Creating a New Scale for the Measurement of Academic Grittiness

Sarah Bonell *

University of Queensland, Australia

*E-mail: sabon22@hotmail.com

Abstract

Defined as persistence and passion for the achievement of long-term goals, grit is considered vital in determining success within academic contexts. Whilst scales do presently exist to measure grit as a holistic construct, no scale currently exists to measure academic-specific grit. The aim of the present study, therefore, was to develop a new scale for the measurement of academic grit. The newly developed scale, titled the Grit Inventory (GI), was assessed for psychometric soundness utilizing a sample of 109 undergraduate psychology students from The University of Queensland, Australia. Results showed the GI to have sound validity and good internal consistency. Nine of the scale's twelve items had acceptable item discrimination indices. It was proposed that the GI be used in conjunction with ability tests to determine university admissions, granted that future research could demonstrate predictive validity for the scale.

Keywords: grit, grittiness, persistence, academic, academic grittiness, academic grit.

1. Introduction

Duckworth, Peterson, Matthews and Kelly (2007) argue that there are two key factors involved in determining an individual's success: ability and grit. Defined as persistence and passion for the achievement of long-term goals, grit is considered especially vital in shaping success within academic-specific contexts. This is because grit has consistently been shown to predict academic success, such as student grade point averages (GPAs), over and above pre-existing effects of ability (Duckworth & Quinn, 2009; Duckworth et al., 2007; Strayhorn, 2014; Wolters & Hussain, 2014). Despite findings such as these, the majority of research within the area of academia has focused solely on ability's role in determining success. This is demonstrated in the multitude of established measures formulated for academic ability, compared to the very few formulated for grit (Duckworth & Quinn, 2009; Bridgeman, McCamley-Jenkins, & Ervin, 2000; Gottfredson, 1997; Hartigan & Wigdor, 1989).

The Short Grit Scale (Grit-S) is perhaps the only established and well-known test specifically for grit. It contains a total of eight items split across two subscales: consistency of interest and perseverance of effort (Duckworth & Quinn, 2009). Several studies conducted by Duckworth and Quinn (2009) have shown the Grit-S to be both a valid and reliable measure of grit. For instance, one of their studies showed that, amongst a group of 279 public high school students, scores on the Grit-S were positively correlated with future GPAs and negatively

correlated with future hours spent watching television, demonstrating predictive validity. In the same study, internal consistency, $\alpha = .82$, and test-retest reliability (one year later), r = .68, p<.001, were shown to be high.

Similar to grit, conscientiousness has regularly been shown to predict academic performance (Barrick & Mount, 1991). Because of this, Duckworth and Quinn (2009) tested the Grit-S against the construct of conscientiousness to demonstrate concurrent validity and found the two measures to be highly correlated, r = .70, p<.001. They concluded conscientiousness and grit to be overlapping constructs, and suggested that the link between conscientiousness and academic performance further supported the idea that grit would be able to predict academic performance (which it subsequently did according to the results of their study).

One limitation for the Grit-S, however, is the fact that it fails to explore domain-specific measures for grit (Duckworth et al., 2007). All items within the Grit-S questionnaire are context-neutral, which assumes that grit can be measured as a universal trait. However, to date no research has been conducted to investigate the reality of this assumption. Therefore, should an individual's grittiness alter cross-contextually (e.g. working extremely hard at their hobby, but not at all at their day job), it is unclear as to whether or not the Grit-S would remain valid.

A number of other measures have been created to assess constructs very similar to grit, almost all of which display sound psychometric properties. For instance, the Delaying Gratification Inventory (DGI) was developed to measure the degree to which one might forego immediate satisfaction in favor of long-term rewards (Hoerger, Quirk & Weed, 2011; Dollinger, 2011). Across a diverse sample of 10 741 individuals, the 35 items composite scale was shown to be internally consistent, a \approx .90, and to have testretest reliability, r = .80 (Hoerger et al., 2011). Concurrent validity was demonstrated for the measure when it significantly correlated with conscientiousness, r = .71, self-discipline, r = .55, and achievement striving, r = .54, amongst other things. The scale's 10 item condensed form was also shown to have sound psychometric properties (Hoerger et al., 2011).

That said, one major weakness for the DGI is the fact that the utilized sample lacked generalizability. The inventory was designed to be universally applicable, yet participants' socioeconomic and educational statuses were significantly higher than the national average at the time of testing. This would not be concerning should the DGI have been tested subsequently across less educated/wealthy groups, but no such research has been conducted to date. Hence, it is unclear as to whether the DGI can be generalized to the broad population for which it was intended (thus limiting its application).

Considering the multitude of limitations present within existing grit (and similar) scales, the aim of the present study will be to develop an improved measurement scale for grit – the Grit Inventory (GI). The GI will improve upon existing measures by: (a) being context-specific for academic grittiness only and (b) being demographic-specific for undergraduate students only. In this way, the GI is designed to be generalizable to a more limited sample/context than previous measures, but to measure this specific sample more accurately than previous measures.

Tests for both reliability and validity will be conducted for the GI in order to ensure its readiness for use. Reliability will be measured as internal consistency, utilizing Cronbach's alpha (a). Individual scale items will also be evaluated via an item discrimination index analysis in order to determine how much each individual scale item contributes to the scale's reliability overall. The GI ought to correlate with existing measures of grittiness and similar constructs (Grit-S and DGI) should it be an accurate measurement of grit. Furthermore, considering that the Grit-S correlates with conscientiousness and is able to accurately predict students' GPAs, the GI ought to do the same. In light of this, four validity hypotheses have been proposed: (a) a positive relationship will be seen between the GI and the Grit-S, (b) a positive relationship will be seen between the GI and the DGI, (c) a positive relationship will be seen between the GI and the International Personality Item Pool Conscientiousness Scale (IPIP-CS), and (d) a positive relationship will be seen between GI scores and students' GPAs.

2. Method

Participants. Participants for this study included 109 undergraduate psychology students from The University of Queensland in Australia. Participants were predominantly female (78%), and ranged between 18 and 57years of age (M = 22.60, SD = 5.90, 18 participants' non-response). Participants were recruited from a core third-year Measurement in Psychology course (PSYC3020). Participation was completely voluntary and anonymous and no penalty was given to students who didn't complete the study.

A. Measures

Demographic Information. Participants were asked to report their age, sex and GPAs (out of 7) prior to completing the body of the questionnaire items.

Grit-S. Developed by Duckworth and Quinn (2009), the Grit-S measured participants' grittiness across an eightitem scale. Four items were measures of perseverance of effort (e.g. I am diligent). The other four were measures of consistency of interest and were all reverse scored (e.g. I often set a goal but later choose to pursue a different one). Responses for all items were measured on a 5-point Likert-type scale (1 = not like me at all, 5 = very much like me). Overall scores were calculated by averaging across all eight items, where a higher score indicated a higher level of grit. The Grit-S was found to have good internal reliability (α = .74).

DGI. Participants' levels of delayed gratification were measured using a ten-item composite version of Hoerger et al.'s (2011) original 35-item DGI (e.g. I can resist junk food when I want to). Four of the ten items were reverse scored (e.g. I would rather take the easy road in life than get ahead). Responses for all items were measured on a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree). Overall scores were calculated by averaging across all ten items, where a higher score indicated a higher level of delayed gratification. The DGI was found to have good internal reliability (α =.71).

IPIP-CS. The IPIP-CS measured participants' conscientiousness across a ten-item scale (e.g. I am always prepared). Five of the ten items were reverse scored (e.g. I waste my time). Responses for all items were measured on a 5-point Likert-type scale (1 = very inaccurate, 5 = very accurate). Overall scores were calculated by averaging across all ten items, where a higher score indicated a higher level of conscientiousness. The IPIP-CS was found to have good internal reliability (α =.76).

GI. The GI measured participants' grittiness across a twelve-item scale (e.g. I don't often procrastinate). A team of researchers at The University of Queensland developed the scale. Initially, the team was split into four groups of approximately five individuals. Each group then proposed approximately five items for the GI, and the team as a whole selected the twelve most favorable items to include. A total of three items were reverse scored or the GI (e.g.

my passion for study decreases the further I am in my course). Responses for all items were measured on a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree). Overall scores were calculated by averaging across all twelve items, where a higher score indicated a higher level of grit. The GI was found to have good internal reliability (α =.74).

Procedure. The survey was administered electronically to all students enrolled in PSYC3020 during Semester Two, 2016. It took approximately one hour to complete. The survey also contained a range of alternate questions unrelated to the present study (used by other researchers). Participants were given approximately one week to complete the survey.

3. Results

Table 1. Descriptive Statistics for Grit-S, DGI, IPIP-CS and GI Scales.

Scale	M	SD	Min.	Max.	Skew**	
Grit-S	3.11	0.57	1.63	4.63	0.59	
DGI	3.18	0.26	2.40	4.90	0.24	
IPIP-CS	3.33	0.59	1.40	4.50	-2.65	
GI	3.08	0.54	1.67	4.17	-0.16	
GPA*	5.66	0.74	3.50	7.00	-2.35	

^{*}after excluding GPAs < 3.5 (impossible scores)

As shown in Table 1, no scale was significantly skewed.

Tests for Validity. All validity tests were conducted using correlations (r). This is because all variables were continuous and no data was significantly skewed (see table 1). Results showed a significant positive correlation between the GI and the Grit-S, r(103) = .53, p < .001. Significant positive correlations were also seen between the GI and DGI, r(104) = .55, p < .001, and the GI and IPIP-CS, r(104) = .59, p < .001. The GI and participants' GPAs were also significantly positively correlated, r(97) = .53, p < .001.

Item Discrimination Indices. Item discrimination indices were used to indicate the extent to which any one given item within the GI would be able to distinguish between high-scoring and low-scoring participants overall. Nine of the scale's twelve items were acceptable (see Table 2). The number of lower-bound participants (lowest-scoring third of participants) who scored highly (four or five) within each item was denoted by L. Similarly, the number of upper-bound participants (highest-scoring third) who scored highly within each item was denoted by U. The number of participants within lower and upper bounds were denoted by nL and nU respectively. Item indices discrimination were denoted

Table 2. Item Discrimination Indices for GI.

Items	U	L	d
Even if a course is hard, I am unlikely to drop out	34	18	.39
I don't often procrastinate	9	0	.24
I commit to study	33	7	.69
I stick to my goals even when I experience failure	33	16	.42
I'm rarely deterred by distractions while I'm working on a task	12	1	.29
I rarely need encouragement from others to persist past obstacles	27	7	.52
I am intrinsically motivated to study	32	6	.69
Failing a piece of assessment has me feeling discouraged from my goals*	9	3	.16
I would be motivated to continue a task that I do not understand much about	27	5	.58
If I did much worse on an assessment than I expected, I would seriously consider dropping the class*	26	18	.17
I can maintain a high level of motivation throughout the semester	20	0	.54
My passion for study decreases the further I am in my course*	25	2	.62

Note. nU = 37 and nL = 34 for all items. Acceptable d> .25

*after reverse scoring

4. Discussion

The aim of the present study was to develop an improved measurement scale for grit (GI). Specifically, the aim was to improve on existing measures by making the GI: (a) context-specific for academic grittiness only and (b) demographic-specific for undergraduate students only. Four hypotheses were proposed for the new scale: (a) a positive relationship would be seen between the GI and the Grit-S, (b) a positive relationship would be seen between the GI and the DGI, (c) a positive relationship would be seen between the GI and the IPIP-CS, and (d) a positive relationship would be seen between GI scores and students' GPAs.

Results demonstrated a significant positive correlation between the GI and Grit-S, and hence the study's first hypothesis was supported. Results also demonstrated a significant positive correlation between the GI and both the DGI and IPIP-CS, and hence the study's second and third hypotheses were supported. The Grit-S is currently considered to be the most established measure for grit across a broad range of domains (Duckworth et al., 2007; Duckworth & Quinn, 2009), and the constructs measured by both the DGI and IPIP-CS have been theoretically linked to grit (Hoerger et al., 2011; Duckworth & Quinn, 2009). Therefore, the fact that individuals scored consistently across these four scales suggests that the GI does indeed measure the same/similar construct as the validating scales.

^{**} significance $\geq \pm 3.29$

Results showed a significant positive correlation between the GI and students' GPAs. A multitude of studies have demonstrated the existence of a relationship between undergraduate students' GPAs and their grittiness, as measured by the Grit-S (Duckworth & Quinn, 2009; Duckworth et al., 2007; Strayhorn, 2013; Wolters & Hussain, 2015). Therefore, the acceptance of this study's fourth hypothesis further suggests that the GI is a valid measure of grit, in the sense that it is able to predict what the existing measure of grit predicts. Internal consistency for the measure was also good, further demonstrating the GI's sound psychometric properties.

All but three items (two, eight and ten) within the GI had acceptable item discrimination indices. Item seven on the scale performed most highly overall, followed by items three and twelve respectively. No one (neither low nor high scorers) identified highly with items two or eight, perhaps due to a sample bias. For instance, for item two, it may be that even the grittiest of university students are likely to identify as procrastinators. This item could be reworded to read "I procrastinate less than the average student," reflecting a tangible difference between participants when compared to a baseline. Item eight could benefit from including a baseline for similar reasons (e.g. compared to my peers, I become particularly discouraged from my goals when I perform poorly on a piece of assessment). The problem with item ten, on the other hand, may be that the item is too extreme. Even score low in grittiness appear students who understandably reluctant to drop an entire class simply because they performed poorly on one piece of assessment. Therefore, perhaps this question could be reworded to be less extreme (e.g. if I performed much worse than expected on assessment in one of my classes, I would consider dropping the class).

One limitation for the present study was potential sample bias. Because the voluntary survey took approximately an hour to complete, it may be the case that students with low grit would not have even bothered to complete the survey. Conversely, considering that the survey was administered mid-semester, it may also be the case that the grittiest students would not have had the time to complete the lengthy survey. Another limitation was the self-report nature of the survey. Despite not completing the survey for course credit, students did complete the survey for their university course. Therefore, despite anonymity, a self-report bias may have existed amongst the students to report high levels of commitment to study.

That said, one strength for the present study was that it focused on a constrained sample — academic undergraduates. In this sense, the study is arguably more generalizable to this specific domain than previous research/scales would have been. Hence, it is proposed that the GI may be useful for undergraduate university admissions testing (where other measures may not have been suitable). Whilst the present study did not demonstrate the GI's ability to predict future academic performance, it did indeed demonstrate that the GI

correlates with students' current GPAs. Hence, should predictive validity be demonstrated for the GI, it could likely be utilized in conjunction with current entry tests to distinguish between candidates who demonstrate near-equal ability.

Considering this, future research ought to assess the GI longitudinally. This would help to confirm the existence of predictive validity for the GI, supporting its use as a university admissions test. It may also demonstrate further psychometric soundness for the GI (e.g. test-retest reliability). To address limitations, studies may also wish to utilize informant-report versions of the GI (in place of self-report), measured independently of other constructs (to shorten time taken to complete the survey). Items two, eight and ten of the inventory could also be reworded in subsequent studies.

5. Conclusions

In conclusion, the present study's aim to create a new and improved measurement scale for grit was achieved. The GI was found to positively correlate with existing measures of grit and its related constructs (Grit-S, DGI and IPIP-CS), as well as with student GPAs. The GI demonstrated sound internal consistency. It was proposed that, should future research confirm predictive validity for the GI, it ought to be considered a part of the undergraduate university admissions process.

References

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44(1), 1–26. doi: 10.1111/j.1744-6570.1991.tb00688.x

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test (College Board Report 2000 –2001). New York: College Entrance Examination Board.

Dollinger, M. J. (2011). Delayed Gratification. Business Horizons, 55(2), 101-103. doi: 10.1016/j.bushor.2011.10.003

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087-1101. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). Journal of Personality Assessment, 91(2), 166-174. doi:10.1080/00223890802634290

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. Intelligence, 24, 79 –132. doi:10.1016/S0160-2896(97)90014-3

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

Hoerger, M., Quirk S. W., & Weed, N. C. (2011). Development and validation of the Delaying Gratification Inventory. Psychological Assessment, 23(3), 725–738. doi:10.1037/a0023286

Strayhorn, T. (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), 1-10. doi:10.1007/s12111-012-9243-0