AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

© 2020 American Psychological Association
ISSN: 0022-0663

2021, Vol. 133, No. 5, 1038–1058
http://dx.doi.org/10.1037/edu0000594

Grit: The Long and Short of It

Monica Morell, Ji Seung Yang, Jessica R. Gladstone, Lara Turci Faust, and Annette R. Ponnock
University of Maryland, College Park

Hyo Jin Lim
Seoul National University of Education

Allan Wigfield
University of Maryland, College Park

Grit is defined as passion and perseverance for achieving long-term goals and consists of two proposed subcomponents: consistency of interests and perseverance of effort. It has become a much-discussed construct even though research on its underlying factor structure has produced inconclusive results. Furthermore, grit as measured by its most frequently used measure, the Grit-S, does not clearly define "long term" or include the word *passion*. In addition, only a few studies have looked at factor structure variation and predictive validity of grit in different age and cultural groups. We examined, using state of the art multidimensional item response models and structural equation models, the factor structure of both the Grit-S and a newly developed scale (LT-Grit scale) that specifies what is meant by long term. Participants included 1,250 U.S. high school students, 600 U.S. college students, and 500 Korean college students. We found varying factor structures for the Grit-S and a stable one-factor structure for LT-Grit, across age and culture. Perseverance of effort as measured by the Grit-S was the strongest predictor of grades in the three samples. LT-Grit predicted grades only in the U.S. high school and Korean college samples. Thus, there is no evidence that a consistent grit factor or factors exist across age and culture.

Educational Impact and Implications Statement

Grit is defined as individuals' passion and perseverance for achieving long-term goals. Over the last 10+ years it has received much attention in the popular press and from education policymakers despite the lack of consensus on its exact meaning or how strongly it relates to academic achievement. We investigated (using state of the art data analytic techniques) how similar or different grit was in high school and college students and in two different cultures (Korea and the U.S) and how strongly its two proposed components, perseverance of effort (PE) and consistency of interests (CI), relate to students' grades. We found differences in the nature of grit in the different groups, and also that PE related more strongly to achievement than did CI. A main implication of our results is that calls to intervene to improve students' overall grit as a way to enhance their achievement are at best premature and at worst a mistake given our findings and those of others.

Keywords: grit, multidimensional item response theory, item factor analysis, cultural differences

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Some psychological constructs become "phenomenon," even relatively shortly after they are introduced. Grit, conceptualized and operationalized by Duckworth and her colleagues as "trait-level perseverance and passion for long-term goals" (Duckworth, Peterson, Matthews, & Kelly, 2007, p. 1087), is one such construct. They and other researchers have found that individuals with higher levels of grit exhibit greater tenacity in achieving certain goals than their less "gritty" counterparts; outcomes assessed in-

This article was published Online First August 24, 2020.

Monica Morell, Ji Seung Yang, Jessica R. Gladstone, Lara Turci Faust, and Annette R. Ponnock, Department of Human Development and Quantitative Methodology, University of Maryland, College Park; Hyo Jin Lim, Graduate School of Education, Seoul National University of Education; Allan Wigfield, Department of Human Development and Quantitative Methodology, University of Maryland, College Park.

Jessica. R. Gladstone is now at the Department of Psychology, New York University. Annette R. Ponnock is now at Center for Emotional Intelligence, Yale University.

This research was supported by the National Science Foundation under Grant 1534846 to Allan Wigfield and Ji Seung Yang. The opinions expressed are those of the authors and do not represent the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Monica Morell, Department of Human Development and Quantitative Methodology, University of Maryland, College Park, 1230 Benjamin Building, 3942 Campus Drive, College Park, MD 20742-1115. E-mail: m.morell003@gmail.com

clude attainment in the Scripps National Spelling Bee, passing boot camp at West Point, and (in some instances) grades in college. They also demonstrated that grit added incremental predictive validity for these outcomes over and above IQ.

Following these studies, grit began receiving much attention in the popular press (e.g., Tough, 2012), and agencies like the Institute for Educational Sciences (IES), where the director called for developing interventions to bolster students' grit (Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013). These developments occurred despite the absence of definitive research findings on crucial basic substantive and methodological issues regarding grit. First, researchers do not yet agree on the factor structure of individuals' responses to the most widely used grit scales, the Grit-O and Grit-S (Duckworth et al., 2007; Duckworth & Quinn, 2009). Because we do not know definitively the nature of grit's factor structure, we do not know substantively what the construct is. To help resolve these inconsistent findings we thoroughly examined grit's factor structure in samples of different ages and two cultures (the United States and Korea), using a state of the art measurement framework called *multidimensional item response theory* (MIRT, described in detail later).

We assessed both age and cultural differences in grit's factor structure because of the relative paucity of information about each. Muenks, Wigfield, Yang, and O'Neal (2017) found that there were differences in grit's factor structure across high school and college samples; however, both samples were relatively small. Researchers also are beginning to find some cultural differences in grit's factor structure. Therefore, a major purpose of the present study was to extend this work in larger samples and to a culture that has only begun to be included in work on grit—Korea. It is necessary to establish the factor structure of grit across ages and cultures before comparing, for example, mean levels of grit in these different groups, or the strength of its relations to other constructs and achievement outcomes.

Findings regarding grit's relations to one long term student outcome, end-of-term grades, are mixed. Duckworth et al. (2007, Study 3) reported that grit related to undergraduate students' GPAs, even when their SAT scores were controlled for; however, the study was done with a small group of undergraduates from an elite private university. By contrast, in their meta-analysis of the research on grit, Credé, Tynan, and Harms (2017) found that grit's relations to GPA and other academic outcomes were variable across studies. They also found that the relations of perseverance of effort (PE) to achievement were much stronger than those of consistency of interests (CI) and achievement. These results led Credé (2018) to wonder recently "what we should do about grit" (p. 1), and to state "at present there is no empirical support for the idea that grit is the combination of perseverance and passion or for the claim that grit adds to our understanding of success and performance" (p. 5). Yet despite these and other critiques (Muenks et al., 2017; Muenks, Yang, & Wigfield, 2018), Duckworth et al. (2019) continue to claim that "grit seems to enable individuals to get going when the going gets tough" (p. 23503).

We therefore also examined how grit relates to students' grades in the different age and cultural groups in our study. We chose grades in science, technology, engineering, and math (STEM) subject areas as the outcome measure for U.S. samples because students often report that math and science subjects are particularly challenging, perhaps making being "gritty" particularly important for STEM performance and continued participation in STEM fields (Linn & Eylon, 2006). In contrast, not all Korean college students are required to take STEM courses, so overall grades were used as the outcome variable.

There are two other "combined" substantive and methodological issues regarding grit that we address in this study. First is the mismatch of Duckworth and colleagues' (Duckworth et al., 2007; Duckworth & Quinn, 2009) definition of grit as perseverance and passion for long-term goals and the items that make up their two measures of grit, the original 12-item grit scale (Grit-O) and the eight-item Grit-S. Their original measures neither clearly defined nor even mentioned "long-term," and the word *passion* was not included in any of the items. We therefore developed a new measure that complements the Grit-S by explicitly including a time frame in each item, specifying what is meant by long term. Second, Duckworth and Quinn (2009) confounded the directionality of the items on the scales measuring CI and PE in the Grit-S; all the CI items are negatively worded, and PE items positively worded. We thus do not know whether CI and PE items factor separately due to being substantively different constructs or because of an item wording effect.

The Factor Structure of Grit: More Questions Than Answers

Duckworth et al.'s (2007) initial self-report instrument to measure grit, the Grit-O, contained 12 items; however, the most frequently used measure in the literature has been the eight item Grit-S developed by Duckworth and Quinn (2009). Four items measure PE and four measure CI. They tested different factor structure models of grit, in a number of samples, and found (using confirmatory factor analyses) that a model including two first order factors (PE and CI) with a higher-order grit factor fit the data better than did a single factor solution. Wolters and Hussain (2015) also reported (in a study of college students) that a higher-order factor model best fit their data. However, both Credé et al. (2017) and Muenks et al. (2017) pointed out that a higher-order factor model with two lower-order factors is not identified without equality constraints on either the lower-order factor loadings or variances. Once the constraints are imposed, the higher-order factor model is mathematically equivalent to a simple two-factor model.

In addition, other researchers have found different factor structures for the Grit-S (see also Credé et al., 2017). Steinmayr, Weidinger, and Wigfield (2018) conducted exploratory factor analyses (EFA) on German adolescents' responses to a German version of the grit scale and found that a correlated two-factor model (also known as simple structure model) best fit the data.[1] Using an item response theory (IRT) approach to factor analyses, Midkiff, Langer, Demetriou, and Panter (2017) found (in a sample of first-generation college students) that a unidimensional model of each subscale separately (CI or PE) was a better representation of grit than a bifactor model (Gibbons & Hedeker, 1992), based on the item parameter estimates and absolute model fits. Muenks et al. (2017), however, found that a bifactor structure

[1] It should be noted that "conventional" exploratory factor analysis does not allow for identification of any hierarchal factor structures, such as the higher-order models; therefore, the different approaches to factor analysis could yield different conclusions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

provided the best fit among college students. Bifactor models presume that there is one factor capturing the relations among the items (thus in this case representing the grit construct), with subsidiary factors that "capture the residual covariation that is independent of the covariation of the general factor" (Reise, Morizot, & Hays, 2007, p. 22). Finding support for a bifactor model of grit thus means that it is best represented as one construct, not a construct with two underlying "substantive" factors. Thus, Duckworth and colleagues' contention that grit is comprised of a higher-order factor with two subfactors is incorrect, and there is variability across studies in what is the best-fitting model.

In most of the studies just mentioned, researchers have treated item responses as continuous when analyzing the Grit-S scale. However, treating the item responses as continuous is not advisable, because Likert scales are ordinal and not interval (i.e., the distance between categories are not equal; Jakobsson, 2004; Jamieson, 2004; Knapp, 1990). Muenks et al. (2017) and Midkiff et al. (2017) were the first to use the item factor analytic approach using multidimensional item response theory (MIRT[2]; Reise, Moore, & Haviland, 2010) to analyze the factor structure of grit. MIRT treats item responses as categorical (i.e., assuming unequal intervals between each response option; Jamieson, 2004) as opposed to continuous (i.e., assuming equal intervals between each response option). Treating categorical data as continuous can produce biased results particularly when the number of categories is smaller than seven (e.g., Cohen, Manion, & Morrison, 2000; Rhemtulla, Brosseau-Liard, & Savalei, 2012), which is the case for the Grit-S scale.

As noted earlier, another problem with the Grit-S is that the items confound construct and item valence. Including negatively valenced items increases the chances that the negatively worded items form an unintended factor (Marsh, 1986). Duckworth et al. (2007) stated that, "we 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 gave no empirical evidence or rationale to support this conviction. When creating the LT-Grit scale we included only one negatively worded item in each proposed subcomponent of LT-Grit.

In summary, the work designed to determine grit's meaning as a construct is plagued with methodological problems, inconsistent results, and incorrect conclusions about grit being a higher-order construct with two subcomponents. The methodological issues of assuming item responses are continuous and the substantive/methodological issue of the confound of the two grit subconstructs with positive and negatively worded items further muddies the grit waters. In addition, whether or not grit functions differently across ages and cultures has not been thoroughly addressed; we turn next to those topics.

Variation in Grit's Factor Structure Across Age

Researchers have used the Grit-S with children, adolescents, and adults, but few studies to date have examined whether the factor structure of grit varies across age. One hypothesis based in interest theory (Hidi & Renninger, 2006) is that college students' grit will show less differentiation than that of high school students because their interests become more clearly defined and relate more closely to their self-regulation. Sansone, Thoman, and Fraughton (2015) suggested that as students mature their interest becomes more embedded in their self-regulation. Thus, the proposed PE and CI components of grit may become more highly correlated as students move through school, until they form one factor.

Muenks et al. (2017) explored whether grit's factor structure varied across high school and undergraduate college students and found that a two-factor model, in which CI and PE are separate but correlated constructs, fit best in the high school sample. By contrast, the bifactor model fit best for a sample of college students, meaning that grit should be conceptualized as a "single factor" construct, as noted earlier. In the present study we assessed the Grit-S's factor structure using the same MIRT models, utilizing larger and more diverse samples of high school and college students than Muenks et al. (2017) had available to them, allowing us to provide more definitive information about grit's factor structure.

Variation in Grit's Factor Structure Across Culture

For many years now social psychologists have described broad cultural differences that characterize Eastern and Western cultures (e.g., Markus & Kitayama, 1991; Triandis, 2001). Eastern cultures are characterized as collectivist and group oriented, and Western cultures as individualistic. Researchers have proposed that these kinds of cultural differences influence the nature of personality in the cultures. Interestingly, however, McCrae (2001) assessed cultural differences in the structure of personality and found that at the scale level the Big Five personality traits are similar across culture. Other researchers have commented that this is less true when looking at item-level responses (Nye, Roberts, Saucier, & Zhou, 2008), and also stated that one reason for the invariance at the scale level is that the measures used to assess the Big Five essentially impose that structure on people's responses (Church, 2001; Van de Vijver & Leung, 2001).

What about cultural differences in more specific personality characteristics like grit? As discussed earlier, Steinmayr et al. (2018) found a two-correlated factor model of grit fit best in their studies of German high school students. Several researchers have examined grit's factor structure in different Asian cultures. Nishikawa, Okugami, and Amemiya (2015) developed a Japanese version of the Grit-S and found support for a two-factor model in their college student sample that was similar to that reported by Duckworth et al. (2007). Interestingly, the items in their measure bear little resemblance to those of the Grit-S so it is difficult to interpret their results. Li, Zhao, et al. (2018) gave a group of Chinese adolescents a translated version of the Grit-S and also found support for a two-factor model.

Datu and colleagues studied grit in Filipino high school and college students. Datu, Valdez, and King (2016) found that a two-factor structure fit best for both of their samples. Based on an interview study with Filipino students Datu, Yuen, and Chen (2017, 2018) developed and tested a measure of "adaptability to situations" that they believe should be included as part of the grit

[2] Note that MIRT models and the confirmatory item factor models/structural equation models with categorical indicators are mathematically equivalent models that often use different parameterizations and default estimation methods that vary across software packages (Finch, 2011; Kamata & Bauer, 2008; Knol & Berger, 1991; Kohli, Koran, & Henn, 2015; Lord, Tukey, & Novick, 1968; McDonald, 1997; Reise, Widaman, & Pugh, 1993; Takane & De Leeuw, 1987; Wirth & Edwards, 2007).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

construct. They gave this measure with their translated version of the Grit-S and found support for a triarchic model of grit (PE, CI, and adaptability).

Most relevant to our work are Kim and Hwang (2015) and Lim's (2017) grit validation studies using South Korean samples. Korea is a particularly interesting "collectivist" country in which to study grit because of its strong focus on education. Shin, Jahng, and Kim (2019) termed this focus as "education fervor," the notion that being admitted to and doing well at a prestigious college is a guarantee of success both economically and socially (see also Song & Chung, 2020). Given this focus on long term educational goals, grit may be an especially important construct in Korea; thus, assessing its factor structure and relations to achievement outcomes will provide much-needed information about the nature of grit in different countries. Lim found that the bifactor model fit best, but Lim chose the two-factor model because of small factor loadings in the bifactor model. However, Lim treated students' responses as continuous rather than categorical and did not examine the predictive validity of grit, something we address in the present study. Kim & Hwang (2015) only reported predictive validity of the two grit subdomain summed scores without testing the factor structure.

The Missing "Long-Term" Aspect of Grit

As noted earlier, although Duckworth and colleagues (2007) defined *grit* as persistence and passion for long-term goals, neither the Grit-O nor the Grit-S captures this long-term aspect well. The items are vague at best about what is meant by "long-term" (e.g., "I have difficulty maintaining my focus on projects that take more than a few months to complete"; see Table 1). Different-aged individuals likely have different interpretations of the phrase "a short time"; and "a few months" may be considered a long time for some groups but not others (i.e., adolescents vs. adults). Thus, we developed and examined a set of items measuring grit that includes wording on individuals' involvement in projects or activities lasting for a relatively longer period of time. Details of the measure development process are in the Methods section.

Grit's Relations to Student's Grades

Duckworth et al. (2007, Study 3) reported that grit related to undergraduate students' GPAs in a study conducted at an elite private university. Students' overall, or combined, grit scores have been found to predict grades in U.S. elementary and middle school students (Rojas & Usher, 2013), undergraduate students in the United States, China, and Korea (Chang, 2014; Duckworth et al., 2007; Hwang, Lim, & Ha, 2018; Lee & Sohn, 2017; Li, Lin, Zhao, Chen, & Wang, 2018; Muenks et al., 2017; Strayhorn, 2014), and doctoral students (Cross, 2014), although the strength of these relationships varies across these studies (Credé et al., 2017). However, Credé and colleagues (2017) concluded that the relationships observed in these studies may likely be mediated, moderated, or both, by other aspects such as related constructs, individual differences, and the specific performance domain. They most likely based this conclusion on their meta-analytic results showing that overall grit had relatively weak relationships with academic performance ($\hat{\rho} = .18$, $k = 39$, $N = 13{,}141$, $SE_{\hat{\rho}} = .11$) and with overall GPA ($\hat{\rho} = .17$, $k = 37$, $N = 12{,}601$, $SE_{\hat{\rho}} = .10$), where $\hat{\rho}$ is the average correlation, k is the number of studies, and N is the total sample size across k studies.

Most studies show that the PE aspect of grit relates more strongly to outcomes than does CI (Credé et al., 2017; Muenks et al., 2017; Steinmayr et al., 2018). In their meta-analysis Credé et al. reported PE correlated at $\hat{\rho} = .26$ (k=11, N= 5,221, $SE_{\hat{\rho}} = .12$) with academic performance, but CI had a correlation of $\hat{\rho}$=.10 (k=11, N= 5,221, $SE_{\hat{\rho}} = .02$), and overall grit correlated at $\hat{\rho} = .18$ ($k = 39$, $N = 13{,}141$, $SE_{\hat{\rho}} = .11$). Credé et al. pointed out that the significant relations of overall grit score to these outcomes are primarily due to the relations of PE to the outcomes.

Muenks et al. (2017) found that PE related more strongly to students' grades than did CI, when examining their predictive validity using a two-stage estimation approach wherein factor scores for the grit constructs were first calculated using a MIRT model and then used in analyses to predict grades in a regression model. Several other authors also have used a two-stage estimation approach when investigating the predictive validity of grit (Bow-

Table 1
All Items for Grit-S and LT-Grit Scales

Construct	Item
Grit (Grit-S; Duckworth & Quinn, 2009)	I often set a goal but later choose to pursue a different one. (R)
	I have been obsessed with a certain idea or project for a short time but later lost interest. (R)
	I have difficulty maintaining my focus on projects that take more than a few months to complete. (R)
	New ideas and projects sometimes distract me from previous ones. (R)
	I finish whatever I begin.
	Setbacks don't discourage me.
	I am diligent.
	I am a hard worker.
Long-term grit (LT-Grit)	I remain passionate about the same activities for 6 months or more.
	I stay interested in the same activities for 6 months or more at a time.
	I lose interest in projects that take more than 6 months to complete. (R)
	I continue to work on my projects even when I feel discouraged.
	I finish projects that I started years ago.
	I continue to work on projects for 6 months or more, even when they are difficult.
	I work hard on the same project for 6 months or more.
	I give up easily on difficult projects after 6 months. (R)

Note. R = reverse coded.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

man, Hill, Denson, & Bronkema, 2015; Datu et al., 2016; Eskreis-Winkler, Shulman, Beal, & Duckworth, 2014). However, this two-stage estimation approach to analyzing the predictive validity of grit may result in misleading conclusions such as underestimated regression slopes (the effect of grit in our case) and standard errors (Bollen, 1989; Skrondal & Laake, 2001), because the uncertainty in the factor scores estimated in the first stage is carried over into the next stage of analysis. Many methodologists state that a better approach is to estimate the relationship between item responses and factor scores at the same time as the regression parameters (i.e., relationship between factor scores and outcomes) in a single stage of estimation (see, e.g., Bartholomew & Knott, 1999; Muthén, 2002; Rabe-Hesketh & Skrondal, 2004). We therefore examined the relations of participants' PE and CI to their achievement in science and math courses using the one-stage estimation approach, the first time this has been done in the literature on grit.

Hypotheses

We addressed several critical substantive and methodological issues regarding grit and its relations to high school and college students' grades in STEM courses to (a) provide more conclusive evidence about the factor structure of the Grit-S by using MIRT-based factor analyses in large and diverse samples; (b) examine the factor structure of grit in the two age groups and in two cultures; (c) develop and test items to measure long-term aspects of grit, LT-Grit; (d) examine the relation of students' grit (measured by both the Grit-S and LT-Grit) to their achievement; and (e) examine the relationship between grit, as measured by the Grit-S, and long-term aspects of grit, as measured by the LT-Grit scale.[3]

The specific hypotheses are

1. The best fitting factor structure model for Grit-S in U.S. high school students will be the correlated two-factor model. The best fitting factor structure model for Grit-S in the U.S. and Korean college students will be the bifactor model.

2. The best fitting factor structure models for LT-Grit will be the same as those hypothesized for the Grit-S in the three samples: The correlated two-factor model will fit best for high school students and the bifactor model will fit best for the U.S. and Korean college students.

3. High school, U.S. college, and Korean college students' grit will significantly predict their grades. We expect that PE will be the stronger predictor, and once PE is controlled for, CI will not significantly predict grades in the three samples.

4. High school, U.S. college, and Korean college students' grit measured by LT-Grit will significantly predict their later grades. We expect that PE will significantly predict grades and, once PE is controlled for, CI will not significantly predict grades in the three samples.

5. High school, U.S. college, and Korean college students' grit as measured by Grit-S will be moderately to strongly correlated with their levels of long-term grit as measured by LT-Grit.

Method

This research received approval from the University of Maryland, College Park and the Seoul National University of Education institutional review boards.

Sample 1 (U.S. High School)

Participants. Participants were 1,246 high school ninth and 10th graders (45.1% female) from six private and one public school in the mid-Atlantic region of the United States. Once schools agreed to participate students were recruited via their math or science classrooms though a partnership between researchers and administrators at each school. In the private schools, passive consent was obtained from parents and passive assent was obtained from students. Active consent and active assent were obtained from parents and students, respectively, in the public school. Each student was offered a $5 gift card for their participation in the study. The mean age was 14.86 (SD = .74 years). The sample was 49% White/European American, 35.7% Black/African American, 6.9% Asian/Asian American, 6.7% Hispanic/Latinx, 0.8% Native American, 0.2% Pacific Islander, and 1.7% Other.

Measures.

Grit. Grit was measured with the Grit-S (Duckworth & Quinn, 2009) scale. The instrument consists of four items which measure CI (e.g., "I often set a goal but later choose to pursue a different one") and four items which measure PE (e.g., "I am diligent"). Items are assessed via a five-category Likert scale with choices ranging from 1 (*not at all like me*) to 5 (*very much like me*).

Long-term grit. The long-term aspect of grit was measured with a newly developed eight-item LT-Grit scale that we developed as follows. First, we consulted with several internationally known experts in the fields of motivation and self-regulation regarding their views on what "long term" means to high school students. We proposed that six months seemed long term for adolescents as it would mean continuing to work on projects across more than one semester. Our consultants unanimously agreed that six months would be long term in the eyes of adolescents. Then, five members of the project team each generated a set of eight items, focusing on writing items that reflected the meaning of each PE and CI item on the Grit-S. The team then scrambled the items and discussed whether each item was meant to measure CI or PE. Any item not correctly identified was discarded. We did not include an item similar to "Setbacks discourage me" because of the problems it caused for Muenks et al. (2017) in a previous study. We also did not include two words we believed would be problematic for respondents to understand: *diligent* and *obsessed*. Going back to

[3] The data used in this study come from a larger longitudinal study of the development of grit and other personality and motivational belief and value constructs during ninth and tenth grade. This is the first report from that project examining the factor structure of grit. Because of the necessity of conducting numerous analyses of both the Grit-S and the LT-Grit measures to test the hypotheses and clarify grit's factor structure we focus only on these two measures in this paper.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the definition of grit as "passion and persistence for long-term goals" we included the word *passionate* in one item. The remaining items were scrambled once again, and to mirror the Grit-S the team met to choose the four items that best reflected the meaning of the items measuring CI on the Grit-S and four items that best reflected the meaning of the items measuring PE. To unconfound the negative-positive wording and content aspect of the Grit-S we included only one negatively worded item in each set of items. Students responded on a five-category Likert scale from 1 (*not at all like me*) to 5 (*very much like me*). The items are presented in Table 1.

Grades. Schools provided students' math and science grades at the end of the semester. The grades were in the form of percentages and on a scale from 1 to 100.

Procedure. Participants completed a questionnaire containing the Grit-S, LT-Grit, six additional cognitive instruments that were not used for the present study, and demographics questions either online or on paper. The Grit-S and LT-Grit items appeared first in all questionnaires. Students completed the questionnaires in 15–30 minutes in one sitting during a school period which was set aside for that purpose. All students were informed that their participation would be voluntary, and their responses would be kept confidential.

Sample 2 (U.S. College)

Participants. Participants were 581 undergraduate students (46.3% female) from a mid-Atlantic university. To recruit participants, a graduate research assistant contacted professors teaching a math or science undergraduate course. The professors who agreed to have their students participate told their students about the study and either offered them extra credit for participating or did not offer extra credit; 94% of the participating students received the extra credit. Professors sent their students a link to the online survey. To participate, students first had to provide active consent via the online questionnaire. Each student who gave consent was entered into a raffle for the chance to win one $50 Amazon gift card. The average age of participants was 19.83 years (SD = 2.21 years). The sample was 49.3% White/European American, 12.5% Black/African American, 5.6% Hispanic/Latinx, 28.5% Asian/Asian American, 7.2% Biracial, and 4.1% Other.

Measures.

Grit. Grit was measured with the same scale as in the high school sample.

Long-term grit. The long-term aspect of grit was measured with the same scale as in the high school sample.

Grades. Researchers collected final semester grades for the math or science course in which the student was recruited from school records. The grades were in the form of letters (e.g., A+, C-, F, etc.) and the authors recoded them on a scale from 1 (F) to 13 (A+).

Procedure. Professors sent a single link out to all students, and students were informed that their participation would be voluntary, and their responses would be kept confidential. The questionnaires were completed on the students' own time and took students approximately 20 minutes to complete. Once the semester was complete, the authors requested the students' grades from the registrar's office.

Sample 3 (Korean College)

Participants. Participants were 529 undergraduate students (57.1% female) from two four-year-universities in South Korea. Both universities were located in metropolitan cities in South Korea, and most of the students in the sample were from the first university (83.1%) from which we recruited. Professors and instructors teaching education-related courses were recruited and asked to administer the survey in their class. Students' majors were education (70.1%), math/natural/engineering science (8.7%), liberal arts/social science (8.5%), arts and physical education (5.1%), and other (3.2%). Fourteen (2.6%) students were excluded from the data set due to entire pages of missing items or single unusual response patterns, yielding a final participant count of 515 students (58.6% female). The average age of participants was 20.02 years (SD = 1.80 years).

Measures. The authors back-translated all survey items to attain consistency of meaning between English and Korean. To do that, the second and sixth authors first translated all items into Korean, and those items were then back-translated into English by two other English-Korean bilingual professors.

Grit. Grit was measured with the translated Grit-S scale, which has items identical to those given to the high school and U.S. college samples.

Long-term grit. The translated LT-Grit scale, which has items identical to those given to the high school and U.S. college samples, was used to measure the long-term aspect of grit.

Grades. Students' overall final semester grade point average (GPA), which could range from 0 (F) to 4.5 (A+) were collected.

Procedure. Students completed the survey in lecture rooms during regular class hours. Professors told their students the purpose of the survey and informed them that their participation would be voluntary, and their responses would be kept confidential. After receiving a consent form and questionnaires, students signed, completed, and submitted them to their professors. Completion of the instructions and survey took approximately 20 min in each participating class.

Analysis

There was less than 1% missing data on both the Grit-S and LT-Grit measures in all three samples. We used full-information maximum likelihood (FIML) estimation, which is a model-based approach to missing data (see Enders, 2010) with the expectation-maximization computation algorithm (Bock & Aitkin, 1981) for all factor analytic models. The models were fit using flexMIRT v3.5 (Cai, 2017). The flexMIRT software package allows for the fitting of multidimensional item factor models with efficient computation algorithms. This is analogous to fitting confirmatory factor analysis (CFA) models using general software packages for latent variable analysis like M*plus* when data are treated as categorical and the estimation method is maximum likelihood. In addition to the computational efficiency, flexMIRT provides M_2 statistic, the uniquely available absolute model fit index for multidimensional item factor models but not available from general SEM software

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

packages such as M*plus* or LISREL.[4] The following[5] goodness of model fit indices were used to determine the model quality and compare models when appropriate: M_2 statistic root mean square error of approximation (RMSEA; Joe & Maydeu-Olivares, 2010), observed log-likelihood, Akaike information criterion (AIC; Akaike, 1981), and Bayesian information criterion (BIC; Schwarz, 1978).

Analyses to test Hypothesis 1. To investigate the best fitting factor structure model for Grit-S, we initially tested three competing item factor models that have been reported in the literature, treating item responses as categorical rather than continuous (e.g., Muenks et al., 2017): (a) a one-factor model in which all eight items measuring grit loaded on one factor; (b) a correlated two-factor model where CI and PE are separate but correlated factors; and (c) a bifactor model in which all eight grit items loaded on one latent factor and two uncorrelated subscale factors capture the residual dependency among the PE and CI items. As discussed earlier, in bifactor models the general factor represents overall grit and the specific factors represent the residual CI and PE that remains after controlling for overall grit.

Also as discussed earlier we did not test a higher-order factor model because mathematically it is the same as the correlated two-factor model for Grit-S. A key assumption of factor analytic models (i.e., unidimensional IRT) is that responses to items are independent from each other after controlling for the latent construct being measured (McDonald, 1982). That is, each item has only one common source of variance. This assumption becomes implausible when an instrument is composed of groups of items with similar content (e.g., items on a subscale). Under such circumstances, it is necessary to use MIRT to properly model (i.e., control for) all the sources of common variance in the item responses. When it is hypothesized that all items measure a common latent construct (i.e., grit), but each item's variance is further influenced by an additional common construct as a result of groups of items tapping similar aspects of the latent construct of interest (i.e., CI and PE), the bifactor model represents relations among items the best (Reise, Moore, & Haviland, 2010). Including bifactor models in the series of models allowed us to assess the possibility that grit is a unidimensional construct when the dependency among subscale items is controlled for. As noted earlier, these models assess the fit of one overarching factor (general factor), with the subscale factors (called residual factors or specific factors) representing the extra dependency among the items. Such models are often used in research on test construction (Cai, Yang, & Hansen, 2011; Reise & Revicki, 2014), as well as on the Big Five scales (Nieto, Abad, & Olea, 2018), and provide a more definitive assessment of whether a given construct is best represented by one factor.

Because the bifactor model fit the best in the U.S. and Korean college samples but had many nonsignificant factor loadings (see the Results section), we fit two additional, alternative types of bifactor models, in which one of the residual factors is omitted from the conventional bifactor model, as a way to examine what exactly the general factor that we would call grit means in these samples. These alternative bifactor models take the omitted residual factor in the conventional bifactor model as a "reference" factor (now the general factor in the alternative model) to which the remaining residual factor is compared (Eid, Geiser, Koch, & Heene, 2017).

Eid et al. (2017), based in stochastic measurement theory, suggested using the alternative bifactor model for estimation stability when there are "irregular" behaviors in the bifactor model. The general factor in alternative bifactor models, called the "*reference factor*," accounts for all the common variation in the items related to residual factor which was omitted from the conventional bifactor model. This reference factor is used as a comparison standard to which the remaining residual factor, which is retained from the conventional bifactor model, is compared. The residual factor then represents the variability in the items that is not shared with the reference factor and can only be interpreted in reference to the general factor. Researchers can use these bifactor models to examine alternative multidimensional factor structures when a residual factor included in a bifactor model "disappears" due to their factor loadings being nonsignificant, which indicates that all of the shared variance in those items is accounted for by the single general (or primary) factor of the conventional bifactor model.

We examined two alternative bifactor models in both college samples. In one of the models the CI residual factor was not specified (we call this the CI:S-1 model) and in the other the PE residual factor was not specified, the PE:S-1 model. In the CI:S-1 model, the only residual factor remaining is the PE factor. Therefore, the general (reference) factor represents the CI construct as that is the only factor where the CI items are specified to load. Because the PE items load on both the general and the residual factor, only the variance shared by the PE and CI items is accounted for in the reference factor; remaining (i.e., unaccounted for by the general factor) common variance in the PE items is accounted for by the residual PE factor.

The structure of the S-1 bifactor model renders the interpretation of the factor scores more challenging and limits the applicability of the model. Thus, although we present the interpretation of factor scores under the CI:S-1 and the PE:S-1 models for completeness, we caution against the use of these models as a way to conceptualize grit. We believe that an S-1 bifactor model in this study functions as a tool to understand the nature of general factor grit in multiple samples. The reference factor score in the CI:S-1 model can be interpreted as an individual's latent scores on CI, with the residual factor scores interpreted as the deviations of the PE factor score from its expected value based on the individual's CI factor scores. A positive residual PE factor score would indicate that the respondent has a higher PE factor score than would be expected given his or her CI factor score. A negative PE factor score would indicate that the respondent has a lower PE factor score than would be expected based on his or her CI factor score. The same logic applies to the PE:S-1 model; in this case the reference

[4] See Han and Paek (2014) for a review of commercial software packages for multidimensional item factor modeling, including flexMIRT.

[5] Note that FIML estimation provides more accurate standard errors compared to weighted least squares (WLS) estimation (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). However, it is also well known that the FIML estimation of multidimensional models with categorical responses data only provides likelihood-based model fit indices that we use here and full information goodness of model fit such as Pearson's residual analysis is not trustworthy due to the small expected frequency per item response pattern. Accordingly, we use M_2 statistics that were developed to assess the model fit for MIRT models.

factor represents the PE construct. The residual factor is CI, with positive and negative factor scores reflecting deviations from expected CI factor scores based on PE factor scores. For more technical details on this interpretation, we refer readers to Eid et. al (2017). Figure 1 shows graphical representations of each model.

It also is important to note that we did not immediately test for measurement invariance across groups because the factor model needs to be first determined to conduct such analysis (see, e.g., Brown, 2015; Bauer, 2017; van de Schoot, Lugtig, & Hox, 2012). That is, configural invariance needs to be established before one can assess measurement invariance (e.g., Horn & McArdle, 1992). Thus, the purpose of testing Hypothesis 1 in this study was to determine more clearly the factor structure of grit using multiple samples. Because, as discussed in the Results section, the factor structures do differ across groups, we could not test for measurement invariance.

Analyses to test Hypothesis 2. To determine if "long-term" grit, as measured by the LT-Grit, is a single construct or two related factors, we fit a one-factor model, the two correlated factor model, and the bifactor model (see Figure 2). We also used a local fit index called *LD* X^2 *(*Chen & Thissen, 1997) to examine the unidimensional IRT model assumption, local independence. The *LD* X^2 compares correlations computed for the model-implied probabilities and the observed proportions of all item pairs to detect local dependency between two items. We examined the pattern of LD X^2 values and made subsequent modifications to the scale.

Analyses to test Hypotheses 3 and 4. To address whether grit as measured by Grit-S and LT-Grit would predict students' grades, we used a structural equation modeling approach to determine whether the latent variable grit from Grit-S and long-term grit from LT-Grit predict observed grades. For the measurement models, the decisions were made based on the item factor analysis results from Hypothesis 1 (see Figure 3) and Hypothesis 2 (see Figure 4), respectively. The authors used M*plus* Version 8 (Muthén & Muthén, 2017) for all of the structural equation analyses with FIML estimation.

Analyses to test Hypothesis 5. To examine the relationship between grit as measured by Grit-S and long-term grit as measured by LT-Grit, we used the best fitting factor structure for grit based on the results from Hypothesis 1 and the best fitting factor structure for long-term grit based on the results from Hypothesis 2 to estimate the latent correlation between the grit constructs. These analyses were done in M*plus* Version 8 (Muthén & Muthén, 2017) using FIML estimation.

Results

Hypothesis 1: The Grit- S's Factor Structure

All of the successfully converged models had adequate absolute fit with RMSEA values ranging from .00 to .06 (see Table 2), indicating relative model fit indices (e.g., AIC or BIC) should be used to choose the best model.

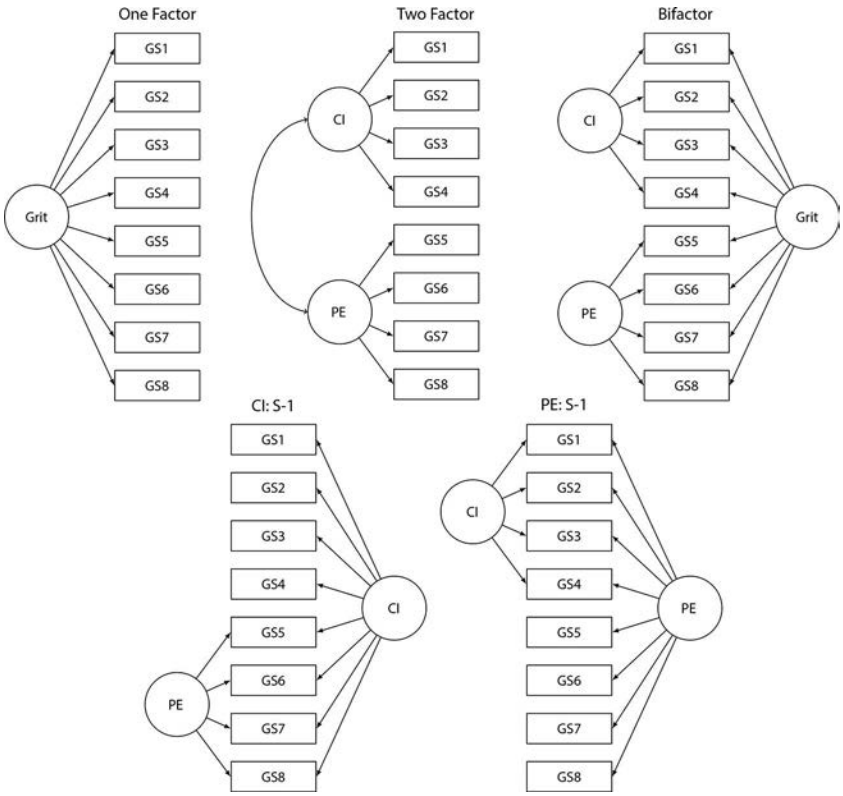


Figure 1. Graphical representations of factor structures for Grit-S. CI = consistency of interests; PE = perseverance of effort; GS = Grit-S item.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

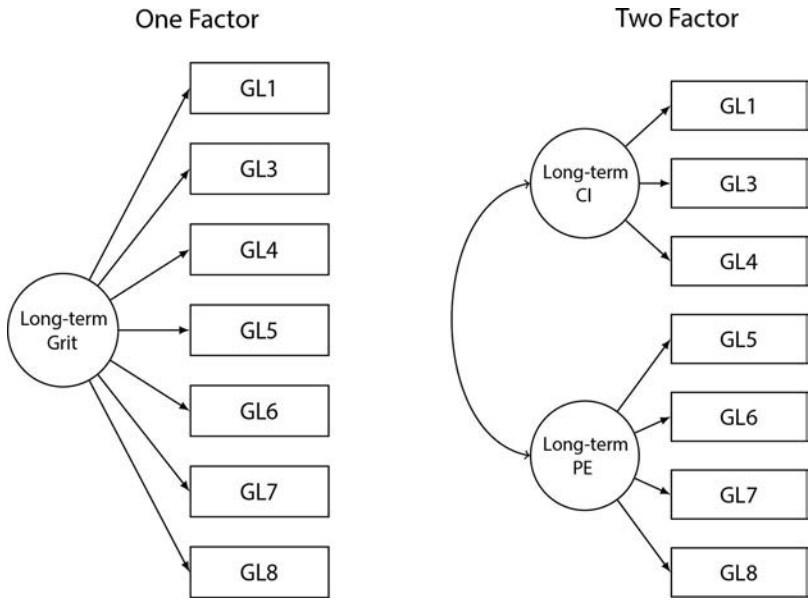


Figure 2. Graphical representations of factor structures for LT-Grit (long-term grit). CI = consistency of interests; PE = perseverance of effort; GL = long-term grit item.

High school sample. The correlated two-factor model yielded the best fit (see Table 2). The likelihood ratio test between the correlated two-factor model and the one-factor model indicated that the two-factor model predicted the observed item responses significantly better (G^2: 26,172.71–25,866.86 = 305.85, *df*: 41–40 = 1, $p < .05$). The CI and PE subfactors correlated at $r = .46$ ($SE = 0.04$). The bifactor model did not properly converge. Therefore, the correlated two-factor model is used in the analyses to address Hypothesis 3.

U.S. college sample. The bifactor model fit better than the one-factor model (G^2: 11,245.20–10,895.76 = 394.44, *df*: 48–40 = 8, $p < .05$) and better than the correlated two-factor model (G^2: 10,990.93–10,895.76 = 95.17, *df*: 48–41 = 7, $p < .05$); in the two-factor model, the latent correlation between CI and PE was $r = .50$ ($SE = 0.05$). However, none of the factor loadings for the residual CI items (the first four items) were significant in the bifactor model. As discussed in the data analysis section, this indicates that there is no residual CI factor because all the common variability in these four items is being accounted for by the general grit factor. Consequently, we also ran the CI:S-1 bifactor model, which poses CI as the reference factor. The conventional bifactor model did not fit significantly better than the CI:S-1 model (G^2: 10,903.95–10,895.76 = 8.19, *df*: 48–44 = 4, $p > .05$). In addition, the latter model had a smaller BIC value and is more parsimonious.

We therefore used both the bifactor and the CI:S-1 models to examine the predictive validity of grit scores. In the bifactor model, the general factor scores represent an individual's level of grit, and the residual PE and CI scores represent an individual's PE

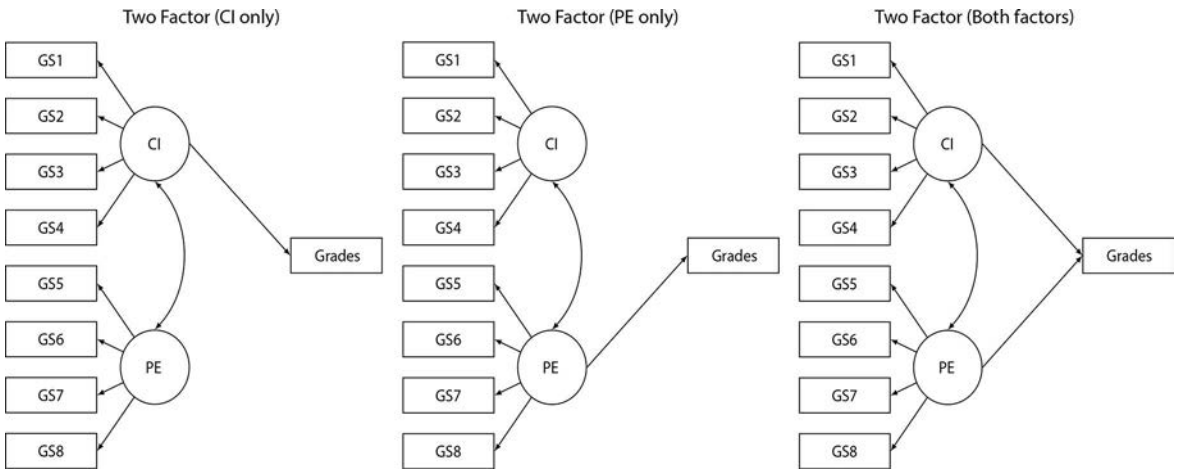


Figure 3. Graphical representations of latent regression models to predict grades in high school sample. CI = consistency of interests; PE = perseverance of effort; GS = Grit-S item.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

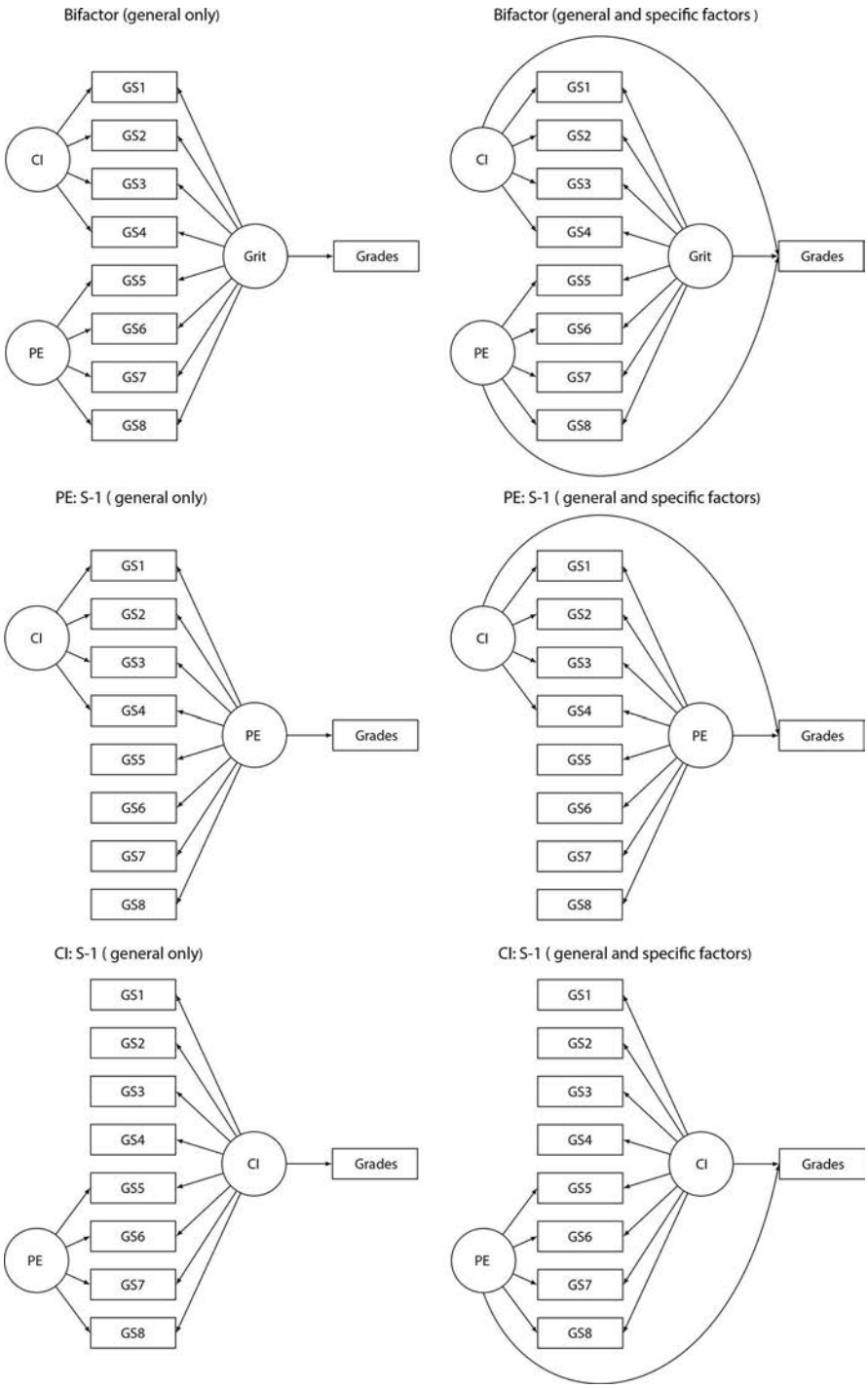


Figure 4. Graphical representation of latent regression models to predict grades in U.S. and Korean college samples. CI = consistency of interests; PE = perseverance of effort; GS = Grit-S item.

or CI level after controlling for that respondent's level of grit. In the CI:S-1 model, the general (reference) factor score represents an individual's CI, and the residual PE scores represents the deviation from the expected PE score based on the individual's CI score. Therefore, comparing how the general (or reference) factor in each of these models predicts student grades provides a clearer understanding of what aspect of grit provides the most predictive power in this sample.

Korean college sample. The bifactor model fit the data significantly better than the one-factor model (G^2: 9983.59–9860.02 = 123.57, *df*: 48–40 = 8, $p < .05$) and the correlated two-factor model (G^2: 9892.40–9860.02 = 32.38, *df*: 48–41 = 7,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Model Fit Indices for the Tested Structural Models of the Grit-S in the High School, U.S. College, and Korean College Samples

Sample	Model	−2LL	AIC	BIC	Pars	RMSEA
High school	One-factor	26,172.71	26,252.71	26,457.67	40	.04
	Two-factors	**25,866.86**	**25,948.86**	**26,158.96**	41	.05
	Bifactor	NC	NC	NC	NC	NC
U.S. college	One-factor	11,245.20	11,325.20	11,499.79	40	.06
	Two-factors	10,990.93	11,072.93	11,251.88	41	.05
	Bifactor	**10,895.76**	**10,991.76**	11,201.27	48	.04
	CI:S-1	10,903.95	10,991.95	**11,184.00**	44	.04
	PE:S-1	10,981.36	11,069.36	11,261.41	44	.05
Korean college	One-factor	9,983.59	10,063.59	1,034.36	40	.05
	Two-factors	9,892.40	9,974.40	10,149.43	41	.05
	Bifactor	**9,860.02**	**9,956.02**	10,160.93	48	.04
	CI:S-1	9,887.75	9,975.75	10,163.59	44	.00
	PE:S-1	9,871.24	9,959.24	**10,147.88**	44	.04

Note. N = 1,242 (high school), 581 (U.S. college), 528 (Korean college). −2LL = −2 * loglikelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion; Pars = number of parameters; RMSEA = root mean square error of approximation based on M_2; NC = not converged; CI = consistency of interests; PE = perseverance of effort. For each sample, the best-fitting model indices are bold.

$p < .05$); in the two-factor model the latent correlation between CI and PE is $r = .64$ ($SE = 0.05$). However, the specific factor loadings of the four items measuring PE were not significant, indicating that there is no residual PE factor because the common variability in these four items is accounted for by the general grit factor. Consequently, we fit the PE:S-1 bifactor model, which poses PE as the reference factor. The conventional bifactor model fit the data significantly better than the PE:S-1 model (G^2: 9871.24–9860.02 = 11.22, *df*: 48–44 = 4, $p < .05$). However, because the specific factor loadings of the four items measuring PE were not significant in the bifactor model, we use both the conventional bifactor and the PE:S-1 models to examine the predictive validity of grit in the analyses testing Hypothesis 3. Doing so provides a clearer understanding of how powerful grit and PE are in predicting students' grades in this sample.

In conclusion, results for the high school sample supported the hypothesis that the correlated two-factor model would fit best for this group. We found partial support that the bifactor model fit best in the college samples; however, a closer examination showed that there are different alternative bifactor models that may be appropriate for the two college samples. In the U.S. sample, grit appears to be represented primarily by CI and in the Korean sample by PE. Taking both the model parsimony and BIC values into account, the CI:S-1 and PE:S-1 fit best in the U.S. college and Korean college samples, respectively. Item parameter estimates for these models appear in the Appendix.

Hypothesis 2: Long Term Grit's Factor Structure

To identify the best fitting factor structure model for long-term grit, we first fit a one-factor model to the eight LT-Grit items and examined local dependence between item pairs; such dependencies could indicate the model is not fitting the data well for certain pairs of items (McDonald, 1982). The standardized *LD* χ^2 values (e.g., larger than 3) identified possible local dependence among items in the three samples, which indicates a violation of the model assumptions that items are unrelated after accounting for the latent factor being measured. In the one-factor models, *LD* χ^2 values between Item 1 and Item 2 were largest (35.4 in the high school sample, 18.4 in the U.S. college sample, and 17.7 in the Korean college sample). Upon review of the item contents, we determined the two items were too similar to each other and one should be removed. We chose to remove the second item from the LT-Grit scale instead of the first item because the *LD* χ^2 values indicated stronger associations between Item 2 and the other items than between Item 1 and the other items. After removing Item 2, the one-factor, two-factor, and bifactor models were fit to the remaining seven items.

High school sample. Both the one-factor and the two-factor models fit the seven-item data well with RMSEA values smaller than 0.01 (see Table 3). Similarly, the bifactor model fit the data well with RMSEA = .02. The likelihood ratio test indicates that the correlated two-factor model did not predict observed item responses significantly better than the one-factor model (G^2: 22,975.31–22,972.41 = 2.90, *df*: 36–35 = 1, $p > .05$). The bifactor model did predict observed item responses better than the one-factor model (G^2: 22,972.41–22,916.93 = 55.48, *df*: 42–35 = 7, $p < .05$). However, none of the specific factor loadings were statistically significant. This finding indicates that although there may be some trivial local dependency among items, the one-factor model is sufficiently capturing the relationship among the items. Thus, we chose to use the one-factor model structure to test Hypothesis 4.

U.S. college sample. Only the one-factor and the bifactor models properly converged, though both had poor absolute fit (RMSEA = .11 and RMSEA = .12). The two-factor model did not converge due to a high correlation between the CI and PE factors ($r = 1.0$); per the likelihood ratio test, the bifactor model predicts observed item response better than the one-factor model (G^2: 9124.80–9106.58 = 18.22, *df*: 42–35 = 7, $p < .05$). However, as in the high school sample, none of the specific factor loadings were statistically significant. Therefore, the one-factor model is used in examining Hypothesis 4.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Model Fit Indices for the Tested Structural Models of the 7-Item LT-Grit for High School, U.S. College, and Korean College Samples

Sample	Model	−2LL	AIC	BIC	Pars	RMSEA
High school	(One-factor	22,972.41	23,042.41	23,221.77	35	.00
	Two-factors	22,975.31	23,047.31	23,231.79	36	.00
	Bifactor	**22,916.93**	**23,000.93**	**23,216.15**	42	.02
U.S. college	One-factor	9,124.80	9,194.80	9,347.56	35	.11
	Two-factors	NC	NC	NC	NC	NC
	Bifactor	**9,106.58**	**9,190.58**	**9,373.90**	42	.12
Korean college	One-factor	7,607.48	**7,677.48**	**7,826.90**	35	.03
	Two-factors	NC	NC	NC	NC	NC
	Bifactor	**7,602.89**	7,686.89	7,866.20	42	.05

Note. $N = 1{,}242$ (high school), 581 (U.S. college), 528 (Korean college). −2LL = −2 * loglikelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion; Pars = number of parameters; RMSEA = root mean square error of approximation based on M_2; NC = not converged; LT-Grit = long-term grit. For each sample, the best-fitting model indices are bold.

Korean college sample. As in the U.S. college sample, the two-factor model did not converge because of the high correlation between the CI and PE factors ($r = 1.0$). The one-factor model and the bifactor model both properly converged and had good fit (RMSEA = .03 and RMSEA = .05). Per the likelihood ratio test, the bifactor model did not predict item responses significantly better than the one-factor model (G^2: 7607.48–7602.89 = 4.59, *df*: 42–35 = 7, $p > .05$), thus the one-factor model is used in examining Hypothesis 4.

Hypothesis 3: Predicting Students' Grades From the Grit-S

High school sample. We used the two-factor structure of Grit-S in the latent regression to predict two separate outcome measures: math and science grades, respectively. The results appear in Table 4. First, to gauge the effect of either CI or PE on each of the subject grades, we regressed each subject grade on each of the two factors (CI and PE). This analysis corresponds to a simple linear regression with a single latent predictor. Both CI ($\hat{\beta} = 0.13$, $SE = 0.04$) and PE ($\hat{\beta} = 0.25$, $SE = 0.04$) predicted math grades; however, the effect size for CI was much smaller. PE was the only significant predictor of science grades ($\hat{\beta} = 0.45$, $SE = 0.04$).

Second, we regressed grades on both factors simultaneously to estimate the effects of CI and PE on the grades while controlling for each other. Students' PE significantly and positively predicted their math grades ($\hat{\beta} = 0.28$, $SE = 0.04$) holding CI constant.

Table 4
Regression Coefficients for Grit-S Prediction of Grades in High School, U.S. College, and Korean College Samples

Sample	Outcome	Model	Factor	$\hat{\beta}$	*SE*
High school	Math grade	Two-factors	CI-only	0.13*	.04
			PE-only	0.25*	.04
			CI	−0.05	.05
			PE	0.28*	.04
	Science grade	Two-factors	CI-only	0.08	.04
			PE-only	0.45*	.04
			CI	−0.13*	.05
			PE	0.34*	.05
U.S. college	Course grade	Bifactor	Grit-only	0.097	.06
			Grit	−0.04	.17
			PE	0.32*	.11
			CI	0.21	.32
		CI:S-1	CI-only	0.08	.05
			CI	0.05	.05
			PE	0.25*	.04
Korean college	GPA	Bifactor	Grit-only	0.21*	.04
			Grit	0.21*	.05
			PE	0.04	.15
			CI	−0.03	.07
		PE:S-1	PE-only	0.21*	.04
			PE	0.22*	.04
			CI	−0.03	.06

Note. N = 1,242 (high school), 581 (U.S. college), 528 (Korean college). $\hat{\beta}$ = standardized regression coefficients; CI = consistency of interests; PE = perseverance of effort; GPA = grade point average. For each model structure, each factor was fit separately and then simultaneously.
* $p < .05$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

However, CI did not significantly predict math grades when controlling for PE. The same pattern occurred for science grades; PE significantly and positively predicted this outcome ($\hat{\beta} = 0.34$, $SE = 0.04$) holding CI constant. It is noteworthy that the path from CI to science grades became significant but negative when we controlled for PE ($\hat{\beta} = -0.13$, $SE = 0.05$). The change in the sign of regression coefficient of CI on the outcome suggests a multicollinearity issue between PE and CI.

We conducted a post hoc analysis to partial out the predictive validity of PE and CI when they are calibrated separately using one-factor models (i.e., the four CI items are used to calculate CI factor scores in one model and the four PE items are used to calculate PE factor scores in a separate model; see Midkiff et al., 2017 for a similar approach). The results, which appear in Table 5, indicate that when CI factor scores are calculated separately from PE factor scores, they do not significantly predict math ($\hat{\beta} = 0.07$, $SE = 0.04$) or science grades ($\hat{\beta} = 0.02$, $SE = 0.04$). Similarly, PE factor scores that have been calculated separately from CI factor scores do significantly predict math ($\hat{\beta} = 0.26$, $SE = 0.04$) and science grades ($\hat{\beta} = 0.29$, $SE = 0.04$). This post hoc analysis demonstrates the predictive power of CI in the correlated two-factor model is a result of the correlation being modeled between CI and PE. Once that correlation is removed and the CI items are modeled separately, CI no longer significantly predicts grades. Thus, overall, PE predicts both math and science grades better than CI.

U.S. college sample. As discussed above, we used both the bifactor and the CI:S-1 models as measurement models in examining predictive validity. First, we fit models predicting grades using only the general factor; then, we fit models predicting grades using the general and the specific factors simultaneously (see Figure 4).

Students' grit did not significantly predict either math or science grades when either was regressed on only the general factor in the conventional bifactor model, ($\hat{\beta} = 0.10$, $SE = 0.06$). When the general grit factor is used in conjunction with residual PE and CI to predict grades, neither general grit ($\hat{\beta} = -0.04$, $SE = 0.17$) nor CI ($\hat{\beta} = 0.21$, $SE = 0.32$) predict either math or science grades. The only significant predictor of course grades was residual PE ($\hat{\beta} = 0.74$, $SE = 0.09$), after controlling for both general grit and CI.

In the alternative CI:S-1 model where the reference factor represents CI scores and the residual factor represents the deviation of individuals' PE scores from their expected score based on their CI score, neither the general nor residual factor predicted students' math or science grades. CI did not significantly predict students' math or science grades when it was the only predictor in the model ($\hat{\beta} = 0.08$, $SE = 0.05$) or after controlling for residual PE ($\hat{\beta} = 0.05$, $SE = 0.05$). As with the conventional bifactor model. The only significant predictor of math and science grades was residual PE ($\hat{\beta} = 0.25$, $SE = 0.04$), after controlling for CI.

These two sets of analyses, using the bifactor model and the alternative bifactor model (CI:S-1), indicate that the variance of the general factor in the bifactor model is dominated by communalities among the CI items (rather than PE items) and therefore does not predict course grades. This conjecture is empirically confirmed by the alternative model CI:S-1, in which CI is not clearly associated with the grades and only residual PE predicts the grades. Accordingly, we can conclude that the role of CI is very minimal in predicting grades in the sample of U.S. college students.

Korean college sample. We used both the conventional bifactor and PE:S-1 models in the predictive analyses. Using the conventional bifactor model, we predicted students' GPAs first using only the general factor and then using both the general and the specific factors (see Figure 4). The general grit factor significantly predicted students' GPAs when it was the only predictor in the model ($\hat{\beta} = 0.21$, $SE = 0.04$) as well as after controlling for the residual PE and CI factors ($\hat{\beta} = 0.21$, $SE = 0.05$). Neither residual PE ($\hat{\beta} = 0.04$, $SE = 0.15$) nor CI ($\hat{\beta} = -0.03$, $SE = 0.07$) were significant predictors of GPA when we controlled for the general factor grit.

The latent predictive analyses based on the PE:S-1 model showed that the general PE factor significantly predicted students' GPAs both on its own ($\hat{\beta} = 0.21$, $SE = 0.04$) and after controlling for the specific CI factor ($\hat{\beta} = 0.22$, $SE = 0.05$). The specific CI factor was not a significant predictor of students' GPAs after controlling for PE ($\hat{\beta} = -0.03$, $SE = 0.06$).

In sum, results supported our prediction that once we controlled for PE, CI does not significantly predict grades, making PE the strongest predictor in both U.S. high school and college samples. However, it is noteworthy that the general factor of grit was the only significant predictor of grades in the Korean college sample. In this sample, however, the general factor is more dominated by the PE items.

Table 5
Regression Coefficients for CI and PE Prediction of Grades in the High School Sample Using One-Factor Models

	Math			Science		
Construct	$\hat{\beta}$	*SE*	*p*	$\hat{\beta}$	*SE*	*p*
CI	0.074	0.041	.069	0.024	0.040	.545
PE	0.260	0.036	<.001	0.286	0.036	<.001

Note. $N = 1{,}242$. $\hat{\beta}$ = standardized regression coefficients; CI = consistency of interests; PE = perseverance of effort.

Hypothesis 4: Predicting Students' Grades From LT-Grit

In the three samples, the unidimensional factor structure was used for long-term grit in the latent regression to predict grades (see Figure 5). The results appear in Table 6. Students' long-term grit significantly predicted their math ($\hat{\beta} = 0.25$, $SE = 0.03$) and science ($\hat{\beta} = 0.20$, $SE = 0.04$) grades in the U.S. high school sample, and significantly predicted GPA in the Korean college sample ($\hat{\beta} = 0.21$, $SE = 0.04$). However, it did not significantly predict either math or science course grades in the U.S. college sample ($\hat{\beta} = 0.05$, $SE = 0.04$). Therefore, Hypothesis 4 is supported only in the U.S. high school and Korean college samples.

Hypothesis 5: Relationship Between Grit and Long-Term Grit

The measurement model used for grit was the correlated two-factor model in the high school sample and the bifactor model in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

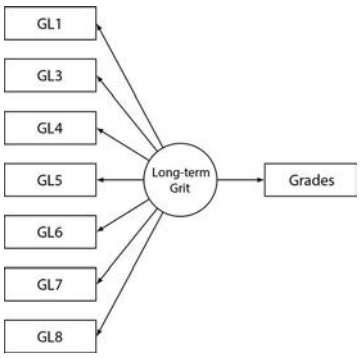


Figure 5. Graphical representation of latent regression model to predict grades from long-term grit. GL = long-term grit item.

the U.S. and Korean college samples. The unidimensional factor structure was used for long-term grit in the three samples. Table 7 shows the latent correlations for the three models. In the high school sample, long-term grit had a moderate positive correlation with CI (r = .65, SE = .04) and a strong positive correlation with PE (r = .83, SE = .03). There was a very strong positive relationship between grit and long-term grit in the U.S. college (r = .99, SE = .02) and Korean college (r = .98, SE = .03) samples.

Additional Correlation Analyses to Understand Grit Scores in the Two College Samples

To illustrate further both the make-up of the general grit factor scores in the U.S. and Korean college samples, as well as the relationship between grit and long-term grit scores, we calculated scores for general grit, CI, and PE using three methods (summed scores, factor scores from the bifactor model, and factor scores from the alternative bifactor models) and scores for general long-term grit using summed scores and factor scores from the one-factor model, and then examined correlations among the scores. The purpose of these correlational analyses is to assess whether the meaning of the total scores for grit based on summing responses to all the items that many researchers have used are inconsistent across samples. If that is indeed the case, it suggests that this inconsistency is one mechanism that helps explain the conflicting predictive validity of total grit scores across samples. We also note that these correlation analyses were possible only after we applied the bifactor model and S-1 bifactor models for each sample.

Table 6
Regression Coefficients for LT-Grit Prediction of Grades in High School, U.S. College, and Korean College Samples

Sample	Outcome	Factor	$\hat{\beta}$	*SE*
High school	Math grade	Grit	0.25*	.03
	Science grade	Grit	0.20*	.04
U.S. college	Course grade	Grit	0.05	.04
Korean college	GPA	Grit	0.12*	.04

Note. N = 1,242 (high school), 581 (U.S. college), 528 (Korean college). $\hat{\beta}$ = standardized regression coefficients; CI = consistency of interests; PE = perseverance of effort; GPA = grade point average; LT-Grit = long-term grit.
* $p < .05$.

Table 7
Latent Correlations Between Grit as Measured by Grit-S and Long-Term Grit as Measured by LT-Grit for the High School, U.S. College, and Korean College Samples

	High school				U.S. college		Korean college	
	CI		PE		Grit		Grit	
Construct	*r*	*SE*	*r*	*SE*	*r*	*SE*	*r*	*SE*
Long-term grit	0.65	.04	0.83	.03	0.99	.04	0.98	.03

We created the summed scores by summing students' responses to the general grit, CI and PI items. Factor scores were calculated using the expected a posteriori (EAP) method. A score was calculated for general grit, residual CI and residual PE using the conventional bifactor model in both samples. Factor scores for general CI and residual PE were calculated using the CI:S-1 model for the U.S. college sample, and factor scores for general PE and residual CI were calculated using the PE:S-1 model for the Korean college sample. Similarly, a summed score was calculated for the seven-item LT-Grit and a factor score was calculated using the EAP method with the one-factor model. The correlations among these scores appear in Tables 8 and 9.

The analyses of the summed scores show the following correlations in the U.S. sample ($r_{Grit,CI}$ = .86; $r_{Grit,PE}$ = .84; $r_{CI,PE}$ = .45) and in the Korean sample ($r_{Grit,CI}$ = .86; $r_{Grit,PE}$ = .84; $r_{CI,PE}$ = .45). These results might lead to the conclusion that the relationships among grit, CI, and PE, are similar in the two samples. However, when examining the relationships between EAP factor scores and summed scores, the difference in the makeup of grit in these two samples becomes clear. In the U.S. college sample, grit factor scores from the bifactor model are very strongly correlated with grit summed scores (r = .94), CI summed scores (r = .96), and general CI factor scores from the CI:S-1 model (r = .99) indicating that these three scores are essentially measuring the same construct. Similarly, factor scores from the CI:S-1 model are also highly correlated with grit summed scores (r = .93) and with CI summed scores (r = .97). Furthermore, the factor scores for residual PE from the bifactor model are highly correlated with PE summed scores (r = .76), which is as we would expect when the general grit factor is only accounting for the variability in items related to CI, leaving the variability in PE items to be measured by the residual PE factor.

By contrast, in the Korean college sample the general grit scores from the bifactor model are very highly correlated with the grit summed scores (r = .91), PE summed scores (r = .93), and with general PE factor scores from the PE:S-1 model (r = .99) indicating these three scores are capturing the same construct. Furthermore, the factor scores for general PE in the PE:S-1 model are strongly correlated with grit summed scores (r = .90) and PE summed scores (r = .93), and factor scores for the residual CI factor in the bifactor model are highly correlated with CI summed scores (r = .85). These correlations indicate general grit is actually measuring PE in the Korean college sample and not capturing the CI aspect of grit.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 8
Correlations Between Grit-S Summed Scores, Bifactor Model Factor Scores, CI:S-1 Model Factor Scores, and LT-Grit Summed Scores and One-Factor Model Factor Scores for the U.S. College Sample

Factor	1	2	3	4	5	6	7	8	9	10
1. $Grit_{Sum}$										
2. CI_{Sum}	.86									
3. PE_{Sum}	.84	.45								
4. LT_{Sum}	.72	.55	.68							
5. $Grit_{BF}$	.91	.63	.93	.66						
6. CI_{BF}	.51	.85	**.01**	.06	.20					
7. PE_{BF}	.23	**.01**	.40	.40	.25	**−.09**				
8. CI_{S-1}	.52	.86	**.01**	.64	.21	.99	−.13			
9. PE_{S-1}	.90	.60	.93	.42	.99	.18	.35	.18		
10. LT_{One}	.70	.53	.65	.94	.65	**.05**	.38	.63	.39	

Note. N = 581. CI =consistency of interests; PE = perseverance of effort; LT-Grit = long-term grit; $Grit_{Sum}$ = summed score for eight Grit-S items; CI_{Sum} = summed score for four CI items; PE_{Sum} = summed score for four PE items; LT_{Sum} = summed scores for seven LT-Grit items; $Grit_{BF}$ = facto score for general grit factor in bifactor model; CI_{BF} = factor score for residual CI factor in bifactor model; PE_{BF} = factor score for residual PE factor in bifactor model; CI_{S-1} = factor score for general CI factor in S-1:CI model; PE_{S-1} = factor score for residual PE factor in S-1:CI model; LT_{One} = factor score for long-term grit factor in one-factor model. Factor scores are calculated using the expected a posteriori (EAP) method. Nonsignificant correlations appear in bold.

Thus, these analyses show that total grit scores, whether calculated by summing individuals' item responses or using factor scores calculated from a bifactor model, are not measuring the same construct in these two college samples. General grit scores are measuring CI in the U.S. college sample and measuring PE in the Korean college sample. Therefore, any attempt to compare these samples on general grit (i.e., mean differences) would be inappropriate and misleading because a U.S. college student's grit score and a Korean college student's grit score do not represent the same construct.

The results of the correlation analyses with LT-Grit scores indicate a similar difference in the makeup of LT-Grit scores across U.S. and Korean college students. In the U.S. college sample, LT-Grit EAP scores were more strongly correlated with the CI factor (r = .63) than with the PE factor (r = .39) in the CI:S-1 bifactor model, whereas in the Korean college sample, LT-Grit EAP scores were more strongly correlated with the PE factor (r = .66) than with the CI factor (r = .27) in the S-1 bifactor model. These results suggest LT-Grit is largely measuring long-term CI in U.S. college students and largely measuring long-term PE in Korean college students. This difference in the makeup of LT-Grit across these two samples may also explain why LT-Grit significantly predicts grades in the Korean but not the U.S. college sample.

Table 9
Correlations Between Grit-S Summed Scores, Bifactor Model Factor Scores, CI:S-1 Model Factor Scores, and LT-Grit Summed Scores and One-Factor Model Factor Scores for the Korean College Sample

Factor	1	2	3	4	5	6	7	8	9	10
1. $Grit_{Sum}$										
2. CI_{Sum}	.86									
3. PE_{Sum}	.84	.45								
4. LT_{Sum}	.67	.57	.58							
5. $Grit_{BF}$	.91	.63	.93	.67						
6. CI_{BF}	.51	.85	**.01**	.26	.20					
7. PE_{BF}	.23	**.01**	.40	.03	.25	**.09**				
8. PE_{S-1}	.52	.86	**.01**	.65	.21	.99	−.13			
9. CI_{S-1}	.90	.60	.93	.27	.99	.18	.35	.18		
10. LT_{One}	.65	.55	.56	.92	.68	.26	.03	.66	.27	

Note. N = 581. CI = consistency of interests; PE = perseverance of effort; LT-Grit = long-term grit; $Grit_{Sum}$ = summed score for eight Grit-S items; CI_{Sum} = summed score for four CI items; PE_{Sum} = summed score for four PE items; LT_{Sum} = summed scores for seven LT-Grit items; $Grit_{BF}$ = factor score for general grit factor in bifactor model; CI_{BF} = factor score for residual CI factor in bifactor model; PE_{BF} = factor score for residual PE factor in bifactor model; CI_{S-1} = factor score for general CI factor in S-1:CI model; PE_{S-1} = factor score for residual PE factor in S-1:CI model; LT_{One} = factor score for long-term grit factor in one-factor model. Factor scores are calculated using the expected a posteriori (EAP) method. Nonsignificant correlations appear in bold.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Discussion

This study provides more definitive answers than previous work to Credé's (2018) question concerning what to do about grit given increasing skepticism about both the nature of the construct and its measurement. First, our results regarding grit's factor structure show that its purported higher-order factor structure reported by Duckworth and colleagues and others (Duckworth & Quinn, 2009; Wolters & Hussain, 2015) is not only incorrect in a statistical sense given its mathematical equivalence to a correlated two-factor structure, but also an inappropriate conceptualization of grit as a construct made up of two distinct subfactors. Second, we found that grit's factor structure differed across age and culture, meaning that the Grit-S is measuring different constructs in different groups. Third, the PE subcomponent of grit predicted students' grades more strongly than the CI subcomponent. The other strong contributions of this study are its methodological advances with respect to the statistical analyses used to answer the hypotheses.

The Factor Structure of the Grit-S: Variability Across Age and Culture

The CFAs showed that in the American high school sample, the correlated two-factor model fit best. By contrast, in both the U.S. and Korean college samples (and in the college sample in Muenks et al.'s (2017) study) bifactor models best represented grit's factor structure. Thus, for college students' grit, as measured by the Grit-S, is one entity. The results provide support for Sansone et al.'s (2015) proposal that self-regulation (analogous to some degree with perseverance of effort) and interest relate more closely to one another as individuals' interests mature, which is more likely in college students than high school students. Researchers should study potential age differences in grit's factor structure in samples including wider age ranges to examine further its developmental course.

Concerning cultural differences, in both the U.S. and Korean college samples the bifactor model fit best. In that sense grit as measured by the Grit-S is similar in its overall factor structure in the two cultures. However, our additional analyses of the bifactor models using Eid et al.'s (2017) approach unearthed some interesting differences with respect to the meaning of the general factor in each culture. In the U.S. sample, the general grit factor was primarily influenced by the CI items, and the PE items made up a secondary factor. By contrast, the general grit factor in the Korean college sample was predominately impacted by PE and not CI. These results suggest that the meaning of the overall grit construct varies across culture. Both general and more specific differences between the two cultures are likely explanations for these differences. At the general level, the different structures could stem in part from Western cultures emphases on individualism and Eastern cultures collectivism (Hofstede, 1980; Markus & Kitayama, 1991). In the former, pursuing one's own interests is a key aspect of development and much research suggests many parents have this as a goal (Wigfield et al., 2015). Thus, developing consistency in one's interests may be a crucial part of grit in western cultures.

Collectivist societies tend to promote the pursuit of group goals (Kwan, Bond, & Singelis, 1997; Markus & Kitayama, 1991). Therefore, as also posited by Datu et al. (2016), CI may not be a contributing factor to a person's "grittiness" because it is more critical that one's interests and values align with one's group. Turning to more specific cultural differences, the "education fervor" focus in Korean culture (Song & Chung, 2020) discussed earlier also provides a further explanation. Korean children are socialized to work very hard and succeed in school, perhaps meaning that PE becomes central to them even irrespective of what subjects at school interest them. Of course, many parents in Western cultures emphasize the importance of school success, but they often do so with respect to finding one's passion rather than just persevering through.

Finally, another explanation is that the word *grit* is defined differently in English and Korean. The English dictionary defines *grit* as the "firmness of mind or spirit," whereas the word *grit* is defined in the Korean dictionary as "patience, perseverance, persistence, or endurance." The Korean definition actually is closer to Duckworth et al.'s (2007) original definition than is the English definition. Such differences in meaning may reflect broader value systems of a country can lead to differences in how individuals in one culture respond to measures developed in another culture. Finding such culturally based configural differences in a specific proposed personality facet measured by a brief scale shows the importance of doing cross-cultural work on the meaning of grit and other constructs defined and studied in the United States in order to better understand how that construct is defined in various cultures.

Factor Structure of the LT-Grit Scale

We designed the LT-Grit measure to improve on the Grit-S's measurement of the long term and "passion" aspects of grit, as well as deal with the confounding of construct and item valence. A one-factor solution fit best for this measure, in all three samples. As further evidence, in all three samples the correlation between long-term CI and long-term PE was almost 1.00. One possible explanation for these findings is that to persevere in reaching a goal over the long term, it is necessary to have interest in that goal over the same long-term period, making CI and PE indistinguishable constructs. We also did not find a negative wording effect in our factor analyses of LT-Grit. Finally, it is possible that introducing "six months" into each item may have produced more common variance among item responses. However, because this phrase was included in both the CI and PE items it still is plausible that a two-factor model could have emerged due to the different foci of each set of items- but it did not. We posit that the two-factor structure of grit found in studies using the Grit-S easily could be due to the item valence rather than construct differences, further weakening Duckworth and colleagues' (Duckworth et al., 2007; Duckworth & Quinn, 2009) supposition that grit is comprised of two subcomponents.

Relations of Grit, PE, CI, and LT-Grit to Students' Grades

Despite Duckworth's recent claims that the grit measures were not meant to be strongly predictive of grades (Barshay, 2019), she and her colleagues found that grit related positively to college students' grades. As we argued above, end of semester grades are critical long-term outcomes in students' lives, which should make being gritty relevant to school success as measured by grades. Like re-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

searchers both in the United States and other cultures, we found that the proposed PE component of grit is a much stronger predictor of grades than is CI (e.g., Credé et al., 2017; Datu et al., 2016; Muenks et al., 2017; Muenks et al., 2018; Steinmayr et al., 2018). This work, coupled with our factor analytic work, suggests that researchers should focus on PE rather than either CI or total grit scores when academic achievement is the outcome of interest; Credé (2018) made the same suggestion.

Our results regarding the predictive validity of long-term grit varied across the three samples. Students' long-term grit significantly predicted both their math and science grades among high school students as well as overall GPA in Korean college students, but not math and science course grades in U.S. college students. In the American college sample, grades were from different math or science courses at different levels of difficulty, whereas the Korean college grades were overall GPA and most of the students were education majors. The different grades available in each sample could have produced the different results. Furthermore, our factor analytic results showed that LT-Grit scores in the U.S. college sample largely reflect CI, which is not predictive of grades.

Methodological Advances and Their Implications for Future Research on Grit and Related Constructs

We believe the inconsistent findings regarding grit's factor structure and relations to outcomes reported in previous studies are at least partially attributable to incorrect and limited analytical approaches different researchers have applied as well as to the different kinds of factor analyses researchers have used. With regard to the first point, as noted earlier the fact that Duckworth and colleagues and others (Datu et al., 2016; Duckworth et al., 2007; Duckworth & Quinn, 2009; Wolters & Hussain, 2015) did not understand, or report, that higher-order models with two subcomponents are mathematically equivalent to correlated two-factor models led to incorrect conclusions about grit's factor structure. Regarding the second point, some researchers (e.g., Steinmayr et al., 2018) used EFA and others used CFA to explore grit's factor structure, and in the CFA models different structures have been proposed (e.g., Midkiff et al., 2017; Muenks et al., 2018; Wolters & Hussain, 2015). These inconsistencies have made it quite difficult to compare results across studies.

Another important methodological takeaway from our analyses is the importance of examining more than just fit indices when comparing models. The factor loadings, their standard errors, and significance helped us to determine whether they were theoretically meaningful. The examination of factor loadings also led us to test alternative bifactor models rather than just the initial one, which provided greater clarity about the general factor grit in the two college samples. Further, researchers often use summed scores to create factor scores; had we done that instead of using the MIRT approach, we would not have been able to demonstrate how the content of overall grit varied across our college samples.

As noted above, most researchers studying grit and other related constructs treat their respondents' data as continuous rather than categorical. The assumption that item responses are continuous is incorrect because Likert scales are ordinal by nature, meaning that the distance between response choices is not necessarily identical. Item responses should be treated as categorial, especially when there are fewer than 11 categories on the response scale (Hodge & Gillespie, 2007; Leung, 2011; Wu & Leung, 2017). Treating item responses as continuous rather than categorical also can mask the appropriate factor structure. For example, Lim (2017) reported in her study of Korean college students that the bifactor model estimation converged when responses were treated as continuous; however, the factor loadings were smaller than what we found when we treated our data as categorical. In fact, based on results from their simulation studies, Rhemtulla et al. (2012) warns against treating Likert-scaled data as continuous when conducting factor analyses. We recommend that researchers test the assumption that the data they obtain from surveys is continuous before treating it as such.

Finally, our study used the "gold standard" simultaneous estimation approach to latent variable modeling (Bartholomew & Knott, 1999; Muthén, 2002; Rabe-Hesketh & Skrondal, 2004). The alternative two-stage approach in which factor scores are estimated using one model and the factor score estimates are done next is problematic because the uncertainty in the factor scores, due to sampling variability estimated in the first stage, is not properly accounted for in the second stage. Furthermore, the factor scores obtained from the first stage (e.g., EAP scores) regress toward the mean and consequently weaken the relation between the factor scores and other variables in multistage estimation. Therefore, we recommend that researchers use the simultaneous estimation approach in future studies of grit and other related variables.

Limitations and Future Directions

Limitations include first that the high school sample predominately consisted of students from private schools, making generalizations to the broader population of high school students perhaps inappropriate. The same point holds true for the two college samples. More research is needed on grit's factor structure in different kinds of samples to see if our results generalize.

Second, we only examined the direct relations of grit, PE, and CI to academic achievement outcomes. However, researchers (e.g., Muenks et al., 2017; Muenks et al., 2018; Steinmayr et al., 2018) found that several motivational and self-regulation variables mediate or moderate the direct relationship of grit to achievement. Future research should include such moderators and mediators when examining how strongly grit, PE, or CI predict student achievement outcomes in different groups.

More research is needed on the LT-Grit scale, to assess its reliability and validity across samples. Researchers also should determine whether the LT-Grit items should be incorporated into the Grit-S, replace it, or whether additional long-term items should be developed. The correlations of the two measures are relatively high, which means merging the two scales could create a more reliable measure. However, the way we defined "long term" may need to be changed in different age groups. The LT-Grit items were developed specifically for use with U.S. high school students in which six months is considered an appropriate definition of "long term." It is likely that for other populations (e.g., college students and adults), six months is not an appropriate way to define "long term". Accordingly, we do not recommend the use of the LT-Grit instrument for populations other than U.S. high school students without further validation work.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In conclusion, we believe this study provides definitive answers to Credé's (2018) question "what to do about grit." We have five recommendations for moving forward. First, grit is not a higher-order construct with two subcomponents and so should not be treated as such. Second, grit has different meanings at different ages and in different cultures. Researchers examining grit in different cultures should look carefully at how grit, PE, and CI are defined in those cultures before attempting to measure them. Third, the PE items are what carry the predictive weight when looking at academic achievement, and so total grit scores should not be the only score used for that purpose in future studies of grit's relations to academic achievement. Fourth, this study began to address the mismatch of grit's definition as "long-term" passion and perseverance with the lack of these terms in the most widely used grit measures; more research is needed to determine what both "long-term" and "passion" mean at different ages and in different cultures. Finally, we hope researchers, educational policymakers, and intervention scientists become increasingly skeptical of Duckworth and colleagues' (e.g., Duckworth et al., 2019) continuing claims about grit's importance in accomplishing challenging tasks like doing well in STEM and other coursework, given the results of the present study and a growing number of others.

References

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16,* 3–14. http://dx.doi.org/10.1016/0304-4076(81)90071-3

Barshay, J. (2019). Research scholars to air problems with using 'grit' at school. *The Hechinger Report, March 11*. Retrieved from https://hechingerreport.org/research-scholars-to-air-problems-with-using-grit-at-school/

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis, Kendalls, Library of Statistics* (Vol. 7). New York, NY: Edward Arnold.

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22,* 507–526. http://dx.doi.org/10.1037/met0000077

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46,* 443–459.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17,* 303–316. http://dx.doi.org/10.1177/0049124189017003004

Bowman, N. A., Hill, P. L., Denson, N., & Bronkema, R. (2015). Keep on truckin' or stay the course? Exploring grit dimensions as differential predictors of educational achievement, satisfaction, and intentions. *Social Psychological and Personality Science, 6,* 639–645. http://dx.doi.org/10.1177/1948550615574300

Brown, T. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Cai, L. (2017). flexMIRT R version 3.51: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods, 16,* 221–248. http://dx.doi.org/10.1037/a0023350

Chang, W. (2014). *Grit and academic performance: Is being grittier better?* (Order No. 3681407; Doctoral dissertation, University of Miami). Open Access Dissertations.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22,* 265–289. http://dx.doi.org/10.3102/10769986022003265

Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality, 69,* 979–1006. http://dx.doi.org/10.1111/1467-6494.696172

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, UK: Routledge.

Credé, M. (2018). What shall we do about grit? A critical review of what we know and what we don't know. *Educational Researcher, 47,* 606–611. http://dx.doi.org/10.3102/0013189X18801322

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

Cross, T. M. (2014). The gritty: Grit and non-traditional doctoral student success. *The Journal of Educators Online*. Advance online publication. http://dx.doi.org/10.9743/jeo.2014.3.4

Datu, J. A. D., Valdez, J. P. M., & King, R. B. (2016). Perseverance counts but consistency does not! Validating the Short Grit Scale in a collectivist setting. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 35,* 121–130. http://dx.doi.org/10.1007/s12144-015-9374-2

Datu, J. A. D., Yuen, M., & Chen, G. (2017). Development and validation of the Triarchic Model of Grit Scale (TMGS): Evidence from Filipino undergraduate students. *Personality and Individual Differences, 114,* 198–205. http://dx.doi.org/10.1016/j.paid.2017.04.012

Datu, J. A. D., Yuen, M., & Chen, G. (2018). The triarchic model of grit is linked to academic success and well-being among Filipino high school students. *School Psychology Quarterly, 33,* 428–438. http://dx.doi.org/10.1037/spq0000234

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

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

Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., & Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *Proceedings of the National Academy of Sciences of the United States of America, 116,* 23499–23504. http://dx.doi.org/10.1073/pnas.1910510116

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22,* 541–562. http://dx.doi.org/10.1037/met0000083

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press press.

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,* 36. http://dx.doi.org/10.3389/fpsyg.2014.00036

Finch, H. (2011). Multidimensional item response theory parameter estimation with nonsimple structure items. *Applied Psychological Measurement, 35,* 67–82.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16,* 625–641. http://dx.doi.org/10.1080/10705510903203573

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57,* 423–436. http://dx.doi.org/10.1007/BF02295430

Han, K. C. T., & Paek, I. (2014). A review of commercial software packages for multidimensional IRT modeling. *Applied Psychological Measurement, 38,* 486–498. http://dx.doi.org/10.1177/0146621614536770

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41,* 111–127. http://dx.doi.org/10.1207/s15326985ep4102_4

Hodge, D. R., & Gillespie, D. F. (2007). Phrase completion scales: A better measurement approach than Likert Scales? *Journal of Social Service Research, 33,* 1–12. http://dx.doi.org/10.1300/J079v33n04_01

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization, 10,* 15–41.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3–4), 117–144. http://dx.doi.org/10.1080/03610739208253916

Hwang, M. H., Lim, H. J., & Ha, H. S. (2018). Effects of grit on the academic success of adult female students at Korean Open University. *Psychological Reports, 121,* 705–725. http://dx.doi.org/10.1177/0033294117734834

Jakobsson, U. (2004). Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences, 18,* 437–440. http://dx.doi.org/10.1111/j.1471-6712.2004.00305.x

Jamieson, S. (2004). Likert Scales: How to (ab)use them. *Medical Education, 38,* 1217–1218. http://dx.doi.org/10.1111/j.1365-2929.2004.02012.x

Joe, H., & Maydeu-Olivares, A. (2010). A general family of limited information goodness-of-fit statistics for multinomial data. *Psychometrika, 75,* 393–419. http://dx.doi.org/10.1007/s11336-010-9165-5

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal, 15,* 136–153.

Kim, H.-M., & Hwang, M.-H. (2015). Validation of the Korean Grit Scale for Children. *Journal of Education for Business, 35,* 63–74.

Knapp, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research, 39,* 121–123. http://dx.doi.org/10.1097/00006199-199003000-00019

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26,* 457–477.

Kohli, N., Koran, J., & Henn, L. (2015). Relationships among classical test theory and item response theory frameworks via factor analytic models. *Educational and Psychological Measurement, 75,* 389–405.

Kwan, V. S., Bond, M. H., & Singelis, T. M. (1997). Pancultural explanations for life satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology, 73,* 1038–1051.

Lee, S., & Sohn, Y. W. (2017). Effects of grit on academic achievement and career-related attitudes of college students in Korea. *Social Behavior and Personality: An International Journal, 45,* 1629–1642. http://dx.doi.org/10.2224/sbp.6400

Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research, 37,* 412–421. http://dx.doi.org/10.1080/01488376.2011.580697

Li, J., Lin, L., Zhao, Y., Chen, J., & Wang, S. (2018). Grittier Chinese adolescents are happier: The mediating role of mindfulness. *Personality and Individual Differences, 131,* 232–237. http://dx.doi.org/10.1016/j.paid.2018.05.007

Li, J., Zhao, Y., Kong, F., Du, S., Yang, S., & Wang, S. (2018). Psychometric assessment of the Short Grit Scale among Chinese adolescents. *Journal of Psychoeducational Assessment, 36,* 291–296. http://dx.doi.org/10.1177/0734282916674858

Lim, H. J. (2017). An exploratory study on grit's factor structure and its validity. *Asian Journal of Education, 18,* 169–192. http://dx.doi.org/10.15753/aje.2017.06.18.2.169

Linn, M. C., & Eylon, B. S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 511–544). Mahwah, NJ: Erlbaum Publishers.

Lord, F. M., Tukey, J. W., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Pub. Co.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224–253. http://dx.doi.org/10.1037/0033-295X.98.2.224

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22,* 37–49. http://dx.doi.org/10.1037/0012-1649.22.1.37

McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons. *Journal of Personality, 69,* 819–846. http://dx.doi.org/10.1111/1467-6494.696166

McDonald, R. P. (1982). Linear versus models in item response theory. *Applied Psychological Measurement, 6,* 379–396.

McDonald, R. P. (1997). Normal-ogive multidimensional model. *Handbook of Modern Item Response Theory* (pp. 257–269). New York, NY: Springer.

Midkiff, B., Langer, M., Demetriou, C., & Panter, A. T. (2017). Measuring Grit among first-generation college students: A psychometric analysis. In L. A. van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas, & W. C. Wang (Eds.), *The annual meeting of the psychometric society* (pp. 407–420). Cham, Switzerland: Springer.

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

Muenks, K., Yang, J. S., & Wigfield, A. (2018). Associations between grit, motivation, and achievement in high school students. *Motivation Science, 4,* 158–176. http://dx.doi.org/10.1037/mot0000076

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29,* 81–117.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Author.

Nieto, M. D., Abad, F. J., & Olea, J. (2018). Assessing the Big Five with bifactor computerized adaptive testing. *Psychological Assessment, 30,* 1678–1690. http://dx.doi.org/10.1037/pas0000631

Nishikawa, K., Okugami, S., & Amemiya, T. (2015). Development of the Japanese Grit Scale (Grit-S). *The Japanese Journal of Personality, 24,* 167–169. http://dx.doi.org/10.2132/personality.24.167

Nye, C. D., Roberts, B. W., Saucier, G., & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality, 42,* 1524–1536. http://dx.doi.org/10.1016/j.jrp.2008.07.004

Rabe-Hesketh, S., & Skrondal, A. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman and Hall/CRC.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92,* 544–559. http://dx.doi.org/10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*(Suppl. 1), 19–31. http://dx.doi.org/10.1007/s11136-007-9183-7

Reise, S. P., & Revicki, D. A. (Eds.). (2014). *Handbook of item response theory modeling: Applications to typical performance assessment*. New York, NY: Routledge. http://dx.doi.org/10.4324/9781315736013

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

conditions. *Psychological Methods*, *17*, 354–373. http://dx.doi.org/10.1037/a0029315

Rojas, J. P., & Usher, E. L. (2013). *Exploring correlations among creativity, grit, and mathematics achievement in socioeconomically diverse schools*. Lexington, KY: University of Kentucky.

Sansone, C., Thoman, D., & Fraughton, T. (2015). The relation between interest and self- regulation in mathematics and science. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in K-16 mathematics and science learning* (pp. 111–132). Washington, DC: American Educational Research Association.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. http://dx.doi.org/10.1214/aos/1176344136

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

Shin, K., Jahng, K. E., & Kim, D. (2019). Stories of South Korean mothers' education fever for their children's education. *Asia Pacific Journal of Education*, *39*, 338–356. http://dx.doi.org/10.1080/02188791.2019.1607720

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–575. http://dx.doi.org/10.1007/BF02296196

Song, J., & Chung, Y. (2020). Reexamining the interaction between expectancy and task value in academic settings. *Learning and Individual Differences*, *78*, 101839. http://dx.doi.org/10.1016/j.lindif.2020.101839

Steinmayr, R., Weidinger, A., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology*, *53*, 106–122. http://dx.doi.org/10.1016/j.cedpsych.2018.02.004

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

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.

Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. New York, NY: Houghton Mifflin Harcourt.

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality*, *69*, 907–924. http://dx.doi.org/10.1111/1467-6494.696169

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486–492. http://dx.doi.org/10.1080/17405629.2012.686740

Van de Vijver, F. J., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality*, *69*, 1007–1031. http://dx.doi.org/10.1111/1467-6494.696173

Wigfield, A., Eccles, J. S., Fredericks, J., Roeser, R., Schiefele, U., & Simpkins, S. (2015). Development of achievement motivation and engagement. In R. Lerner (Series Ed.) & C. Garcia Coll & M. Lamb (Volume Eds.). *Handbook of child psychology: Vol. 3. Social and emotional development* (7th ed., pp. 657–700). New York, NY: Wiley. http://dx.doi.org/10.1002/9781118963418.childpsy316

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79. http://dx.doi.org/10.1037/1082-989X.12.1.58

Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relation with college students' self-regulation learning and academic achievement. *Metacognition and Learning*, *10*, 293–311. http://dx.doi.org/10.1007/s11409-014-9128-9

Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales? A simulation study. *Journal of Social Service Research*, *43*, 527–532. http://dx.doi.org/10.1080/01488376.2017.1329775

Appendix

Item Parameter Estimates for Grit-S and LT-Grit

Slope and Intercept Parameter Estimates for Grit-S

Sample	Item	Slopes			Intercepts			
		Grit	CI	PE	1	2	3	4
U.S. high school two-factor model	Item1		0.97 (.08)		2.58 (.11)	0.87 (.08)	−0.79 (.07)	−2.67 (.12)
	Item 2		1.19 (.10)		3.77 (.18)	1.87 (.10)	−0.27 (.07)	−2.39 (.11)
	Item 3		1.38 (.10)		2.88 (.13)	1.44 (.09)	−0.67 (.08)	−3.00 (.15)
	Item 4		1.51 (.12)		2.48 (.13)	0.51 (.09)	−1.77 (.11)	−3.45 (.16)
	Item 5			1.25 (.09)	5.06 (.28)	3.01 (.14)	0.90 (.08)	−0.95 (.08)
	Item 6			0.79 (.07)	2.92 (.13)	1.36 (.08)	−0.18 (.07)	−1.63 (.09)
	Item 7			1.91 (.14)	6.47 (.44)	3.90 (.21)	1.07 (.10)	−1.13 (.11)
	Item 8			2.51 (.23)	7.37 (.55)	5.55 (.37)	2.56 (.19)	−0.26 (.11)

(*Appendix continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix (*continued*)

Sample	Item	Slopes			Intercepts			
		Grit	CI	PE	1	2	3	4
U.S. college bifactor model	Item 1	1.71 (.17)	0.28 (.31)		4.58 (.34)	1.65 (.16)	−0.81 (.14)	−3.99 (.29)
	Item 2	2.16 (.32)	0.91 (.54)		4.21 (.53)	1.96 (.29)	−1.24 (.21)	−4.04 (.47)
	Item 3	1.86 (.28)	−0.23 (.52)		4.82 (.47)	2.01 (.25)	−0.52 (.15)	−4.11 (.46)
	Item 4	1.14 (.18)	0.74 (.42)		2.62 (.26)	0.17 (.12)	−2.45 (.24)	−5.32 (.55)
	Item 5	1.25 (.15)		0.72 (.13)	5.93 (.57)	2.81 (.20)	0.55 (.12)	−1.99 (.16)
	Item 6	0.81 (.11)		0.29 (.11)	3.94 (.30)	0.93 (.11)	−0.57 (.10)	−2.80 (.20)
	Item 7	2.04 (.46)		3.52 (.86)	11.50 (2.51)	6.68 (1.27)	2.38 (.51)	−2.24 (.50)
	Item 8	1.44 (.39)		3.56 (.92)	10.63 (2.09)	7.91 (1.63)	3.48 (.75)	−0.67 (.28)
Korean college bifactor model	Item 1	1.04 (.14)	1.03 (.17)		2.57 (.11)	0.86 (.08)	−0.79 (.07)	−2.67 (.12)
	Item 2	1.17 (.17)	1.51 (.22)		3.75 (.18)	1.87 (.10)	−0.27 (.07)	−2.39 (.11)
	Item 3	1.19 (.15)	0.79 (.15)		2.88 (.13)	1.45 (.09)	−0.66 (.08)	−3.00 (.15)
	Item 4	0.67 (.14)	1.60 (.25)		2.47 (.13)	0.51 (.09)	−1.77 (.11)	−3.43 (.16)
	Item 5	1.60 (.24)		−0.20 (.36)	5.08 (.28)	3.01 (.14)	0.90 (.08)	−0.96 (.08)
	Item 6	1.27 (.16)		−0.02 (.24)	2.93 (.13)	1.36 (.08)	−0.18 (.07)	−1.63 (.09)
	Item 7	1.29 (.40)		1.24 (1.08)	6.48 (.43)	3.89 (.21)	1.07 (.10)	−1.14 (.11)
	Item 8	2.01 (.33)		0.70 (.52)	7.32 (.55)	5.56 (.37)	2.58 (.19)	−0.25 (.11)

Note. *N*s = 1,242 (high school), 581 (U.S. college), and 528 (Korean college). CI = consistency of interests; PE = perseverance of effort.

Slope and Intercept Parameter Estimates for LT-Grit

Sample	Item	Slope	Intercepts			
		Long-term grit	1	2	3	4
U.S. high school	Item 1	1.19 (.08)	4.33 (.21)	2.28 (.11)	0.21 (.07)	−1.65 (.09)
	Item 3	1.16 (.07)	2.94 (.13)	1.51 (.09)	−0.31 (.07)	−2.56 (.12)
	Item 4	1.08 (.07)	3.79 (.17)	2.01 (.10)	0.20 (.07)	−1.79 (.09)
	Item 5	1.27 (.08)	1.72 (.09)	−0.03 (.08)	−1.51 (.09)	−2.85 (.13)
	Item 6	1.70 (.10)	3.58 (.16)	1.72 (.10)	−0.31 (.09)	−2.80 (.14)
	Item 7	2.09 (.12)	4.42 (.22)	1.72 (.12)	−1.12 (.11)	−3.87 (.19)
	Item 8	1.09 (.08)	4.18 (.20)	2.92 (.13)	0.92 (.08)	−1.09 (.08)
U.S. college	Item 1	1.74 (.16)	6.07 (.53)	3.60 (.25)	0.91 (.14)	−1.75 (.15)
	Item 3	1.70 (.13)	5.12 (.37)	2.55 (.17)	0.39 (.13)	−3.25 (.23)
	Item 4	1.52 (.12)	5.55 (.45)	2.52 (.17)	0.14 (.12)	−2.84 (.21)
	Item 5	1.72 (.14)	3.87 (.24)	1.12 (.14)	−0.91 (.14)	−3.27 (.23)
	Item 6	2.75 (.21)	6.60 (.41)	2.83 (.23)	−0.31 (.17)	−4.19 (.30)
	Item 7	2.68 (.22)	7.17 (.55)	2.99 (.25)	−0.81 (.18)	−4.96 (.35)
	Item 8	1.85 (.16)	5.56 (.42)	3.67 (.25)	1.07 (.15)	−2.24 (.18)
Korean college	Item 1	2.01 (.16)	6.91 (.70)	2.55 (.19)	−0.33 (.15)	−4.08 (.30)
	Item 3	1.95 (.16)	5.48 (.46)	2.17 (.18)	−0.41 (.15)	−3.38 (.24)
	Item 4	1.58 (.13)	5.51 (.50)	2.63 (.18)	−0.12 (.13)	−4.09 (.30)
	Item 5	1.25 (.11)	4.49 (.35)	1.39 (.13)	−0.64 (.12)	−3.55 (.28)
	Item 6	3.34 (.32)	9.51 (1.19)	4.27 (.41)	−0.48 (.21)	−6.77 (.61)
	Item 7	2.76 (.22)	8.18 (.78)	4.00 (.30)	−0.89 (.20)	−6.06 (.58)
	Item 8	2.09 (.17)	5.81 (.50)	2.55 (.21)	−0.20 (.16)	−4.03 (.27)

Note. *N*s = 1,242 (high school), 581 (U.S. college), and 528 (Korean college). LT-Grit = long-term grit.

Received October 11, 2019
Revision received May 15, 2020
Accepted May 18, 2020 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.