

Measuring Grit Among First-Generation College Students: A Psychometric Analysis

Brooke Midkiff, Michelle Langer, Cynthia Demetriou, and A. T. Panter

Abstract The concept of grit is of interest in the field of education, particularly as it pertains to persistence to a 4-year college degree. This study offers an IRT analysis of the Grit Scale when used among first-generation college students (FGCSs) as well as recent first-generation college graduates and non-FGCS recent graduates. The Grit Scale was included in surveys administered as part of an array of other research projects within The Finish Line Project—a US Department of Education First in the World grant-funded project that seeks to improve FGCS access to, persistence in, and completion of postsecondary education through rigorous research into various programs and supports for FGCSs. The reliability and validity of the Grit Scale have not yet been analyzed for use with FGCS or overall with students at large, research universities. By comparing enrolled students and recent graduates, the psychometric analysis in this study offers insight into the measurement of student grit for use in program development and policy-making to improve student retention. Item response theory (IRT) analyses, analysis of differential item functioning (DIF), reliability analyses, convergent and discriminant validity analyses, and known groups validity analyses were used to examine the Grit Scale.

Keywords Grit • First-generation college students • Item response theory • Differential item functioning

1 Factor Structure and Uses of the Grit Scale

The latent construct of grit is reported to be comprised of two elements—perseverance of effort and consistency of interest in the original research into grit (Duckworth et al. 2007). Grit has been shown to be an effective predictor of success and retention in a variety of contexts such as the national spelling bee, military,

B. Midkiff (✉) • C. Demetriou • A.T. Panter
University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: bmidkiff@email.unc.edu; cyndem@email.unc.edu; panter@email.unc.edu

M. Langer
American Institutes for Research, Washington, DC 20007, USA
e-mail: mlanger@air.org

workplace, school, and marriage (Duckworth et al. 2011; Eskreis-Winkler et al. 2014; Strayhorn 2014). At the same time, first-generation college students (FGCSs) experience lower likelihoods of completing a 4-year degree when compared to their continuing-generation peers (Chen and Carroll 2005; D’Amico and Dika 2013; Engle and Tinto 2008; Vaughan et al. 2014). Therefore, the measurement of grit among this population has the potential to uncover underlying issues that impact FGCS retention, which could lead to improved interventions and supports. Additionally, within the extant literature on grit, a recent meta-analytic study states that “the grit literature may benefit from a refinement of the Grit Scale using methods based on Item Response Theory (IRT)” (Credé et al. 2016). In this chapter, we examine the Grit Scale (Duckworth et al. 2007; Duckworth and Quinn 2009) when used among FGCSs, a population as yet unstudied in conjunction with grit, compared to non-FGCSs, and offer an IRT analysis useful for the emergent area of research into refinements of the Grit Scale.

1.1 Predictive Validity of Grit in the Extant Literature

Previous research has provided evidence of the predictive validity of the Grit Scale on metacognition (Arslan et al. 2013), the retention of first-year military cadets (Maddi et al. 2012), educational attainment among adults, and grade point average (GPA) among Ivy League undergraduates (studied by Duckworth et al. 2007). However, these populations and scenarios differ in important ways from FGCSs pursuing a baccalaureate degree at a large, public research university. For example, the distribution and variance of undergraduate GPAs among students at Ivy League universities are likely to differ from those of FGCSs at a public university, as students in the research conducted by Duckworth et al. (2007) are positioned in the most advantageous university settings, with student populations, largely derived from the most advantaged high school students in the United States. Similarly, research done by Arslan et al. (2013) studied grit and metacognition among college students, yet is contextually different in that it focuses on Turkish university students—a group likely to have nontrivially different cultural norms than the average FGCS at a public university in the United States. Other studies of grit, such as those around the retention of military cadets and spelling bee champions, present radically different contexts from that of FGCSs completing a baccalaureate degree. Further, while the previous studies mentioned here demonstrate successful use of the Grit Scale, none offer an IRT analysis that can more deeply examine the psychometric properties of the scale.

1.2 Factor Structure of the Grit Scale

In line with research by Credé et al. (2016) which calls for IRT analyses of grit, the research presented here offers insight into the existing scholarly disagreement over grit as a latent construct—whether it is substantively different from conscientiousness, its incremental validity, and even its factor structure. For example, Duckworth and Quinn (2009) used confirmatory factor analysis (CFA) to determine that grit has a higher-order structure with two first-order factors and one second-order factor. Yet this finding is tempered by the fact that, using CFA, a higher-order model would exhibit identical fit to a model using two correlated first-order factors and no higher-order factor, making the analysis of limited utility (Credé et al. 2016). In fact, such a factor structure was examined by Duckworth et al. (2007) and found to have poor fit based on comparative fit index (0.83) and root mean square error of approximation (RMSEA) (0.11). Credé et al. (2016) suggest that a more meaningful way to assess the factor structure would be to examine the correlation between the two theoretical components of grit—perseverance of effort and consistency of interest. However, empirical estimates of this correlation summarized in the meta-analysis by Credé et al. (2016) find that the strength of the correlation has wide variation, with the correlation dropping as low as zero in some empirical studies (Chang 2014; Datu et al. 2016; Jordan et al. 2015). IRT analysis offers insights into the factor structure of the scale to shed light on the contradictory findings from previous CFA analyses.

1.3 Grit and FGCSs

Despite substantial debate over grit as an important noncognitive factor related to student success, grit has become an important buzzword in education as both an explanation for student achievement and as an intervention (Anderson et al. 2016). As the role of grit in education gains popular attention, it is important to understand if the measurement of grit is both reliable and valid among populations whose retention has historically been at greater risk. This research seeks to assess the validity and reliability of the Grit Scale among one such group—FGCSs—and to answer the overarching research question: What are the psychometric properties of the Grit Scale when used among FGCSs? To answer this question, we examine the reliability and factor structure of the scale and test for local dependence and differential item functioning.

2 Methods

IRT analysis was used to examine the psychometric properties of the items on the Grit Scale. IRT also allows for the assessment of local dependence between items. Analysis of DIF through logistic regression was used to test the validity of the scale for FGCSs and non-FGCSs, as well as by race/ethnicity and gender.

2.1 Data

A total of 648 participants completed a version of the Grit Scale. The sample consisted of 190 undergraduates who completed a survey at the end of their first year of college that contained the 12-item Grit Scale (Duckworth et al. 2007) in Spring 2015. An additional 458 recent graduates completed a survey in Fall 2015 that contained 9 items taken from the 12-item Grit Scale.

2.1.1 Differences in Scale Administration

The 9-item scale is a subset of the 12 items from the published 12-item scale (Duckworth et al. 2007). However, the 9-item scale used response options “very much like me” (1) to “not like me at all” (5), while the 12-item scale used response options “strongly disagree” (1) to “strongly agree” (5). In addition to the different response wording, responses also differed in direction. The 12-item scale contained a neutral option “neither disagree nor agree,” although the 9-item scale did not. Lastly, different prompts were used between the two administrations. The 9-item scale was preceded by the following prompt: “Please indicate how true the following statements are for you. Rate each statement.” The 12-item scale was preceded by the following prompt: “Here are a number of statements that may or may not apply to you. For the most accurate score, when responding think of how you compare to most people—not just the people you know well, but most people in the world. There are no right or wrong answers, so just answer honestly!” Relevant items were reverse coded so that higher scores reflect more grit across both administrations.

2.1.2 Descriptive Statistics

Of those who answered at least one Grit Scale item, <1% (four students) skipped one or more items. For the IRT analyses, only participants who did not respond to any Grit Scale item were removed from the analyses. For DIF analyses, participants with any missingness on the Grit Scale were listwise deleted. The scale items, along with associated sample size and mean scores, are given in Table 1. Differences in sample sizes are due to the combination of different surveys as explained previously.

2.1.3 Demographics

Of the sample who completed at least one Grit Scale item, 155 FGCSs completed the 12-item scale, 182 FGCS recent graduates completed the 9-item scale, and 254 non-FGCS recent graduates completed the 9-item scale. Participants included 189 men and 402 women. There were 361 non-Hispanic, White participants compared to 234 other races and ethnicities. When disaggregated by FGCS status, there were 333 FGCSs compared to 268 non-FGCSs.

Table 1 Combined scale sample item means

Item	N	Mean	Std. Dev.	Minimum	Maximum
I have overcome setbacks to conquer an important challenge	609	4.05	0.90	1	5
New ideas and projects sometimes distract me from previous ones	610	2.74	0.98	1	5
My interests change from year to year	174	2.73	1.16	1	5
Setbacks do not discourage me	609	3.15	1.01	1	5
I have been obsessed with a certain idea or project for a short time but later lost interest	610	2.98	1.03	1	5
I am a hard worker	608	4.45	0.75	1	5
I often set a goal but later choose to pursue a different one	610	3.23	0.96	1	5
I have difficulty maintaining my focus on projects that take more than a few months to complete	174	3.01	1.11	1	5
I finish whatever I begin	610	3.82	0.89	1	5
I have achieved a goal that took years of work	609	4.10	1.03	1	5
I become interested in new pursuits every few months	609	2.85	1.01	1	5
I am diligent	174	4.25	0.77	1	5

2.2 Analytic Strategy

2.2.1 IRT

Structural validity was evaluated with a factor analytic approach using IRT, implemented using the software IRTPRO (Cai et al. 2011). Each subscale of the Grit Scale was examined in a unidimensional confirmatory factor analysis (CFA) with the graded response IRT model (Samejima 2010) as well as within a bifactor graded response IRT model to examine if one underlying factor explained most of the variability in the two subscales. For the bifactor model, the explained common variance (ECV) of the general factor was examined and unidimensionality considered for values greater than 0.85. In addition to inspection of item content to assess local independence, the LD χ^2 proposed by Chen and Thissen (1997) wherein values larger than 10 are considered evidence of local dependence and values between 5 and 10 may indicate either local dependence or sparseness in the underlying table of frequencies was used as the statistical criteria for determining local dependence.

Item fit was assessed based on the SS- χ^2 fit statistics proposed by Orlando and Thissen (2000, 2003), for which a nonsignificant result ($p > 0.05$, adjusted for multiple comparisons) was an indicator of adequate model fit. Model fit was determined piecemeal through item fit, as well as the M_2 statistic proposed by

Maydeu-Olivares and Joe (2005, 2006) and its associated RMSEA. Additionally, the -2 log likelihood, the Akaike information criteria (AIC) (Akaike 1974), and the Bayesian information criterion (BIC) (Schwarz 1978) were also examined.

2.2.2 DIF

Analysis of potential DIF by FGCS status, gender, and race/ethnicity was conducted using ordinal logistic regression (OLR) of summed scores. For each item within a domain, an OLR model was used to examine whether item responses were significantly associated with group membership after controlling for students' summed score on the measure. Uniform DIF was detected by a likelihood ratio test comparing an OLR model with one predictor, summed score, to an OLR model with an additional predictor, group membership, representing a shift in the use of the response options due to group membership. Nonuniform DIF was detected by a likelihood ratio test comparing the OLR model with two predictors, summed score and group membership, to an OLR model with an additional interaction term, representing a difference in how strongly the item is related to the underlying construct due to group membership. With each paired-group analysis, an initial OLR model was run to identify a clean anchor group of items without DIF. For each sequential OLR model, any items previously identified as having DIF were removed from the summed score computation. The final OLR model used a summed score computed with only the DIF-free anchor items to test for DIF. Subsequent OLR models used a summed score computed with only the clean anchor items to test for DIF. The Benjamini-Hochberg procedure was used to make inferential decisions in the context of the multiple comparisons. In addition to examining the significance ($p < 0.05$), magnitude of DIF was further evaluated by examining the expected item scores and estimating the effect sizes ($\Delta R^2 > 0.02$ indicative of salient DIF).

2.2.3 Reliability

Internal consistency was evaluated by Cronbach's α for both versions of the scale, as well as for both subscales within each version using the software Mplus (Muthén and Muthén 1998). Alpha values of 0.70 or greater are an acceptable minimum for group-level assessment.

2.2.4 Convergent and Discriminant Validity

Participants who completed the Grit Scale also completed the Growth Mindset Scale (Dweck 2008) and the 5-item Guilt Proneness Scale (Cohen et al. 2011; Cohen et al. 2014). To assess convergent validity, for both grit subscales, the correlation between the mean item score and the mean item score of the Growth Mindset Scale was computed. To assess divergent validity, for both grit subscales, the correlation

between the mean item score and the mean item score of the Guilt Proneness Scale was computed. Growth Mindset response items ranged from “strongly disagree” (1) to “strongly agree” (5) and were scored such that higher mean scores reflect more growth mindset. Guilt Proneness response items ranged from “extremely unlikely” (1) to “extremely likely” (5) and were scored such that higher mean scores reflect more guilt proneness.

2.2.5 Known Groups Validity

The validity of the Grit Scale was examined by assessing the extent to which it could discriminate between several known groups that should, in theory, differ. These groups included FGCSs who were current students, FGCS recent graduates, non-FGCS recent graduates, race and ethnicity, race and ethnicity interacted with gender, and participants grouped by their reported use of university resources. The use of university resources was measured by a list of resources and the frequency with which students used them while in college, with responses ranging from “never” (1) to “ten or more times” (6). We compared means across all groups using a one-way analysis of variance (ANOVA). Statistical significance was defined at the 0.05 alpha level for evaluation of convergent, discriminant, and known groups validity.

3 Results

3.1 Item Response Theory

Two items on the perseverance of effort subscale had low cell counts in the extreme categories, and thus for those two items, categories were collapsed for IRT analysis. The factor structure was further examined by estimating the graded response model (Samejima 2010), on the item scores for each grit subscale, and then modeling all scores using a bifactor model. Table 2 shows the fit of the graded response IRT models fit to the Grit Scale. The bifactor model fit well, although one item had a high factor loading (0.73) on the overall factor and a low loading (0.25) on the perseverance of effort subscale. The ECVs for this model were 0.45 for consistency of interest and 0.63 for perseverance of effort, suggesting these two subscales do not support one underlying factor. Therefore, unidimensional IRT models were fit to each of the Grit Scale subscales, including both a 5-item and 4-item (dropping the problematic item identified in the bifactor model) version of the perseverance of effort subscale. As Table 2 indicates, these models also fit the data well.

The IRT parameters of the bifactor model as well as the final two unidimensional models for the Grit Scale subscales are presented in Table 3. For the consistency of interest subscale, all items have high IRT a parameters; the item “I have been

Table 2 Comparison of IRT graded response models

Model	Consistency of interest subscale	Perseverance of effort subscale	Perseverance of effort subscale	Bifactor model
Number of items	4	4	5	9
<i>Item fit</i>				
Local dependence	None	None	None	None
SS- χ^2 item fit	All nonsignificant*	All nonsignificant*	One item does not fit well*	All nonsignificant*
<i>Model fit</i>				
M ₂	231.25	165.58	277.25	1081.81
RMSEA	0.05	0.05	0.05	0.05
-2 log likelihood	6287.33	5596.55	6974.20	13121.35
AIC	6327.33	5632.55	7018.20	13223.35
BIC	6415.60	5712.00	7115.29	13448.43

Notes: (*) indicates the use of Benjamini-Hochberg to control multiple comparisons

obsessed with a certain idea or project for a short time but later lost interest” has the strongest relationship with the underlying construct. For the perseverance of effort subscale, the items also have strong IRT a parameters. However, the item “I am a hard worker” has a slightly weaker relationship to the underlying construct, although it also measures the lower end of the Grit Scale ($b_I = -5.16$).

3.1.1 Overall Test Information Curves

Figure 1 shows the overall test information curve and standard error for the consistency of interest subscale. This figure indicates that test information is high across the range of consistency of interest, providing adequate measurement from -3 to 3 SDs below and above the mean. Figure 2 shows the same information for the 4-item version of the perseverance of effort subscale, after dropping the poorly fitting item “I finish whatever I begin.” However, test information is only high for the perseverance of effort subscale at the lower end of the scale, indicating that the Grit Scale measures perseverance of effort well for those with levels at 1 SD above the mean and below.

3.2 Differential Item Functioning

Analysis of potential DIF resulted in only one item being flagged. The item, “I have overcome setbacks to conquer an important challenge,” had significant uniform DIF ($p.03$) with a large effect size for FGCSs ($n = 333$) and non-FGCSs ($n = 268$). Figure 3 displays the item means by summed score for each of these groups. FGCSs have slightly higher item means across the score range, indicating they are slightly more likely to endorse this item than non-FGCSs, resulting in a higher overall score on the Grit Scale.

3.3 Reliability

Cronbach’s alphas are adequate for both administered versions of the scale: 0.72 for the 12-item scale and 0.65 for the 9-item scale. The perseverance of effort subscale had lower alphas (0.60, 0.57) in both versions of the scale administered than did the consistency of interest subscale (0.69 in both versions).

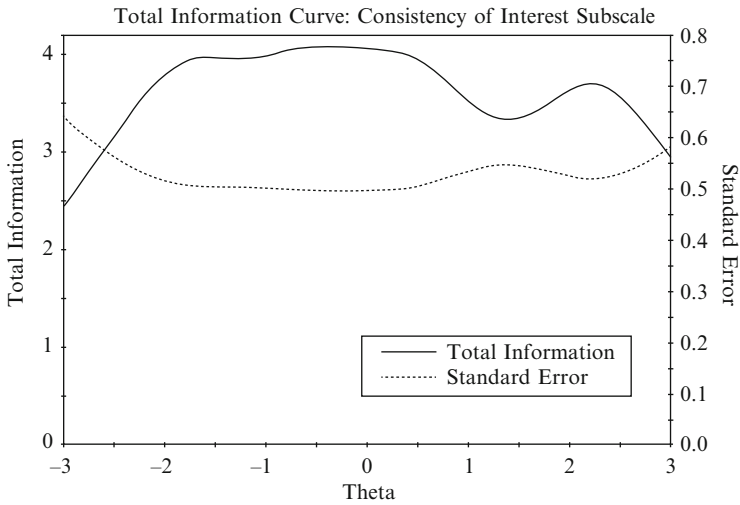


Fig. 1 Total information curve for consistency of interest subscale

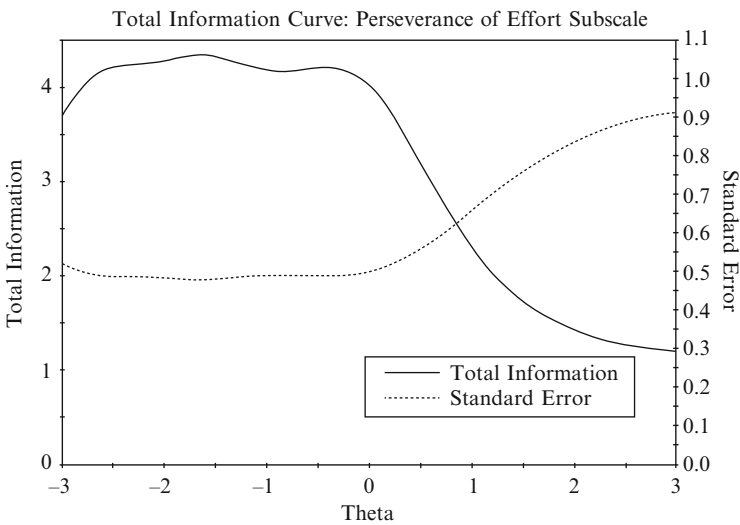


Fig. 2 Total information curve for perseverance of effort subscale

3.4 Convergent and Discriminant Validity

The mean item score on growth mindset was 3.55 (s.d. 0.84). It was significantly correlated with perseverance of effort ($r = 0.20, p = 0.001$), but not with consistency of interest ($r = 0.01, p = 0.73$). This indicates grit and growth mindset are moderately related; however, it also provides evidence of the difference between

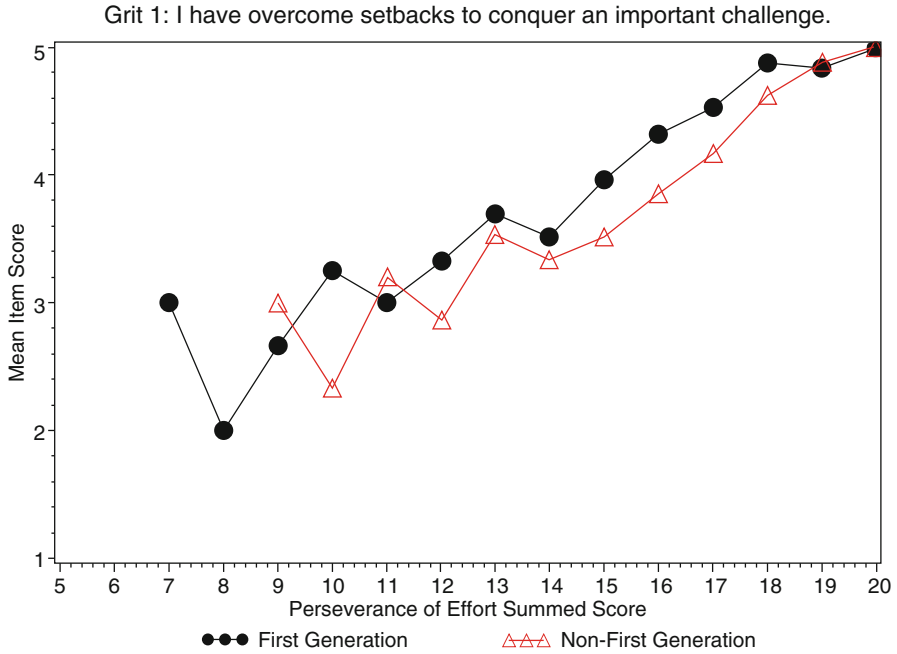


Fig. 3 DIF between FGCSs and non-FGCSs

the two Grit Scale subscales. The mean item score on guilt proneness was 4.12 (s.d. 0.74) and was significantly correlated with both subscales. Perseverance of effort correlated at $r = 0.19$ ($p < 0.001$) and consistency of interest correlated at $r = 0.12$ ($p = 0.002$).

3.5 Known Groups Validity

The results from the known groups analysis are given in Table 4. Of the groups analyzed (see Sect. 2.2.5), only two were statistically significantly different from one another. On the perseverance of effort subscale, men showed less perseverance of effort than females (mean = 3.85 vs. 3.97). In consistency of interest, the interaction of gender, race/ethnicity, and FGCS status resulted in statistically significant differences. Women who were White, non-Hispanic, and not FGCSs scored the highest in consistency of interest (mean = 3.13), while men who were non-White and FGCSs scored the lowest (mean = 2.80). Across all gender, race/ethnicity, and FGCS groupings, non-FGCSs (means range from 3.13 to 2.94) scored the highest in consistency of interest than FGCS (means range from 2.92 to 2.80).

Table 4 ANOVA results of known groups

Groups	DF	Sum of squares	Mean square	F	Pr > F
<i>Perseverance of effort subscale</i>					
Male, White, non-Hispanic, FGCS	7	8.25	1.18	2.29	0.0261
<i>Consistency of interest subscale</i>					
Gender	1	1.64	1.64	4.16	0.0418

Notes: Benjamini-Hochberg procedure used to control for Type I error rate in multiple comparisons. Only statistically significant differences after correction are provided; no other group differences were significant after the use of Benjamini-Hochberg procedure

4 Discussion

The IRT analysis suggests that the 4-item consistency of interest subscale fits well with no local dependence, as does the 4-item perseverance of effort subscale, after dropping the poorly fitting item “I finish whatever I begin.” The item “I have overcome setbacks to conquer an important challenge” exhibits uniform DIF among FGCS and non-FGCS with a large effect size. Consistent with previous literature (see Credé et al. 2016 for a comprehensive overview), our findings indicate that the higher-order factor structure suggested by Duckworth and Quinn (2009) is not supported. IRT analysis provided here demonstrates that two unidimensional subscales fit better than the bifactor model. Factor loadings from the IRT analysis suggest that there is little evidence of a higher-order construct, when using the data analyzed in this research. Given the discrepancies in previous research about the factor structure of the Grit Scale, along with the recent notation that IRT analysis is needed, this research contributes the important finding that IRT analysis, conducted using these data, does not support a factor structure wherein consistency of interest and perseverance of effort load onto the higher-order construct grit.

References

- H. Akaike, A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**(6), 716–723 (1974)
- C. Anderson, A.C. Turner, R.D. Heath, C.M. Payne, On the meaning of grit . . . and hope . . . and fate control . . . and alienation . . . and locus of control . . . and . . . self-efficacy . . . and . . . effort optimism . . . and . . . *Urban Rev.* **48**(2), 198–219 (2016). doi:[10.1007/s11256-016-0351-3](https://doi.org/10.1007/s11256-016-0351-3)
- S. Arslan, A. Akin, N. Çitemel, The predictive role of grit on metacognition in Turkish university students. *Stud. Psychol.* **55**(4), 311 (2013)
- L. Cai, D. Thissen, S.H.C. DuToit, *ITPRO 3 (Version 3) [Windows]* (Scientific Software International, Lincolnwood, IL, 2011)
- W. Chang, Grit and academic performance: is being grittier better?, Ph.D. Dissertation, University of Miami, Miami, FL, 2014
- X. Chen, C.D. Carroll, First-generation students in postsecondary education: a look at their college transcripts. Postsecondary Education Descriptive Analysis Report. NCEES 2005-171. National Center for Education Statistics (2005)

- W.-H. Chen, D. Thissen, Local dependence indexes for item pairs using item response theory. *J. Educ. Behav. Stat.* **22**(3), 265–289 (1997)
- T.R. Cohen, S.T. Wolf, A.T. Panter, C.A. Insko, Introducing the GASP scale: a new measure of guilt and shame proneness. *J. Pers. Soc. Psychol.* **100**(5), 947–966 (2011). doi:[10.1037/a0022641](https://doi.org/10.1037/a0022641)
- T.R. Cohen, Y. Kim, A.T. Panter, *The Five-Item Guilt Proneness Scale (GP-5)* (Carnegie Mellon University, Pittsburgh, PA, 2014), p. 1
- M. Credé, M.C. Tynan, P.D. Harms, Much ado about grit: a meta-analytic synthesis of the grit literature. *J. Pers. Soc. Psychol.* (2016). doi:[10.1037/pspp0000102](https://doi.org/10.1037/pspp0000102)
- M.M. D'Amico, S.L. Dika, Using data known at the time of admission to predict first-generation college student success. *J. Coll. Stud. Retent. Res. Theory Pract.* **15**(2), 173–192 (2013)
- J.A.D. Datu, J.P.M. Valdez, R.B. King, Perseverance counts but consistency does not! Validating the short grit scale in a collectivist setting. *Curr. Psychol.* **35**(1), 121–130 (2016)
- A.L. Duckworth, P.D. Quinn, Development and validation of the short grit scale (Grit-S). *J. Pers. Assess.* **91**(2), 166–174 (2009). doi:[10.1080/00223890802634290](https://doi.org/10.1080/00223890802634290)
- A.L. Duckworth, C. Peterson, M.D. Matthews, D.R. Kelly, Grit: perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* **92**(6), 1087 (2007)
- A.L. Duckworth, T.A. Kirby, E. Tsukayama, H. Berstein, K.A. Ericsson, Deliberate practice spells success why grittier competitors triumph at the national spelling bee. *Soc. Psychol. Personal. Sci.* **2**(2), 174–181 (2011)
- C.S. Dweck, *Mindset: The New Psychology of Success*, Ballantine Books Trade Pbk edn. (Ballantine Books, New York, NY, 2008)
- J. Engle, V. Tinto, Moving beyond access: college success for low-income, first-generation students. Pell Institute for the Study of Opportunity in Higher Education (2008)
- L. Eskreis-Winkler, E.P. Shulman, S.A. Beal, A.L. Duckworth, The grit effect: predicting retention in the military, the workplace, school and marriage. *Front. Psych.* **5** (2014). doi:[10.3389/fpsyg.2014.00036](https://doi.org/10.3389/fpsyg.2014.00036)
- M.H. Jordan, T. Gabriel, R. Teasley, W.J. Walker, M. Schraeder, An integrative approach to identifying factors related to long-term career commitments: a military example. *Career Dev. Int.* **20**(2), 163–178 (2015)
- S.R. Maddi, M.D. Matthews, D.R. Kelly, B. Villarreal, M. White, The role of hardiness and grit in predicting performance and retention of USMA cadets. *Mil. Psychol.* **24**(1), 19 (2012)
- A. Maydeu-Olivares, H. Joe, Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables: a unified framework. *J. Am. Stat. Assoc.* **100**(471), 1009–1020 (2005)
- A. Maydeu-Olivares, H. Joe, Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika* **71**(4), 713–732 (2006)
- L.K. Muthén, B.O. Muthén, *Mplus User's Guide*, 7th edn. (OECD, Los Angeles, CA, 1998)
- M. Orlando, D. Thissen, Likelihood-based item-fit indices for dichotomous item response theory models. *Appl. Psychol. Meas.* **24**(1), 50–64 (2000)
- M. Orlando, D. Thissen, Further investigation of the performance of S-X2: an item fit index for use with dichotomous item response theory models. *Appl. Psychol. Meas.* **27**(4), 289–298 (2003)
- F. Samejima, in *The general graded response model*, ed. by M.L. Nering, R. Ostini. *Handbook of Polytomous Item Response Theory Models*, (Routledge, New York, NY, 2010), pp. 77–108
- G. Schwarz, Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
- T.L. Strayhorn, What role does grit play in the academic success of black male collegians at predominantly White institutions? *J. Afr. Am. Stud.* **18**(1), 1–10 (2014)
- A. Vaughan, J. Parra, T. Lalonde, First-generation college student achievement and the first-year seminar: a quasi-experimental design. *J. First-Year Exp. Stud. Trans.* **26**(2), 51–67 (2014)