RESEARCH PAPER

Grit Across Nations: The Cross-National Equivalence of the Grit-O Scale

Llewellyn E. van Zyl[1,2,3,4] · **Babet Heijenk**[1] · **Jeff Klibert**[5] · **Rebecca Shankland**[6] ·
Nicolas B. Verger[7] · **Sebastiaan Rothmann**[2] · **Vincent Cho**[8] · **Katherine Feng**[8] ·
Eric W. K. See-To[9] · **Lara C. Roll**[2,10] · **Leander van der Meij**[1]

Abstract

Despite its popularity in practice, the Grit-O Scale has shown inconsistent factorial structures and differing levels of internal consistency in samples outside the USA. The validity of the Grit-O Scale in different contexts is, therefore, questionable. As such, the purpose of this paper was to determine whether the Grit-O Scale could be used as a valid and reliable measure to compare grit across different nations. Specifically, the aim was to investigate the factorial validity, reliability, and concurrent validity of the Grit-O Scale and to investigate measurement invariance across three national cohorts (Europe, the USA, and Hong Kong). Data were gathered from 1888 respondents stemming from one USA- ($n=471$), two Hong Kong- ($n=361$) and four European ($n=1056$) universities. A series of traditional CFA and less restrictive ESEM models were estimated and systematically compared to determine the best factorial form of the Grit-O Scale. The results showed that a bifactor ESEM model, with one general factor of overall grit and two specific factors (consistency of interest and perseverance of effort), fitted the data best, showed strong measurement invariance across the three samples, and showed itself to be a reliable measure. Furthermore, concurrent validity was established by showing that the three grit factors were directly and positively related to task performance. Meaningful latent comparisons between the three cultural cohorts could therefore be made. The results imply that cross-national comparisons of grit may only be problematic when traditional CFA approaches are favoured. In contrast, ESEM modelling approaches may compensate for cross-national differences in understanding grit and control for differences in the interpretation of the scale's items. Therefore, the bifactor ESEM approach may be more appropriate for cross-cultural and cross-national comparison studies, as it allows for these differences to be meaningfully captured, modelled, and controlled for.

Keywords Grit · Consistency of interest · Perseverance of effort · Invariance testing · Task performance · Psychometric properties

✉ Llewellyn E. van Zyl
   llewellyn101@gmail.com

Extended author information available on the last page of the article

1 Introduction

Worldwide, university students are burdened with increasing study-related demands, including overwhelming study loads and challenging course content/assignments (Macfarlane, 2015). Despite these increases, study-related resources (e.g., lecturer support, information availability) are severely limited (Cattaneo et al., 2019; Gaotlhobogwe, 2017), requiring students to lean heavily on their own personal resources as a primary coping mechanism (Lesener et al., 2020; Van Zyl, 2021). From the study demands-resources framework, personal resources are seen as inner traits or abilities (e.g., resilience, self-efficacy, grit) that students access to control or positively alter their study environment (Lesener et al., 2020; Van Zyl et al., 2021c). When study-related demands are high and support is low, students' personal resources buffer against the negative impact of the academic burden on psychological well-being and educational performance outcomes. In effect, personal resources help students cope with increasing institutional demands (Krifa et al., 2021; Mokgele & Rothmann, 2014). However, the availability of personal resources varies among students; some may not have sufficient access to effectively cope with the academic pressures at university (Van Zyl, 2021). When students experience high pressure and cannot access or activate their personal resources, their intention to switch to an easier degree or drop out of university increases significantly (Krifa et al., 2022; Van Zyl, 2021). This, in turn, negatively affects student well-being, academic performance, and student throughput (Alexander, 2000). It is, therefore, not surprising that universities are exploring and implementing various personal and professional development programs designed to help students develop and access impactful personal resources to support academic coping efforts and increase high throughput (Macfarlane, 2015; Saeed et al., 2016; Stelnicki et al., 2015). One major personal resource that these programs are actively targeting is the development of 'grit' (Duckworth et al., 2007).

Grit is a non-cognitive trait that supports students to work toward long-term challenging, but achievable, goals despite setbacks in pursuing goal attainment (Duckworth et al., 2007). Gritty students are happier and healthier, report higher levels of psychological well-being, and perform better than their non-gritty counterparts in different performance tasks (Credé et al., 2017; Duckworth et al., 2007; Jachimowicz et al., 2018). Gritty students are also more likely to complete their academic degrees than their non-gritty counterparts (Duckworth, 2016). Gritty individuals tend to adopt a growth mindset, where failures are seen as valued opportunities from which to learn and grow, to support their achievement endeavors (Duckworth & Eskreis-Winkler, 2013). This, in turn, naturally results in higher levels of personal and academic performance over time (Hill et al., 2016). Research suggests that grit plays a more critical role in academic achievement than 'natural talent' and fluid intelligence (Duckworth, 2016). Considering that grit can be developed by helping students discover their passions and aligning these more completely with their study programs (Van der Vaart et al., 2021), it is not surprising that developing grit among students is a central facet of major personal/professional development interventions hosted by universities (Duckworth, 2016). Therefore, it is essential to ensure that grit can be accurately and reliably measured in different student populations to track the efficacy of these interventions.

One of the most popular instruments used in practice to measure grit is the original Grit Scale (Grit-O; Duckworth et al., 2007). The Grit-O Scale is a 12-item self-report measure that measures grit as a function of consistent interest in long-term goals and perseverance in the effort one exerts in goal attainment (Duckworth et al., 2007). The practical use of the

scale was popularized as a result of its being published in Duckworth's (2016) best-selling book *Grit: The power of passion and perseverance*. Duckworth's (2016) book was ranked as number one in the top 100 educational development books category, number five in the popular psychology category, and number seven in the cognitive psychology category (Amazon, 2018). Given the popularity of Duckworth's TED Talk (2013), her book has been translated into over 50 languages and distributed across six continents. Therefore, grit and its accompanying Grit-O Scale have a massive global appeal and impact (cf. Amazon, 2018; Duckworth & Eskreis-Winkler, 2013). It is, therefore, safe to assume that the Grit-O Scale has become a globally popular instrument to measure grit in practice (Van der Vaart et al., 2021). However, limited scientific studies exist outside the United States (US) that provide insight into its structural validity and reliability (Van Zyl et al., 2020; Van Zyl & Rothmann, 2022).

Only a few studies have investigated the psychometric properties of the Grit-O Scale outside the USA. These studies reported conflicting findings concerning the factorial validity, reliability (internal consistency), measurement invariance, and concurrent validity of the scale (cf. Disabato et al., 2019; Kim & Lee, 2015; Tyumeneva et al., 2019; Van Zyl et al., 2020). Most of the contradictory evidence was found in diverse international samples (Van Zyl et al., 2020). Specifically, studies highlighted different factor structures of the Grit-O, ranging from a single first-order factor structure, two-factor structure, three-factor structure, and bifactor structure through to less restrictive exploratory structural equation models (Areepattamannil & Khine, 2018; Disabato et al., 2019; Kim & Lee, 2015; Van Zyl et al., 2020). Variation in structure might imply that grit is interpreted differently across nationalities, which brings its validity into question. However, these findings are in direct contrast to the argument made by Duckworth et al. (2007) that the experience and manifestation of grit are consistent across different cultural groups and nationalities. Therefore, the question remains whether grit is experienced and expressed in the same way or form across different contexts and whether the Grit-O Scale is a viable measure for assessing and comparing grit within, between, and across different national contexts.

Therefore, the purpose of this study was to examine the psychometric properties of the Grit-O Scale in a cross-cultural sample consisting of university students from Europe, the USA, and Hong Kong. Specifically, this study investigated the factorial validity, internal consistency, and measurement invariance of the Grit-O Scale across different national cohorts of students. It, furthermore, aimed to determine concurrent validity by associating grit with an important performance-related metric ('task performance') within the global academic environment. Should the Grit-O Scale be seen as a valid and reliable measure, meaningful comparisons between national groups can be made.

2 Literature Review

2.1 Conceptualizing Grit

Grit is a non-cognitive personality trait or character strength that encompasses great effort, deep commitment, and long-term interest in achieving challenging goals despite setbacks or adversity (Duckworth & Quinn, 2009). More specifically, grit is characterized as a trait-like attribute that aids individuals in achieving long-term goals and is comprised of two

components: (a) consistency of interest in long-term goals and (b) perseverance of effort exerted in achieving these goals[1] (Duckworth et al., 2007). These subdomains work in tandem to support goal achievement. Notably, sustained interest leads to the development of self-efficacy or mastery, which, in effect, fuels perseverance when confronted with challenges and barriers to goal attainment (Duckworth, 2016). Within different contexts, grit can vary slightly in terms of its conceptualization and expression. For example, in collectivistic settings, grit is a mechanism to support students' ability to adapt to their changing academic environments (Datu et al., 2017). In particular, grit represents the ability to respond flexibly in overcoming new difficulties arising through unexpected circumstances. In this vein, grit is, furthermore, associated with a flexible form of perseverance, which includes demonstrable evidence of persistence, readiness to experience/accept failure, determination, and hard work (Datu et al., 2018). Overall, grit is not merely "just resilience in the face of failure"; it is also a mechanism to support the creation and maintenance of deep commitments needed to achieve personal goals (Perkins-Gough, 2013, p. 14).

Datu et al. (2018) write that, in an academic environment, being able to graduate within the allotted time requires students to show enduring levels of perseverance and a deep level of personal interest in, or connection to, either the content of the degree or longer-term goals where the degree is merely a subgoal that needs to be achieved. From this perspective, perseverance asks students to show persistence, readiness to experience/accept failure, determination, and hard work. In contrast, interest demands long-term focus, enduring passion, and an inherent ability to prioritize (Datu et al., 2018). When students can embody these principles, they are more likely to perform and less likely to quit their academic programs (Van Zyl et al., 2021c).

Findings on the performance-related returns of those who report high levels of grit span a plethora of domains and contexts. For example, grittier individuals are more likely to win spelling bee competitions and report higher grade point averages in secondary school and at university compared to their lower-level gritty counterparts (Duckworth et al., 2007; Mason, 2018). Furthermore, gritty individuals outperform their less gritty peers despite comparable levels of intelligence (Duckworth et al., 2007). Grit also predicts student retention rates at different academic institutions (Credé et al., 2017). However, a recent meta-analysis showed that the perseverance component related more strongly to academic performance and student retention than the interest component (Credé et al., 2017). In this regard, Credé et al. (2017) argue that this might be due to the validity and reliability of the instruments used to measure grit in the contexts in which they were applied.

2.2 The Measurement of Grit

Given the popularity of grit as an alternative to traditional achievement theories, at least three psychometric instruments measuring grit have become apparent within the literature: (a) Grit-O (Duckworth et al., 2007), (b) Grit-S (Duckworth & Quinn, 2009) and (c) the Triarchic Model of Grit Scale (Datu et al., 2017).

---

[1] Subsequently referred to as *interest* and *perseverance*, respectively.

2.2.1 Grit-O Scale

The original Grit Scale (Grit-O) was developed by Duckworth et al. (2007) to measure overall grit and its two components. This twelve item self-report instrument, measured on a five-point Likert type scale, aimed to measure perseverance of effort (six items: e.g., "I finish what I begin") and consistency of interest (six items: e.g., "I often set a goal but later choose to pursue a different one") as essential elements or 'building blocks' of grit. The scale showed to be a valid and reliable instrument during its initial validation. However, Duckworth and Quinn (2009) argued that the original validation study failed to show the differential predictive functioning (i.e. 'predictive validity') of the perseverance and interest factors on important achievement outcomes, that some items in the scale may be problematic and that the scale was rather lengthy.

2.2.2 Grit-S Scale

To address these challenges and prior critiques about its overlap with certain personality characteristics (such as conscientiousness), Duckworth and Quinn (2009) developed and validated the short form Grit-S scale. The Grit-S scale was comprised of a subset of items derived from the Grit-O scale and attempted to maintain the originally conceptualized two-factor structure of grit (Van Zyl et al., 2020). Four items from the original Grit-O scale were eliminated due to small factor loadings and poor inter-item correlations. An eight-item self-report scale (rated on a five-point Likert scale) remained, which measured perseverance and consistency of interest with four items each (Duckworth & Quinn, 2009). The scale showed acceptable levels of internal consistency, test–retest reliability, consensual validity (with informant reported versions of the scale) and predictive validity in terms of grade point average (Duckworth & Quinn, 2009). However, the Grit-S scale's factorial validity, internal consistency, and predictive capacity have shown to vary significantly in contexts outside of the USA; even more so in eastern contexts and collectivistic cultural settings (Datu et al., 2017).

2.2.3 Triarchic Model of Grit Scale

To address these issues, Datu et al. (2017) developed and validated the Triarchic Model of Grit Scale (TMGS) as means to measure grit in more collectivistic contexts. The ten-item self-report TMGS borrows- and slightly reformulates six items from the original Grit-O scale and introduces an additional component to form a triarchic model of grit appropriate for collectivistic cultures. The scale measures three components of grit: (a) perseverance of effort (three items: e.g., "I am a hard worker"), (b) consistency of interest (three items: e.g., "New ideas and projects sometimes distract me from previous ones") and (c) adaptability to situations (four items: e.g., "I am able to cope with the changing circumstances in life") (Datu et al., 2017). This scale introduced 'adaptability to situations' as an additional component required for success in more collectivistic cultures (Datu et al., 2017, 2018). Adaptability is seen as an essential element of longterm success within collectivistic cultures as it aids individuals to expect/accept new challenges, facilitates flexibility and indicates an inner drive to overcome difficulties as they arise (Datu et al., 2018). The scale showed to be factorially valid, produced high levels of reliability in various studies and was shown to be concurrently valid (Datu et al., 2017, 2018).

Despite these different measures to assess grit, the Grit-O scale is still the most widely used within practice. This is largely due to Duckworth's (2016) bestselling book and its mention in more mainstream popular psychology outlets such (Van Zyl et al., 2020). However, despite its popularity in practice, academic studies investigating the validity and reliability of the scale is scarce. In studies where the scale's psychometric properties have been explored, differences in factor structures, internal consistencies and predicted outcomes were found.

2.3 The Factorial Validity of the Grit-O Scale

Duckworth et al. (2007) developed the 12-item self-report Grit-O Scale to measure grit. The scale was initially developed in a secondary school context and later expanded to other academic domains. The Grit-O Scale measures perseverance and interest as functional components of an individual's overall level of grit. In initial psychometric evaluations, grit was conceptualized and measured as a first-order factorial model comprised of two correlated factors (Van Zyl et al., 2020). Within Duckworth's research group, there has been consistent empirical evidence for the Grit-O Scale as a valid and reliable instrument in various USA-based contexts (Duckworth et al., 2007; Van Zyl et al., 2021a, 2021b, 2021c). However, in studies not originating from Duckworth's laboratory, the Grit-O Scale produced different factor structures and varying ranges of internal consistency (Credé et al., 2017; Disabato et al., 2019). In addition, Credé et al. (2017) indicate that the Grit-O Scale has predominantly been used in mono-specific contexts. This implies that the Grit-O Scale may not be sensitive to account for contextual nuances in experiences and expressions of grit (Van Zyl et al., 2020). Given that grit is also deemed a psychological strength (which is naturally embedded in the values and beliefs of a population), it is inherently culture-bound (Van Zyl et al., 2021a, 2021b, 2021c). As the Grit-O Scale was developed in a Western, industrial, educated, resourced, and democratic context, the way the construct is perceived and measured may be different for students with different nationalities (Templin & Henson, 2010). This has been evident in the differences in factor structures, internal consistencies, and outcomes of grit when the Grit-O has been applied to different national-based samples.

Although grit was conceptualized as a dynamic interaction between two first-order latent constructs (i.e., perseverance and interest), the Grit-O Scale produced different factorial models in contexts outside the USA (Van Zyl et al., 2020). Van Zyl et al. (2020) argue that various traditional independent clustering confirmatory factory analytical (CFA) and exploratory structural equation modelling (ESEM) approaches have been relied on to investigate the factorial validity of the Grit-O Scale. In their paper, Van Zyl et al. (2020) highlight that at least 10 different factorial permutations of the Grit-O Scale are possible, which range from various first-order CFA structures (e.g., one-, two-, or three-factor models) to hierarchical models (where grit is a function of an interaction between two or three first-order factors) or even bifactor models. Van Zyl et al. (2020) also argue that these restrictive CFA models may not adequately fit the data and that these models are not in line with the original theoretical construction of the grit factor. For this reason, ESEM versions of the original CFA structures may be more appropriate. These ESEM models are less restrictive, as items are allowed to cross-load, but constrained to be as close to zero as possible. This is more in line with the original tenet that grit is a function of a dynamic interaction between perseverance and interest (Credé et al., 2017; Disabato et al., 2019).

Several CFA models are supported in the literature for use. In their original investigation, Duckworth et al. (2007) found support for a correlated two-factorial model of grit

comprised of perseverance and interest. However, this model could not be reproduced in further studies. In their second study, Duckworth et al. (2007) found that a unidimensional model (or a single first-order factorial model) of grit fitted the data best, and this model was better able to predict academic performance than a two-factor model. This finding was also confirmed in an Arabic sample, where Areepattamannil and Khine (2018) confirmed the unidimensional nature of grit. However, these authors found that Item 11 ("I become interested in new pursuits every few months") posed problems for model fit and showed non-significant loading on the associated factor. They argued that Item 11 had to either be excluded in totality in future studies or substantially rephrased in order to ensure that it accurately measured the interest construct (Areepattamannil & Khine, 2018).

In addition, the original two-factor structure was supported in studies conducted by Christensen and Knezek (2014) and Tyumeneva et al. (2019) in the Russian context. However, in both studies, the authors needed to make several modifications to the original instrument to ensure good data-model fit. In Christensen and Knezek's (2014) study, Item 9 ("I finish what I begin") was removed, since it showed cross-loadings on both the interest and perseverance factors. Tyumeneva et al. (2019), who also employed a Rasch analytical approach, found support for the two-dimensional factor structure; however, Item 4 ("Setbacks do not discourage me") from the perseverance subscale loaded significantly on interest and, therefore, had to be removed subsequently. Christensen and Knezek (2014), in contrast, did not find a high factor loading of Item 4 on perseverance (.37), supporting evidence to question the factor this item theoretically belongs to.

A three-factor first-order structure for the Grit-O Scale was reported in a South Korean sample (Kim & Lee, 2015). Kim and Lee (2015) found that a three-factor structure consisting of perseverance, interest, and industriousness fitted their data the best. Perseverance consisted of four items (i.e., Items 1, 4, 9, and 10), interest represented five items (i.e., Items 3, 5, 7, 8, and 11), and industriousness consisted of the two remaining items (i.e., Items 6 and 12). However, Item 2 ("New ideas and projects sometimes distract me from previous ones") had to be eliminated due to poor communalities, a weak factor loading, and poor item-total correlation. None of the items showed factor loadings below .30, and therefore, no further issues were mentioned. Kim and Lee (2015) argued that there was a cultural difference in how perseverance was perceived, indicating that perseverance was a function of one's ability to push through challenges and that the tendency toward diligence was inherent in South Koreans.

Another traditional CFA structure that was found in the literature was grit as a bifactor model (Disabato et al., 2019; Van Zyl et al., 2020). In a Dutch sample, Van Zyl et al. (2020) found that a bifactor model with one general (or overall) grit factor and two specific factors (perseverance and interest) also fitted the data. All items loaded statistically significantly onto the general grit factor and the two specific factors. However, similar to the studies mentioned above, Item 4 again posed problems. In this sample, Item 4 loaded just above the cut-off value of .30 (i.e., 0.31) on the interest subscale. Disabato et al. (2019) also found evidence for the bifactor structure of the Grit-O Scale. Here, however, the factor loading for Item 9 rounded off to 0.30 for the perseverance subscale and, therefore, was again not convincing of its explanatory power. For both studies, three corresponding items had factor loadings onto the general grit factor below .30 (i.e., Items 2, 3, and 11), thus also raising questions regarding their explanatory power (Disabato et al., 2019).

If psychometric instruments do not meet the required model fit and measurement quality criteria of traditional CFA models, ESEM is likely to be more accurate and reliable in producing parameter estimates (Morin et al., 2020; Van Zyl & Ten Klooster, 2022). For this reason, Van Zyl et al. (2020) also investigated whether these less restrictive ESEM models

could better explain the factorial structure of the Grit-O Scale. ESEM models incorporate elements of exploratory factor analysis (EFA) and CFAs by allowing items to cross-load between factors. These cross-loadings are constrained to be as close to zero as possible (Morin et al., 2020; Van Zyl & Ten Klooster, 2022). ESEM models require the construction of a priori theoretical models of an instrument, where items are permitted to cross-load. ESEM models provide a compromise between the mechanistic and iterative process employed to find optimal factorial solutions of EFAs and the highly restrictive modelling approach employed by CFAs where cross-loadings are constrained to zero (Morin et al., 2020). Therefore, ESEM models also compensate for the inflated parameter estimates and interfactor correlations between first-order factors, thus providing more 'accurate' estimations (Morin et al., 2020). ESEM is, therefore, better suited for investigating the factorial structure of multidimensional measuring instruments such as the Grit-O Scale.

Van Zyl et al. (2020) tested three ESEM models based on the a priori theoretical structure of the Grit-O Scale: a two-factor, a three-factor, and a bifactor ESEM model. Their results showed that all three ESEM models fitted the data excellently; however, the bifactor ESEM model did not produce a statistically significant general factor. The two-factor ESEM showed better model fit than the traditional CFA bifactor model (with one general and two specific factors), but this model could not establish measurement invariance between genders. Nevertheless, this could be due to the small male sample size ($n=92$) in the invariance assessment. Van Zyl et al. (2020), furthermore, found that, unlike the CFA models, the ESEM models did not inflate factor saturation and reliability. Therefore, the ESEM models showed promise.

Taking all the findings together, it is clear that differences exist in how the Grit-O Scale factorially manifests within the various cultural contexts. Therefore, it is not clear how grit might manifest within different contexts or how the instrument would function in diverse national contexts. It would not be possible to compare the grit of cultural groups, as the groups see grit differently. Given these different factorial permutations, the internal consistency of the instrument is also called into question (Van der Vaart et al., 2021).

2.4 Internal Consistency

Although different factorial models were found, previous studies have shown support for the notion that the Grit-O Scale is a reliable measure in different contexts (Disabato et al., 2019; Duckworth et al., 2007; Kim & Lee, 2015; Van Zyl et al., 2020). However, the extent to which the instrument produced consistent levels of internal consistency differed between studies and sample populations. For example, in two USA studies, Duckworth et al. (2007) found a Cronbach's alpha value for grit as a general factor of $\alpha=0.85$, for perseverance of $\alpha=0.78$, and for interest of $\alpha=0.84$. In their study in South Korea, where the three-factor model of grit was found, Kim and Lee (2015) reported acceptable, yet lower, levels of Cronbach's alphas for perseverance ($\alpha=0.76$), interests ($\alpha=0.79$), and industriousness ($\alpha=0.84$). For the bifactor structure in the Dutch context, Van Zyl, Arijs, et al. (2021), Van Zyl, Rothmann, et al. (2021), Van Zyl, Olckers, et al. (2021)) reported upper- (rho) and lower-bound (Cronbach's alpha) coefficients for reliability testing with the following values: the overall factor grit had values of $\alpha=0.79$ and $\rho=0.77$; perseverance was reported to have coefficients of $\alpha=0.76$ and $\rho=0.78$; and for interest, the reliability coefficients were equal to $\alpha=0.79$ and $\rho=0.79$. All values reported exceeded the cut-off values of $\alpha>0.70$ and $\rho>0.70$ (Wong & Wong, 2020); however, differences were apparent. Due to

the different reliability coefficients among cultures, reviewing these using a cross-national sample is essential.

2.5 Measurement Invariance

In order to ensure that the Grit-O Scale can be reliably and consistently used in diverse national contexts to discriminate and compare their concept of grit, measurement invariance across nations needs to be established (He & Van de Vijver, 2012, 2013). In its original conceptualization, Duckworth et al. (2007) argued that grit transcended cultural bounds and was experienced in the same way by collectivistic and individualistic cultures. However, researchers outside the USA found various factorial structures that contradicted this assumption (e.g., Kim & Lee, 2015; Tyumeneva et al., 2019). Since grit is positioned as a psychological strength, it might be culture-bound, as strengths are firmly embedded in the norms, beliefs, and values of a culture (Disabato et al., 2019; Kim & Lee, 2015). For example, Disabato et al. (2019) found that grit (as measured by the Grit-O Scale) was highly valued in individualistic countries, as goals were usually highly individualistic. Here, individuals strove toward the accomplishment of personal goals that were separate from their relational responsibilities (Disabato et al., 2019). However, in collectivistic contexts, social environments and relationships tended to supersede individual needs/goals, and perseverance and interests were associated with group or 'collectivistic' goals (He & Van de Vijver, 2013). The personal goals of collectivistic individuals were more relation-oriented than those of individualistic individuals. Individuals from collectivistic cultures might be more motivated to show grit as a way of contributing to their family or communities (Disabato et al., 2019). Similarly, Disabato et al. (2019) found that items on the interest subscale might fail to incorporate collectivistic conceptualizations of passion for long-term goals, thus indicating potential cultural bias. Therefore, the difference between these two cultures might influence how items on the Grit-O Scale are interpreted and, consequently, how grit is seen or experienced between the cultures. Disabato et al., (2019, p. 13) support this notion with their finding that the Grit-O Scale "does a poorer job at assessing overall grit in collectivistic cultures".

These assumptions and findings are in contrast to how Duckworth et al. (2007) conceptualized grit as a meta-level concept that would not differ across nations or cultures. For this reason, in order to compare findings across different national contexts with different cultures, it is necessary to estimate whether the Grit-O Scale is invariant across nations. In essence, measurement invariance needs to be estimated with nationality or culture as an indicator. If invariance is not established, it will be impossible to compare grit between different national contexts. However, due to the inconsistency between the findings of Duckworth et al. (2007) and others, it is assumed that invariance may not be present.

2.6 Concurrent Validity

In order to further explore the validity of the Grit-O Scale, concurrent validity should be established through its relationship with performance or achievement. Given that grit is positioned to be a precursor to individual success and performance, the proficiency of students to make the right choices and take the initiative to perform the most

important or core/substantive tasks central to their studies could be seen as an essential outcome (i.e., task performance). Previous research has found considerable support for the grit-task performance thesis in contexts ranging from business (Jordan et al., 2019; Steuber et al., 2019; Webster-Wright, 2019) and sports (Cazayoux & DeBeliso, 2019) through to university (Duckworth & Quinn, 2009; Jachimowicz et al., 2018; Nelson & Baltes, 2019). Koopmans et al. (2012) argue that task performance is a direct consequence of an individual's interest in the nature of the task and the ability to push through or manage challenges appropriately. Gritty individuals tend to perform better because they can utilize their capabilities more effectively when performing tasks that are aligned with their interests (Vogelsang, 2018). Gritty students have been shown to be more effective in prioritizing the completion of their short-term assignments and tasks, as these are directly related to their long-term goals (Duckworth et al., 2007). They are also less likely to be affected by setbacks at university (Van Zyl, Arijs, et al., 2021; Van Zyl, Olckers, et al., 2021; Van Zyl, Rothmann, et al., 2021). Therefore, establishing a positive relationship between grit and task performance could support the concurrent validity of the Grit-O Scale.

2.7 The Present Study

The purpose of this study was to determine whether the Grit-O Scale could be used as a valid and reliable measure for assessing and comparing grit across different national contexts. Specifically, the aim was to investigate the factorial validity, reliability, and measurement invariance of the Grit-O across three diverse cohorts (Europe, the USA, and Hong Kong). The study, furthermore, aimed to extend concurrent validity by associating grit with an important performance-related metric (task performance) in the global academic environment. Finally, latent mean comparisons could be made if the Grit-O were to be a valid and reliable measure that was invariant between diverse national cohorts. Given that grit is theoretically positioned as a psychological strength that is invariant across cultures/nations, as well as to avoid biases arising from contextual or linguistic variation, the objectives of this study were tested for the overall sample.

3 Method

3.1 Research Design

The study employed a cross-sectional, electronic survey-based research design to determine the factor validity, internal consistency, measurement invariance, and concurrent validity of the Grit-O Scale across Europe, the USA, and Hong Kong. This provided a means of gathering data at a single point in time.

3.2 Research Procedure

The data obtained for this paper was drawn from a large-scale cross-national student well-being project. The original data was gathered by using a convenience sampling strategy, where invitations to participate were sent to university students from collaborative institutions. Representatives from universities in the Netherlands, France, Belgium, the United

Table 1 Demographic characteristics of participants ($N=1888$)

Item	Category	Frequency (f)	Percentage (%)
Gender	Male	748	39.60
	Female	1133	60.00
	Other	7	0.40
Age (years)	18–20 years	343	18.20
	21–25 years	1072	56.80
	26–30 years	165	8.70
	31–35 years	80	4.20
	36 years and older	83	4.40
	Missing	145	7.70
Nationality	European	1056	55.90
	USA	471	24.90
	Hong Kong	361	19.10
Home language	English	486	25.70
	Dutch	535	28.30
	German	155	8.20
	French	224	11.90
	Chinese	361	19.10
	Other European	127	6.80

States, and Hong Kong collected the data. Students were asked questions about their demographics, biographical characteristics, language proficiency, grit, and task performance in the electronic surveys. Before participation, participants were informed about the nature of the study, and their rights/responsibilities were highlighted. Participants were informed that participation was voluntary and that the right to confidentiality/anonymity was ensured. The survey was distributed through Qualtrics (www.qualtrics.com). Data management procedures were in line with the requirements of the General Data Protection Regulation (GDPR).

3.3 Participants

A convenience sampling strategy was employed to draw 1888 respondents stemming from one US- ($n=471$), two Hong Kong- ($n=361$) and four European ($n=1056$) universities. Table 1 provides an overview of the demographic characteristics of the participants. The majority of the participants were Dutch-speaking (28.3%) women (60%) between the ages of 21 and 25 (56.8%), who resided in Europe (59.59%).

3.4 Measuring Instruments

The following instruments were used to gather data for this study:

A *biographical questionnaire* was used to gather basic descriptive information about participants relating to their gender, age, nationality, and home language.

The *Grit-O Scale* (Duckworth et al., 2007) was used to measure grit. This 12-item self-report scale measures grit as a non-cognitive trait that is a function of two factors: (a)

consistency of interest (six items, e.g., "My interests change from year to year") and perseverance (six items, e.g., "Setbacks do not discourage me"). Each item was rated on a five-point Likert-type scale, ranging from 1 (*not like me at all*) to 5 (*very much like me*). The Grit-O Scale showed acceptable levels of internal consistency, with McDonald's omegas ranging from .77 to .79 on the consistency of interest and perseverance subscales in Europe (Van Zyl et al., in press). Duckworth et al. (2007), furthermore, found that the overall factor grit (α = .85) and the two specific factors, perseverance (α = .78) and interest (α = .84), were reliably measured in the USA.

The *Task Performance subscale* of the *Individual Work Performance Scale* developed by Koopmans et al. (2012) was used to measure overall task performance. Task performance was measured by seven items on a six-point Likert scale, ranging from 1 (*never*) to 6 (*always*). An example of an item is "I knew how to set the right priorities". Magada and Govender (2017) found acceptable levels of internal consistency for the instrument, with a Cronbach's alpha level of .86. In Europe, the scale also produced acceptable levels of internal consistency, as represented by a McDonald's omega of .84 (Van Zyl et al., 2022).

3.5 Statistical Analyses

Data was processed by using both JASP 0.15 (JASP, 2021) and Mplus 8.6 (Muthén & Muthén, 2021) through structural equation modelling (SEM). Missing data was managed through the full information maximum likelihood estimation method (FIML), the default approach in Mplus 8.6. A stepwise, sequential strategy was employed.

Firstly, the presence of common method bias was examined. Both Harman's single-factor test (Harman, 1976) and the common latent factor approach (Podsakoff et al., 2003) were employed. For the Harman's single-factor approach, all observed indicators for both the Grit-O Scale and task performance subscale of the Individual Work Performance Scale were entered into an unrotated exploratory factor analysis, and a single factor was specified to be extracted from the data. Fuller et al. (2016) indicate that the extracted variance of the single factor should not exceed 50%. After that, a series of common latent factor approaches through SEM were employed. Through a confirmatory factor analytical (CFA) approach, all items were specified to load onto a unidimensional first-order factorial model. Such a unidimensional model should not produce a good data-model fit, and the unstandardized variance should be below 50% (Tehseen et al., 2017). Finally, an a priori measurement model was specified, and a further single latent factor was additionally estimated, comprised of all observed indicators. The variance of this common latent factor was constrained to one, and the factor loadings were constrained to be equal. This model was then compared, based on chi-square, to a model where no common latent factor was specified. There should be no statistically significant chi-square difference between these two models (Podsakoff et al., 2003). If all of these conditions were met, common method bias might not be a concern in the current study.

Secondly, the factorial validity of the overall sample was explored by employing a competing measurement modelling approach with the robust maximum likelihood estimation method (MLR) in Mplus. Both traditional confirmatory factor analytical (CFA) models and exploratory structural equation models (ESEM) based on the a priori factorial structure of the Grit-O Scale were estimated and systematically compared to find the best-fitting model for the data. The CFA models were specified according to the independent cluster model assumptions, where items were only permitted to load onto their a priori theoretical factor, and cross-loadings were constrained to zero. For the

bifactor CFA models, factors were specified as orthogonal, and a target rotation was employed. Here, a general grit factor (G-Factor) was specified to comprise all the Grit-O Scale items. Furthermore, specific factors (S-Factors), corresponding to the a priori theoretical dimensions of the Grit-O, were specified and items targeted to load onto their respective factors. The ESEM models were specified in line with the guidelines proposed by Van Zyl and Ten Klooster (2022). Factors were again specified in line with their a priori theoretical factorial structures; however, cross-loadings between items and non-target factors were permitted, but constrained to be as close to zero as possible (Brown, 2006). A target rotation was employed. For the bifactor ESEM model, a single G-Factor of overall grit and two S-Factors were specified where cross-loadings between specific factors were permitted, but targeted to be as close to zero as possible. The code for all the ESEM models can be generated with the De Beer and Van Zyl (2019) ESEM code generator for Mplus. The best-fitting measurement model for the data was determined by contrasting and comparing models based on both the goodness-of-fit criteria of Hu and Bentler (1999) (cf. Table 2) and measurement quality. Measurement quality for the traditional CFA and ESEM models was evaluated through standardized factor loadings (e.g., $\lambda > 0.35$), the item uniqueness (e.g., > 0.10, but < 0.90), and levels of tolerance for cross-loadings (Kline, 2010). For both the CFA and ESEM bifactor models, measurement quality was assessed through a well-defined general factor (with significant factor loadings) and with relatively well-defined specific factors (allowing for lower factor loadings). Only models meeting both criteria were retained for further analysis. After finding the best model, the separate measurement models were also estimated for each national group separately (published in Appendix 2).

Thirdly, the Grit-O factorial equivalence was investigated by estimating measurement invariance between the three cultural groups (European, US, Hong Kong). Here, a series of increasingly restrictive invariance models were estimated: (a) configural invariance (similar factor structures), (b) metric invariance (similar factor loadings), and (c) scalar invariance (similar intercepts). Invariance was determined through comparing these different models based on the following criteria: changes in RMSEA ($\Delta < 0.015$), SRMR ($\Delta < 0.02$ for configural versus metric/scalar; $\Delta < 0.01$ for metric versus scalar), CFI ($\Delta < 0.01$), and TLI ($\Delta < 0.01$) (Chen, 2007; Wong & Wong, 2020). Traditionally, a non-significant difference in chi-square between cohorts ($p > 0.05$) is also required; however, chi-square is sensitive to sample size, where it reaches significance with minimal deviations from perfect fit in larger samples (Morin et al., 2020; Wong & Wong, 2020). Therefore, chi-square and chi-square differences are reported for transparency, but were not used as criteria.

In the *fourth* place, the standardized factor loadings, item uniqueness, explained common variance (ECV), item-level explained common variance (IECV), and McDonald's omega were estimated for the model, which showed invariance between culturally diverse national cohorts. Here, standardized factor loadings and the item uniqueness were reported (Kline, 2010). Morin et al. (2020) indicate that, for bifactor ESEM models, the general factor should be well defined (where all items load significantly on it) and specific factors reasonably well defined (cross- and non-significant loadings are permitted). The mean average of target loadings should be greater than 0.30 (Wong & Wong, 2020). The ECV was estimated to assess the proportion of common variance explained by each of the three factors in both the CFA and ESEM bifactor models. The IECV "provides the extent to which an item's responses are accounted for by variation on the latent general dimension alone, and thus acts as an assessment of unidimensionality at the individual item level" (Stucky et al., 2013, p. 51). If IECVs exceed 0.80, it indicates the unidimensionality of the general factor. Therefore, IECVs lower than 0.80 are preferred to provide support for the

Table 2 Model fit statistics

Fit indices	Cut-off criterion	Sensitive to N	Penalty for model complexity
Absolute fit indices			
Chi-square (χ^2)	• Lowest comparative value between measurement models • Non-significant chi-square ($p > 0.01$)	Yes	No
Approximate fit indices			
Root mean square error of approximation (RMSEA)	• 0.06 to 0.08 (marginally acceptable); 0.01 to 0.05 (excellent) • Not-significant ($p > 0.01$) • 90% confidence interval (CI) range should not include zero	No	Yes
Standardized root mean square residual (SRMR)	• 0.06 to 0.08 (marginally acceptable); 0.01 to 0.05 (excellent)	Yes	No
Incremental fit indices			
Comparative fit index (CFI)	• 0.90 to 0.95 (marginally acceptable fit); 0.96 to 0.99 (excellent)	No	No
Tucker–Lewis index (TLI)	• 0.90 to 0.95 (marginally acceptable fit); 0.96 to 0.99 (excellent)	No	Yes
Akaike information criterion (AIC)	• Lowest value in comparative measurement models	Yes	Yes
Bayesian information criterion (BIC)	• Lowest value in comparative measurement models	Yes	Yes

multidimensionality of the Grit-O Scale. To determine the level of reliability of the Grit-O Scale, McDonald's omega ($\omega > 0.70$; Hayes & Coutts, 2020) was estimated.

In the *fifth* place, to establish concurrent validity, a structural model was employed. Here, the best-fitting measurement model of the Grit-O Scale was used as an exogenous factor and regressed on task performance (as an endogenous factor). A statistically significant, positive relationship between grit and task performance was required to establish concurrent validity ($p < 0.05$). Furthermore, relative weight analysis was used to determine the unique variance contribution of each of the grit factors in task performance.

Finally, if the factorial validity, internal consistency, measurement invariance, and concurrent validity of the Grit-O could be established, latent mean differences between culturally diverse national cohorts could be estimated. Here, the European group was used as a reference point and compared to the US and Hong Kong groups.

4 Results

4.1 Common Method Bias

Common method bias was assessed through both Harman's single-factor test (Harman, 1976) and the Podsakoff et al. (2003) common latent factor approach. Firstly, all observed indicators were combined into one general factor, resulting in common shared variance of 31.34%. This was below the suggested 50% (Fuller et al., 2016). Secondly, the single latent factor approach suggested by Podsakoff et al. (2003) was used. Here all items were specified to load onto a single latent factor. The results showed that a single factor fitted the data poorly ($\chi^2_{(1888)} = 5032.165$; $df = 135$; CFI $= 0.64$; TLI $= 0.59$; RMSEA $= 0.14$ [.135, .142], $p < 0.01$; SRMR $= 0.12$). Furthermore, the unstandardized variance of this single-factor model was 17.7%. Finally, the common latent factor approach was employed. Initially, all observed items were structured according to their factor grit or task performance. Then, an 'empty' unmeasured latent factor was added, with regression lines to each observed item. After that, all paths from the unmeasured latent factor were constrained to be equal, and the variance of this factor was set to be one. The results showed that variance explained by the common latent factor was low ($< 50\%$) and that no significant difference in chi-square was apparent between the different models ($p > 0.05$). Therefore, common method bias might not be an issue in the current model, and we could proceed to the next step.

4.2 Factorial Validity: Competing CFA and ESEM Measurement Models

To determine the best-fitting measurement model for the data, a competing measurement modelling strategy was employed. Ten theoretically informed factorial models were estimated and subsequently compared. Measured items were treated as observed indicators for first-order factorial models. Item parceling was not permitted and error terms were left uncorrelated. No modifications were made to any of the factorial structures to improve fit. However, Item 4 (GRIT_4: "Setbacks do not discourage me") showed a non-significant factor loading on all factorial models and was, therefore, removed. The following factor analytical models were estimated:

1. **Model 1**: a unidimensional first-order CFA model was specified where all 11 items loaded directly onto one general grit factor (cf. Appendix 1a).

2. **Model 2**: a two first-order factor CFA model was constructed where five items (i.e., 1, 6, 9, 10, and 12) loaded directly on one specific factor, perseverance. The remaining six items (i.e., 2, 3, 5, 7, 8, and 11) were constrained to load directly on the other specific factor, interest (cf. Appendix 1).

3. **Model 3**: a second-order CFA model with two first-order factors (perseverance and interest) and a general grit factor was estimated based on Model 2. Here, the two specific factors loaded onto one second-order general grit factor (cf. Appendix 1).

4. **Model 4**: a three first-order factor CFA model was constructed where six items loaded onto interest (i.e., 2, 3, 5, 7, 8, and 11), three items on perseverance (i.e., 1, 9, and 10), and two items (i.e., 6 and 12) on industriousness (cf. Appendix 1).

5. **Model 5**: a second-order CFA model comprised of three specific factors was estimated based on Model 4. Interest, perseverance, and industriousness were specified to load onto a second-order factor called overall grit. To establish convergence, all first-order factors were freely estimated, the factorial variance of the second order was constrained to one, and the first-order paths were constrained to equal (cf. Appendix 1).

6. **Model 6**: a bifactor CFA model with one general grit factor (all 11 items loaded directly) and two specific factors (interest and perseverance) was estimated. These factors were specified as orthogonal (i.e., uncorrelated) (cf. Appendix 1).

7. **Model 7**: a bifactor CFA model with one general grit factor (on which all 11 items loaded directly) and three specific factors (interest, perseverance, and industriousness) was estimated. These factors were specified as orthogonal (i.e., uncorrelated). This model failed to converge due to a lack of parsimony (cf. Appendix 1).

8. **Model 8**: an ESEM model with two first-order factors was specified (based on Model 2). In this model, items were specified to load onto their a priori factorial first-order factors; cross-loadings were permitted, but targeted to be close to zero (cf. Appendix 1).

9. **Model 9**: a hierarchal ESEM model with two first-order factors and one second-order factor was specified. Model 8 was used as input. Here, the perseverance and interest first-order factors were specified to load onto a higher-order factor called grit. Once more, items were specified to load directly onto their a priori first-order factors. Cross-loadings were again permitted, but constrained to be as close to zero as possible (cf. Appendix 1).

10. **Model 10**: a bifactor ESEM model was estimated with one general grit factor and two specific factors (perseverance and interest). All 11 items loaded directly onto the general grit factor. Items were then targeted to load onto their respective specific factors. Factors were specified as orthogonal. However, cross-loadings were again permitted, but constrained to be as close to zero as possible (cf. Appendix 1).

The results, summarized in Table 3, showed that the less restrictive *two-factor ESEM model* (Model 8) and its *hierarchal ESEM* counterpart (Model 9: $\chi^2_{(1888)} = 218.62$; $df = 34$; CFI $= 0.97$; TLI $= 0.95$; RMSEA $= 0.05$ [.047, .061], $p > 0.05$; SRMR $= 0.02$; AIC $= 54113.811$; BIC $= 54352.172$), as well as the *bifactor ESEM model* (Model 10: $\chi^2_{(1888)} = 127.827$; $df = 25$; CFI $= 0.98$; TLI $= 0.97$; RMSEA $= 0.05$ [.039, .055], $p > 0.05$; SRMR $= 0.02$; AIC $= 54041.02$; BIC $= 54329.271$), fitted the data best. From the more restrictive CFA models, only the *bifactor CFA model* (Model 6: $\chi^2_{(1888)} = 264.984$;

Table 3 Competing measurement models for the overall sample

Model	Type	χ^2	df	CFI	TLI	RMSEA		SRMR	AIC	BIC	ABIC	Meets model-fit criteria	Meets measurement quality criteria
Model 1	Unidimensional model	2592.59	44	0.58	0.48	0.18	[.169–.181]	0.16	56467.78	56650.71	56545.87	No	No
Model 2	Two first-order factor model	635.91	43	0.90	0.88	0.09	[.080–.091]	0.08	54513.10	54701.57	54593.56	No	Yes
Model 3	Second-order model with two first-order factors	635.91	43	0.90	0.88	0.09	[.080–.091]	0.08	54513.10	54701.57	54593.56	No	Yes
Model 4	Three first-order factor model	620.94	41	0.91	0.87	0.09	[.081–.093]	0.08	54502.13	54701.69	54587.32	No	Partially
Model 5	Second-order model with three first-order factors*	766.96	43	0.88	0.85	0.09	[.089–.100]	0.09	54644.16	54832.63	54724.61	No	Partially
Model 6	Bifactor model (one general and two specific factors)	264.98	33	0.96	0.94	0.06	[.054–.068]	0.04	54162.18	54406.08	54266.30	Partially	Yes
Model 7	Bifactor model (one general and three specific factors)	–	–	–	–	–	–	–	–	–	–	–	–
Model 8	ESEM model with two first-order factors	218.62	34	0.97	0.95	**0.05**	[.047–.061]	0.02	54113.81	54352.17	54215.56	Yes	Yes
Model 9	H-ESEM model with two first-order factors	218.62	34	0.97	0.95	**0.05**	[.047–.061]	0.02	54113.81	54352.17	54215.56	Yes	Partially
Model 10	Bifactor ESEM model (one general and two specific factors)	127.83	25	0.98	0.97	**0.05**	[.039–.055]	0.02	54041.02	54329.27	54164.07	Yes	Yes

χ^2 chi-square; *df* degrees of freedom; *TLI* Tucker–Lewis index; *CFI* comparative fit index; *RMSEA* root mean square error of approximation [90% CI]; *SRMR* standardized root mean square residual; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *ABIC* adjusted Bayesian information criterion; bold=non-significant $p > 0.01$; * for convergence, first-order factor loadings were constrained to be equal

$df = 35$; CFI $= 0.96$; TLI $= 0.94$; RMSEA $= 0.06$ [.054, .068], $p < 0.05$; SRMR $= 0.04$; AIC $= 54162.18$; BIC $= 54406.08$) showed adequate data fit; however, it did produce a significant RMSEA. As such, none of the CFA models were retained.[2]

The next step was to inspect the measurement quality of the four best-fitting models. The results showed that all models produced adequate factor loadings and that the item uniquenesses were in acceptable ranges. However, for the hierarchal ESEM model (Model 9), only the perseverance factor loaded significantly ($\lambda = 0.55$; $SE = 0.09$) onto the higher-order grit factor. Interest showed both a low and non-significant factor loading ($\lambda = 0.06$; $SE = 0.09$; $p = 0.47$). Therefore, Model 9 did not given further consideration.

4.3 Measurement Invariance

Measurement invariance across the three culturally diverse national cohorts (European, US, and Hong Kong) was, therefore, estimated for both the ESEM model with two first-order factors (Model 8) and the bifactor ESEM model with one general and two specific factors (Model 10).

Table 4 summarizes the results of the measurement invariance estimation for Model 8 across the three culturally diverse national cohorts. The results showed that measurement invariance could not be established, since there were significant differences in chi-square between the configural, metric, and scalar models. Furthermore, differences in CFI ($\Delta < 0.01$), TLI ($\Delta < 0.01$), RMSEA ($\Delta < 0.015$), and SRMR ($\Delta < 0.02$ for configural versus metric; $\Delta < 0.01$ for metric versus scalar) exceeded the suggested cut-off scores (Chen, 2007). Therefore, the ESEM model with two first-order factors was not equivalent across these three diverse national cohorts. For this reason, the model was excluded from further consideration.

Next, measurement invariance was estimated for the bifactor ESEM model with one general and two specific factors. The results, summarized in Table 5, showed a non-significant difference in chi-square between the configural, metric, and scalar models as well as changes in CFI ($\Delta < 0.01$), TLI ($\Delta < 0.01$), RMSEA ($\Delta < 0.015$), and SRMR ($\Delta < 0.02$ for configural versus metric; $\Delta < 0.01$ for metric versus scalar) below the maximum cut-offs (Chen, 2007). Therefore, measurement invariance was established for Model 10, and it was retained for further analysis and mean comparison.

4.4 Item-Level Parameters: Standardized Factor Loadings and Internal Consistencies

The item-level parameters and reliability estimates were computed for the model which both fitted the data best and showed to be invariant between nations: Model 10—The bifactor ESEM Model (cf. Figure 1).

The results, summarized in Table 6, showed a well-defined general grit factor and two well-defined specific factors. The mean target loadings for grit (mean $\lambda = 0.33$), interest (mean $\lambda = 0.55$), and perseverance (mean $\lambda = 0.59$) exceeded 0.30. All items loaded significantly on their target, and a priori factors and cross-loadings were low. Only GRIT_3 showed a non-significant loading on the general grit factor ($\lambda = 0.09$, $p > 0.05$), which implied that it described interest ($\lambda = 0.65$, $p < 0.05$) better than overall experiences of grit. Additionally, all items showed stronger loading onto the two specific factors than the overall general factor.

---

[2] All theoretical models were subsequently tested for each population group separately. Similar trends were found. The bifactor ESEM model consistently showed better data fit than any other theoretical model in each of the three population groups (cf. Appendix 2).

Table 4 Measurement invariance between national cohorts of model 8: the ESEM model with two first-order factors

Model	χ^2	df	CFI	TLI	RMSEA		SRMR	Model comparison	$\Delta\chi^2$	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR	Meets model-fit criteria	Meets invariance criteria
M1 Configural invariance	290.53	102	0.97	0.95	**0.054**	[.047–.062]	0.027							Yes	No
M2 Metric invariance	379.26	138	0.96	0.95	**0.053**	[.046–.059]	0.046	M2 vs M1	88.74	− 0.01	0.00	0.019	0.019	Yes	No
M3 Scalar invariance	540.09	156	0.93	0.93	0.063	[.057–.068]	0.058	M3 vs M2	160.83	− 0.02	− 0.02	0.012	0.012	Yes	No
								M3 vs M1	249.56	− 0.03	− 0.02	0.009	0.031		

χ^2 chi-square; df degrees of freedom; TLI Tucker–Lewis index; CFI comparative fit index; $RMSEA$ root mean square error of approximation [90% CI]; $SRMR$ standardized root mean square residual; AIC Akaike information criterion; BIC Bayesian information criterion; $ABIC$ adjusted Bayesian information criterion; bold = non-significant $p > 0.01$

Table 5 Measurement invariance between national cohorts of model 10: the bifactor ESEM model (one general and two specific factors)

Model	χ^2	df	CFI	TLI	RMSEA		SRMR	Model comparison	$\Delta\chi^2$	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR	Meets model-fit criteria	Meets Invariance Criteria
M1 Configural invariance	149.13	75	0.98	0.97	**0.040**	[.030–.049]	0.026	M2 vs M1	132.79	−0.01	−0.01	0.015	0.015	Yes	Yes
M2 Metric invariance	281.92	123	0.97	0.96	**0.045**	[.038–.052]	0.041	M3 vs M2	47.38	−0.01	0.00	0.002	0.004	Yes	Yes
M3 Scalar invariance	329.30	139	0.97	0.96	**0.047**	[.040–.053]	0.045	M3 vs M1	180.17	−0.01	−0.01	0.007	0.019	Yes	Yes

χ^2 chi-square; *df* degrees of freedom; *TLI* Tucker–Lewis index; *CFI* comparative fit index; *RMSEA* root mean square error of approximation [90% CI]; *SRMR* standardized root mean square residual; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *ABIC* adjusted Bayesian information criterion; bold = non-significant $p > 0.01$
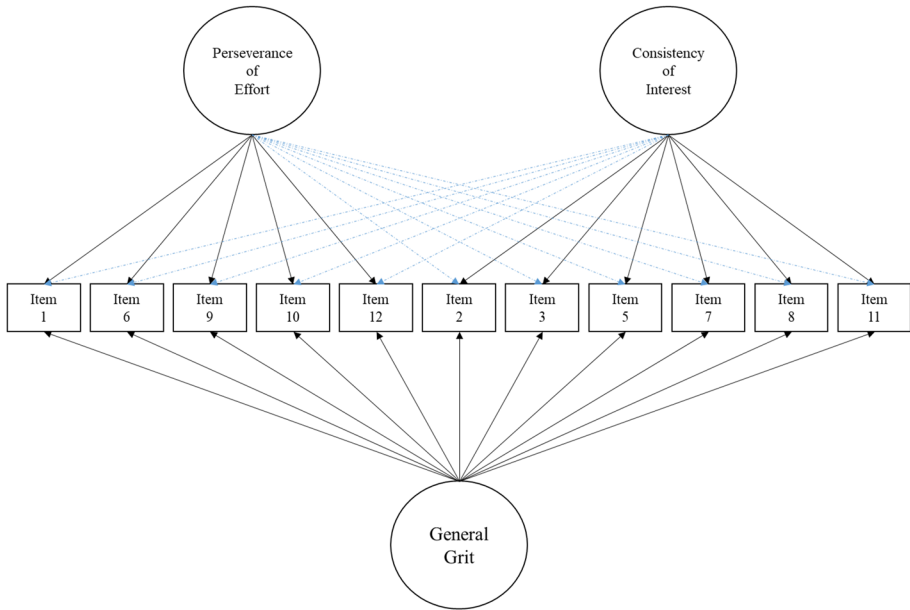
Fig. 1 Bifactor ESEM model

Furthermore, the two specific factors explained the majority of the common variance in item scores. The two specific factors produced higher levels of ECV (interest: 77%; perseverance: 71%) than the general grit factor (26%). Similarly, the IECV showed evidence of the multidimensionality of the Grit Scale, as each item shared variance with both the specific and general factor. Finally, all three factors produced McDonald's omegas exceeding the suggested cut-off criteria (omega < 0.70), indicating that these were reliable measures.[3]

4.5 Concurrent Validity: The Relationship Between Grit and Task Performance

To determine concurrent validity, a structural model was estimated with the bifactor ESEM model and task performance. Here, task performance was specified as a unidimensional construct comprised of seven observed indicators.[4] Grit, perseverance, and interest were regressed directly onto task performance. This model produced

---

[3] Omegas, based on the bifactor ESEM model, were also computed for each cultural cohort independently. The results showed acceptable levels of internal consistency of the three factors for all three cohorts, with omegas > 0.70 (general grit: Europe = 0.86; USA = 0.86; Hong Kong = 0.76; interest: Europe = 0.78; USA = 0.83; Hong Kong = 0.73; and perseverance: Europe = 0.84; USA = 0.83; Hong Kong = 0.79).

[4] The relationship between the bifactor ESEM factorial structure of the Grit-O Scale and task performance was also estimated separately for each cultural cohort. The results showed similar trends to the overall sample. A positive relationship between overall grit and task performance was found in Europe ($\beta = 0.65$; $SE = 0.04$; $p < 0.05$), the USA ($\beta = 0.38$; $SE = 0.06$; $p < 0.05$), and Hong Kong ($\beta = 0.47$; $SE = 0.09$; $p < 0.05$). Interest was also found to positively relate to task performance in Europe ($\beta = 0.10$; $SE = 0.04$; $p < 0.05$), the USA ($\beta = 0.13$; $SE = 0.06$; $p < 0.05$), and Hong Kong ($\beta = 0.21$; $SE = 0.08$; $p < 0.05$). Similarly, perseverance and task performance were positively related in Europe ($\beta = 0.22$; $SE = 0.08$; $p < 0.05$), the USA ($\beta = 0.51$; $SE = 0.06$; $p < 0.05$), and Hong Kong ($\beta = 0.30$; $SE = 0.07$; $p < 0.05$). Grit, interest, and perseverance additionally explained a significant amount of variance in task performance in Europe ($R^2 = 0.48$), the USA ($R^2 = 0.39$), and Hong Kong ($R^2 = 0.46$).

Table 6 Standardized factor loadings, item uniqueness, and IECV of the bifactor ESEM model

Item	Description	Overall grit		Consistency of interest		Perseverance of effort		δ	IECV
		λ_G	SE	λ_{S1}	SE	λ_{S2}	SE		
GRIT_2	New ideas and projects sometimes distract me from previous ones	0.18	0.06	0.50	0.03	−0.12	0.03	0.70	0.12
GRIT_3	My interests change from year to year	0.09	0.08	0.65	0.03	0.03	0.04	0.57	0.02
GRIT_5	I have been obsessed with a certain idea of a project for a short time, but later lost interest	0.34	0.07	0.57	0.04	−0.04	0.03	0.56	0.27
GRIT_7	I often set a goal, but later choose to pursue a different one	0.36	0.05	0.56	0.03	0.09	0.02	0.55	0.29
GRIT_8	I have difficulty maintaining my focus on projects that take more than a few months to complete	0.51	0.05	0.45	0.04	−0.01	0.04	0.53	0.56
GRIT_11	I become interested in new pursuits every few months	0.24	0.07	0.69	0.06	−0.02	0.03	0.48	0.11
GRIT_1	I have overcome setbacks to conquer an important challenge	0.32	0.06	−0.14	0.02	0.57	0.03	0.55	0.24
GRIT_6	I am a hard worker	0.37	0.05	−0.03	0.02	0.70	0.03	0.37	0.22
GRIT_9	I finish what I begin	0.46	0.03	0.13	0.02	0.60	0.03	0.41	0.37
GRIT_10	I have achieved a goal that took years of work	0.35	0.05	−0.04	0.02	0.41	0.04	0.71	0.42
GRIT_12	I am diligent	0.38	0.04	0.03	0.02	0.65	0.03	0.43	0.26
	Explained common variance (ECV)	0.26	–	0.77	–	0.71	–	–	–
	McDonald's omega	0.84	–	0.80	–	0.83	–	–	–

λ standardized factor loading; *SE* standard error; δ item uniqueness; *IECV* item-level explained common variance

Table 7 Concurrent validity: the relationship between grit, interest, effort, and task performance

Regression path			Standardized				R^2	Rescaled relative weight (%R^2)	Validity established
			β	SE	t-value	p			
Overall grit	→	Task performance	0.53	0.04	14.92	0.00	0.45	68.68	Yes
Interest	→	Task performance	0.10	0.03	2.93	0.00		1.77	Yes
Perseverance	→	Task performance	0.41	0.04	11.55	0.00		29.55	Yes

→ regression; *β* standardized beta; *SE* standard error; *p* statistical significance; R^2 variance

Table 8 Estimated latent mean (standard errors) comparisons with Europe as reference

No	Factor	Europe	USA		Hong Kong	
			Mean (*SE*)	95% CI	Mean (*SE*)	95% CI
1	Overall grit	-	− 0.52 (.10)	[− 0.73, − 0.32]	− 0.31 (.12)	[− 0.54, − 0.08]
2	Interest	-	− 0.21 (.09)	[− 0.39, − .003]	**− 0.09 (.08)**	[− 0.25, 0.08]
3	Perseverance	-	1.36 (.14)	[1.09, 1.64]	− 0.20 (.10)	[− 0.41, 0.00]

Mean estimated latent mean; *SE* standard error; bold = *p* > 0.05

adequate fit ($\chi^2_{(1888)} = 775.28$; *df* = 25; CFI = 0.95; TLI = 0.93; RMSEA = 0.06 [.052, .060]; SRMR = 0.03; AIC = 91058.92; BIC = 91485.75).

The results, summarized in Table 7, showed that overall grit (β = 0.53; *SE* = 0.04; *p* < 0.05), interest (β = 0.10; *SE* = 0.03; *p* < 0.05), and perseverance (β = 0.41; *SE* = 0.04; *p* < 0.05) were all positively associated with task performance. All three factors explained 45% of the total variance in task performance. A relative weight analysis was conducted as well to determine the relative 'importance' of each factor for task performance. Here, the purpose was to determine how much unique variance each factor contributed to the overall explained variance in task performance. The rescaled relative weight showed that overall grit (68.68%) and perseverance (29.55%) contributed the most unique variance in task performance. Interest, however, only contributed 1.77% unique variance in task performance. Taken together, the results provided support for the concurrent validity of the bifactor ESEM model of the Grit-O Scale.

4.6 Latent Mean Comparisons Across Diverse National Cohorts

Given that Model 10 showed itself to be invariant between culturally diverse national cohorts and showed concurrent validity, it was used to determine the latent mean differences in grit, perseverance, and interest between Europe, the USA, and Hong Kong. Table 8, with Europe as the reference group, shows that respondents from the USA showed significantly lower levels of overall grit (ΔM = − 0.52; *SE* = 0.10; *p* < 0.05) and interest (ΔM = − 0.21; *SE* = 0.09; *p* < 0.05), yet higher levels of perseverance (ΔM = 1.36; *SE* = 0.14; *p* < 0.05), than Europeans.

Respondents in Hong Kong showed significantly lower levels of overall grit (ΔM = − 0.31; *SE* = 0.12; *p* < 0.05) and perseverance (ΔM = − 0.20; *SE* = 0.10; *p* < 0.05)

than those in Europe. However, no differences in interest ($\Delta M = -0.09$; $SE = 0.08$; $p > 0.05$) between those in Europe and Hong Kong were reported.

5 Discussion

The principle aim of this study was to determine whether the Grit-O Scale could be used as a valid and reliable measure to compare grit across different national cohorts. Specifically, the aim was to investigate the factorial validity, reliability, and concurrent validity of the Grit-O Scale and to examine measurement invariance across three national cohorts (Europe, the USA, Hong Kong). A series of traditional CFA and less restrictive ESEM models were estimated and systematically compared to determine the best factorial form of the Grit-O Scale. The results showed that a bifactor ESEM model, with one general factor of overall grit and two specific factors (consistency of interest and perseverance of effort), fitted the data the best, showed strong measurement invariance across the three national cohorts, and showed itself to be a reliable measure. Furthermore, concurrent validity was established by showing that the three grit factors were directly and positively related to task performance. As such, meaningful comparisons between the three national cohorts could be made. Latent mean comparisons showed that students from the USA showed lower overall grit and consistency in interest than their European counterparts. However, those from the USA reported higher levels of perseverance. Similarly, students from Hong Kong also showed slightly lower levels of overall grit and perseverance than those from Europe. However, no differences in their interest were revealed.

5.1 Psychometric Properties of the Grit-O Scale

When considering the factorial validity of the Grit-O Scale, the results showed that none of the traditional and restrictive CFA models fitted the data adequately. However, the bifactor CFA model did show a good fit, but did not produce a non-significant RMSEA. To some extent, this provided support for the multidimensionality of the Grit-O Scale. These results are contrary to other studies where various forms of the traditional CFA models were shown to fit the data well in different countries (e.g., Disabato et al., 2019; Duckworth et al., 2007; Kim & Lee, 2015; Tyumeneva et al., 2019). These contrasting findings imply that the different national cohorts in this study may see grit differently and that the restrictive way in which the Grit-O Scale measures/models grit does not adequately compensate for these interpretative differences. This means that when the Grit-O Scale is used to assess grit in different national contexts, the results may not be reliable and also that comparisons between national cohorts cannot be made. Researchers should, therefore, be cautious about using the total and sub-scores of the Grit-O Scale when using CFA approaches in different nations (Disabato et al., 2019).

In contrast, the less-restrictive ESEM models seemed to fit the data significantly better than their CFA counterparts. An ESEM model with two first-order factors and a bifactor ESEM model (with one general grit factor and two specific factors) met both model fit and measurement quality criteria. This implies that allowing for cross-loadings (but constraining these to be as close to zero as possible) provides a more accurate representation of how grit is seen and measured in samples comprised of different nations. Therefore, these cross-loadings can potentially compensate for both wording and interpretation effects that may occur between national cohorts (Morin et al., 2020). This ESEM approach is also more in

line with Duckworth's (2016) original theoretical conceptualization of grit, where grit is seen as a function of (but also separate from) interest and perseverance. Duckworth et al. (2007) argue that perseverance and interest are dynamically tied together—where perceptions of the one affect the performance of the other. This, however, cannot adequately be modelled or captured by traditional CFA approaches where the interactions (or 'factor loadings') between items are constrained to be zero. Furthermore, the ESEM approach also provides some evidence that supports Duckworth et al.'s (2007) idea that grit is a universal construct, through compensating for cultural differences in the interpretation of some of the items. As such, ESEM models seem to be a more viable modelling strategy when assessing grit across nations.

In more practical terms, the bifactor ESEM model fitted the data significantly better than the ESEM model with two first-order factors. This implies that grit (as measured by the Grit-O Scale) is more adequately explained by a general factor of overall grit, accompanied by two specific factors (interest and perseverance), where cross-loadings are permitted. This model also ignores the hierarchical superiority of constructs through the expression of the cross-loadings (Morin et al., 2020). Therefore, each specific factor measured by the Grit-O Scale has unique explanatory power, over and above an overall grit factor (Van Zyl et al., 2020). The items of the Grit-O Scale, thus, represent some shared variance between the general factor and the specific factors. In this study, reliability estimates further confirmed that these factors were reliably measured within the current sample. The results, therefore, supported the multidimensionality of the Grit-O Scale, and the subscores of the ESEM bifactor model could be meaningfully used to describe individuals' grit, interest, and perseverance.

To determine whether these sub-scores could be used to compare grit between different national cohorts, the two best-fitting ESEM models (i.e., the ESEM model with two first-order factors and the bifactor ESEM model) were subjected to invariance testing. The results showed that configural, metric, and scalar invariance could only be established for the bifactor ESEM model. In other words, only the bifactor ESEM model showed itself to be invariant between the European, US, and Hong Kong cohorts. The configural invariance showed that the factorial structure of the bifactor ESEM model was equivalent across cohorts. This implies that participants from Europe, the USA, and Hong Kong view grit, perseverance, and interest similarly when such is expressed through cross-loadings between the two specific factors. Metric invariance suggested that the strength of the relationships between the items and the three factors was similar for those from Europe, the USA, and Hong Kong. Therefore, the relationship between latent factors and other factors could be reliably compared (Wong & Wong, 2020). Finally, scalar invariance was also established. This, therefore, implies that the intercepts were similar between the three cohorts. Taken together, the results showed that the differences in scores on items/factors of the bifactor ESEM model could be meaningfully interpreted and that comparisons between groups could be made.

5.2 Concurrent Validity: Grit and Task Performance

Before mean comparisons could be made, the final step was to establish the concurrent validity of the bifactor ESEM model of the Grit-O Scale by relating it to task performance. Positive relationships between grit, perseverance, interest, and task performance were identified. This implies that gritty individuals with high levels of interest and perseverance are more likely to perform better in the completion of their educational tasks. However, relative

weight analysis did show that overall grit and perseverance were the most substantial contributors to task performance. Interest only explained 1.77% of the rescaled variance in task performance. This means that overall experiences of grit and perseverance seem to be the most important facets facilitating students' performance and that interest plays a minor role. Given that grit is positioned as a non-cognitive indicator of both achievement and performance, these results were not surprising (Duckworth et al., 2017). Furthermore, Van Zyl et al. (2020) found that results were similar with a more traditional bifactor CFA model, showing that grit and perseverance seemed to have stronger relationships to performance than interest. Taken together, concurrent validity could, therefore, be established.

5.3 Cross-National Differences

Since the factorial validity, measurement invariance, and concurrent validity could be established for the bifactor ESEM model, meaningful cross-national comparisons could be made. With Europe as the reference group, latent mean comparisons indicated that participants from the USA showed lower levels of overall grit and interest, but scored higher on perseverance. Similarly, participants from Hong Kong showed lower levels of grit and perseverance than those from Europe. This implies that European students in this sample may be grittier than their US and Hong Kong peers; however, they were less likely to show drive and perseverance in pursuing their goals than those from the USA. These findings are broadly in line with those of Disabato et al. (2019).

5.4 Limitations and Recommendations

Several limitations of the current study are worth mentioning. First, the sample was broadly cross-sectional and relied on convenience sampling. This implies that the results might not adequately represent the sentiments of the entire population and that positive bias could be present. Gritty individuals were possibly more likely to respond to the questionnaire, which might also have skewed results. Future research should aim for a more representative sample of the overall population to make meaningful inferences about the generalizability of the results.

Second, grit is seen as being a relatively stable trait, but has been shown to be higher in older adults (Duckworth, 2016). Our sample was made up of primarily young university students who were still in the process of discovering their core interests. For this reason, our results may not be comparable to other cohorts of adults. Therefore, it would be interesting to investigate whether the general and specific factors found in this study are stable over time. Furthermore, given that the factorial structure of grit has varied significantly between different studies, it is suggested that the bifactor ESEM model be replicated in other contexts to determine whether this is, indeed, the best way to view or model grit. Third, concurrent validity was only established by a self-report measure on task performance. It is suggested that future research employ more objective measures (such as hard academic performance metrics or academic throughput) as well as measures pertaining to personal and professional accomplishment. Last, in the current sample, Item 4 on the perseverance subscale ("Setbacks do not discourage me") was problematic and was, thus, removed. This item may, therefore, not be a good indicator of grit. This specific item has also been shown to be problematic in other studies (cf. Christensen & Knezek, 2014; Tyumeneva et al., 2019; Van Zyl et al., 2020). That being the case, it is suggested that the

item either be removed in future research or be reformulated to be more in line with local values, norms, and beliefs.

6 Conclusion: Clinical and Theoretical Implications

In conclusion, the results of the study showed that researchers should be cautious when employing the Grit-O Scale in cross-cultural samples. Modelling grit through the traditional independent cluster modelling CFA approach does not allow researchers to make meaningful cross-national comparisons. Forcing items to load onto specific factors and constraining cross-loadings to zero may produce biased parameters, therefore limiting the straightforward use of the instrument. This is in line with Disabato et al.'s (2019) argument that researchers interested in studying grit should rather report the scores separately for each facet for each cultural or national cohort, and that meaningful comparisons between cultures are cumbersome. However, our results support the notion that cross-national comparisons may only be problematic when traditional CFA approaches to estimate the factorial structure of grit is favoured. In contrast, ESEM modelling approaches may compensate for cross-cultural differences in the experience of the concept, and also control for cross-national differences in the interpretation of items. Therefore, the ESEM approach may be more appropriate for cross-cultural and cross-national comparison studies, as it allows for these differences in the experience/interpretation of grit to be meaningfully captured, modelled, and compensated for. As such, the bifactor ESEM model (with one general grit factor and two specific factors) may overcome the Grit-O Scale's limitations to measure grit accurately and reliably across nations. Evidence for this model appears increasingly important as professionals are beginning to integrate grit assessments and interventions into different clinical and academic services (Griffin et al., 2016). Our findings provide a foundation by which grit, as measured by the Grit-O scale, can be employed from a culturally-responsive framework to ensure these services are clinical meaningful and affirming given clients' goals and values. Moreover, these findings provide an interesting platform by which grit can be evaluated from different sociocultural perspectives to support greater cultural competence among different service providers.

Finally, the bifactor ESEM model found in this study also has practical implications for the theoretical understanding of grit. Unlike in Duckworth et al.'s (2007) original conceptualization of grit as a unidimensional- or higher-order factor, our results show that grit could perhaps be seen an auxiliary factor that is similar yet distinct from perseverance and long-term interest. This implies that when one "controls" for an individual's grittiness (as a general trait), the "left over" experiences of perseverance and long-term interest (as specific factors) could be seen as unique factors that are unrelated to one's overall experience of grit.

Appendix 1: Graphical Representations of the Different Factorial Models
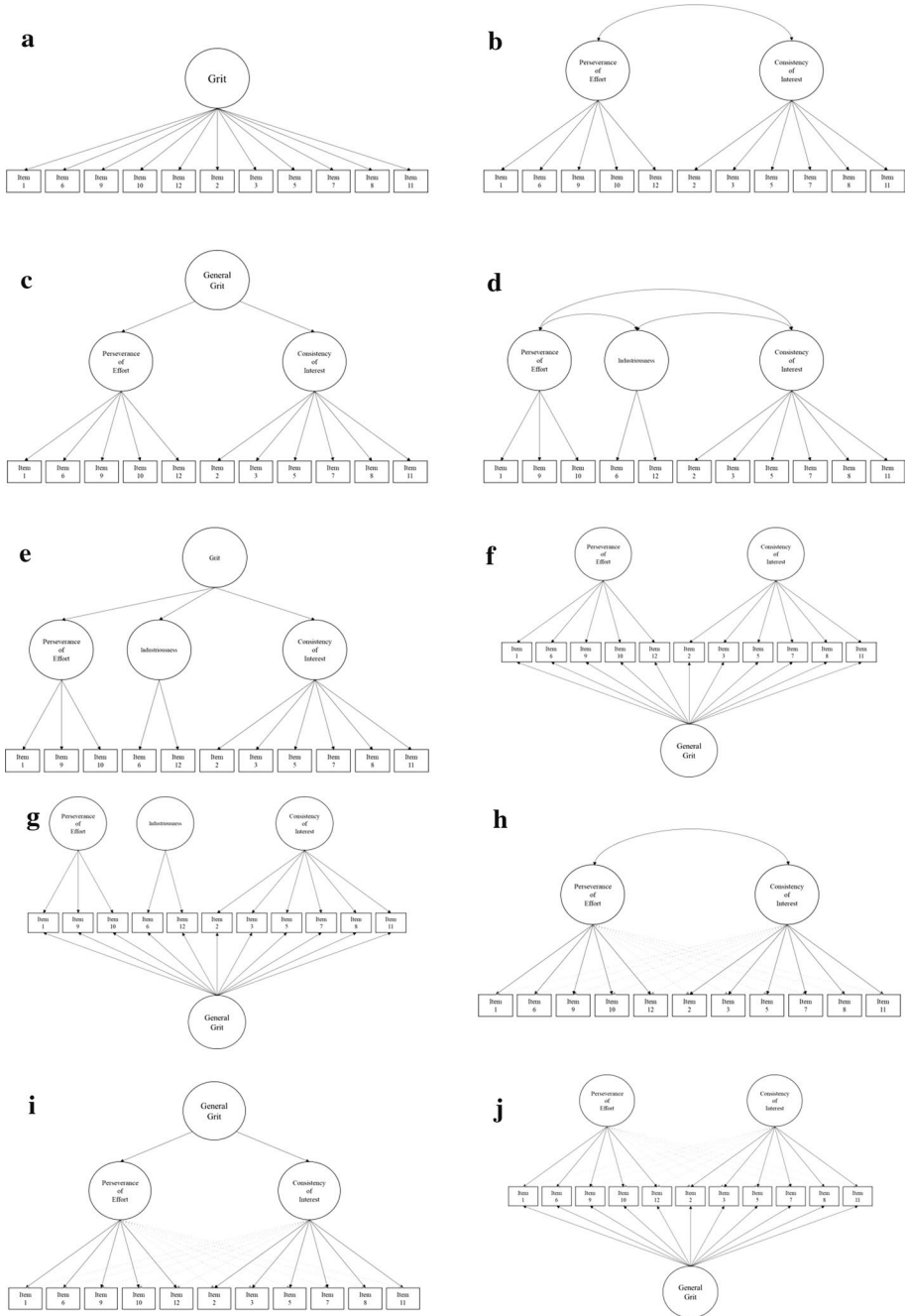
Appendix 2: Competing Measurement Models for Separate Countries

See Tables 9, 10, 11.

Table 9 Competing measurement models for Europe

Model	Type	χ^2	df	CFI	TLI	RMSEA		SRMR	AIC	BIC	ABIC	Meets model-fit criteria	Meets measurement quality criteria
Model 1	Unidimensional model	1149.57	44	0.66	0.58	0.15	[.147–.162]	0.12	31669.13	31832.89	31728.07	No	No
Model 2	Two first-order factor model	497.27	43	0.86	0.82	0.10	[.092–.108]	0.09	31018.83	31187.54	31079.56	No	Yes
Model 3	Second-order model with two first-order factors	497.27	43	0.86	0.82	0.10	[.092–.108]	0.09	31018.83	31187.54	31079.56	No	Yes
Model 4	Three first-order factor model	482.90	41	0.87	0.82	0.10	[.093–.109]	0.09	31008.46	31187.10	31072.76	No	Partially
Model 5	Second-order model with three first-order factors*	539.45	43	0.85	0.81	0.11	[.093–.113]	0.09	31061.01	31229.72	31121.73	No	Partially
Model 6	Bifactor model (one general and two specific factors)	167.63	33	0.96	0.93	**0.06**	[.053–.073]	0.06	30709.19	30927.52	30787.77	Yes	Yes
Model 7	Bifactor model (one general and three specific factors)	–	–	–	–	–	–	–	–	–	–	–	–
Model 8	ESEM model with two first-order factors	123.33	34	0.97	0.96	**0.05**	[.041–.060]	0.02	30662.89	30876.26	30739.69	Yes	Yes
Model 9	H-ESEM model with two first-order factors	123.33	34	0.97	0.96	**0.05**	[.041–.060]	0.02	30662.89	30876.26	30739.69	Yes	Yes
Model 10	Bifactor ESEM model (one general and two specific factors)	68.42	25	0.99	0.97	**0.04**	[.029–.052]	0.02	30625.98	30884.02	30718.86	Yes	Yes

χ^2 chi-square; *df* degrees of freedom; *TLI* Tucker–Lewis index; *CFI* comparative fit index; *RMSEA* root mean square error of approximation [90% CI]; *SRMR* standardized root mean square residual; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *ABIC* adjusted Bayesian information criterion; bold = non-significant $p > 0.01$; * for convergence, first-order factor loadings were constrained to be equal

Table 10 Competing measurement models for the USA

Model	Type	χ^2	df	CFI	TLI	RMSEA		SRMR	AIC	BIC	ABIC	Meets model fit criteria	Meets measurement quality criteria
Model 1	Unidimensional model	818.39	44	0.53	0.41	0.19	[.182–.205]	0.17	13946.35	14083.46	13978.72	No	No
Model 2	Two first-order factor model	160.42	43	0.93	0.91	0.08	[.064–.089]	0.06	13290.37	13431.64	13323.73	Yes	Yes
Model 3	Second-order model with two first-order factors	160.42	43	0.93	0.91	0.08	[.064–.089]	0.06	13290.37	13431.64	13323.73	Yes	Yes
Model 4	Three first-order factor model	159.86	41	0.93	0.90	0.08	[.066–.091]	0.06	13293.81	13443.39	13329.13	Yes	Partially
Model 5	Second-order model with three first-order factors*	98.47	43	0.93	0.91	**0.06**	[.044–.075]	0.06	9538.27	9670.50	9562.63	Yes	Partially
Model 6	Bifactor model (one general and two specific factors)	59.40	33	0.98	0.97	**0.04**	[.024–.058]	0.03	13209.36	13392.17	13252.52	Yes	Yes
Model 7	Bifactor model (one general and three specific factors)	–	–	–	–	–	–	–	–	–	–	–	–
Model 8	ESEM model with two first-order factors	100.91	34	0.96	0.93	**0.07**	[.050–.079]	0.03	13248.87	13427.53	13291.05	Yes	Yes
Model 9	H-ESEM model with two first-order factors	100.91	34	0.96	0.93	**0.07**	[.050–.079]	0.03	13248.87	13427.53	13291.05	Yes	Partially
Model 10	Bifactor ESEM model (one general and two specific factors)	**47.32**	26	0.99	0.98	**0.04**	[.022–.060]	0.02	13211.27	13423.17	13261.31	Yes	Yes

χ^2 chi-square; *df* degrees of freedom; *TLI* Tucker–Lewis index; *CFI* comparative fit index; *RMSEA* root mean square error of approximation [90% CI]; *SRMR* standardized root mean square residual; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *ABIC* adjusted Bayesian information criterion; bold=non-significant $p>0.01$; * for convergence, first-order factor loadings were constrained to be equal

Table 11 Competing measurement models for Hong Kong

Model	Type	χ^2	df	CFI	TLI	RMSEA	[90% CI]	SRMR	AIC	BIC	ABIC	Meets model-fit criteria	Meets measurement quality criteria
Model 1	Unidimensional model	382.27	44	0.56	0.45	0.15	[.133–.160]	0.13	9820.08	9948.41	9843.72	No	No
Model 2	Two first-order factor model	96.21	43	0.93	0.91	**0.06**	[.043–.074]	0.06	9536.01	9668.23	9560.37	Yes	No
Model 3	Second-order model with two first-order factors	96.21	43	0.93	0.91	**0.06**	[.043–.074]	0.06	9536.01	9668.23	9560.37	Yes	No
Model 4	Three first-order factor model	88.25	41	0.94	0.92	**0.06**	[.040–.073]	0.06	9532.05	9672.05	9557.84	Yes	No
Model 5	Second-order model with three first-order factors*	98.47	43	0.93	0.91	**0.06**	[.044–.075]	0.06	9538.27	9670.50	9562.63	Yes	No
Model 6	Bifactor model (one general and two specific factors)	47.70	33	0.98	0.97	**0.04**	[.004–.056]	0.03	9507.51	9678.62	9539.02	Yes	Yes
Model 7	Bifactor model (one general and three specific factors)	–	–	–	–	–	–	–	–	–	–	–	–
Model 8	ESEM model with two first-order factors	66.28	34	0.96	0.93	**0.05**	[.033–.070]	0.03	9524.09	9691.31	9554.89	Yes	Yes
Model 9	H-ESEM model with two first-order factors	66.28	34	0.96	0.93	**0.05**	[.033–.070]	0.03	9524.09	9691.31	9554.89	Yes	Partially
Model 10	Bifactor ESEM model (one general and two specific factors)	**41.93**	26	0.98	0.97	**0.04**	[.015–.063]	0.03	9515.74	9714.07	9552.27	Yes	Yes

χ^2 chi-square; *df* degrees of freedom; *TLI* Tucker–Lewis index; *CFI* comparative fit index; *RMSEA* root mean square error of approximation [90% CI]; *SRMR* standardized root mean square residual; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *ABIC* adjusted Bayesian information criterion; bold = non-significant $p > 0.01$; *For convergence, first-order factor loadings were constrained to be equal

References

Alexander, F. K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *The Journal of Higher Education, 71*(4), 411–431.

Amazon. (2018). Grit: The power of passion and perseverance. Retrieved from https://www.amazon.com/Grit-Passion-Perseverance-Angela-Duckworth/dp/1501111116?SubscriptionId=AKIAI2KQ2FKXUWTESZCA&tag=rt50004-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1501111116#nav-top. Accessed on 31 May 2022.

Areepattamannil, S., & Khine, M. S. (2018). Evaluating the psychometric properties of the original grit scale using Rasch analysis in an Arab adolescent sample. *Journal of Psychoeducational Assessment, 36*(8), 856–862.

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

Cattaneo, M., Horta, H., Malighetti, P., Meoli, M., & Paleari, S. (2019). Universities' attractiveness to students: The Darwinism effect. *Higher Education Quarterly, 73*(1), 85–99.

Cazayoux, M., & DeBeliso, M. (2019). Effect of grit on performance in Crossfit in advanced and novice athletes. *Turkish Journal of Kinesiology, 5*(1), 28–35.

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.

Christensen, R., & Knezek, G. (2014). Comparative measures of grit, tenacity and perseverance. *International Journal of Learning, Teaching and Educational Research, 8*(1), 16–30.

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.

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

Datu, J. A. D., Yuen, M., & Chen, G. (2018). Exploring determination for long-term goals in a collectivist context: A qualitative study. *Current Psychology, 37*(1), 263–271.

De Beer, L.T. & Van Zyl, L.E. (2019). ESEM code generator for Mplus. Retrieved from 10.6084/m9.figshare.8320250

Disabato, D., Goodman, F., & Kashdan, T. (2019). Is grit relevant to wellbeing and strengths? Evidence across the globe for separating perseverance of effort and consistency of interests. *Journal of Personality.* https://doi.org/10.1111/jopy.12382

Duckworth, A. L. (2016). *Grit: The power of passion and perseverance* (Vol. 234). Scribner.

Duckworth, A. L., & Eskreis-Winkler, L. (2013). True grit. *Aps Observer, 26*(4), 1–10.

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

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.

Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research, 69*(8), 3192–3198.

Gaotlhobogwe, M. (2017). The impact of lack of resources on declining students' enrolments in design and technology in Botswana junior secondary schools. *Design and Technology Education: An International Journal, 17*(1), 10–17.

Griffin, M. L., McDermott, K. A., McHugh, R. K., Fitzmaurice, G. M., & Weiss, R. D. (2016). Grit in patients with substance use disorders. *The American Journal on Addictions, 25*(8), 652–658. https://doi.org/10.1111/ajad.12460

Harman, H. H. (1976). *Modern factor analysis*. Chicago, IL. University of Chicago press.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24.

He, J., & Van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*(2), 2307–2919.

He, J., & Van de Vijver, F. (2013). Methodological issues in cross-cultural studies in educational psychology. *Advancing cross-cultural perspectives on educational psychology: A festschrift for Dennis McInerney* (pp. 39–56). Information Age Publishing.

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.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1–55.

Jachimowicz, J. M., Wihler, A., Bailey, E. R., & Galinsky, A. D. (2018). Why grit requires perseverance and passion to positively predict performance. *Proceedings of the National Academy of Sciences USA, 115*(40), 9980–9985.

JASP. (2021). Jeffery's Amazing Stats Program (v.0.14.1). Accessed via the World Wide Web: https://jasp-stats.org

Jordan, S. L., Ferris, G. R., Hochwarter, W. A., & Wright, T. A. (2019). Toward a work motivation conceptualisation of grit in organisations. *Group & Organization Management, 44*(2), 320–360. https://doi.org/10.1177/1059601119834093

Kim, Y. J., & Lee, C. S. (2015). Effects of grit on the successful aging of the elderly in Korea. *Indian Journal of Science and Technology, 8*(S7), 373–378.

Kline, R. B. (2010). *Principles and practices of structural equation modelling* (3rd ed.). Gilford Press.

Koopmans, L., Bernaards, C., Hildebrandt, V., Van Buuren, S., Van der Beek, A. J., & De Vet, H. C. (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

Krifa, I., Hallez, Q., Van Zyl, L. E., Braham, A., Sahli, J., Nasr, S. E., & Shankland, R. (2021). Effectiveness of an online positive psychology intervention among Tunisian healthcare students on mental health and study engagement during the Covid-19 pandemic. *Applied Psychology: Health and Wellbeing*. https://doi.org/10.1111/aphw.12332

Krifa, I., Van Zyl, L. E., Braham, A., Nasr, S. E., & Shankland, R. (2022). Mental health during COVID-19 pandemic: The role of optimism and emotional regulation. *International Journal of Environmental Research and Public Health, 19*(3), 1413. https://doi.org/10.3390/ijerph19031413

Lesener, T., Pleiss, L. S., Gusy, B., & Wolter, C. (2020). The study demands-resources framework: An empirical introduction. *International Journal of Environmental Research and Public Health, 17*, 5183. https://doi.org/10.3390/ijerph17145183

Macfarlane, B. (2015). Student performativity in higher education: Converting learning as a private space into a public performance. *Higher Education Research & Development, 34*(2), 338–350.

Magada, T., & Govender, K. K. (2017). Culture, leadership and individual performance: A South African Public Service Organization study. *Journal of Economics and Management Theory, 1*(1), 1–2. https://doi.org/10.22496/jemt20161104109

Mason, H. D. (2018). Grit and academic performance among first-year university students: A brief report. *Journal of Psychology in Africa, 28*(1), 66–68. https://doi.org/10.1080/14330237.2017.1409478

Mokgele, K. R., & Rothmann, S. (2014). A structural model of student wellbeing. *South African Journal of Psychology, 44*(4), 514–527.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory exploratory structural equation modeling 19 structural equation modeling and bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (4th ed., Vol. 2, pp. 1044–1073). Wiley Publishers.

Muthén, L. K., & Muthén, B. O. (2021). *Mplus (Version 8.4). [Statistical software]*. Muthén and Muthén.

Nelson, S., & Baltes, B. (2019). For the culture: Grit, student engagement, and academic performance at a historically black community college. *Journal of Applied Research in the Community College, 26*(1), 89–102.

Perkins-Gough, D. (2013). The significance of grit: A conversation with Angela Lee Duckworth. *Educational Leadership, 71*(1), 14–20.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879.

Saeed, A. A., Bahnassy, A. A., Al-Hamdan, N. A., Almudhaibery, F. S., & Alyahya, A. Z. (2016). Perceived stress and associated factors among medical students. *Journal of Family & Community Medicine, 23*(3), 166.

Stelnicki, A. M., Nordstokke, D. W., & Saklofske, D. H. (2015). Who is the successful university student? An analysis of personal resources. *Canadian Journal of Higher Education, 45*(2), 214–228.

Steuber, T. D., Nisly, S. A., & Gillette, C. M. (2019). Grit in pharmacy faculty: a pilot analysis focused on productivity measures. *Currents in Pharmacy Teaching and Learning, 11*(10), 1029–1034.

Stucky, B. D., Thissen, D., & Orlando Edelen, M. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement, 37*(1), 41–57.

Tehseen, S., Ramayah, T., & Sajilan, S. (2017). Testing and controlling for common method variance: A review of available methods. *Journal of Management Sciences, 4*(2), 142–168.

Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Tyumeneva, Y., Kardanova, E., & Kuzmina, J. (2019). Grit: Two related but independent constructs instead of one. Evidence from item response theory. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000424

Van der Vaart, L., Van Zyl, L. E., & Van Wingerden, J. (2021). Developing gritty job seekers: A need supportive approach to grit interventions. In L. E. van Zyl, C. Olckers, & L. van der Vaart (Eds.), *Multidisciplinary perspectives on grit: contemporary theories, assessments, applications and critiques.* Springer.

Van Zyl, L. E. (2021). Social study resources and social wellbeing before and during the intelligent COVID-19 lockdown in The Netherlands. *Social Indicator Research*. https://doi.org/10.1007/s11205-021-02654-2

Van Zyl, L. E., Arijs, D., Cole, M. L., Glinska, A., Roll, L. C., Rothmann, S., Shankland, R., Stavros, J., & Verger, N. (2021a). The strengths use scale: Psychometric properties, longitudinal invariance and criterion validity. *Frontiers in Psychology, 12*, 676153. https://doi.org/10.3389/fpsyg.2021.676153

Van Zyl, L.E., Klibert, J., Shankland, R., See-To, E.W.K., & Rothmann, S. (2022). The general academic self efficacy scale: Psychometric properties, longitudinal invariance and criterion validity. *Journal of Psychoeducational Assessment*. https://doi.org/10.1177/07342829221097174

Van Zyl, L. E., Olckers, C., & Roll, L. C. (2020). The psychometric properties of the Grit-O scale within the Twente Region in The Netherlands: An ICM-CFA vs. ESEM approach. *Frontiers in Psychology, 11*, 796. https://doi.org/10.3389/fpsyg.2020.00796

Van Zyl, L. E., Olckers, C., & van der Vaart, L. (2021b). *Multidisciplinary perspectives on grit: Contemporary theories, assessments, applications and critiques*. Springer. https://doi.org/10.1007/978-3-030-57389-8

Van Zyl, L. E., & Rothmann, S. (2022). Grand challenges for positive psychology: Future perspectives and opportunities. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.833057

Van Zyl, L. E., Rothmann, S., & Zondervan-Zwijnenburg, M. A. J. (2021c). Longitudinal trajectories of study characteristics and mental health before and during the COVID-19 lockdown. *Frontiers in Psychology, 12*, 1–13. https://doi.org/10.3389/fpsyg.2021.63353

Van Zyl, L. E., & Ten Klooster, P. M. (2022). Exploratory structural equation modelling: Practical guidelines and tutorial with a convenient online tool for mplus. *Frontiers in Psychiatry, 12*, 795672. https://doi.org/10.3389/fpsyt.2021.795672

Vogelsang, K. (2018). *The Relationship between Person-Environment fit, Grit, and Task performance: A mediation analysis* (Bachelor's thesis, University of Twente).

Webster-Wright, A. (2019). Grace and grit: the politics, poetics and performance of ageing as a woman. *Life Writing, 16*(1), 97–111.

Wong, J., & Wong, X. (2020). *Structural equation modelling: Applications using Mplus* (2nd ed.). Wiley.

Authors and Affiliations

Llewellyn E. van Zyl[1,2,3,4] · **Babet Heijenk**[1] · **Jeff Klibert**[5] · **Rebecca Shankland**[6] ·
Nicolas B. Verger[7] · **Sebastiaan Rothmann**[2] · **Vincent Cho**[8] · **Katherine Feng**[8] ·
Eric W. K. See-To[9] · **Lara C. Roll**[2,10] · **Leander van der Meij**[1]

[1]   University of Eindhoven, Human Performance Management, Eindhoven, The Netherlands

[2]   Optentia Research Unit, North-West University (Vaal Triangle Campus), Vanderbijlpark,
        South Africa

[3]   Department of Human Resource Management, University of Twente, Enschede, The Netherlands

[4]   Department of Social Psychology, Goethe University, Frankfurt am Main, Germany

[5]   Department of Psychology, Georgia Southern University, Statesboro, USA

[6]   Laboratoire Développement, Individu, Processus, Handicap, Éducation (DIPHE), Université
        Lumière Lyon 2, Lyon, France

[7]   Department of Psychology and Allied Health Sciences, Glasgow Caledonian University, Glasgow,
        UK

[8]   Department of Management and Marketing, Hong Kong Polytechnic University, Hung Hom,
        Hong Kong

[9]   Department of Computing and Decision Sciences, Lingnan University, Tuen Mun, Hong Kong

[10]   Work, Organisational, and Personnel Psychology, KU Leuven, Leuven, Belgium