi:10.1007/s11336-003-0974-7

Vita

Caihong R. Li

EDUCATION

2015 **Master of Science, Educational Psychology,** University of Kentucky

2009 **Bachelor of Science, Educational Technology,** Tianjin Foreign Studies University, (TFSU, Tianjin, China)

PROFESSIONAL POSITIONS

2013-2015 Researcher, P20 Motivation and Learning Lab, University of Kentucky

2014-2015 Researcher, Robinson Scholars Program, University of Kentucky

SCHOLASTIC AND PROFESSIONAL HONORS

Fall 2014-Spring 2015	International Student Tuition Scholarships, International Center, UK
Fall 2014- Spring 2015	Ashland Graduate Fellows

MANUSCRIPT

Fedewa, A. L., Toland, M. D., Usher, E. L, & **Li, C. R** (2014). *Elementary school students' health-related self-beliefs*. Paper in preparation for Journal of School Health.

Mamaril, N.A., Usher, E.L., **Li, C. R.**, Economy, D.R., & Kennedy, M. S. (2015). *Measuring Undergraduate Students' Engineering Self-Efficacy: A Scale Validation.* Paper resubmitted for Journal of Engineering Education.