The Psychometric Properties of the Short Grit Scale (Grit-S) within an International Context

Theres Friederike Ackermann

Master Thesis in Psychology, 30 ECTS

Specialization: Positive Psychology and Technology

Faculty of Behavioural, Management and Social Sciences

1st Supervisor: dr. P. M. ten Klooster

2nd Supervisor: dr. S. M. Kelders

Academic Year: 2017/2018

Date: 28.06.2018

UNIVERSITY OF TWENTE.

Acknowledgements

First, I would like to thank dr. Peter ten Klooster for his very helpful feedback and support, especially regarding structural equation modelling and Mplus. I would also like to thank dr. Mirjam Radstaak and dr. Llewellyn van Zyl for supporting me during the earlier stages of my thesis and dr. Saskia Kelders for her feedback. Moreover, I would like to thank Flo for always having my back and for always believing in me throughout the time of my bachelor and master. I would also like to express my gratitude to my family for always encouraging me to follow my dreams. Finally, I would like to thank my close friends for the motivational talks and support.

Abstract

During the past decades, noncognitive traits and their contributions to a satisfied and fulfilled life have been extensively studied. A noncognitive trait which attracted much attention during more recent years is grit. Defined as *passion and perseverance for long-term goals*, grit plays an important role in numerous areas of life from academic achievement and career endorsement to athleticism and mental health. Grit can be assessed by means of the Grit-S, an 8-item self-report questionnaire which comprises two subscales representing consistency of interest and perseverance of effort. While research has demonstrated the importance of grit, the validity and reliability of the Grit-S remain disputed. The present study investigated the psychometric properties of the Grit-S in an international sample (N = 272) by means of confirmatory factor analysis and correlational analysis using Mplus and SPSS. Results support the Grit-S as a valid and reliable measurement instrument of grit, providing evidence for its factorial validity, internal consistency, convergent and criterion validity. Moreover, the present study found support for the multidimensionality of the Grit-S, highlighting the importance of its subscales. However, results show the Grit-S to be non-invariant across different nationalities and age groups. Hence, care should be taken when using the Grit-S in cross-cultural and cross-age studies as scores from different nationalities and age groups may not be directly comparable.

Content

Acknowledgements II

Abstract III

1. Introduction 1

1.1. The Potential of Grit 2

1.2. Measuring Grit 7

2. Methodology 10

2.1. Procedure 10

2.2. Participants 11

2.3. Materials 12

2.4. Data Analysis 14

3. Results 20

3.1. Factorial validity 20

3.2. Measurement invariance 21

3.3. Reliability 22

3.4. Convergent validity 23

3.5. Criterion validity 24

4. Discussion 25

5. Bibliography 33

1. Introduction

Since the earliest days of philosophy, people have been concerned with the key to a successful and happy life (Aristotle, trans. 2011). To this day, the topic continues to attract much attention. As shown in the success of books like Seligman's (2002) *Authentic Happiness: Using the new positive psychology to realize your potential for lasting fulfillment*., the interest in a fulfilled life is as up-to-date in the 21st century as it was more than 2000 years ago. As a result, the topic has been extensively studied.

Over the past decades, many studies have shown that cognitive ability is of vital importance for occupational success and happiness (Enkvist, Ekström, & Elmståhl, 2013; Hunter, 1986). The term cognitive ability is closely related to intelligence and describes mental abilities like verbal communication, abstract reasoning and logical, numerical and spatial thinking. Our cognitive abilities enable us to communicate with each other and to learn and profit from experience. They are needed to adjust to new situations, solve problems and achieve goals (Wechsler, 1950). Although good cognitive abilities are important for a happy and successful life, they do not explain why some individuals achieve more than other equally intelligent individuals. As cognitive abilities alone do not fully explain individual differences when it comes to happiness and success, a focus on noncognitive abilities may contribute to a better understanding of these differences (Duckworth, Peterson, Matthews, & Kelly, 2007). Consequently, much study in recent years has explored the role of noncognitive abilities in extraordinary achievement, academic performance and orientations to happiness and well-being (see, for example Duckworth & Gross, 2014; Dumfart & Neubauer, 2016; Richardson, Abraham, & Bond, 2012; Vainio & Daukantaitė, 2016; Von Culin, Tsukayama, & Duckworth, 2014).

Noncognitive abilities are intertwined with cognitive components since the application of noncognitive abilities involves cognition. Yet, cognitive and noncognitive abilities differ in important aspects (Messick, 1979). Noncognitive ability is often used as an umbrella term to describe attributes which cannot be assessed through tests of cognitive ability (West et al., 2016). In contrast to cognitive abilities, noncognitive abilities are “dependent upon temperament and personality” (Wechsler, 1950). Examples of noncognitive abilities are personality traits such as conscientiousness, self-control and grit (West et al., 2016). The power of noncognitive abilities is demonstrated in numerous investigations. It was found that employees who score higher on conscientiousness develop higher levels of job knowledge and

conscientiousness has also been identified as a significant predictor of school achievement (Dumfart & Neubauer, 2016; Schmidt & Hunter, 1998). Moreover, self-control correlates positively with grade point average (GPA), good adjustment, interpersonal success and negatively with psychopathology (Tangney, Baumeister, & Boone, 2004). Grit, defined as consistency of interest and perseverance of effort for long-term goals (Duckworth et al., 2007), is strongly related to well-being and also associated with higher levels of education, higher GPA and fewer career changes (Duckworth et al., 2007; Vainio & Daukantaitė, 2016).

As noncognitive abilities play a key role in a considerable number of diverse areas of a successful and happy life, there is a growing interest in ways to foster them. Especially the noncognitive ability grit has recently attracted much attention. It was found that positive affect and purpose commitment are predictors of grit and that grit is fostered through the development of a growth mindset (Duckworth, 2013; Hill, Burrow, & Bronk, 2016). In contrast to individuals with a fixed mindset, who believe that skills are set in stone, individuals with a growth mindset believe that they can develop them. They do not see failure as a setback, but rather as an opportunity to learn (Dweck, 2017). The reason for grit's growth in popularity might be rooted in this connection with a growth mindset, as it shows grit as an ability which can be developed (Duckworth, 2013). Additionally, grit suggests that goals are reachable via passion and perseverance, which is a positive and empowering message implying that people forge their own future (Duckworth, 2013; Duckworth et al., 2007; Stokas, 2015).

1.1. The Potential of Grit

"Many of life's failures are people who did not realize
how close they were to success when they gave up."
Thomas Edison

While the noncognitive abilities conscientiousness and self-control have been extensively studied for many years (see for example Rosenbaum, 1989; Wiggins, 1996) researchers' interest in the concept of grit is relatively new. Since the paper *Grit: Passion and Perseverance for Long-Term Goals* of Duckworth, Matthews, Kelly and Peterson (2007), the concept of grit has grown in popularity. The gritty individual is one who does not give up in the face of adversity or failure. Moreover, the gritty individual stays committed to its goal, even if this implies to invest in it for many years. As a result, grit is considered to be an essential attribute that all high-achievers have in common (Duckworth et al., 2007). This idea is in line with a

quote of Thomas Edison, one of the most renowned businessmen and inventors of the 19th and 20th century. What he identifies as the key to success is essentially what grit is about: failure itself is just an obstacle to overcome, therefore the only thing considered to be a failure is giving up. This idea does not only provide an interesting starting point for research, but is also an empowering message, freeing people from the notion that their skills are predetermined.

Grit was found to be positively associated with academic achievement (Duckworth et al., 2007). Furthermore, grit was found to be negatively associated with procrastination, indicating that grittier individuals are less prone to delay difficult tasks (Wyszyńska, Ponikiewska, Karaś, Najderska, & Rogoza, 2017). These findings are supported by research showing that students who scored high on grit in their junior year, were more likely to graduate from high school on time (Eskreis-Winkler, Shulman, Beal, & Duckworth, 2014). Moreover, grit plays an important role in athleticism and achievement in sports, as it is positively associated with deliberate practice and skill group membership and negatively associated with thoughts about quitting (Tedesqui & Young, 2018).

It was also observed that grittier individuals are more likely to pursue their chosen career paths, as individuals who score high on grit are less prone to change careers frequently (Duckworth et al., 2007). This is in line with a study that explored the influence of grit on the summer training program at the United States Military Academy West Point. Results showed that grit was positively associated with the completion of West Point's summer training program (Duckworth et al., 2007). In addition, grit appears not only to be linked to career change and career endorsement, but also to work engagement (Suzuki, Tamesue, Asahi, & Ishikawa, 2015).

Another domain grit is associated with is happiness. Grit relates positively to emotional, psychological and social well-being as well as to life satisfaction and harmony in life (Salles, Cohen, & Mueller, 2014; Vainio & Daukantaitė, 2016; Wyszyńska et al., 2017). According to the theory of authentic happiness (Seligman, 2002, 2011), there are three different orientations towards happiness: pleasure, engagement and meaning. An individual oriented towards pleasure seeks happiness through positive emotions like comfort, warmth or ecstasy. In contrast, individuals oriented towards engagement seek happiness through absorbing activities that consume their whole attention and consciousness. An example would be a musician who fully merges with the music and his instrument while playing. The third orientation towards happiness is meaning. It is defined as the pursuit of purpose in life. The focus lies on something

bigger than oneself, like advocating for human rights, belonging to a family or being religious (Seligman, 2011).

Research found that grit correlates significantly with each of the orientations to happiness, highlighting its importance for a happy and fulfilled life. Studies showed positive correlations between grit and orientations towards meaning and engagement. The same studies also found negative correlations between grit and an orientation towards pleasure (Suzuki et al., 2015; Von Culin et al., 2014). As the pursuit of happiness via pleasure may be defined as an approach towards "immediately hedonically positive activities" (Von Culin et al., 2014 ; Seligman, 2002), a negative correlation illustrates the focus of gritty individuals on long-term goals.

While the importance of grit for happiness and other life domains has been established, many gaps in our knowledge about grit remain. As grit highly correlates with conscientiousness and self-control, questions whether grit adds value beyond these two factors arose. It was argued that grit might not be a new construct, but rather a facet of conscientiousness (Credé et al., 2017; Duckworth & Quinn, 2009). Traditionally, the Big-5 personality trait conscientiousness combines qualities like order, dutifulness and competence (Wiggins, 1996). Similar to grit, conscientiousness is associated with academic achievement and job performance (Dumfart & Neubauer, 2016; Schmidt & Hunter, 1998; Suzuki et al., 2015). Nevertheless, conscientiousness differs from grit regarding long-term goals. While the conscientious individual focuses on short-term intensity goals, the gritty individual concentrates on valued long-term goals. Since a focus on long-term goals is vital for long-term achievement, it is assumed that grit, more than conscientiousness, is the quality which defines exceptionally successful individuals (Duckworth & Gross, 2014; Duckworth et al., 2007). Research about sport achievement supports this assumption, showing that grit is more strongly associated with sport expertise development than conscientiousness (Tedesqui & Young, 2018). Moreover, it was shown that grit significantly predicted long-term intensity exercise, while conscientiousness did not (Reed, Pritschet, & Cutton, 2013).

Another trait which correlates highly with grit is self-control (Duckworth et al., 2007; Duckworth & Gross, 2014). Self-control is the ability to control one's behavior and emotions. It is needed to inhibit impulses and resist temptations (Diamond, 2013). Self-control shares elemental processes with grit. Like grit and conscientiousness, self-control is associated with high levels of job performance as well as with academic performance (de Ridder, Lensvelt-Mulders, Finkenauer, Stok, & Baumeister, 2012; Duckworth et al., 2007; Dumfart & Neubauer, 2016; F. L. Schmidt & Hunter, 1998; Suzuki et al., 2015). Nevertheless, grit and self-control

relate to different forms of achievement. While self-control is important for everyday success, grit is essential for accomplishments that take years to reach (Duckworth & Gross, 2014).

Grit and self-control can be illustrated by means of a hierarchical goal framework (Duckworth & Gross, 2014). The hierarchical goal framework demonstrates the differences between the two abilities while simultaneously providing theoretical knowledge about the way grit works. Within this framework, a main goal is associated with several subordinate high-level goals. Those high-level goals are in turn associated with numerous low-level goals, which are linked to action tendencies (Figure 1). While rising within the hierarchical pyramid, the importance and complexity of the goals increase. Additionally, goals of a higher order require more endurance. Moreover, individuals can have multiple goal hierarchies at the same time. The differences between self-control and grit are reflected in the hierarchies they are linked to.

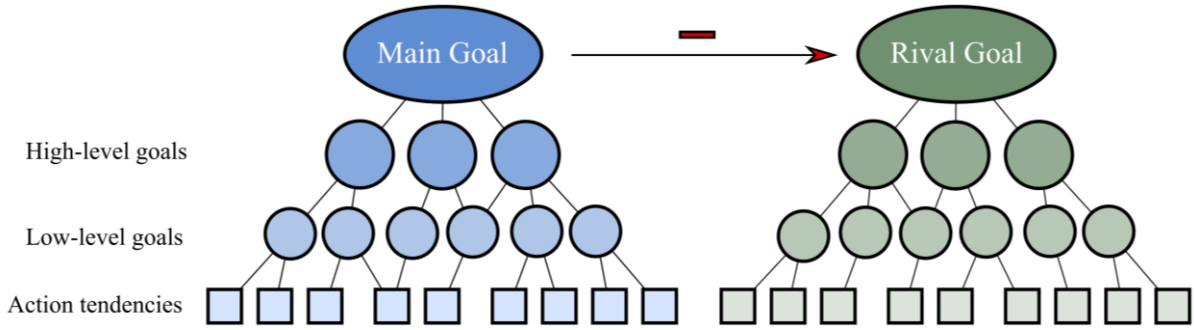

Figure 1. The main goal suppressing the rival goal. Adapted from Duckworth and Gross (2014).

Grit is needed to reach the main goal at the top of a hierarchy. To successfully reach the main goal, grit is required to suppress rival main goals from other hierarchies (Figure 1). For example, a student has two main goals of two different hierarchies: To become a doctor and to do a round-the-world trip. To suppress the rival goal of going on a world trip, she needs grit's passion and perseverance for her goal of becoming a doctor. Furthermore, grit is needed when confronted with setbacks (Duckworth & Gross, 2014). An example would be when a lower-order goal is needed to attain the main goal but turns out to be impossible to reach. Instead of giving up, the gritty individual would then search and establish a new lower-order goal or action tendency which corresponds to the main goal (Figure 2). For example, our future doctor had planned to get high grades and get accepted at med school. Unfortunately, her grades were average, and she got rejected. Instead of giving up, she stays gritty and looks for another way to become a doctor. She finds another med school which accepts students with average grades if they hold

a nurse certificate. She then generates the new goal of getting a nurse certificate. Thereupon, she gets her nurse certificate and afterwards applies for med school again.

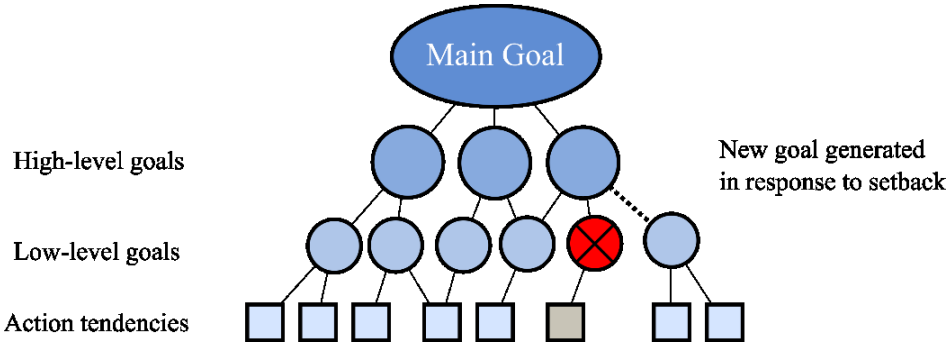

Figure 2. Generation of new goal in response to setback.
Adapted from Duckworth and Gross (2014)

Self-control takes place at the level of action tendency, which is at the lowest level of the hierarchical pyramid (Duckworth & Gross, 2014). When one action tendency connects to a less valued, but temporally stronger goal while a second action tendency connects to an enduringly valued goal, the individual will experience a conflict between the two actions. At those moments, self-control is required to decide on the momentarily weaker, but ultimately more valued goal.

An example of a situation when self-control is needed could be when a person has two different low-level goals (Duckworth & Gross, 2014): A student has the low-level goals to get good grades at a test and the low-level goal to spend time with her friends. The action tendency corresponding with the lower-level goal of achieving good grades would be to stay at home and study. In contrast, the action tendency for the low-level goal to spend time with friends would be to call them. Spending time with her friends is more fun and therefore has a momentarily higher value than getting good grades at the test. Nevertheless, she knows that her goal to get good grades helps her to achieve an important high-level goal of her: Passing the biology module she follows. In contrast, the high level-goal linked to the lower goal of spending time with her friends would be to go out more frequently. She values passing her biology module above her other goal of going out more frequently. Yet, she longs to call her friends instead of studying. In this situation, she needs self-control to suppress the urge to call her friends and instead study. Through recalling that her ultimately valued goal is to pass the biology module, she suppresses the rival goal's action tendency to call her friends.

1.2. Measuring Grit

Like self-control (see for example Patton, Stanford, & Barratt, 1995; Tangney et al., 2004), grit is frequently assessed via self-report questionnaires. It was first measured via the Grit Scale (Grit-O), a self-report questionnaire expected to show incremental predictive validity for high accomplishment over and beyond self-control (Duckworth et al., 2007). As there was a need to improve the model fit of the Grit-O, it was re-examined, resulting in the now widely used Grit Short Scale (Grit-S) (Duckworth & Quinn, 2009). To date, the Grit-O, the Grit-S and translated or adapted versions of the two questionnaires are the only measurement instruments of grit (see, for example Duckworth & Quinn, 2009; Duckworth et al., 2007; Schmidt, Fleckenstein, Retelsdorf, Eskreis-Winkler, & Möller, 2017; Wyszyńska et al., 2017). Moreover, all theoretical groundwork concerning grit relies on research of Duckworth et al. (2007) and Duckworth and Quinn (2009), who were the first to describe grit and who also developed the Grit-O and the Grit-S. They define grit as an ability comprised of "perseverance of effort" and "consistency of interest". While perseverance of effort refers to the quality to work hard and progressively towards a set goal despite adversity, consistency of interest implies a persistency of focus and passion for a set goal (Credé et al., 2017; Duckworth et al., 2007). Accordingly, Duckworth et al. (2007) present the Grit-S as multidimensional, with grit as a higher-order dimension, comprising the two lower-order dimensions consistency of interest (CI) and perseverance of effort (PE).

To gain insight into the structure and dimensionality of the Grit-S, various factor models have been investigated by means of confirmatory factor analysis (CFA). Initially, Duckworth and Quinn (2009) proposed a second-order factor model for grit, with CI and PE as first-order latent factors and grit as second-order latent factor. They reported adequate fit for this model. The second-order model also showed better fit to the data than a unidimensional model with grit as single factor, omitting CI and PE. Although these findings provide evidence for a higher-order structure of grit best represented by a second-order factor model, the structure of grit remains unclear. Findings of other research contradicted the results of Duckworth and Quinn (2009), showing that a modified second-order model with an error correlation between items 7 and 8 described the data better than the original model (Schmidt, Fleckenstein, Retelsdorf, Eskreis-Winkler, & Möller, 2017). Another studies suggest that CI and PE are two related but different dimensions on their own, rather than two lower-order dimensions combined by one of higher-order (see for example Tyumeneva, Kardanova, & Kuzmina, 2017). Furthermore, a study found that opposed to PE, CI did not load onto the higher-order factor, suggesting that CI is not related

to grit (Datu, Valdez, & King, 2016). Overall, these findings reject the assumption that grit is best represented by the original second-order factor model. Hence, more research regarding the best fitting model is needed.

Additionally, CI might have weaker predictive power than claimed. For example, one study reports a significant negative association of the CI subscale and freshmen's first year GPA, indicating that better performing students score lower on CI. PE, on the other hand, showed a significant positive association with first year GPA (Chang, 2014). Furthermore, another study reports that PE predicts key psychological outcomes like academic engagement and subjective well-being better than CI (Datu et al., 2016). Those results conflict with the research of Duckworth et al. (2007) who state that neither PE, nor CI predicts outcomes better than the other.

Moreover, few studies have investigated the measurement invariance of the Grit-S regarding different nationalities and age-groups. So far, samples of many studies consist either of students or young adults (see for example Datu et al., 2016; Schmidt et al., 2017; Tyumeneva et al., 2017; Wyszyńska et al., 2017). Even more studies are based on national samples (see for example Duckworth et al., 2007), missing out on additional insight more culturally diverse samples could bring about factors like item interpretation or the interpretation of grit in general. Consequently, little is known about the comparability of grit scores from different nationalities and different age-groups. To ensure that grit scores in future cross-cultural and cross-age studies are compared correctly, it is necessary to conduct more research regarding the measurement invariance of the Grit-S.

Lastly, the convergent validity of the Grit-S has previously been assessed based on conscientiousness and self-control (see, for example Schmidt et al., 2017), as they are the constructs most obviously related to grit (Duckworth & Gross, 2014). However, an alternative approach to analyze the convergent validity of the Grit-S could provide additional insight. Research found grit to also be positively associated with happiness and well-being (Suzuki et al., 2015; Von Culin et al., 2014), showing that they might be related to grit. In order to investigate the convergent validity of the Grit-S based on different constructs than consciousness and self-control, it will be based on measurements of well-being of the Mental Health Continuum Short Form (Keyes, 2002) and measurements of happiness of the Orientations to Happiness Scale (Peterson, Park, & Seligman, 2005; Seligman, 2002).

As former research shows, our knowledge about the measurement of grit is limited. To this date, there is no consensus about the structure and dimensionality of the Grit-S. Studies have investigated its dimensionality through comparison of different factor models. Unfortunately, many findings contradict each other. As a result, the connections and possible interactions between the assumed dimensions of grit, PE and CI, remain unclear. In particular, findings have raised questions about the connection between grit and CI and the validity of the CI subscale. Moreover, more research on the measurement invariance of the Grit-S is needed. Furthermore, alternative approaches to examine the convergent validity of the Grit-S could provide new insight. The present study aims to address the described problems through investigation of the psychometric properties of the Grit-S. The factorial structure and dimensionality, measurement invariance, reliability and convergent as well as criterion validity of the Grit-S will be examined based on an international sample consisting of adults of different age-groups.

.

2. Methodology

2.1. Procedure

The current study was part of a broader line of research of the University of Twente. Data was collected through a web-based online survey in October and November 2017. The survey included the informed consent, a demographic questionnaire, the Grit-S, the Orientations to Happiness Scale, the Mental Health Continuum Short Form and the Task Performance Scale of the Individual Work Performance Questionnaire (IWPQ 0.3) (Duckworth & Quinn, 2009; Keyes, 2002, 2005; Koopmans et al., 2014; Peterson, Park, & Seligman, 2005). To start the survey, participants had to indicate that they read and accepted the informed consent. The demographic questionnaire included questions on age, gender, occupation, work, education, language and nationality.

The survey was conducted in English and compiled via the research platform Qualtrics (www.qualtrics.com). After the ethics commission of the faculty of behavioral, management and social sciences approved the survey, distribution was started. The chosen sampling method was convenience sampling. The survey was distributed through the test subject pool systems SONA (of the University of Twente) and SurveyCircle (www.surveycircle.com). Additionally, it was distributed via email and through social media accounts (Facebook). An invitation letter was included on the sites of the subject pool systems, in the emails and on social media stating the purpose, aim and possible relevance of the survey. Moreover, the invitation letter assured the confidentiality and anonymity of participant's biographical details and survey results. To ensure that a broad range of employed participants from different fields was reached, a rural district in northern Germany was approached as well. The district consists of several towns and municipalities and is responsible for their administration. It employs 894 people working in several fields including legal, social, organizational, management and health departments as well as departments responsible for public order, education, culture, construction, nature conservation and water and waste management. The invitation letter to the survey was sent to all employees via their work-email address. All participants took part in the survey for free. However, three Amazon giftcards (à 20 Euros) were raffled among them. The total survey took approximately 12 minutes to complete (*Mdn* = 11.78 min).

2.2. Participants

The sample consisted of young, middle-aged and older adults ($N = 272$). Inclusion criteria were average or higher proficiency in English and age of 18 years or older. Of the original 317 participants, 45 did not complete the survey (14%). The remaining sample consisted of 272 participants of whom 186 (68.4%) were female, 82 (30.1%) were male and one was of other gender (0.4%). Participants were between 18 and 74 years old and the mean age was 31 ($SD = 13.1$) years. As the researchers of the study had different backgrounds and the sample was obtained via convenience sampling, the different backgrounds were reflected in the participants' nationalities. The predominant nationality of participants was German, followed by South African, Dutch, British and American (USA). The remaining participants were from Asian, Central American, North American or other African and European countries. The first language of most participants was German, followed by English, Afrikaans, Dutch and other languages. Furthermore, 52.6% of participants were either employed full time or part time. The remaining participants were mostly students. A relatively small percentage of participants was either retired or unemployed/ looking for work. Considering that 40.5% of the participants were between 18 - 24 years old, participants' overall level of education was relatively high with 97.1% holding at least a high school degree and 58.5% completing higher education programs.

Table 1.
Demographic characteristics of participants.

Demographics						
Nationality	German	South African	Dutch	Other	British	American (USA)
	59.9%	20.6%	8.1%	7.4%	2.2%	1.8%
First Language	German	English	Afrikaans	Dutch	Other	
	59.9%	14.7%	13.6%	7.7%	4.0%	
Occupation	Student	Employed full-time	Employed part-time	Retired	Un-employed	
	44.1%	43.8%	8.8%	1.8%	1.5%	
Level of Education	High-school	Master	Bachelor	PhD/ graduate work	Some college	No degree
	31.2%	22.4 %	19.9%	16.2%	7.4%	2.9%

Note. $N = 272$

2.3. Materials

Grit-S: The Grit-S (Duckworth et al., 2007) consists of eight items: four representing the CI subscale and four representing the PE subscale (Table 2). An example of an item of the CI subscale is "I have difficulty maintaining my focus on projects that take more than a few months to complete.". The PE subscale is represented by items such as "I finish whatever I begin.". Items are scored on a 5-point Likert scale from 1 = "not at all like me" to 5 = "very much like me" (Duckworth & Quinn, 2009). Former studies report acceptable to good reliability for the Grit-S, with Cronbach's α values ranging from .73 to .83. Cronbach's α values for the CI subscale were slightly better than the ones for the PE subscale (α values ranging from .73 to .79 and .60 to .78 respectively) (Duckworth & Quinn, 2009).

Table 2.
Grit-S with subscales and items.

Scale/item	
Grit-S	
Consistency of interest (CI)	
1	New ideas and projects sometimes distract me from previous ones.
3	I have been obsessed with a certain idea or project for a short time but later lost interest.
5	I often set a goal but later choose to pursue a different one.
6	I have difficulty maintaining my focus on projects that take more than a few months to complete.
Perseverance of effort (PE)	
2	Setbacks don't discourage me.
4	I am a hard worker.
7	I finish whatever I begin.
8	I am diligent.

Orientation to Happiness Scale: This scale comprises three subscales representing three different approaches towards happiness: orientation towards meaning, pleasure and engagement (Peterson et al., 2005; Seligman, 2002). Each subscale comprises six items resulting in 18 items for the overall scale. Items are rated on a 5-point Likert scale ranging from "1= not like me at all" to "5=very much like me".

The subscales focus on different attitudes and behavior regarding a respondent's individual well-being. For example, the subscale of meaning is represented by items such as "My life serves a higher purpose", whereas the subscale of pleasure is represented by items such as "Life is too short to postpone the pleasures it can provide". An example of an item of the engagement

subscale is “I am always very absorbed in what I do.” (Peterson et al., 2005; Seligman, 2002). In the present study, the Orientations to Happiness Scale showed acceptable reliability overall (α =.79). Observed internal reliabilities for the subscales of meaning (α = .78) and pleasure (α = .78) were also acceptable. However, the engagement subscale showed lower internal reliability (α= .64). These measures are in accordance with the ones in previous research, where acceptable internal consistency for the subscales meaning (α = .77) and pleasure (α = .74) and a lower reliability for the subscale engagement (α = .66) was found as well (G.-H. Chen, 2010).

Mental Health Continuum Short Form*:* The MHC-SF (adult version) consists of 14 items assessing three different types of well-being (Keyes, 2002, 2005): Emotional (three items), social (five items) and psychological well-being (six items). Participants rate items according to the frequency they experience the feelings described by the items (1= “never”, 2= “once or twice”, 3= “about once a week”, 4= “about two or three times a week”, 5= “almost every day” or 6 = “every day”).

All items are related to the occurrence of these feelings during the past month. For example, an item assessing emotional well-being is “During the past month, how often did you feel satisfied with life?”, which the participant scores via the described scale. Social well-being is measured via items such as “…how often did you feel that you belonged to a community (like a social group, your school, or your neighborhood?”. Lastly, psychological well-being is measured via items such as “…how often did you feel that you liked most parts of your personality?”. Studies report good internal consistency (α >.80) and construct validity for the adult version of the MHC-SF as well as moderate test-retest reliability. Furthermore, its 3-factor structure was confirmed and its convergent as well as discriminant validity were supported (Lamers, Westerhof, Bohlmeijer, ten Klooster, & Keyes, 2011). Likewise, the present study confirmed good internal consistency for the MHC-SF with a Cronbach’s α of .88 for the total scale. Reliabilities for the three subscales were acceptable, with an α of .79 for emotional well-being, an α of .77 for social well-being and an α of .79 for psychological well-being.

Task Performance Scale: The Task Performance Scale is a subscale of the Individual Work Performance Questionnaire (Koopmans et al., 2012). The scale is designed to measure task performance, defined as an individual’s performance on assignments and duties related to a specific job or function (Lievens, 2011). The task performance scale comprises seven items such as “I was able to separate main issues from side issues at work” and was scored on a 6-point rating scale ranging from 1= “seldom”, 2= “sometimes”, 3= “frequently”, 4= “often” to 5= “always” (Koopmans et al., 2014). The Task Performance Scale showed good internal

consistency with a Cronbach's α value of .84. Former research also supports the scale's good reliability with a Person Separation index (PSI) of .82 (Koopmans et al., 2014). The PSI can be interpreted similarly to Cronbach's α (Linacre & Wright, 1993).

2.4. Data Analysis

Factor analysis of the Grit-S was conducted by means of Mplus 7.11 (Muthén& Muthén, 2010). Other psychometric properties were investigated using SPSS Version 25 (IBM, 2017). All negative items were reverse coded.

Factorial validity: Factorial validity was determined through confirmatory factor analysis (CFA). Three nested models were analyzed: A unidimensional model with all items loading on one single factor, a second-order factor model with CI and PI as first order latent factors, loading on the second-order latent factor grit and a bifactor model with grit as general factor and CI and PE as group specific factors (Figure 3).

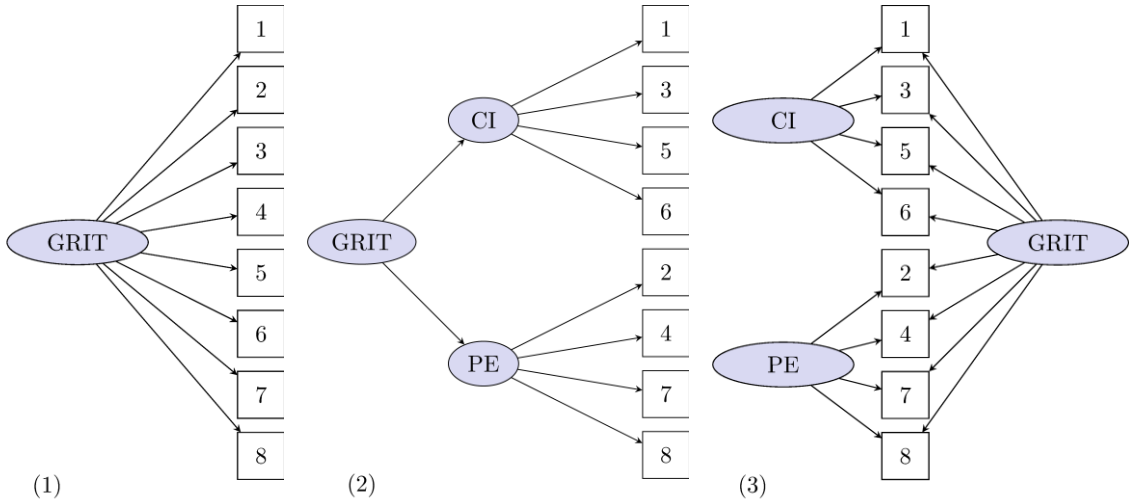

Figure 3. The analyzed models. The (1) unidimensional, (2) second-order and (3) bifactor model.

As item responses were measured on an ordinal level, it was opted for the robust weighted least squares estimator (WLSMV) (Byrne, 2012; Muthén & Muthén, 2010). Common statistical indices of CFA to evaluate model fit are the chi-square test, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root means square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) (Bialosiewicz, Murphy, & Berry, 2013; Byrne, 2012). For a better estimation of the model fit, it is recommended to examine several indices, as each index has individual advantages and disadvantages (Berkout, Gross, & Young, 2014). In case of the chi-square test, non-significant p-values and a small chi-square indicate good model fit (Bialosiewicz et al., 2013; Byrne, 2012). As a cutoff, a value of $p > .05$ was chosen.

However, results of the chi-square test should be interpreted carefully as it is sensitive to sample size and tends to reject models for large samples and often fails to reject models for small samples (Cochran, 1952). The CFI and TLI range from 0 to 1 and are interpreted similarly, with values greater than .90 indicating an acceptable (Brown, 2006), and values close to .95 indicating good model fit (Hu & Bentler, 1999). The smaller the RMSEA, the better the model fits. Values of .08 and lower may be viewed as indicating acceptable (Schermelleh-Engel, Moosbrugger, & Müller, 2003) and values below .06 as indicating good model fit (Hu & Bentler, 1999). As the Weighted Root Mean Square Residual (WRMR) tends to perform better with categorical data than the more common SRMR (Yu, 2002) and Mplus uses the WRMR for categorical data (Muthén & Muthén, 2010), WRMR was used instead of SRMR. Acceptable model fit is indicated by a WRMR value of 1.00 and lower, while good model fit is indicated by a value of .95 and lower (Yu, 2002). Comparison of different models was conducted via chi-square difference tests. As chi-square model fit indices calculated by means of WLSMV cannot be compared via a regular chi-square difference test, models were compared via Mplus' chi-square difference test ("difftest") for WLSMV and MLMV estimators (Muthén & Muthén, 2010).

Duckworth et al. (2007) originally proposed a second-order model for grit, emphasizing its multidimensionality with grit as general factor and PE and CI as secondary factors, rejecting a unidimensional model with only one factor. An increasingly popular method to gain further in-depth knowledge about multidimensionality is the computation of a bifactor model (Chen, West, & Sousa, 2006; Rodriguez, Reise, & Haviland, 2016). Hence, bifactor models have also been used to explore the multidimensionality of grit (see for example Li, 2015). Bifactor models consist of a general factor representing the overall construct to be measured and group factors representing particular subdomains of the overall construct (Rodriguez et al., 2016). Although bifactor models seem similar to second-order models in this regard, they hold several advantages over them. An important advantage of bifactor models is that the factor loadings can be used to determine the strength between group factors and their corresponding items, as they are independent of the general factor (Chen, West, & Sousa, 2006). Due to this advantage, strengths between items and group factors can be compared to strengths between items and the general factor via the standardized factor loadings. In general, high factor loadings indicate a strong connection between a latent factor and its corresponding items. Within bifactor models, the strengths between items and their corresponding factor display information about the dimensionality of the measurement instrument and accordingly provide insight into the

structure of the construct being measured. High loadings on group factors emphasize the group factor's importance for the overall model, indicating multidimensionality. Low factor loadings on the group factors with high loadings on the general factor on the other hand indicate unidimensionality (Chen et al., 2006; Rodriguez et al., 2016). Unidimensionality can be divided into strict and essential unidimensionality. Essential unidimensionality means that a model is mainly unidimensional but may hold minor secondary latent variables. Strict unidimensionality on the other hand does not allow secondary latent variables (Slocum-Gori, Zumbo, Michalos, & Diener, 2009). Within psychological research, strict unidimensionality is often too difficult to attain, as psychological constructs usually comprise (minor) secondary latent variables. Therefore, essential unidimensionality is often used as the sufficient condition required to declare unidimensionality (Slocum-Gori et al., 2009).

In contrast to the factor loadings of bifactor models, standardized factor loadings of second-order models cannot be used to directly study dimensionality, as their factor loadings are linked to the first order (general) factor (Chen et al., 2006; Rodriguez et al., 2016). Another advantage bifactor models hold over second-order models is that they can be used to determine the relative strength of an underlying general factor (Rodriguez et al., 2016). The relative strength can be calculated by means of the estimated common variance (ECV), which gives a more precise measure of multidimensionality than the mere evaluation of standardized factor loadings. ECV values indicate the percentage of common variance of data explained by the general factor, with high ECV values supporting essential unidimensionality of a construct (Rodriguez et al., 2016). To date, there are no strict cutoff values for ECV. Nevertheless, ECV values of .90 and greater may be viewed as indicating a strong general factor, affirming unidimensionality. ECV values below .70 may be viewed as indicating multidimensionality (O'Connor Quinn, 2014).

Although bifactor models provide valuable insight into the dimensionality of measurement instruments, most studies in present research do not make use of this advantage. Studies often aim to solely identify the best fitting model, the investigation of dimensionality is secondary (Rodriguez et al., 2016). Accordingly, bifactor models are frequently treated similar to other multidimensional models. This might be problematic, as research found bifactor models to be prone to "overfitting". That is, they not only model significant trends in the data but also unwanted noise, accommodating nonsense response patterns (Bonifay, Lane, & Reise, 2017). As a result, they tend to outperform other models in model comparison studies, although they might not always accurately represent the structure of the true population model (Bonifay et al., 2017). Consequently, comparisons of model fit between bifactor and other models should be

met with caution. Therefore, the primary use of the bifactor model in the present study was to assess dimensionality.

Measurement invariance: Measurement invariance indicates whether a measurement tool assesses the same underlying construct across different groups or across time. Measurement invariance is anchored in factor analysis, which is based on the theory that underlying (unobserved) constructs can be measured through indicator variables, like scale items or subscales. When participants with identical scores on the underlying construct have the same chance for the same scores on the observed indicator variables, regardless of their group or the time of the measurement, a measure is invariant (Bialosiewicz et al., 2013; Byrne, 2012; Byrne et al., 1989; Millsap, 2011; Yoon & Lai, 2018). The present study examines measurement invariance across groups, also referred to as multi-group invariance (Byrne, 2012). Measurement invariance was tested based on age (age 18-30 versus age ≥ 31) and nationality (European versus non-European). It was investigated on different levels which build upon each other in the following order: configural invariance, metric invariance and scalar invariance (van de Vijver & Tanzer, 2004). For configural invariance, it is necessary to investigate whether the same items measure the same construct across groups (Bialosiewicz et al., 2013; Byrne, 2012). To assess configural invariance, the overall model fit was tested using CFA. Thereupon the invariance of factor loadings across groups was assessed. This form of invariance is known as metric invariance. Metric invariance was evaluated through comparison of the configural model with the metric model by means of a chi-square difference test. Lastly, scalar invariance requires item intercepts to be the same across groups. To test for scalar invariance, the fit of the scalar model was contrasted with the one of the metric model. Again, this was carried out using a chi-square difference test. Scalar invariance is established if there is no significant discrepancy in model fit (Bialosiewicz et al., 2013; Byrne, 2012; van de Vijver & Tanzer, 2004).

To show that models did not differ significantly, a $p > .05$ for the chi-square difference test was aimed for. As the chi-square difference test, like the chi-square test (Cochran, 1952), is overly sensitive to sample size (Chen, 2007; Cheung & Rensvold, 2002) it is recommended to additionally look at other cut-off criteria for measurement invariance. Research shows that CFI, RMSEA and SRMR complement each other. Therefore, it is recommended to link cutoff criteria for them at different levels of measurement invariance (Chen et al., 2006). At the metric level, a shift of ≤ -.005 in CFI complemented by a shift of ≥ .010 in RMSEA or ≥ .025 in SRMR are recommended as cutoff criteria indicating non-invariance. At the scalar level, a shift of ≥ -.005 in CFI complemented by a shift of ≥ .010 in RMSEA or ≥ .005 in SRMR are recommended

as cutoff criteria indicating non-invariance. It should be noted that data was treated as continuous instead of ordered categorical to assess metric invariance of the second-order model. When treating data as ordered categorical, Mplus aborted the computation, as the full range of response categories was not used for some items within the non-European group. Also, MLR was used as estimator as WLSMV is not available for measurement invariance testing with continuous data. Hence, WRMR could also not be used and SRMR was used instead. Overall, SRMR values below .08 indicate acceptable, and values below .05 good model fit (Berkout et al., 2014; Hu & Bentler, 1999).

Reliability: Measurements of all questionnaires were tested for their reliability. Cronbach's α (Cronbach, 1951) was chosen as estimator. As a general rule, a Cronbach's α above .70 is interpreted as adequate (Rammstedt, 2004). An α value of .80 is commonly interpreted as good (Nunnally & Bernstein, 1994) and a value greater than .90 as very good (Weise, 1975).

Convergent validity: Concepts associated with grit are well-being and happiness (e.g. Salles et al., 2014; Suzuki et al., 2015). Grit correlates with different approaches to happiness (Seligman, 2011), showing that gritty individuals seek happiness through engaging activities, meaningful purposes and positive relationships with others (Suzuki et al., 2015; Von Culin et al., 2014). Frequently used instruments to assess well-being and happiness are the Orientations to Happiness Scale and the MHC-SF (Keyes, 2002, 2005; Peterson et al., 2005), which were chosen in the present analysis as well. To estimate the convergent validity of the Grit-S, measurements of the overall scale and its two subscales CI and PE were correlated with the three subscales meaning, pleasure and engagement of the Orientations to Happiness Scale. Furthermore, measurements of the Grit-S (overall, CI and PE) were correlated with the three subscales emotional, psychological and social well-being of the MHC-SF. Previous research shows grit to share weak significant as well as moderate significant correlations with all orientations towards happiness. While correlations with an orientation towards meaning and engagement were positive, correlations with an orientation towards pleasure were negative (Suzuki et al., 2015; Von Culin et al., 2014). Based on the previous research, it is expected that grit overall and the PE subscale will show weak to moderate significant, positive correlations with happiness orientations towards meaning and engagement. The CI subscale is expected to be positively associated with orientations towards engagement and meaning as well, showing weak but significant correlations. While correlations with an orientation towards pleasure are expected to be weak, significant and negative for the overall grit scale and CI, the correlation with PE is expected to be negligible. Furthermore, grit is positively linked to well-being,

showing weak to moderate significant correlations with all three subscales of the MHC-SF (Wyszyńska et al., 2017). Based on previous research, it is expected that the CI and PE subscale will show weak but significant positive correlations with emotional and social well-being. Correlations with psychological well-being were expected to also be positive and somewhat higher, with significant moderate correlations for CI and PE.

Criterion Validity: A prominent focus in research about grit is its connection with high achievement and gritty people are expected to accomplish higher levels of performance in their careers (Duckworth et al., 2007; Suzuki et al., 2015). A form of performance frequently associated with high achievement is task performance (see for example, Senko & Harackiewicz, 2005). Task performance is defined as "the effectiveness with which job incumbents perform activities that contribute to the organization's technical core either directly... or indirectly..." (Borman & Motowidlo, 1997; 1993). The Task Performance Scale of the IWPQ (Koopmans et al., 2012) was chosen to measure task performance as items are not specifically related to an organization and may be interpreted to task performance related to studying as well as working. As the sample in the present study consisted of young, middle-aged and older adults, the scale was considered appropriate. To investigate criterion validity of the Grit-S, Grit, CI and PE were correlated with task performance. To this date, few studies have examined the relation between grit and task performance. However, the relation between grit and task performance might resemble the one of grit and academic performance. Given that academic and task performance both reflect an individuals' effectiveness and engagement regarding an occupation (Smith, 1988), correlation between task performance and grit might be similar to the one between grit and academic performance. Overall, grit is positively associated with academic performance, showing Pearson correlation coefficients (r) to be small but significant (Duckworth et al., 2007). Based on these findings, grit is expected to positively correlate with task performance as well. Correlation coefficients were expected to also be small but significant.

As item responses were measured on an ordinal scale, Spearman's rho (r_s) was chosen as an appropriate correlation coefficient. While correlations of +1.00 indicate a perfect positive correlation, correlations of -1.00 indicate a perfect negative correlation. A correlation of .00 indicates that there is no relationship between the observed variables (Goodwin, 2010). A level of 95% ($p \leq .05$) was chosen as cutoff value for statistical significance. The interpretation of correlation coefficients was based on guidelines of Cohen (1988, 1992): values of .30 (-.30) and lower were interpreted as weak correlations, .30 to .50 (-.30 to -.50) as moderate and values greater than .50 (-.50) as strong correlations.

3. Results

3.1. Factorial validity

Similar to previous research (Duckworth et al., 2007), the unidimensional model showed poor fit on all indices: χ^2 (20) = 250.862 (p < .001), CFI = .767, TLI = .673, RMSEA = .206, WRMR = 1.919. Model fit of the originally proposed second-order model of Duckworth and Quinn (2009) was adequate. The CFI (.962), TLI (.944) and WRMR (.796) indicated good model fit, only the RMSEA (.085) was somewhat higher than the recommended cutoff for acceptable fit. The chi-square test of model fit (χ^2 (19) = 56.439, p < .001) indicated poor fit, which might be attributed to its high sensitivity to sample size. In comparison to the other items, item one ("New ideas and projects sometimes distract me from previous ones.") and two ("Setbacks don't discourage me.") had relatively low factor loadings, indicating they might be less representative for their respective subscale than the other items (Figure 4). A comparison of the two models by means of a chi-square difference test confirmed that the initial second-order model fitted the data better than the unidimensional model (χ^2 (1) = 65.524, p < .001).

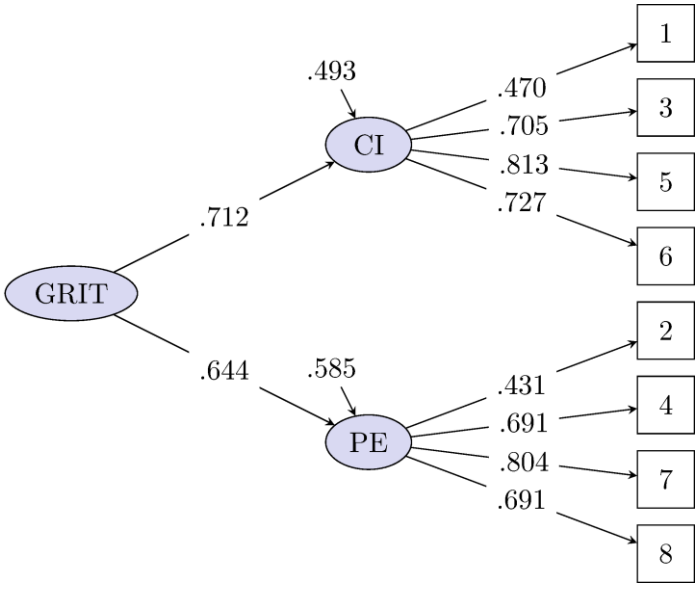

Figure 4. Second-order model with standardized factor loadings.

As expected, the bifactor model displayed good model fit overall: χ^2 (12) = 19.292 (p = .0817), CFI = .993, TLI = .983, RMSEA = .047, WRMR = .378. Chi-square difference tests between the bifactor model, the unidimensional and the second order model indicated that of the three models, the bifactor fit the data best: χ^2 (8) = 169.145 (p < .001) and χ^2 (7) = 32.848

($p < .001$) respectively. This comes as no surprise, as it was also the least restrictive model with fewer degrees of freedom than the unidimensional and the second-order model. Given the nature of bifactor models to overfit and since the second-order model fit the data adequately, even though it put more constraints on the data, further investigations of model fit were based on the second-order model. Hence, the bifactor model was only used to further investigate the dimensionality of grit.

Analysis of the bifactor model showed that the item covariances of six out of eight items of the Grit-S were better explained by CI and PE than the general factor grit (*Figure 5*), as indicated by the higher factor loadings. Only covariances of items seven ("I finish whatever I begin.") and eight ("I am diligent.") of the PE subscale were better explained by the general factor. Surprisingly, item seven loaded negatively on PE (-.005), whereas all other loadings in the model were positive. Out of all items, item seven also loaded highest on the general factor grit (.966). Overall, the results indicate that grit is, as assumed, multidimensional. The findings were also supported by an ECV value of .48, showing that the general factor grit explained only 48% of the common variance, while the group specific factors explained 52%.

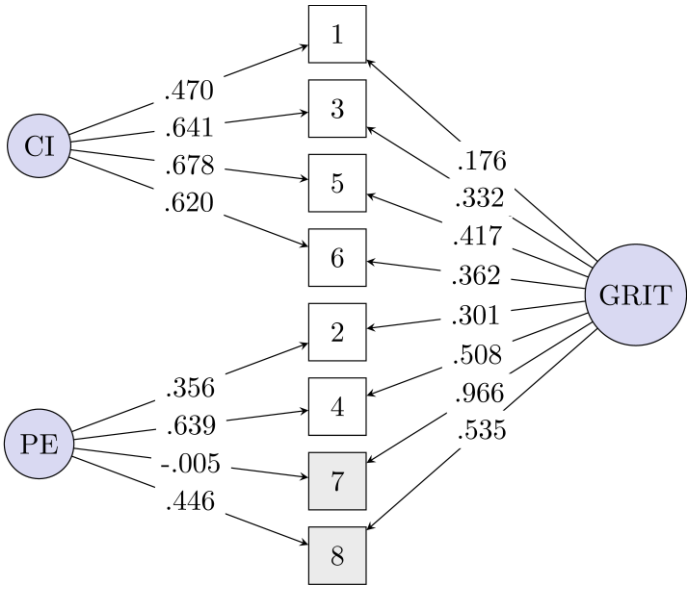

Figure 5. Bifactor model with standardized factor loadings.

3.2. Measurement invariance

Measurement invariance was assessed for the second-order model. When testing for configural invariance across age, the model showed acceptable fit overall (χ^2 (38) = 69.154, p = .0015; CFI = .917; TLI = .877; RMSEA = .080; SRMR = .063. As can be seen in Table 2, the Grit-S showed strong measurement invariance at the metric level across both age groups.

However, a significant p-value for scalar against metric invariance indicated non-invariance at the scalar level. A change of -.021 in CFI supplemented by a change of .003 in RMSEA and .009 in SRMR confirmed scalar noninvariance as well. Across both nationality groups, the second-order model showed strong configural invariance (χ^2 (38) = 42.767, p = .274; CFI = .988; TLI = .982; RMSEA = .030; SRMR = .048). Yet, the second order model was noninvariant for nationality at the metric as well as the scalar level.

Table 3.
Invariance testing for the second-order model based on age and nationality.

Second-order model	**CFI** (Δ)	**RMSEA** (Δ)	**SRMR** (Δ)	**χ2**	***df***	***p***
Age						
configural model	.917	.080	.063	69.154	38	.0015
metric model	.917 (.000)	.074 (-.006)	.071 (.008)	74.828	44	.0026
scalar model	.896 (-.021)	.077 (.003)	.080 (.009)	88.654	50	.0006
metric against configural				5.077	6	.5340
scalar against metric				14.020	6	.0294
Nationality						
configural model	.988	.030	.048	42.767	38	.2738
metric model	.966 (-.022)	.048 (.018)	.071 (.023)	57.458	44	.0839
scalar model	.943 (-.023)	.057 (.009)	.088 (.017)	72.263	50	.0214
metric against configural				15.171	6	.0190
scalar against metric				14.166	6	.0278

3.3. Reliability

Reliability measurements as well as means and standard deviations for the items of the Grit-S are provided in Table 4. The Grit-S showed acceptable internal consistency for the overall scale (α = .738) and the CI subscale (α = .739). The PE subscale showed inadequate internal consistency (α = .679).

Table 4.
Items of the Grit-S with Cronbach's alphas, means and standard deviations.

Scale	Item	Item keywords	α	M	SD
Grit-S			.738		
Consistency of interest (CI)			.739		
	1	"Ideas"		2.84	0.939
	3	"Interest"		3.42	1.066
	5	"Goal"		3.47	0.990
	6	"Focus"		3.42	1.104
Perseverance of effort (PE)			.697		
	2	"Setbacks"		3.16	1.007
	4	"Worker"		3.90	0.923
	7	"Finish"		3.79	0.922
	8	"Diligent"		3.84	0.919

3.4. Convergent validity

Correlation coefficients of measurements of the Grit-S with measurements of the Orientations to Happiness Scale, the MHC-SF and the Task Performance Scale are displayed in Table 5. Correlations between the Grit-S and the subscales of the Orientations to Happiness Scale were weak but significant ($p < .01$). As hypothesized, grit correlated positively with orientations towards meaning and engagement and negatively with an orientation towards pleasure. Correlation coefficients between grit and the subscales of the MHC-SF were also as expected, showing small to moderate, significant positive correlations. While correlations between grit and emotional well-being and grit and social well-being were small, correlations between grit and psychological well-being were moderate.

Table 5.
Correlation coefficients of the Grit-S with the other questionnaires.

	MHC-SF E	MHC-SF P	MHC-SF S	OTH M	OTH P	OTH E	TP
Grit	.206**	.316**	.238**	.283**	-.177**	.240**	.449**
CI	.136*	.180**	.172**	.083	-.263**	.007	.283**
PE	.193**	.355**	.203**	.399**	-.016	.398**	.479**

Note. Spearman's rhos for the Grit-S and the different subscales of the MHC-SF (emotional "E", psychological "P" and social "S" well-being), the Orientations to Happiness Scale (orientation towards meaning "M", pleasure "P" and engagement "E") and for the Task Performance Scale "TP"; significance is flagged*(p < .01**; p < .05*)

Investigation of the subscales CI and PE of the Grit-S gave a more detailed picture about the associations between grit and measurements of the other questionnaires. PE correlated stronger than CI with each of the different subscales, except for orientation to pleasure. In particular correlations between PE and orientations to meaning and engagement differed a lot with the ones of CI and these subscales. While PE and the subscales showed moderate, significant correlations, correlations for CI were negligible. Both, CI and PE, were negatively associated with orientation to pleasure, but the correlation with PE seemed negligible as it was close to zero and not significant. Correlation between CI and orientation to pleasure was stronger and significant. Correlations between CI, PE and the MHC-SF were similar to the one for the overall scale. An exception was psychological well-being, which shared a stronger, moderate correlation with PE.

Overall, correlations between grit and the different subscales were as expected: the correlations were positive, small to moderate and significant. The correlation between grit and the subscale of orientation towards pleasure also met expectations, showing a small, but significant negative correlation. The results support convergent validity of the Grit-S, providing evidence that grit is related to happiness and well-being.

3.5. Criterion validity

Grit was associated with task performance, showing a significant positive correlation just below the cutoff for moderate correlations (Table 5). The correlation between grit and task performance exceeded expectations as it was even stronger than the correlations between grit and academic performance in previous studies (Duckworth et al., 2007). CI and PE both had significant positive correlations with task performance as well. The correlation between PE and task performance was the strongest. Similar to the overall grit scale, its value was somewhat below the cutoff for moderate correlations. Results indicate that grittier people also have higher levels of task performance, supporting the criterion validity of the Grit-S.

4. Discussion

Overall, the results regarding the structure and multidimensionally of the Grit-S compare well with findings of Duckworth and Quinn (2009) and Duckworth et al. (2007). CFA showed the originally proposed second-order model to explain the data adequately. By nature, second-order models are multidimensional, comprising one or more second-order latent factors and at least two first-order factors (Chen et al., 2006). Hence, the results provide more evidence for the multidimensionality of grit. The better fit of the second-order model over a unidimensional model was in good agreement with findings obtained during analysis of the bifactor model.

Investigation of the factor loadings of the bifactor model showed that CI, PE and the general grit factor all added value independent of their interactions. Those findings were further supported by the ECV, which showed that the group specific factors CI and PE explained more than half of the common variance, while the rest was explained by the general factor. This highlights the importance of distinguishing between the group specific factors and the general factor, supporting the fit of a second-order model including all three factors. At large, the results support the factorial validity of the Grit-S. They provide evidence that grit is best represented by an overall grit scale, comprising one subscale measuring consistency of interest and another one measuring perseverance of effort. The items of the Grit-S seem to be a good representation of consistency of interest and perseverance of effort as well as of grit overall.

The results also give a more detailed picture about the interactions between grit, CI and PE addressed by critics. A previous study found that CI did not load on grit, indicating that CI might not be a subscale of grit but rather a scale on its own (Datu et al., 2016). These findings were contradicted in the present study since CI had a high factor loading on grit. Yet, the bifactor model showed that CI explained all item covariances better than grit. CI shared high factor loadings with items three, five and six, whereas all items of the CI subscale had relatively low factor loadings on the overall grit factor. In comparison, variance of most items of the PE subscale was about equally as well explained by PE as by grit. While these findings do not support that CI and grit are unrelated, they do indicate that the general factor grit explains less variance in the CI subscale than in the PE subscale.

The only item of the PE subscale which was considerably better explained by the general factor than the group specific factor was item seven. Item seven had a very high positive factor loading on grit while having a very weak negative factor loading on PE. This might indicate that item seven is not a good representation of perseverance of effort. Instead, item seven might

be representative of a yet unknown factor. This might also be reflected in the wording of the item ("I finish whatever I begin."). Taken literally, the sentence implies that upon agreeing with the item, a respondent indicates to finish every action he or she has ever started or will start. This might not be compatible with PE, as putting effort into important long-term goals is not to the same as putting effort into "whatever" one has ever begun or will begin.

Apart from item seven, items one and two also differed from the others in that they had the lowest factor loadings in both the second-order as well as the bifactor model. In contrast to item seven, they did not load highly on the general grit factor either. This suggests that items one and two are less representative of their respective subscales and grit overall than the other items. Previous research has identified problems with item two as well. It was pointed out that item two "Setbacks don't discourage me" is a double negative (Li, 2015), in that both "discourage" and "don't" are negative. Furthermore, the word "setback" also has a negative connotation. Negatively phrased items can lead to biased answers, as they are more challenging for respondents to process than positive ones (Goodwin, 2010). Hence, item two might have been misinterpreted by some respondents, resulting in lower factor loadings for item two. The problem with item one ("New ideas and projects sometimes distract me from previous ones.") might be unclear phrasing. The part "previous ones" could either be interpreted as former projects being (successfully) completed or not being completed. The Developers of the Grit-S (Duckworth & Quinn, 2009; Duckworth et al., 2007) most likely meant that previous projects were not completed due to new projects distracting the respondent. However, if respondents interpreted previous projects to already be completed when starting a new project, the item would be illogical since one cannot be distracted from a project which is not of importance anymore. Interestingly, item one is not the only complex item of the CI subscale. In fact, all items of the CI subscale are more complex and longer than the ones of the PE subscale. Moreover, CI's items describe negative actions or attributes (e.g. item eight "I have been obsessed with a certain idea or project for a short time but later lost interest."), while PE's items describe positive ones (e.g. "I am a hard worker"). Future research could explore the influence of negatively framed items of CI by compiling a positively framed CI subscale and comparing it to the original one. For example, the negatively phrased item eight could be replaced by a positive one: "I have been obsessed with a certain idea or project for a long time and not lost interest". An even shorter and less complex version could be "I stay passionate about certain things for a long time".

The analysis of measurement invariance showed the Grit-S to be noninvariant for age at the scalar level. Hence, item intercepts for respondents aged 18 to 30 were unequal to item intercepts for respondents aged 31 and older. This indicates that one group systematically gave higher or lower responses than the other (Byrne, 2012; Vandenberg & Lance, 2000; Xu, n.d.). Comparison of the different nationality groups not only showed the groups to be noninvariant at the scalar level, but also on the less strict metric level. Europeans and non-Europeans had unequal factor loadings, suggesting the groups might have interpreted the meaning of items differently (Byrne, 2012; Vandenberg & Lance, 2000; Xu, n.d.). Like for the age groups, noninvariance at the scalar level indicated that item intercepts of Europeans were unequal to item intercepts of non-Europeans. Accordingly, one group systematically gave higher or lower responses than the other (Byrne, 2012; Vandenberg & Lance, 2000; Xu, n.d.).

Measurement invariance could also have been impacted by sample size. The sample of respondents aged 30 and older consisted of only 91 people and within nationality, only 71 respondents were non-European. Evaluation of measurement invariance in small samples is highly complex as small samples have less statistical power than bigger samples (Meade, 2005) and measurement invariance is evaluated based on chi-square tests which are sensitive to sample size (Cheung & Rensvold, 2002; Cochran, 1952). The fact that non-invariance was detected nonetheless might indicate that it has been underestimated. Hence, non-invariance might be even more prominent in larger samples. As results show the Grit-S to be non-invariant across different nationalities and age groups, care should be taken when using the Grit-S in cross-cultural and cross-age studies. Scores from different nationalities and age groups may not be directly comparable, since item interpretation seems to differ across groups. In order to verify whether non-invariance is indeed more prominent in larger samples, future studies should include larger samples when testing for measurement invariance of the Grit-S.

Differences between the younger and older age group could also have been due to the diversity of the older age group: while the age gap between the youngest and the oldest respondents of the young age group (18-30) was 12 years, the youngest and oldest respondent of the older age group (30+) was 43 years. Likewise, the non-European group was more heterogenous than the European group, as it included respondents from all over the world. To pinpoint the exact differences between groups, the groups themselves should be more homogenous. For example, a comparison of individual countries, like Germany and South Africa could add more value than comparing large, heterogenous groups like Europeans and non-Europeans. Cultures as well as different generations are often characterized by different norms, values and attitudes.

Individuals of one country are more likely to share the same culture than individuals of different countries or even different continents. Likewise, people of one generation are more likely to share the same values and attitudes than younger or older generations. If differences between groups are substantial, but differences of participants within the respective groups are small, possible factors leading to non-invariance could be identified more easily. The Grit-S could then be adapted accordingly, and different nationalities or cultures could be compared more accurately.

Overall, expectations regarding the associations of the Grit-S with the subscales of the orientation to happiness scale were met and thus support convergent validity for the Grit-S. As hypothesized, grit showed positive, weak and significant correlations with happiness orientations towards meaning and engagement. PE shared somewhat stronger positive correlations with orientations towards meaning and engagement, which were moderate and significant. Correlations between CI and orientations towards meaning and engagement were positive but lower as expected. Against expectations, they were also not significant. However, a correlation between CI and a happiness orientation towards pleasure was, as expected, negative, small and significant. Although the correlations of CI with orientations towards engagement and meaning did not meet the specific expectations, the correlations between CI and the subscales of the orientation to happiness scale supported an expected trend: CI was positively correlated with orientations towards meaning and engagement and negatively with an orientation towards pleasure. Correlations of grit and PE with an orientation towards pleasure were as expected. While the correlation between grit and pleasure was negative, small and significant, the correlation between PE and pleasure was negligible.

A more in-depth analysis of CI and PE indicates that CI and PE impact happiness differently. While a persistency of effort might contribute to the seeking of happiness through meaning and engagement, a lack of consistency of interest might contribute to the seeking of happiness through pleasure. Not only is PE positively associated with meaning and engagement, but it seems not to be connected to an orientation towards pleasure at all. CI on the other hand, is negatively associated with pleasure but not connected to meaning and engagement. Seeking happiness through meaning or engagement differs from seeking happiness through pleasure in that it requires an individual to invest time and effort in the long term (Seligman, 2011). For example, someone is only able to engage in playing the piano if he or she is willing to put time and work into it. Likewise, individuals will not find meaning if they are not committed to the purpose in life they chose for themselves. This differs from an orientation towards pleasure,

which is characterized by the seeking of immediate positive emotions, for example through going shopping, taking drugs or watching a movie (Seligman, 2002, 2011). Accordingly, PE seems to be connected to the fulfillment of long-term goals which are crucial for meaning and engagement. On the other hand, CI seems to be important for the oppression of undesirable action tendencies, inhibiting the seeking of pleasure. This might indicate a connection between CI and self-control, which is also involved in the oppression of action tendencies (Duckworth & Gross, 2014). According to the hierarchical goal frame work of grit proposed by Duckworth and Gross (2014), grit is responsible for the oppression of rival main goals and the generating of new goals in response to setbacks. This seems to be important when seeking happiness through meaning, since the experience of meaning seems to be the main goal for individuals choosing this path. Moreover, happiness through meaning and engagement cannot be reached if one cannot handle setbacks. Anyone putting effort into something will experience setbacks at some point, individuals seeking happiness through engagement, too. For example, someone training for a marathon will sometimes run slower than aimed for during a training or get sick for some time, throwing him or her off schedule. The positive association between PE and meaning and engagement might indicate that perseverance of effort is the dimension responsible for dedication to long-term goals and the handling of setbacks.

Grit was also positively associated with well-being, providing additional evidence for the convergent validity of the Grit-S. The positive association of grit with well-being is consistent with grit also being positively associated with happiness, as well-being and happiness are closely related (Seligman, 2011). Of all three forms of well-being, grit was strongest associated with psychological well-being. Psychological well-being impacts the way an individual functions in all aspects of life. It is characterized by self-acceptance and development, positive relationships with others, autonomy and purpose and determination (Keyes, 2002). Like seeking happiness through meaning, psychological well-being correlated strongest with PE. Psychological well-being seems to overlap with the seeking of happiness through meaning regarding purpose and determination in life, which might account for the shared connection with PE.

Associations with the overall grit scale as well as associations with CI and PE were weaker for emotional and social well-being than for psychological well-being. A reason for this might be that social well-being is more dependent on other people than psychological well-being. While social well-being is greatly impacted by relationships with others, psychological well-being depends on someone's own view and evaluation of themselves (Keyes, 2002). Although

all forms of well-being are connected (Keyes, 2002) and grit was shown to be positively associated with all of them, the stronger connection between grit and psychological well-being seems coherent. Like psychological well-being, grit seems to be an ability primarily depending on the mind-set and actions of the individual and less on the social environment. Emotional well-being on the other hand might differ from psychological well-being in that it is more reflective of feelings and affect and less about functioning in life (Keyes, 2002). Similar to psychological well-being, grit is about the way an individual functions and handles certain aspects of life and less about emotions. This might account for the stronger positive association between grit and psychological well-being and the weaker association between grit and emotional well-being.

Lastly, grit was also positively associated with task performance, confirming the criterion validity of the Grit-S. PE had a stronger correlation with task performance than CI, suggesting that perseverance of effort might be more crucial for a good performance than consistency of interest. PE embodies the ability to work hard and progressively towards a set goal, even when faced with setbacks, whereas consistency of interest refers to the quality to maintain enthusiasm for a set goal even when faced with distractions (Duckworth et al., 2007). Task performance might be more closely related to PE as it refers to an individual's performance on assignments and duties related to a specific job (Lievens, 2011). Hence, task performance reflects the effort put into assignments and duties. Enthusiasm and interest for the duties and assignments might be secondary, which could explain the weaker association between task performance and CI. Although the positive association between grit and task performance supports the criterion validity of the Grit-S, future research should aim to find additional measurement instruments of task performance other than a self-report questionnaire. In this way, the probability for the occurrence of method effects could be reduced. Method effects can occur when two different constructs are assessed via the same or similar methods (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Roth, Hearp, & Switzer, 1999), like grit and task performance. Both are assessed via self-report questionnaires, they are even assessed via the same survey and at the same time. Measuring constructs in the same context and via the same method could artificially inflate covariances (Podsakoff et al., 2003), resulting in higher correlations of constructs than they actually have. This might especially apply for self-report questionnaires, as their response sets are prone to social desirability (Goodwin, 2010; Larsen, Buss, & Wismeijer, 2013). For future research, a stronger indicator of the criterion validity of the Grit-S could for example be its association with task performance measured based on observations or actual achievements.

Overall, the results of the present study challenge findings of Duckworth et al. (2007) who stated that neither PE, nor CI predicts outcomes better than the other. Results clearly show that PE is more strongly associated with happiness orientations towards meaning and engagement, with all forms of well-being and with task performance than CI. Although results are based on a correlational analysis and do not mirror the predictive validity of CI and PE, they suggest that PE and CI differ in their relations with external constructs. Hence, they are in good agreement with previous studies which found that PE was a better predictor of academic engagement and subjective well-being than CI (Datu et al., 2016).

While CFA and correlational analysis support the validity of the Grit-S overall, results also raise questions regarding the validity of the CI subscale. Results indicate that CI might have weaker predictive power than PE. A possible explanation for these findings could be that consistency of interest is indeed a component of grit, but a less prominent one than perseverance of effort. Another explanation could be rooted in the items of the CI subscale. As described previously, long and negatively phrased items of the CI subscale could have been too complex and have led to misinterpretations. Furthermore, people might be more likely to agree with positive statements about themselves than with negative ones due to social desirability (Goodwin, 2010; Larsen et al., 2013). Although the present study provides evidence for important differences between the PE and CI subscale, more research is needed to investigate them on a more in-depth level. An adapted version of the CI subscale with less complex and negative items could bring new insight into the causes of the differences and potentially increase the validity of the CI subscale.

In conclusion, results show the Grit-S to be non-invariant across different nationalities and age groups, indicating that scores from different groups may not be directly comparable. Future studies should consider possible non-invariance between different nationalities and age groups and adapt comparisons between groups accordingly. Additionally, it is necessary to conduct more research about the measurement invariance of the Grit-S in general, in order to examine whether characteristics of participants other than nationality and age impact the measurement of grit. Moreover, analysis of the Grit-S suggests problems regarding the items and the validity of the CI subscale, which should be addressed in future studies. Future work could also investigate the criterion validity of the Grit-S based on a measurement instrument of task performance other than a self-report questionnaire. Ideally, grit should be assessed first and task performance at a later point in time (Podsakoff et al., 2003).

Nevertheless, results show that the Grit-S is a valid and reliable measurement instrument for the assessment of grit. The present study provides evidence for the factorial validity, internal consistency, convergent and criterion validity of the Grit-S in an international context. Moreover, results support the fit of a second-order model and confirm the multidimensionality of the Grit-S. Furthermore, the present study demonstrates the importance of distinguishing between the subscales consistency of interest and perseverance of effort.

5. Bibliography

Aristotle, & Kenny, A. (2011). *The Eudemian Ethics*. Oxford: St. John's College.

Berkout, O. V., Gross, A. M., & Young, J. (2014). Why So Many Arrows? Introduction to Structural Equation Modeling for the Novitiate User. *Clinical Child and Family Psychology Review*, *17*(3), 217–229. https://doi.org/10.1007/s10567-014-0165-3

Bialosiewicz, S., Murphy, K., & Berry, T. (2013). *An Introduction to Measurement Invariance Testing: Resource Packet for Participants: Do our Measures Measure up? The Critical Role of Measurement Invariance*. Presented at the Demonstration Session American Evaluation Association, Washington DC. Retrieved from http://comm.eval.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=63758fed-a490-43f2-8862-2de0217a08b8

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*(1), 184–186.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In *Personnel selection in organiratlons*. San Francisco: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance*, *10*(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3

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

Byrne, B. M. (2012). *Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York: Routledge Academic.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456.

Chang, W. (2014). *Grit and academic performance: Is being grittier better?* (PhD Thesis). University of Miami.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F., West, S., & Sousa, K. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Chen, G.-H. (2010). Validating the orientations to happiness scale in a Chinese sample of university students. *Social Indicators Research*, *99*(3), 431–442.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Cochran, W. G. (1952). The χ2 test of goodness of fit. *The Annals of Mathematical Statistics*, 315–345.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. USA: Lawrence Erlbaum Associates.

Cohen, J. (1992). A Power Primer, *Psychological Bulletin [PsycARTICLES]*(112), 155–159.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

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*, *35*(1), 121–130. https://doi.org/10.1007/s12144-015-9374-2

de Ridder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review*, *16*(1), 76–99.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Duckworth, A., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science*, *23*(5), 319–325.

Duckworth, A. L. (2013, April). *Der Schlüssel zum Erfolg? Durchhaltevermögen.* TED Talks Education. Retrieved from https://www.ted.com/talks/angela_lee_duckworth_grit_the_power_of_passion_and_perseverance?language=de

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

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

Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness Is the Most Powerful Noncognitive Predictor of School Achievement in Adolescents. *Journal of Individual Differences*, *37*(1), 8–15. https://doi.org/10.1027/1614-0001/a000182

Dweck, C. S. (2017). *Mindset: Changing The Way You think To Fulfil Your Potential*. London: Robinson, an imprint ofConstable and Robinson Ltd.

Enkvist, Ekström, & Elmståhl, S. (2013). Associations between cognitive abilities and life satisfaction in the oldest-old. Results from the longitudinal population study Good Aging in Skåne. *Clinical Interventions in Aging*, 845. https://doi.org/10.2147/CIA.S45382

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*. https://doi.org/10.3389/fpsyg.2014.00036

Goodwin, C. J. (2010). *Research in Psychology: Methods and Design*. USA: John Wiley & Sons.

Hill, P. L., Burrow, A. L., & Bronk, K. C. (2016). Persevering with Positivity and Purpose: An Examination of Purpose Commitment and Positive Affect as Predictors of Grit. *Journal of Happiness Studies*, *17*(1), 257–269. https://doi.org/10.1007/s10902-014-9593-5

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–362.

IBM. (2017). *IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.*

Keyes, C. L. M. (2002). The Mental Health Continuum: From Languishing to Flourishing in Life. *Journal of Health and Social Behavior*, *43*(2), 207. https://doi.org/10.2307/3090197

Keyes, C. L. M. (2005). Mental Illness and/or Mental Health? Investigating Axioms of the Complete State Model of Health. *Journal of Consulting and Clinical Psychology*, *73*(3), 539–548. https://doi.org/10.1037/0022-006X.73.3.539

Koopmans, L., Bernaards, C., Hildebrandt, V., van Buuren, S., van der Beek, A. J., & de Vet, H. C. W. (2012). Development of an individual work performance questionnaire. *International Journal of Productivity and Performance Management*, *62*(1), 6–28. https://doi.org/10.1108/17410401311285273

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Van Buuren, S., Van der Beek, A. J., & De Vet, H. C. (2014). Improving the individual work performance questionnaire using rasch analysis. *J Appl Meas*, *15*(2), 160–175.

Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., Klooster, P. M. ten, & Keyes, C. L. M. (2011). Evaluating the psychometric properties of the mental health Continuum-Short Form (MHC-SF). *Journal of Clinical Psychology*, *67*(1), 99–110. https://doi.org/10.1002/jclp.20741

Larsen, R., Buss, D., & Wismeijer, A. (2013). *Personality Psychology: Domains of Knowledge about Human Nature*. Berkshire: McGraw-Hill Higher Education.

Li, C. (2015). The Short Grit Scale: A Dimensionality Analysis.

Lievens, F. (2011). *Human Resource Management: Back to Basics*. Den Haag: Lannoo nv Tielt & Academic Service.

Linacre, J. M., & Wright, B. D. (1993). *A user's guide to BIGSTEPS: Rasch-model computer program*. Chicago: Mesa Press.

Meade, A. W. (2005). Sample size and tests of measurement invariance. In *Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA*.

Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology*, *71*(3), 281.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York: Routledge Taylor & Francis Group.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus: Statistical analysis with latent variables: User's guide*. Muthén & Muthén Los Angeles.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (Vol. 3. Ed.). New York: McGraw-Hill.

O'Connor Quinn, H. (2014). *Bifactor Models, Explained Common Variance (ECV), and the Usefulness of Scores from Unidimensional Item Response Theory Analyses*. University of North Carolina at Chapel Hill, Chapel Hill.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *John Wiley & Sons Inc.*, *Journal of Clinical Psychology*(6), 768–774.

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: the full life versus the empty life. *Journal of Happiness Studies*, *6*(1), 25–41. https://doi.org/10.1007/s10902-004-1278-z

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Rammstedt, B. (2004). Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung.

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

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosenbaum, M. (1989). Self-control under Stress: The Role of learned Resourcefulness, (Vol. 11), pp. 249–258.

Roth, P. L., Hearp, C., & Switzer, F. S. (1999). The effect of method variance on relationships between the work ethic and individual difference variables. *Journal of Business and Psychology*, *14*(1), 173–186.

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

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262.

Schmidt, F. T. C., Fleckenstein, J., Retelsdorf, J., Eskreis-Winkler, L., & Möller, J. (2017). Measuring Grit: A German Validation and a Domain-Specific Approach to Grit.

European Journal of Psychological Assessment, 1–12. https://doi.org/10.1027/1015-5759/a000407

Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment.* New York: Free Press.

Seligman, M. E. P. (2011). Retrieved from https://www.authentichappiness.sas.upenn.edu/learn/wellbeing

Senko, C., & Harackiewicz, J. M. (2005). Achievement Goals, Task Performance, and Interest: Why Perceived Goal Difficulty Matters. *Personality and Social Psychology Bulletin*, *31*(12), 1739–1753. https://doi.org/10.1177/0146167205281128

Slocum-Gori, S. L., Zumbo, B. D., Michalos, A. C., & Diener, E. (2009). A Note on the Dimensionality of Quality of Life Scales: An Illustration with the Satisfaction with Life Scale (SWLS). *Social Indicators Research*, *92*(3), 489–496. https://doi.org/10.1007/s11205-008-9303-y

Smith, D. M. (1988). On academic performance. *Area*, 3–13.

Stokas, A. G. (2015). A genealogy of grit: Education in the new gilded age. *Educational Theory*, *65*(5), 513–528.

Suzuki, Y., Tamesue, D., Asahi, K., & Ishikawa, Y. (2015). Grit and Work Engagement: A Cross-Sectional Study. *PLOS ONE*, *10*(9), e0137501. https://doi.org/10.1371/journal.pone.0137501

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324.

Tedesqui, R. A. B., & Young, B. W. (2018). Comparing the contribution of conscientiousness, self-control, and grit to key criteria of sport expertise development. *Psychology of Sport and Exercise*, *34*, 110–118. https://doi.org/10.1016/j.psychsport.2017.10.002

Tyumeneva, Y., Kardanova, E., & Kuzmina, J. (2017). Grit: Two Related but Independent Constructs Instead of One. Evidence From Item Response Theory. *European Journal of Psychological Assessment*, 1–10. https://doi.org/10.1027/1015-5759/a000424

Vainio, M. M., & Daukantaitė, D. (2016). Grit and Different Aspects of Well-Being: Direct and Indirect Relationships via Sense of Coherence and Authenticity. *Journal of Happiness Studies*, *17*(5), 2119–2147. https://doi.org/10.1007/s10902-015-9688-7

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research, *Organizational Research Methods*(3), 4–70.

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

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

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist*, *5*(3), 78.

Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148–170.

Wiggins, J. S. (1996). *The Five-Factor Model of Personality: Theoretical Perspectives*. New York: The Guilford Press.

Wyszyńska, P., Ponikiewska, K., Karaś, D., Najderska, M., & Rogoza, R. (2017). Psychometric Properties of the Polish Version of the Short Grit Scale. *Polish Psychological Bulletin*, *48*(2). https://doi.org/10.1515/ppb-2017-0026

Xu, K. (n.d.). *Multiple group measurement invariance analysis in Lavaan*. University of Cambridge.

Yoon, M., & Lai, M. H. C. (2018). Testing Factorial Invariance With Unbalanced Samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Vol. 30). University of California, Los Angeles Los Angeles.